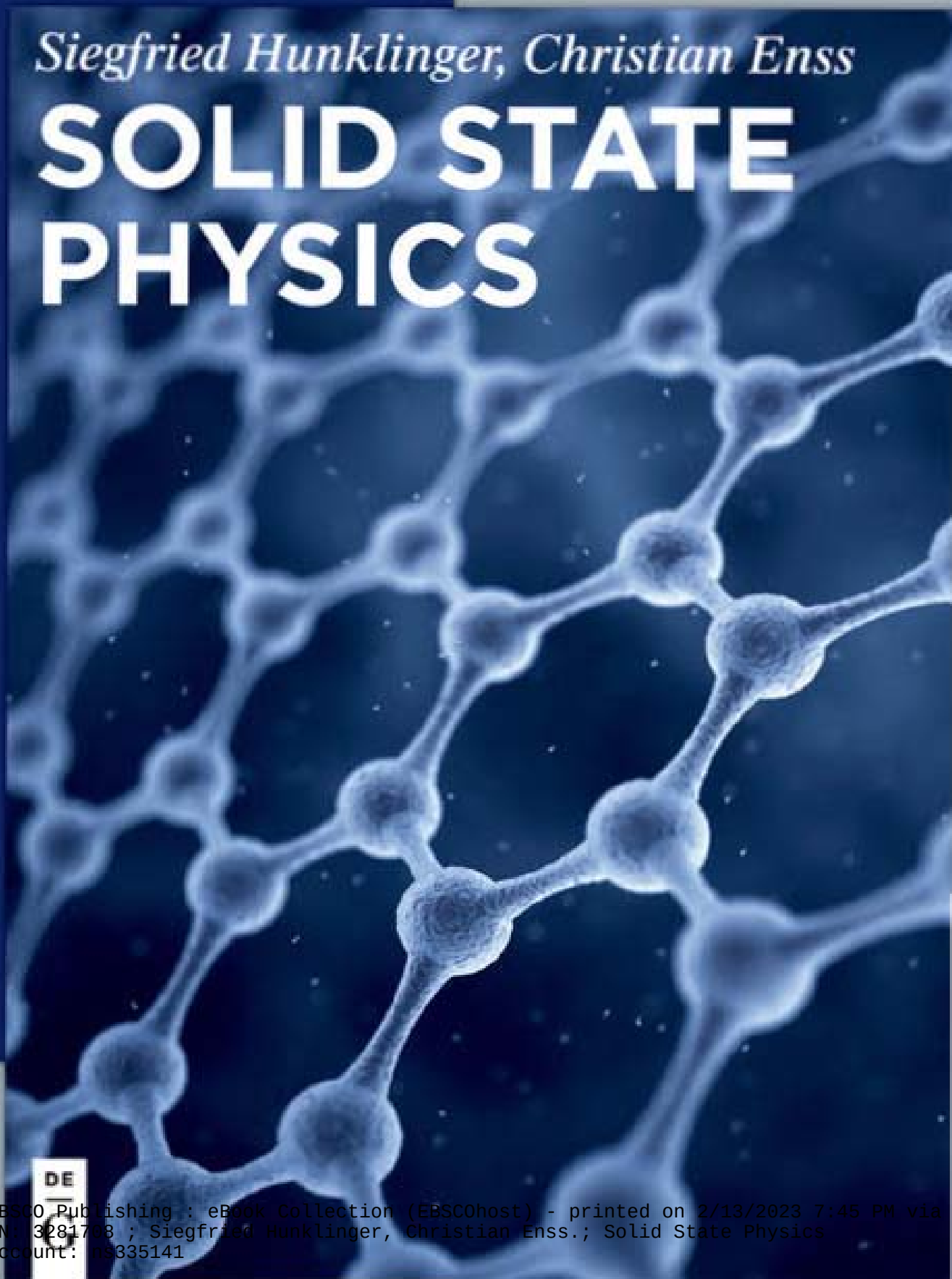
DE GRUYTER

GRADUATE

Siegfried Hunklinger, Christian Enss

SOLID STATE PHYSICS



Copyright 2022. De Gruyter. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

EBSCO Publishing : eBook Collection (EBSCOhost) - printed on 2/13/2023 7:45 PM via
AN: 3281708 ; Siegfried Hunklinger, Christian Enss.; Solid State Physics
Account: ns335141

Siegfried Hunklinger, Christian Enss

Solid State Physics

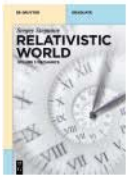
EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Also of interest

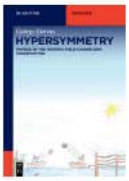
Classical Mechanics
Hiqmet Kamberaj, 2021
ISBN 978-3-11-075581-7, e-ISBN (PDF) 978-3-11-075582-4

De Gruyter Studies in Mathematical Physics
Edited by Michael Efroimsky, Leonard Gamberg, Dmitry Gitman, Alexander Lazarian, Boris M. Smirnov
ISSN 2194-3532, e-ISSN 2194-3540

Quantum Mechanics
An Introduction to the Physical Background and Mathematical Structure
Gregory L. Naber, 2021
ISBN 978-3-11-075161-1, e-ISBN (PDF) 978-3-11-075194-9



Relativistic World
Volume 1 Mechanics
Serhii Stepanov, 2018
ISBN 978-3-11-051587-9, e-ISBN (PDF) 978-3-11-051588-6



Hypersymmetry
Physics of the Isotopic Field-Charge Spin Conservation
György Darvas, 2020
ISBN 978-3-11-071317-6, e-ISBN (PDF) 978-3-11-071318-3

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Siegfried Hunklinger, Christian Enss

Solid State Physics

DE GRUYTER

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Mathematics Subject Classification 2010
Primary: 74N05, 82D30; Secondary: 37K60, 74E15, 80A17, 82D37, 82D55

Authors
Prof. Dr. Siegfried Hunklinger
s.hunklinger@physik.uni-heidelberg.de

Prof. Dr. Christian Enss
christian.enss@kip.uni-heidelberg.de

Kirchhoff Institute for Physics
Heidelberg University
Im Neuenheimer Feld 227
D-69120 Heidelberg
Germany

ISBN 978-3-11-066645-8
e-ISBN (PDF) 978-3-11-066650-2
e-ISBN (EPUB) 978-3-11-066708-0

Library of Congress Control Number: 2022935818

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

© 2022 Walter de Gruyter GmbH, Berlin/Boston
Cover image: Gettyimages / Andrey Prokhorov
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Contents

Preface — XI

1 Introductory Remarks — 1

2 Bonding in Solids — 5
2.1 Types of Bonds — 5
2.1.1 Binding Energy — 7
2.2 The Van der Waals Bond — 9
2.2.1 The Lennard-Jones Potential — 10
2.2.2 The Binding Energy of Noble Gas Crystals — 11
2.3 The Ionic Bond — 13
2.3.1 The Determination of the Binding Energy — 13
2.3.2 The Madelung Energy — 14
2.4 The Covalent Bond — 17
2.4.1 H_2^+ Molecular Ion — 17
2.4.2 The Hydrogen Molecule — 21
2.4.3 Types of Covalent Bonds — 23
2.5 The Metallic Bond — 26
2.6 The Hydrogen Bond — 29
2.7 Exercises and Problems — 31

3 The Structure of Solids — 33
3.1 The Production of Crystalline and Amorphous Solids — 33
3.1.1 Single Crystal Growth — 33
3.1.2 Production of Alloys — 36
3.1.3 Glass Production — 42
3.2 Order and Disorder — 43
3.3 The Structure of Crystals — 48
3.3.1 Translational Symmetry and Crystal Systems — 48
3.3.2 Clusters and Quasicrystals — 55
3.3.3 Notation and the Influence of the Basis — 57
3.3.4 Simple Crystal Lattices — 61
3.3.5 The Wigner-Seitz Cell — 67
3.3.6 Nanotubes — 68
3.3.7 Surfaces of Solids — 70
3.4 The Structure of Amorphous Solids — 71
3.4.1 The Pair Correlation Function — 72
3.5 Exercises and Problems — 75

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

4 Structure Determination — 79
4.1 General Remarks — 79
4.2 Diffraction Experiments — 81
4.2.1 Scattering Amplitude — 82
4.3 The Fourier Description of Point Lattices — 84
4.3.1 The Reciprocal Lattice — 84
4.3.2 The Brillouin Zone — 87
4.3.3 Miller Indices — 90
4.4 Determination of the Crystal Structure — 93
4.4.1 Ewald Sphere and Bragg Condition — 95
4.4.2 The Structure Factor — 97
4.4.3 Atomic Structure Factor — 101
4.4.4 Surfaces and Thin Layers — 103
4.4.5 The Phase Problem and the Diffraction Peak Width — 104
4.5 Scattering Experiments on Amorphous Solids — 106
4.6 Experimental Methods — 111
4.6.1 Measurement Procedures — 113
4.6.2 Measurements on Surfaces and Thin Films — 117
4.7 Exercises and Problems — 120

5 Structural Defects — 123
5.1 Point Defects — 123
5.1.1 Vacancies — 124
5.1.2 Color Centers — 128
5.1.3 Interstitials — 131
5.1.4 Impurities — 132
5.1.5 Atomic Transport — 132
5.2 Extended Defects — 138
5.2.1 Mechanical Strength — 138
5.2.2 Dislocations — 141
5.2.3 Grain Boundaries — 149
5.3 Defects in Amorphous Solids — 151
5.4 The Order-Disorder Transition — 154
5.5 Exercises and Problems — 157

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

6 Lattice Dynamics — 159
6.1 Elastic Properties — 159
6.1.1 Stress and Deformation — 160
6.1.2 Elastic Constants — 162
6.1.3 Sound Propagation — 165
6.2 Lattice Vibrations — 170
6.2.1 Lattice with a Monatomic Basis — 171
6.2.2 Lattice with a Polyatomic Basis — 175
6.2.3 The Equation of Motion — 180
6.3 Experimental Determination of Dispersion Curves — 182
6.3.1 Dynamic Scattering, Phonons — 182
6.3.2 Coherent Inelastic Neutron Scattering — 186
6.3.3 Debye-Waller Factor — 188
6.3.4 Experimentally Determined Dispersion Curves — 190
6.3.5 Light Scattering — 193
6.4 The Specific Heat Capacity — 198
6.4.1 Phonon Density of States — 199
6.4.2 The Specific Heat in the Debye Approximation — 205
6.4.3 The Specific Heat in Low-dimensional Systems — 210
6.4.4 The Zero-point Energy and Number of Excited Phonons — 212
6.5 Vibrations in Amorphous Solids — 213
6.5.1 Heat Capacity of Glasses — 215
6.6 Exercises and Problems — 219

7 Lattice Anharmonicity — 223
7.1 Thermal Expansion and the Equation of State — 223
7.2 Phonon-Phonon Scattering — 228
7.2.1 Three Phonon Processes — 228
7.2.2 Ultrasonic Absorption in Crystals — 229
7.2.3 Spontaneous Phonon Decay — 234
7.2.4 Ultrasound Absorption in Amorphous Solids — 235
7.3 Heat Transport in Dielectric Crystals — 238
7.3.1 Ballistic Propagation of Phonons — 239
7.3.2 Thermal Conductivity — 240
7.3.3 Phonon-Phonon Scattering — 242
7.3.4 Influence of Defects — 245
7.3.5 Heat Transport in One-dimensional Samples — 247
7.4 The Thermal Conductivity of Amorphous Solid — 250
7.5 Exercises and Problems — 252

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

8 Electrons in Solids — 255
8.1 Free Electron Gas — 255
8.1.1 Density of States — 257
8.1.2 The Fermi Energy — 261
8.2 Specific Heat — 265
8.3 Collective Phenomena in the Electron Gas — 269
8.3.1 Screened Coulomb Potential — 269
8.3.2 Metal-Insulator Transition — 272
8.4 Electrons in a Periodic Potential — 273
8.4.1 The Bloch Function — 274
8.4.2 Quasi-Free Electrons — 277
8.4.3 Tightly Bound Electrons — 284
8.5 Energy Bands — 290
8.5.1 Metals and Insulators — 290
8.5.2 Brillouin Zones and Fermi Surfaces — 292
8.5.3 Density of States — 296
8.5.4 Graphene and Nanotubes — 299
8.6 Exercises and Problems — 304

9 Electronic Transport Properties — 307
9.1 Equation of Motion and Effective Mass — 307
9.1.1 Electrons as Wave Packets — 307
9.1.2 Electron Motion in Bands — 312
9.1.3 Electrons and Holes — 314
9.2 Transport Properties — 316
9.2.1 Sommerfeld Theory — 317
9.2.2 Boltzmann Equation — 318
9.2.3 Electric Charge Transport — 320
9.2.4 The Scattering of Conduction Electrons — 322
9.2.5 Temperature Dependence of the Electrical Conductivity — 327
9.2.6 One-dimensional Conductors — 330
9.2.7 Luttinger Liquid — 334
9.2.8 Quantum Dots — 337
9.2.9 Heat Transport in Metals — 340
9.2.10 Fermi Function in Stationary Equilibrium — 343
9.3 Electrons in Magnetic Fields — 346
9.3.1 Cyclotron Resonance — 346
9.3.2 Landau Levels — 351
9.3.3 Density of States in Magnetic Fields — 355
9.3.4 De Haas-van Alphen Effect — 358
9.3.5 Hall Effect — 361
9.3.6 Quantum Hall Effect — 364

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

9.3.7 Quantum Hall Effect in Graphene — 371
9.4 Exercises and Problems — 372

10 Semiconductors — 375
10.1 Intrinsic Crystalline Semiconductors — 376
10.1.1 Band Structure, Band Gap and Optical Absorption — 376
10.1.2 The Effective Mass of Electrons and Holes — 380
10.1.3 Charge Carrier Density — 383
10.2 Doped Crystalline Semiconductors — 387
10.2.1 Doping — 387
10.2.2 Charge Carrier Density and Fermi Level — 391
10.2.3 Mobility and Electrical Conductivity — 397
10.3 Amorphous Semiconductors — 400
10.3.1 Electrical Conductivity — 402
10.3.2 Defect States — 405
10.4 Inhomogeneous Semiconductors — 410
10.4.1 p-n Junction — 410
10.4.2 Metal-semiconductor Junction — 418
10.4.3 Semiconductor Heterostructures and Superlattices — 419
10.5 Devices — 424
10.5.1 Technical Application of the p-n Junction — 424
10.5.2 Transistors — 429
10.5.3 The Semiconductor Laser — 432
10.6 Exercises and Problems — 434

11 Superconductivity — 437
11.1 Phenomenological Description — 437
11.1.1 Meissner-Ochsenfeld Effect, London Equations — 439
11.1.2 Thermodynamic Properties — 447
11.2 Microscopic Description — 451
11.2.1 Cooper Pairs — 451
11.2.2 BCS Theory — 456
11.2.3 Experimental Evidence for an Energy Gap — 462
11.2.4 Current Transport Through Interfaces — 467
11.2.5 Critical Current and Critical Magnetic Field — 468
11.3 Macroscopic Wave Function — 471
11.3.1 Flux Quantisation — 471
11.3.2 Josephson Effect — 474
11.4 Ginzburg-Landau Theory and Type-II Superconductors — 479
11.4.1 Ginzburg-Landau Theory — 479
11.4.2 Type-II Superconductors and Interface Energy — 482
11.5 Unconventional Superconductors — 487

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

11.5.1 High-temperature Superconductors — 487
11.5.2 Heavy Fermion Systems — 493
11.5.3 Technical Application of Superconductivity — 496
11.6 Exercises and Problems — 497

12 Magnetism — 499
12.1 General Remarks on Magnetic Quantities — 499
12.2 Dia- and Paramagnetism — 500
12.2.1 Diamagnetism — 500
12.2.2 Paramagnetism — 502
12.3 Ferromagnetism — 509
12.3.1 Meanfield Approximation — 510
12.3.2 Exchange Interaction — 514
12.3.3 Band Ferromagnetism — 518
12.3.4 Spin Waves – Magnons — 522
12.3.5 Temperature Dependence of Magnetization — 524
12.3.6 Ferromagnetic Domains — 526
12.4 Ferri- and Antiferromagnetism — 527
12.4.1 Ferrimagnetism — 527
12.4.2 Antiferromagnetism — 528
12.4.3 Giant Magnetoresistance — 532
12.5 Spin Glasses — 535
12.6 Exercises and Problems — 539

13 Dielectric and Optical Properties — 541
13.1 Dielectric Function and Optical Quantities — 541
13.2 Local Electrical Field — 543
13.3 Dielectric Polarization — 547
13.3.1 Electronic Polarizability — 548
13.3.2 Ionic Polarization — 551
13.3.3 Optical Phonons in Ionic Crystals — 552
13.3.4 Dielectric Function of Ionic Crystals — 554
13.3.5 Phonon Polaritons — 556
13.3.6 Orientation Polarization — 561
13.3.7 Ferroelectricity — 568
13.3.8 Excitons — 573
13.4 Optical Properties of Free Charge Carriers — 575
13.4.1 Electromagnetic Waves in Metals — 577
13.4.2 Plasmons — 580
13.5 Exercises and Problems — 584

Index — 587

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Preface

The understanding of the physics of solids has made enormous progress in the last century, mainly due to the development of quantum mechanics, and has revolutionized our daily lives in many ways. A prime example is the development of semiconductor-based electronics, which has become a constant companion and now controls many aspects of our lives. Clearly, very few scientific and technological developments have had as great an impact in the last century as the spectacular advances in our understanding of the properties of solids. This story is still unfolding and there is very little doubt that modern solid state physics is needed to address some of the challenges we face today.

The present book is based on material from lectures given by both authors over many years at Heidelberg University. A first German edition of this book was published in 2007 and has been expanded over the years to the 5th edition in 2017. This first English edition is a revised and supplemented version of the German textbook. It provides a general introduction to solid state physics. The book covers a wide range of physical phenomena occurring in solids and discusses basic concepts for describing them. However, solid state physics is such a vast subject that it is not even remotely possible to present it in its entirety in a single textbook. We have selected the material so that all relevant thematic subfields are included and together provide a comprehensive overview. A special feature of this book is that the physics of disordered solids, normally presented only as an appendix, is treated throughout the book and complements the physics of ideal crystals. Although disordered solids are ubiquitous in our daily lives and play an important role in a large number of technological applications, many of their properties are still not fully understood and the concepts describing them are not as advanced as the models based on the symmetry of ideal crystals. Another aspect that is important to us is the use of original data in plots rather than smoothed schematic curves, since we believe that the original measured data show to some extent the experimental difficulties involved in obtaining them. Therefore, we have used original data in as many figures as possible.

The textbook is intended for students who have completed their undergraduate studies and wish to study solid state physics at the graduate level. Basic knowledge of atomic physics and quantum mechanics as well as thermodynamics and statistical physics is assumed. At Heidelberg University, it is used as lecture material for a weekly four-hour course. For further studies, problems and exercises related to the material discussed are given at the end of each chapter. They are designed in such a way that every student with knowledge of the material presented can solve them without difficulty. SI units are used throughout this book, with two exceptions: Ångströms and electron volts. The latter are very common in physics literature and therefore belong to basic physics education.

This book would never have appeared without the help of many colleagues and coworkers. G. Pickett has read major parts of the manuscript and has given invaluable

https://doi.org/10.1515/9783110666502-203

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

advice. We are deeply thankful for the enormous time he has committed to identify errors and shortcomings and to make this book much more readable. For many suggestions and discussions over the years, we would like to express our sincere gratitude to G. Weiss, M. von Schickfus, A. Pucci, R. Klingeler, A. Fleischmann, A. Reiser and A. Halfar. Their contributions have certainly improved the quality of the book and we are most grateful for their help. Special thanks go to R. Weis for his support and help with the editing process.

Heidelberg, March 2022 S. Hunklinger, C. Enss

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

1 Introductory Remarks

Solid state physics deals with the structure and properties of solid materials. In the simplest approximation, solids are a collection of nuclei and electrons that interact with each other via electrostatic forces. In contrast to cosmology, astrophysics or high-energy physics, where the physical laws are not completely known, in solid state physics the relevant fundamental laws are very well known. In principle, almost all the properties of solids can be derived from the Schrödinger equation. Nonetheless, since solid state physics is in no way complete, there will always be unexpected discoveries and surprises in this area of physics due to the complexity arising from the many interacting constituents.

Solid materials, as they surround us in daily life, typically have volumes of a few cubic centimeters and thus consist of 10^{23} or more atoms. This very large number might suggest that a quantitative description of the properties of solids is hardly possible. This is all the more true if one considers the diverse range of solid state phenomena. We thus need to explain that certain materials are conducting, others insulating that there are transparent and opaque, hard and soft, ductile and brittle solids, and that while some solids respond strongly to magnetic fields, others hardly do. However, it turns out that in many cases it is precisely the large number of atoms that enables the development of models for a quantitative description. Of course, not all properties can be treated with a single approach, since the various classes of solids, such as insulators, semiconductors, metals or superconductors, are subject to different macroscopic laws and react differently to external fields.

Once the underlying principles are known, a further, technically-important step follows, the development of new materials or components with properties tailored to specific applications. In many cases, optimized materials and their targeted modification form the basis for new technologies, impressive examples being in information and communication technology, whose development is based on a comprehensive knowledge of solid state physics. One example of the practical application of superconductivity is that of the coils used to generate high magnetic fields for magnetic resonance imaging or in future fusion reactors. Extremely small magnetic fields can be detected with magnetometers that are based on the Josephson effect with a range of uses from medical diagnostics and ground exploration in geology. Further well-known applications are semiconductor lasers, used in every living room, and semiconductor detectors, which are used in the large high-energy physics experiments.

Which fundamental concepts come into play in solid state physics? As in many areas of physics, the Schrödinger equation certainly plays a central role. We will also encounter again and again the Pauli principle which has a significant influence on many properties of solids. Furthermore, we will find that Maxwell's equations and the concepts of thermodynamics and statistical mechanics are very important. As already mentioned, the Coulomb interaction between nuclei and electrons dominates, whereas

https://doi.org/10.1515/9783110666502-001

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the magnetic interaction between the building blocks in non-magnetic materials is practically insignificant. As far as the application of thermodynamics is concerned, it is very important that the solids under consideration are either in equilibrium or near enough to thermodynamic equilibrium.

While in liquids or gases the position of atoms changes over time, their spatial arrangement in solids remains largely the same. In terms of their atomic order, solids can be roughly divided into two groups: A strictly periodic sequence of the atomic building blocks is typical of **ideal crystals**, whereas a completely disordered arrangement of the atoms characterizes **ideal amorphous solids**. The structures of real solids lie somewhere between these two limits, crystals having defects in their structure, and amorphous solids exhibiting a certain amount of local order. Most of the basic treatments of solid state physics assume the periodic structure of crystals and are therefore strictly speaking only applicable to this class of materials. They will be the focus of our considerations below. However, since interest in the properties and peculiarities of complex structures and irregularly-built solids has grown strongly in recent years and theoretical concepts for the description of such systems have increasingly been developed, we also include amorphous solids in our discussion, albeit to a lesser extent.

In the solid state, atoms are to be found in local minima of the potential energy and therefore occupy well-defined positions. Further, when a single atom is deflected, all other neighboring atoms begin to vibrate since they are all coupled to each other in the solid. On the other hand, we can start by considering the collective motion of all the atoms, which can be broken down into harmonic oscillations of the entire solid. An interesting aspect is that we find that the oscillation energies and thus also the amplitudes of these normal vibrations are quantized. In analogy with the energy quantum of electromagnetic radiation being named the photon, the energy quantum of these elastic vibrations is named the phonon. These atomic vibrations have a considerable influence on the elastic, thermal, electrical and optical properties of solids and are therefore an essential component of solid state theory.

Electrons are subject to the Pauli principle which states that two or more identical fermions cannot simultaneously occupy the same quantum state within a quantum system. For multi-electron systems, this means that as the number of electrons increases, they are forced to occupy states of higher and higher energy. In solids we can roughly distinguish between core electrons in states with low energy and valence electrons in states with higher energy. The electrons in the core are relatively firmly bound and are only slightly influenced by neighboring atoms or external fields. Apart from the magnetic behavior, the characteristic properties of solids are primarily determined by the valence electrons. They originate from the s- and p-states of the atoms involved, participate in the interatomic bond and react sensitively to external fields.

Of course, in practice there are no perfect crystals nor perfect amorphous solids. Each crystal has defects that manifest themselves as local deviations from the rest of the structure. Defects have a strong influence on the properties of real solids. They alter

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the mechanical and thermal properties, increase the electrical resistance and impair the optical transparency of the materials. Examples of such defects are missing atoms in crystals or impurity atoms which are incorporated during production, or extended defects which usually also occur during sample production. With amorphous solids, the characterization of defects is much more difficult owing to their irregular structure. Unsaturated chemical bonds provide typical defects in amorphous materials held together by covalent atomic bonds and do not occur in crystals in this form. Another important consideration is that in disordered structures local rearrangements of atoms are possible which, of course, cannot take place in the ordered crystals. Unsaturated bonds and structural rearrangements crucially influence many properties of amorphous solids.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

2 Bonding in Solids

In the following sections we will see that the diversity of solid state properties is based on the interplay of various factors. Clearly, the atomic properties play a crucial role, but the arrangement of the atoms and the bonds between them are at least as important. In order to gain a deeper understanding of the properties of solids, it is essential to study their structures and bonding mechanisms. Therefore, we will first take a closer look at the bonds between the atomic building blocks and then deal with the ensuing structures of solids and the methods for their determination.

2.1 Types of Bonds

In both crystalline and amorphous solids, five basic types of bonds occur differing primarily in the spatial distribution of the electrons involved in the bond. In general, however, these basic bonding types are not found in pure form, but several act in combination. The **hydrogen bond**, which is found in hydrogen-containing substances, is a special case. This type of bond is discussed at the end of this chapter.

An important consideration for the properties of solids is whether the valence electrons of the constituent atoms are in closed shells or not. The closed-shell case is relevant to solids consisting of molecules or noble gas atoms. Here, only the relatively weak **Van der Waals force** provides the binding force between the atoms or molecules. In the case of molecular crystals, we can distinguish between *intra*molecular and *inter*molecular interactions, depending on whether we are interested in the forces within or between the molecules. In ionic crystals, an electron transfer takes place between the binding partners, to create closed shells. In this case oppositely-charged ions are held together by the relatively strong Coulomb forces. Tis type of bond is known as an **ionic bond** .

A **covalent bond** is a type of bond that results from atoms sharing electron pairs in order to complete a closed common outer shell. In such bonds, two valence electrons are shared and located between the atoms involved. In many insulators or semiconductors, the atoms are held together by this mechanism. Such covalent bonds play a central role in molecular chemistry, but the treatment of this aspect would be far beyond the scope of this book.

Completely different is the **metallic bond**, in which the valence electrons are largely evenly distributed in the solid. The strong "smearing" of the electrons not only enables charge transport and thus electrical conductivity, but also provides the binding forces between the atomic cores.

Figure 2.1 shows schematically the electron distribution of these various bonding types. Starting with the Van der Waals bond in solid neon, the series moves to the ionic crystal NaCl, the covalently bonded carbon, ending with metallic sodium with

https://doi.org/10.1515/9783110666502-002

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

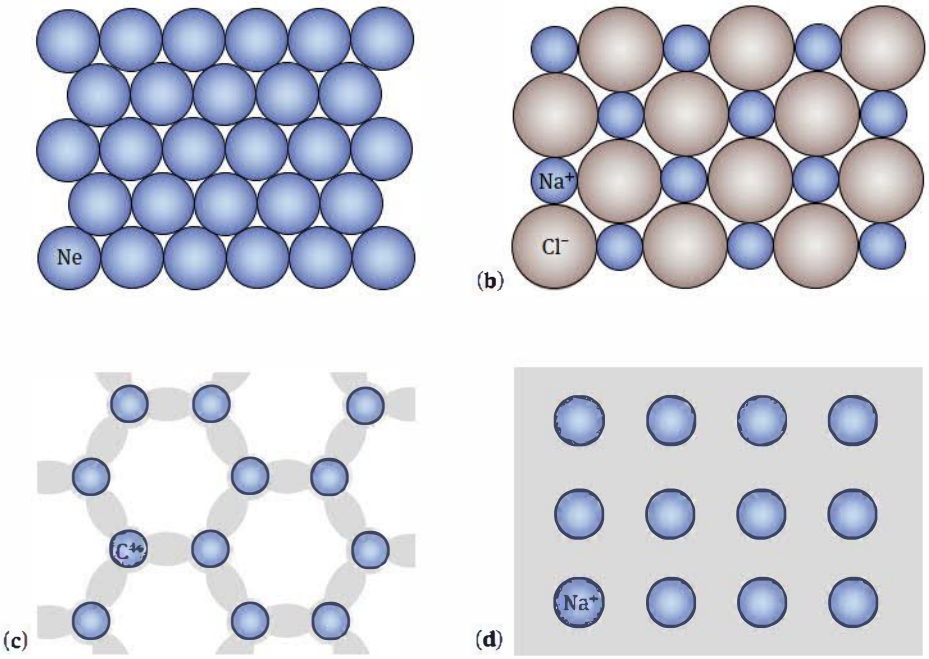


Fig. 2.1: Schematic representation of the distribution of binding electrons in the different types of bonds. The light grey areas indicate the preferred locations of the binding electrons. As typical representatives for the respective binding types **(a)** neon, **(b)** sodium chloride, **(c)** graphite and **(d)** sodium were chosen. The series leads from the Van der Waals and ionic bond via the covalent to the metallic bond.

completely delocalized valence electrons. Since neon only has closed electron shells, no binding electrons are shown in this picture.

In the case of sodium and chlorine, an electron is transferred from the sodium to the chlorine leading to the two ions both having closed shells and the bond is based on the electrostatic interaction between them. Therefore also here no binding electrons are shown. In the case of graphite, which is made up of carbon atoms, the bond between the three in-plane neighbors is formed by a shared pair of electrons whose probability density is particularly high between the atomic cores. As we will see in Section 2.4, in graphite the fourth electron of the L-shell of the carbon atoms does not form a localized bond, its wave function lying perpendicular to the drawing plane of the figure. In sodium, the atomic cores with closed electron shells "are emerged" in a "sea of electrons".

In general, the above binding types do not occur in their pure forms, but rather in combination. For example, considering the elements found in the fourth period of the periodic table, we find a gradual transition as we proceed from KBr, with a purely ionic bond via CaSe and GaAs to tetravalent germanium with a purely covalent bond.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

There is also the possibility that the same atoms in a substance form different bonds. However, when discussing the individual bond types here, we will limit ourselves to the simplest cases, i.e. we will only look at systems in which one bonding mechanism dominates.

2.1.1 Binding Energy

Electrostatic forces act between the atoms and molecules of solids. A measure of the strength of interaction and the number of interacting atoms is the **binding energy** or **lattice energy**. The binding energy is defined as the energy which has to be applied for the full decomposition of the solid into its constituent atoms or ions. It is assumed that both the solid and the resulting atomic building blocks are in their ground state before and after the decomposition. It is important that the zero-point energy is included in the energy balance of the solids as this can noticeably affect the binding energy in certain cases as we will see in Section 2.2.2.

Table 2.1 lists the binding energy per atom for the elements of the second period of the periodic table. As we progress through the period from left to right, the metallic character of the bonds decreases and the covalent character increases. Since the covalent bond is generally stronger than the metallic bond, the binding energy also initially increases. While the bond in lithium is exclusively metallic, in boron the covalent nature already contributes noticeably to the bond energy, and finally the carbon atoms of diamond in the fourth main group of the periodic table are bound exclusively by covalent forces. The remaining elements from nitrogen to fluorine exist as molecular crystals in their solid form and the relatively high binding energies of these materials stems almost entirely from the energy released during the formation of the molecules. The weak Van der Waals interaction, which contributes comparatively little to the binding energy, acts between the molecules. It is therefore not surprising that nitrogen and oxygen both melt at relatively low temperatures despite their relatively high binding energies. Finally, at the end of the group, the atoms of the noble gas neon are bound together by Van der Waals force alone leading to a very low binding energy in this case.

Tab. 2.1: Binding energies and melting temperatures of the elements of the second period of the period table. The data for carbon refer to diamond. (The data were taken from various sources.)

	Li	Be	B	C	N	O	F	Ne
Binding energy (eV/atom)	1.64	3.32	5.81	7.37	4.91	2.60	0.84	0.020
Binding energy (kJ/mol)	158	320	561	711	474	251	81.0	1.92
Melting temperature (K)	453	1560	2348	4765	63.2	54.4	53.5	24.6

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

We should note that crystals and amorphous solids of the same chemical composition have slightly different binding energies. In amorphous materials, the absence of a long-range order causes a reduction in mass density, and since not only the directly adjacent atoms, but also those further away, contribute, this can lead to a reduction in the binding energy, the magnitude of the reduction depending both on the range of the binding forces, and also on the local order.

Repulsive Forces. When two neutral atoms approach each other, they are subject to forces which are attractive at relatively large distances, but are repulsive at short distances. Figure 2.2 shows the typical profile of the interaction potential. The actual profile depends on the type of bond. Here, the **Lennard-Jones potential**[1] φ is sketched. This potential acts between neutral atoms or molecules with closed electron shells. We will discuss this special potential in detail below.

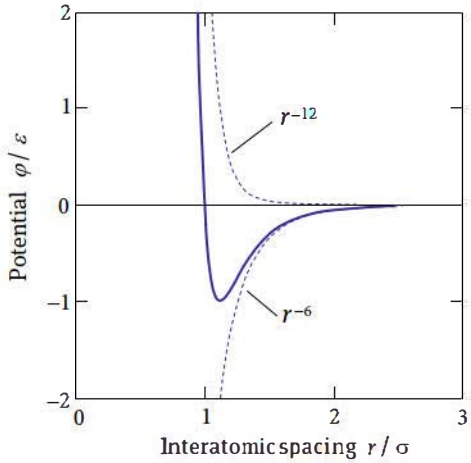


Fig. 2.2: The Lennard-Jones potential. The attracting part of the potential is proportional to r^{-6}, whereas the repulsive part is proportional to r^{-12}. The parameters ε and σ determine the depth at the minimum and the zero crossing of the potential.

For our discussion we need to note that the attraction changes into a repulsion at short distances, preventing the interpenetration of the atoms. Initially, we might well believe that the classical Coulomb repulsion of the electron clouds is responsible for this effect. However, as a simple calculation shows, this Coulomb repulsion leads to a relatively weak effect, depending only slowly on the distance. If it were responsible for the repulsion, atoms would behave like relatively soft spheres. In fact, it is the *Pauli principle*[2] which is crucial here, because at small distances the wave functions of the atoms overlap. In this case we must treat the interacting atoms as a single unit where the electrons take up the available common states. An overlap of the wave functions

1 John Edward Lennard-Jones, *1894 Leigh, †1954 Stokes-on-Trent
2 Wolfgang Pauli, *1900 Vienna, †1958 Zurich, Nobel Prize 1945

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

of electrons in closed shells therefore forces transitions of a fraction of the electrons to unoccupied states of higher energy. This increases the total energy of the system, which results in a repulsive force strongly dependent on distance. The Pauli principle causes the atoms to behave almost like hard spheres.

Various analytical expressions are used to describe the repulsion empirically. Often, as in Figure 2.2, the repulsive potential $\varphi(r)$ is described by

$$\varphi(r) = \frac{\mathcal{A}}{r^{12}}\,, \tag{2.1}$$

where $\mathcal{A}$ is a positive constant and the exponent describes the steepness of the potential drop. Since the exact course of the curve is usually not important, a representation of the potential with a simple, steep enough function is sufficient. There is no deeper reason for the choice of exponent twelve. Often, also an exponential dependence of the form

$$\varphi(r) = \mathcal{A}' \mathrm{e}^{-r/\varrho} \tag{2.2}$$

for the repulsive potential is also used. Here $\mathcal{A}'$ is again a positive constant. The range of the potential is determined by ϱ.

2.2 The Van der Waals Bond

We begin the discussion of the binding forces with the simplest example, namely the noble gas crystals. Noble gas atoms with their closed electron shells normally do not form a chemical bond with their neighbors. They keep their spherical shape also in the solid state and form crystals with high symmetry.[3]

The attractive forces between the atoms are given entirely by the weak Van der Waals forces. The same applies to the intermolecular forces of many molecule crystals. Although the assumption of spherical lattice building blocks in this case is often only an approximation, much of the knowledge gained from noble gas crystals can be transferred directly to the case of molecular crystals.

Van der Waals Forces. Van der Waals[4] forces exist between all atoms. They are based on the electrical *dipole-dipole interaction*. Such interactions between noble gas atoms is initially surprising, since spherically symmetric atoms have no (permanent) dipole moments. However, it should be kept in mind that atoms are not rigid structures and that their charge distribution is subject to fluctuations. A shift of the electron cloud relative to the nucleus causes a dipole moment p_1, which induces a charge displacement

3 In fact, *pure* noble gases in the solid state only occur in crystalline form. However, certain solid noble gas mixtures can also be produced in amorphous form.

4 Johannes Diderik van der Waals, *1837 Leiden, †1923 Amsterdam, Nobel Prize 1910

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

in neighboring atoms. The induced dipole moment is directed in such a way that the force attracts. In qualitative terms, a dipole p_1 causes an electric field $\mathcal{E}_1 \propto p_1/r^3$ at a distance r if we neglect the directional dependence. This field induces a dipole moment $p_2 \propto \mathcal{E}_1 \propto p_1/r^3$ in a neighboring atom. Thus, the interaction potential is given by $\varphi \propto -p_1 p_2/r^3 \propto 1/r^6$, so that the Van der Waals potential can be written as

$$\varphi(r) = -\frac{\mathcal{B}}{r^6} \,. \tag{2.3}$$

Here $\mathcal{B}$ is a positive constant characteristic of the atoms involved. With the help of simple quantum mechanical perturbation calculations, it can be shown that the interaction described always leads to an energy reduction, i.e. to an attracting force. The magnitude of the attracting force is determined by the polarizability of the two atoms involved.

2.2.1 The Lennard-Jones Potential

If one adds to the repulsive potential (2.1) the attractive Van der Waals potential, one obtains the **Lennard-Jones potential**

$$\varphi(r) = \frac{\mathcal{A}}{r^{12}} - \frac{\mathcal{B}}{r^6} \equiv 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] , \tag{2.4}$$

the form of which can be seen in Figure 2.2. The parameter ε indicates the depth of the potential minimum, σ determines the zero crossing of the potential. The potential minimum occurs at the distance $r_0 = 2^{1/6}\sigma = 1.122\,5\,\sigma$.

Table 2.2 lists material parameters of noble gas atoms and noble gas crystals. The first two lines contain the values of the parameters ε and σ. These values can be obtained, for example, from the angular dependence of the scattering intensity of noble

Tab. 2.2: Material parameters of noble gas crystals. The quantities and magnitudes are explained in the text. (The data were taken from various sources.)

	Ne	Ar	Kr	Xe
ε (meV)	3.10	10.4	14.1	20.0
σ (Å)	2.74	3.40	3.65	3.98
R_0 (Å)	3.16	3.76	3.99	4.34
R_0/σ	1.15	1.11	1.09	1.09
U_B/N (meV)	−26	−89	−127	−174
U_0/N (meV)	8	9	7	6
$(U_B + U_0)/N$ (meV)	−18	−80	−120	−168
U_B^{exp}/N (meV)	−20	−81	−116	−166

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

gases in atomic beam experiments. Another possibility is measuring thermodynamic state variables from which the desired parameters can be obtained with the help of a slightly modified Van der Waals equation. The parameters can also be deduced from crystal properties such as the equilibrium distance R_0 of the atoms (which is not identical to the distance r_0 of the potential minimum), the sublimation heat and the compressibility. It is noteworthy, that the parameters obtained by these various methods differ only slightly.

2.2.2 The Binding Energy of Noble Gas Crystals

We now calculate the binding energy of noble gas crystals. For simplicity we only consider the potential energy, although the zero-point energy can cause a noticeable reduction of the binding energy. Since the zero-point energy increases with decreasing mass, it implies considerable corrections for light noble gas crystals, which we estimate at the end of this section. Helium, the lightest noble gas, remains liquid even at absolute zero due to its high zero-point energy and the weak bonding forces between atoms. Solidification of helium can only be achieved when exposed to an external pressure exceeding 2.5 MPa.

In order to calculate the binding energy U_B of N atoms, we first look at the binding energy φ_m of a single atom m, which interacts with all other atoms n of the crystal. We sum these contributions and write $\varphi_m = \sum_{n \neq m} \varphi_{mn}$, where φ_{mn} denotes the interaction energy of the pair of atoms m, n given by the Lennard-Jones potential (2.4). The summation over all atoms m leads to the result

$$U_B = \frac{1}{2}\sum_m \varphi_m = \frac{N}{2}\varphi_m = 2N\varepsilon \sum_{n \neq m} \left[\left(\frac{\sigma}{r_{mn}}\right)^{12} - \left(\frac{\sigma}{r_{mn}}\right)^{6} \right] . \tag{2.5}$$

The factor 1/2 arises since the contribution of each atom pair would otherwise be counted twice. We now express the interatomic distance r_{mn} in units of R, the spatial separation between two directly adjacent atoms, and write $r_{mn} = p_{mn}R$, the value of p_{mn} depending on the structure of the crystal under consideration. For noble gas crystals with face-centered cubic (fcc) structure (cf. Figure 2.1 and Section 3.3) we find $p_{mn} = 1, \sqrt{2}, 2, \ldots$ for the nearest, second-nearest neighbor and so forth. Furthermore, we have to consider the number of atoms at these respective distances. For noble gas crystals with fcc lattice there are 12 nearest neighbors, 6 second-nearest neighbors, and so on. We can now write (2.5) in the following form

$$U_B = 2N\varepsilon \left[\left(\frac{\sigma}{R}\right)^{12} \underbrace{\left(\frac{12}{1^{12}} + \frac{6}{(\sqrt{2})^{12}} + \ldots \right)}_{\approx 12.1319} - \left(\frac{\sigma}{R}\right)^{6} \underbrace{\left(\frac{12}{1^{6}} + \frac{6}{(\sqrt{2})^{6}} + \ldots \right)}_{\approx 14.4539} \right] . \tag{2.6}$$

For these very rapidly converging series the values 12.1319 and 14.4539 are obtained, as indicated.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

With this expression we can determine the ratio between the range of the repulsive potential and the equilibrium distance of the atoms using the fact that the binding energy has a minimum at the equilibrium distance R_0. From the conditions $[dU_B/dR]_{R_0} = 0$ and $[d^2U_B/dR^2]_{R_0} > 0$ it follows that

$$R_0 = 1.090\,2\,\sigma\,. \tag{2.7}$$

Apart from helium (which is a liquid under normal pressure), this relationship should apply to all noble gases, since all noble gas crystals have the same crystal structure at normal pressure.

In crystals, the equilibrium distance of the atoms is slightly smaller than the distance resulting from the minimum of the Lennard-Jones potential (2.4) for the pair interaction. There we had that $r_0 = 1.122\,5\,\sigma$. The experimentally determined values for the crystals are listed in the fourth row of Table 2.2. While small deviations can be found for light noble gases, the agreement for the heavy noble gases is perfect.

If one inserts (2.7) into (2.6), then one finds for the binding energy

$$U_B(R_0) = -8.61\,N\varepsilon\,. \tag{2.8}$$

The values for the binding energy U_B given in Table 2.2 were calculated with the values for ε from the second row. They are higher than the experimental values in the last row. The reason for this discrepancy and for the deviations of R_0/σ from the ideal value lies in the omission of the quantum mechanical zero-point motion of the atoms. This occurs in solids because the atoms reside in a cage confined by their neighbors. Here we will only briefly consider the dependence of the zero-point energy U_0 (their values listed in the sixth row of the table) for the various noble gases, on the potential parameters. The absolute values are not discussed until Chapter 6, since an exact calculation requires knowledge of the atomic vibrational spectra.

We assume the Lennard-Jones potential and approximate the potential minimum by a parabola. The oscillation frequency of an atom of mass M can easily be calculated, from which the zero-point energy follows directly. If we insert the corresponding quantities into the resulting relationship $U_0 \propto (\varepsilon/\sigma^2\,M)^{1/2}$, then we find a good qualitative agreement with the values in the table. Quantitative deviations are not surprising, since we have only considered interacting atomic pairs and ignored the three-dimensionality of the zero-point motion. We need not to refine the estimate at this point because, as already mentioned, a quantitative calculation of the zero-point energy will be made in Section 6.4. A comparison of the theoretical values from the seventh row of the table with the experimental binding energies of the last row shows the good agreement between the theoretical and the experimental results.

Owing to the zero-point motion, the interatomic distance even at absolute zero is not exactly given by the minimum of the potential energy. Since the interaction potential is not exactly parabolic, the equilibrium distance increases with the zero-point energy and thus with decreasing atomic mass. A direct consequence is that the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

nearest neighbor distance will show an isotope effect. Thus we find that $R_0 = 3.156\,8$ Å for ^{20}Ne, while $R_0 = 3.150\,8$ Å for ^{22}Ne.

2.3 The Ionic Bond

Between certain atoms, for example, between potassium and chlorine atoms an electron transfer occurs, leading to the formation of ions and thus giving rise to an **ionic** or **heteropolar bond**. The strong Coulomb interaction between ions is independent of direction and thus, as in case of the Van der Waals bond, allows close packing of ions. We can take sodium chloride as representative of ionic crystals. In this case the ions carry just one (positive or negative) elementary charge. Figure 2.3 illustrates the electron density distribution derived from X-ray diffraction experiments (cf. Section 4.4), showing that the charge distribution of the ions actually has the expected spherical symmetry. Only at the edge of the Cl^- ions at very low electron densities do we observes deviations from the spherical shape. When evaluating the experimental data, it should be noted that the spacing between the contour lines follow a logarithmic scale.

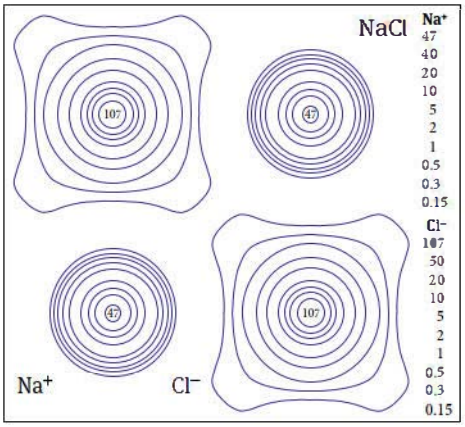


Fig. 2.3: Contour lines of the electron density distribution in NaCl. The numbers on the right side of the figure give the electron density per Å^3 of the individual contour lines. As expected, the charge distribution in the ions is almost spherically symmetrical. The deviation in case of the Cl^- ions only occurs at very low electron densities. (After G. Schoknecht, Z. Naturforschung **12**, 983 (1957).)

2.3.1 The Determination of the Binding Energy

If we take the ions as charged spheres, then we find for the binding energy of an isolated ion pair $\varphi = -e^2/4\pi\varepsilon_0 R_0 = -8.2 \times 10^{-19}$ J $= -5.1$ eV using $R_0 = 2.8$ Å for NaCl crystals. While the ions in the crystal interact with all other neighboring ions, we expect in the solid state a slightly stronger bond, but the order of magnitude should be right. The high value of the expected binding energy makes it easy to understand why

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the weak Van der Waals forces and the zero-point energies are of secondary importance for ionic crystals.

The binding energies of ionic crystals cannot be measured directly, since according to the definition, it is the energy that must be applied to decompose the solid into its individual components. In the case of NaCl, that would mean producing a dilute gas of Na^+ and Cl^- ions, which of course is not possible. In order to still determine the binding energy, one can conceive **cyclic processes** in which as a thought experiment the NaCl crystal is broken down into its components and then reassembled, with all the energies for each step being known, yielding finally the unknown binding energy.

We will briefly describe the necessary steps using the *Born-Habercycle*[5,6]. We start with an NaCl crystal and use the unknown binding energy U_B to produce the required diluted gas from Na^+- and Cl^- - ions. In the next step, the ions are neutralized by the capture or release of a single electron. When an electron is captured by the sodium ion, the ionization energy, known from optical spectroscopy experiments is released. Energy must be applied to neutralize the chlorine ions. This is the electron affinity, which can also be determined spectroscopically. The next step consists of condensing the sodium atoms into solid sodium and forming the Cl_2 molecules. The sublimation energy of sodium can be determined thermochemically. The transition $2Cl \dashrightarrow Cl_2$ releases the dissociation energy of the Cl_2 molecule, which can also be measured thermochemically or spectroscopically. In the final step, the two materials are allowed to react with each other and the heat of formation is registered. A NaCl crystal is thus formed and the thought experiment cycle is closed. Adding the contributions, we find $U_B/N = -8.18\,\text{eV}$ per ion pair.[7] As expected, this value is somewhat higher than our estimate of $-5.1\,\text{eV}$ for the binding energy of a single ion pair.

2.3.2 The Madelung Energy

The binding energy of ionic crystals can be calculated in a similar way to that of noble gas crystals above. For the repulsive potential we use again (2.1), although for ions an exponential potential (2.2) reflects the experimentally observed situation somewhat better. We also have to take into account that the Coulomb interaction has different signs depending on the charges of the interacting ion pairs. For the binding energy per ion pair we obtain:

$$\varphi_m = \sum_{n\neq m}\left[\frac{\mathcal{A}}{r_{mn}^{12}} \pm \frac{e^2}{4\pi\varepsilon_0 r_{mn}}\right] \approx z\frac{\mathcal{A}}{R^{12}} - \sum_{n\neq m}\frac{\pm e^2}{4\pi\varepsilon_0 p_{mn} R} = z\frac{\mathcal{A}}{R^{12}} - \alpha\frac{e^2}{4\pi\varepsilon_0 R}\,. \tag{2.9}$$

5 Max Born, *1882 Breslau (now Wroclaw), †1970 Göttingen, Nobel Prize 1954

6 Fritz Haber, *1886 Breslau (now Wrocław), †1934 Basel, Nobel Prize 1918

7 The corresponding numerical values are: Ionization energy 5.14 eV, electron affinity 3.61 eV, sublimation energy 1.13 eV, dissociation energy 1.23 eV and heat of formation 2.26 eV.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The second expression follows under the assumption that due to the difference in the ranges of the repulsive force (short) and the Coulomb force (long), only the z nearest ions at distance R need to be considered for the repulsive force. Furthermore, we express the distance between the ions in units of the distance to the directly adjacent ions and therefore write $r_{mn} = p_{mn}R$, as already done for the noble gas crystals. In the final expression of (2.9) we have introduced the dimensionless **Madelung constant**[8] α. It is defined by the relation

$$\alpha \equiv \sum_{n \neq m} \frac{\pm 1}{p_{mn}} \tag{2.10}$$

standing for the summation of all the Coulomb terms.

In the simple case of a linear chain consisting of alternating positive and negative ions, the Madelung constant can easily be determined as:

$$\alpha = 2\left(1 - \frac{1}{2} + \frac{1}{3} - + \ldots\right) = 2 \ln 2 \approx 1.386 , \tag{2.11}$$

where the factor 2 takes into account the symmetry of the problem, namely that the interaction with neighbors on both sides of a given starting point contribute equally to the binding energy.

A serious problem occurs for three-dimensional lattices: Although the Coulomb potential decreases with r^{-1}, the number of ions to be taken into account increases quadratically with distance and the values obtained for α are found to be quite different depending on where the summation is truncated. The physical reason for this is that charges occur on the surface of ionic crystals, the distribution of which depends on the crystal shape. An electrical polarization is associated with this surface charge, giving rise to a field energy comparable to the binding energy. This aspect will be discussed in detail in Chapter 13.

The problem for the calculation of the binding energy can be solved in the following way: The crystal is divided into hypothetical so-called *Evjen cells*, which are chosen so as to carry no resultant charge while retaining the symmetry of the crystal. This is only possible if ions are "sliced", resulting in fractions of the electron charge located on the faces, edges and corners of the cell, as shown in Figure 2.4 for the cubic cell of an NaCl crystal. The starting crystal concerned is then reassembled by adding together multiple Evjen cells. The electrostatic interaction between these neutral cells is a rapidly decreasing multipole interaction, so that the sum converges relatively quickly. The same effect can be achieved by systematically extending the original Evjen cell. If the cell shown in Figure 2.4 is selected for NaCl, the electrostatic interaction energy already decreases with the inverse fifth power of the distance. If we make the calculation, we obtain numerical values differing only little from one crystal structure to another. Thus for the structures described in Chapter 3 we find the values

$$\alpha_{NaCl} = 1.747\,56, \qquad \alpha_{CsCl} = 1.76267 \qquad \text{and} \qquad \alpha_{ZnS} = 1.63806 . \tag{2.12}$$

8 Erwin Madelung, *1881 Bonn, †1972 Frankfurt am Main

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

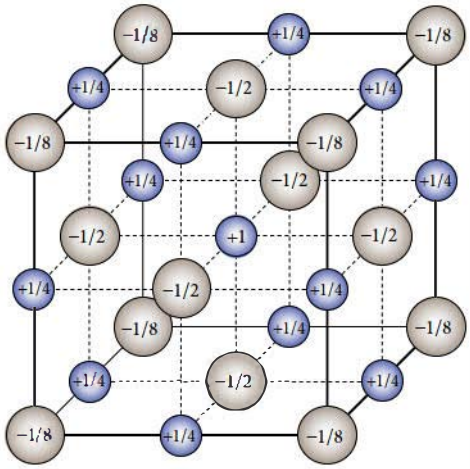


Fig. 2.4: Evjen cell of NaCl. The Na^+-ion in the center is surrounded by six Cl^--ions. As they are located on the cube sides, only half of their charge is taken into account. Only a fourth of the charge of the twelve positive ions on the cube edges is counted, and only an eighth of that of the eight negative ions at the cube corners. (After H. M. Evjen, Phys. Rev. **39**, 675 (1932).)

To calculate the binding energy of the entire solid, we multiply (2.9) by the number of molecules N, in this case the number of ion pairs, and obtain the result

$$U_B = N\varphi_m = N\left(\frac{z\mathcal{A}}{R^{12}} - \frac{\alpha e^2}{4\pi\varepsilon_0 R}\right) . \tag{2.13}$$

Noting, as for noble gas crystals, that in equilibrium the first derivative of the potential energy with respect to distance goes to zero, we can eliminate the parameter $z\mathcal{A}$ to find:

$$U_B = -\frac{N\alpha e^2}{4\pi\varepsilon_0 R_0}\left(1 - \frac{1}{12}\right) . \tag{2.14}$$

The first term represents the *Madelung energy* caused by the electrostatic interaction, the second the repulsion by the nearest neighbors. This result shows that the repulsion offsets about 10% of the Madelung energy. We thus find the value $U_B/N = -8.25$ eV per ion pair for the binding energy of sodium chloride. The agreement with the experimental value $U_B^{exp}/N = -8.15$ eV is satisfactory.

Obviously, the repulsive part of the binding energy is determined by the range of the repulsive potential. One might therefore think that it would be more reasonable to first leave the exponent in (2.1), referred to here as n, open and then fix it by adjusting (2.14). However, it turns out that this way of determining the exponent results in rather large errors.

Better results can be obtained by determining the exponent with the help of another measurable parameter. For this one can use, for example, the compressibility κ or the compression modulus $B = 1/\kappa$, which are connected to the binding energy via the second derivative with respect to volume. With $B = V(\partial^2 U/\partial V^2)$ we find:

$$B = \frac{\alpha e^2(n-1)}{72\pi\varepsilon_0 R_0^4} \tag{2.15}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

and thus

$$n = 1 + \frac{72\pi\varepsilon_0 R_0^4 B}{\alpha e^2} \,. \tag{2.16}$$

Inserting the data for the ionic crystals into this equation, we then find values of n, which are a little smaller than 10 and thus point to a somewhat lesser steepness of the repulsive potential. As already mentioned, the agreement improves if we assume an exponential dependence of the potential. In Table 2.3, the compression modulus B and the distance R_0 of the nearest neighbor ions are given for a number of alkali halide crystals. As can easily be verified, the exponents of the repulsive potential, obtained in this way have values well below 12.

Tab. 2.3: The compression modulus B and ion spacing R_0 of some alkali halide crystals. (The data are taken from various sources.)

	LiF	LiCl	NaCl	NaBr	KCl	KI	CsCl	CsBr	CsJ
B (GPa)	62.0	29.8	24.4	19.9	17.4	11.9	22.3	16.7	12.7
R_0 (Å)	2.01	2.56	2.80	2.99	3.15	3.53	3.57	3.72	3.96

Comparing the strength of the repulsive part of the interaction potential for NaCl and the noble gas crystals, we would initially expect the reduction in the binding energy caused by repulsion to be comparable. However, if we insert the corresponding parameters, it is surprising that the repulsion in case of NaCl has a much stronger effect than for solid krypton, for example. The explanation is quite simple: the interatomic distance is much greater for noble gas crystals than for NaCl, so that the rapidly decreasing repulsion acts less strongly there.

2.4 The Covalent Bond

In many solids, **covalent bonding** is important. Here the electrons are not distributed spherically around the nucleus, but are predominantly distributed between the atoms. This type of bond occurs not only in solids, but also in molecules. We will discuss the peculiarities of the covalent bond only briefly here and refer to the relevant literature of atomic and molecular physics for further study.

2.4.1 H_2^+ Molecular Ion

The physics of this bond is particularly simple in case of the H_2^+ *molecular ion*, which consists of two protons and *one* binding electron. We first look at this relatively simple

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

system and then briefly discuss the bond in the neutral hydrogen molecule and covalent bonds in various solids.

Starting point of the quantum mechanical treatment is the Hamiltonian[9]

$$H = -\frac{\hbar^2}{2m}\Delta - \frac{e^2}{4\pi\varepsilon_0 r_a} - \frac{e^2}{4\pi\varepsilon_0 r_b} + \frac{e^2}{4\pi\varepsilon_0 R_{AB}} \,. \tag{2.17}$$

The definitions of the geometric quantities can be found in Figure 2.5. The first term represents the kinetic energy of the electron, the two following terms the attraction between the electron and the protons and the last term the repulsion between the protons A and B. This latter term depends only on the distance R_{AB} and not on the wave function of the electron, and thus plays no role in the discussion of the attractive interaction. We therefore leave it out for the time being and add it again at the end of our calculation.

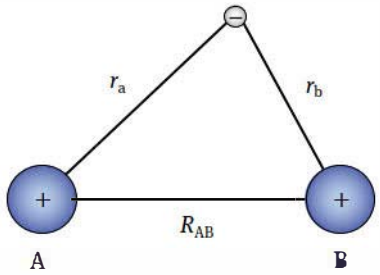


Fig. 2.5: Scheme of a H_2^+ molecular ion with the denotation used in (2.17).

Despite the simplicity of the Hamiltonian, the Schrödinger[10] equation can only be solved approximately. If the two protons are far away from each other, the electron is located either at proton A or at proton B and the associated wave functions φ_a and φ_b are identical to those of a hydrogen atom in the ground state. At relatively small distances we can represent the wave function ψ of the molecular ion in first approximation as a superposition of these two wave functions. This procedure is called the *LCAO method* (**L**inear **C**ombination of **A**tomic **O**rbitals). Here we use the *linear combination*

$$\psi = c_1\,\varphi_a + c_2\,\varphi_b \,. \tag{2.18}$$

We still have to determine the real-valued constants, c_1 and c_2. With a known wave function, the expectation value E of the energy can be calculated using the stationary Schrödinger equation $H\psi = E\psi$, since the following applies

$$E = \frac{\int \psi^* H\,\psi \mathrm{d}V}{\int \psi^*\psi\,\mathrm{d}V} = \frac{c_1^2 H_{aa} + c_2^2 H_{bb} + 2c_1c_2H_{ab}}{c_1^2 + c_2^2 + 2c_1c_2S} \,, \tag{2.19}$$

9 William Rowan Hamilton, *1805 Dublin, †1865 Dunsink

10 Erwin Rudolf Josef Alexander Schrödinger, *1887 Vienna, †1961 Vienna, Nobel Prize 1933

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

where the abbreviations $H_{aa} = \int \varphi_a^* H \varphi_a \, dV$, $H_{ab} = \int \varphi_a^* H \varphi_b \, dV = \int \varphi_b^* H \varphi_a \, dV$, $H_{bb} = \int \varphi_b^* H \varphi_b \, dV$, and $S = \int \varphi_b^* \varphi_a \, dV$ were used. Further, we have set $H_{ba} = H_{ab}$, because the problem is symmetrical regarding the indices a and b. The quantity S is called the *overlap integral*.

At large distances R_{AB}, H_{ab} and S vanish since there is no overlap between φ_a and φ_b and the problem is reduced to that of two isolated hydrogen atoms. We can then write $H_{aa} = H_{bb} = E_0 = -13.60\,\text{eV}$ which is just given by the ground state energy E_0 of the hydrogen atom. At shorter distances R , the electric field at site of the proton A is modified by the presence of the proton B. This leads to a slight decrease of H_{aa} and H_{bb}.

The crucial factor for the bond is the energy H_{ab}, which is exclusively of quantum mechanical origin. A short calculation leads to the result:

$$H_{ab} = \int \varphi_a^* H \varphi_b \, dV = E_0 S + \int \varphi_a^*(r_a) \left(-\frac{e^2}{4\pi\varepsilon_0 r_a} \right) \varphi_b(r_b) \, dV \,. \tag{2.20}$$

In this expression, instead of the usual electron charge density $-e|\varphi_a|^2$ the so-called *exchange density* $-e\varphi_a^* \varphi_b$ appears. The corresponding integral is known as *exchange integral* and the associated energy is called the **exchange energy**. It is negative, i.e. attractive, and is based on the fact that the electron is partly in the state φ_a and partly in the state φ_b.

Now the values of the constants c_1 and c_2 have to be determined. Bearing in mind that the exact wave function always leads to smaller energy eigenvalues than any approximate solution. We come particularly close to the actual solution by looking for the minimum value of E with respect to the constants c_1 and c_2. We use $\partial E/\partial c_1 = 0$ and $\partial E/\partial c_2 = 0$ as equations for determining c_1 an c_2. By calculating the derivatives using (2.19) we find

$$\begin{aligned} c_1 (H_{aa} - E) + c_2 (H_{ab} - ES) &= 0 \,, \\ c_1 (H_{ab} - ES) + c_2 (H_{bb} - E) &= 0 \,. \end{aligned} \tag{2.21}$$

These equations have non-trivial solutions when the determinant of coefficients is zero, i.e. for

$$(H_{aa} - E)(H_{bb} - E) - (H_{ab} - ES)^2 = 0 \,. \tag{2.22}$$

Since the two protons are equivalent, we set $H_{aa} = H_{bb}$ and solve this equation for E. At this point we include the repulsive interaction of the two protons that we neglected at the beginning of our calculation and obtain the following result:

$$E_\pm = \frac{H_{aa} \pm H_{ab}}{1 \pm S} + \frac{e^2}{4\pi\varepsilon_0 R_{AB}} \,. \tag{2.23}$$

The spatial overlap of the wave functions φ_a and φ_b and the resulting negative exchange energy H_{ab} lifts the degeneracy of the energy levels, as indicated in Figure 2.6. The two new states with energies E_+ and E_- are called *bonding* or *antibonding* states. At

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

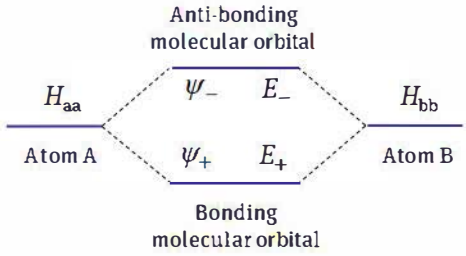


Fig. 2.6: Splitting of the degenerate energy levels of the isolated atoms into a bonding and an anti-bonding state.

short distances R_{AB} the Coulomb energy $e^2/4\pi\varepsilon_0 R_{AB}$ rises rapidly and thus prevents the protons from coming closer.

Using equations (2.21), the constants c_1 and c_2 can be determined for the two different eigenvalues. For the bonding state with energy E_+ one finds $c_1 = c_2$, and for the antibonding state with energy E_- the relation $c_1 = -c_2$ holds. Therefore, for the symmetrical wave function of the bonding state and for the anti-symmetrical wave function of the antibonding state we can write:

$$\psi_+ = c\,(\varphi_a + \varphi_b) \qquad \text{and} \qquad \psi_- = c\,(\varphi_a - \varphi_b)\,. \tag{2.24}$$

The constant c is used to normalize the wave functions ψ_+ and ψ_-. These functions are schematically shown in Figs. 2.7 and 2.8. For the bonding it is crucial that the

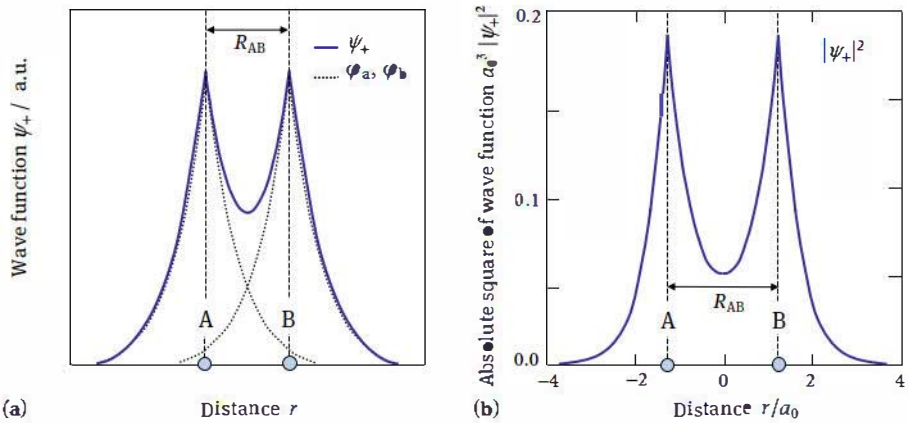


Fig. 2.7: Sectional view of the wave function and its absolute square value for the H_2^+ molecular ion in the *symmetrical* state along the axis joining the nuclei. **a)** Schematic representation of the wave function ψ_+. The thin dotted lines show for comparison the ground state wave functions φ_a and φ_b of isolated hydrogen atoms scaled to match ψ_+ at the maximum. **b)** The absolute square $|\psi_+|^2$ of the symmetrical wave function of the H_2^+ molecular ion. Here a_0 denotes the Bohr[11] radius. The separation of the nuclei corresponds to the real situation $R_{AB} \simeq 2.45\,a_0$.

11 Niels Henrik David Bohr, *1885 Copenhagen, †1981 Copenhagen, Nobel Prize 1922

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

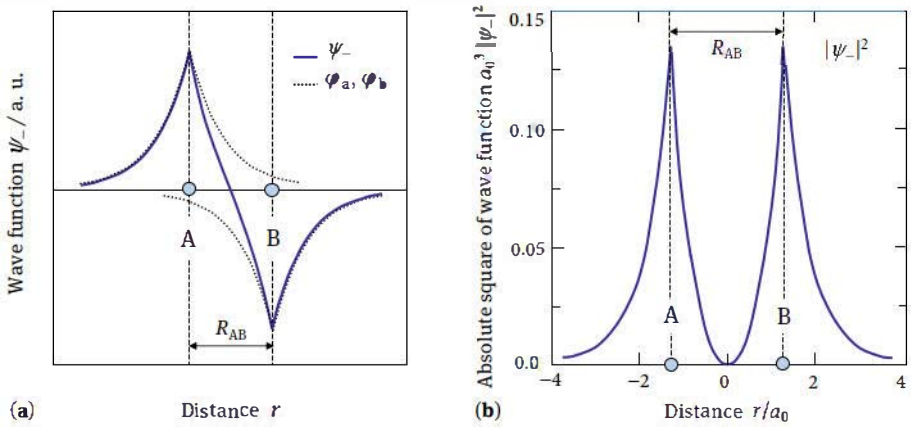


Fig. 2.8: Sectional view of the wave function and its absolute square value for the H_2^+ molecular ion in the *anti-symmetric* state along the axis joining the nuclei. **a)** Schematic representation of the wave function ψ_-. The thin dotted lines show for comparison the ground state wave functions φ_a and $-\varphi_b$ for isolated hydrogen atoms scaled to match ψ_+ at the maximum and the minimum respectively. **b)** The absolute square $|\psi_-|^2$ of the antisymmetric wave function of the H_2^+ molecule. Here a_0 denotes the Bohr radius. The separation of the nuclei corresponds to the real situation $R_{AB} \simeq 2.45\,a_0$.

amplitudes φ_a and φ_b of the wave functions and not the square of their absolute values are added together. Thus, the probability of finding the electron between the two protons in the binding state is substantially larger. Energetically the electron benefits from the Coulomb attraction of both protons. In the anti-bonding state, the electron density between the protons is greatly reduced, falling to zero at the center point between the two protons. This can be seen clearly on the right side of the two figures in which the absolute squares $|\psi_+|^2$ and $|\psi_-|^2$ for the corresponding wave functions are shown for the H_2^+ molecular ion.

If one inserts the numerical values into the solution of the LCAO approximation, one finds 1.77 eV for the binding energy at a proton separation of 0.13 nm. The experimental value of the binding energy is 2.79 eV. In view of the vast simplifications we have made in the calculation of the binding energy, the agreement is quite satisfactory. Significant improvements of the quantitative description can be obtained by allowing the wave function φ_a and φ_b to take a non-spherical shape in the molecular ion.

2.4.2 The Hydrogen Molecule

Let us briefly discuss the question which changes the presence of the second electron causes in the *hydrogen molecule*. We first look at the corresponding Hamiltonian. Besides the contributions we already know from equation (2.17) for the H_2^+ molecule, we have to consider further terms, namely the kinetic energy of the sec-

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

ond electron, its interaction with the two protons and the repulsion between the two electrons. First of all it should be noted that the two electrons are described by a *two-particle wave function* $\Psi(\mathbf{r}_1, \mathbf{r}_2)$. The symmetrical wave function, for example, has the form $\Psi_+(\mathbf{r}_1, \mathbf{r}_2) \propto [\varphi_a(\mathbf{r}_1) + \varphi_b(\mathbf{r}_1)][\varphi_a(\mathbf{r}_2) + \varphi_b(\mathbf{r}_2)]$. This product also contains terms describing the state at which both electrons are in the vicinity of one of the two nuclei. The high Coulomb repulsion between the two electrons makes this state less likely. The mathematical treatment of the problem is greatly simplified if these terms are not taken into account when calculating the ground state energy. In the approach proposed by *W. Heitler*[12] and *F. London*[13] the terms that are less important for the ground state are neglected leading to

$$\Psi_\pm(\mathbf{r}_1, \mathbf{r}_2) \propto [\varphi_a(\mathbf{r}_1)\varphi_b(\mathbf{r}_2) \pm \varphi_b(\mathbf{r}_1)\varphi_a(\mathbf{r}_2)] \,. \tag{2.25}$$

The wave functions Ψ_+ with the positive sign and Ψ_- with the negative sign have even or odd parities with respect to the exchange of the local coordinates of the electrons.

According to the Pauli principle, the total wave function of a system of several fermions must be anti-symmetrical. If, for example, the two electrons involved in the bond are in the anti-symmetric orbital state $\Psi_-(\mathbf{r}_1, \mathbf{r}_2)$, the spin function must be symmetric. In this case, the total spin must have the value one. Since here there are three possible orientations of the spins with respect to an external field, this state is described as a *triplet state*. On the other hand, if the orbital wave function is the symmetric form $\Psi_+(\mathbf{r}_1, \mathbf{r}_2)$, then the spin function must be anti-symmetric. In this case the two spins are aligned anti-parallel, thus with total spin zero, and we describe this as the *singlet state*. From (2.25) and the complete Hamiltonian the expectation value of the energy can be calculated. The result does not differ much qualitatively from the case of the hydrogen molecular ion.

The very simple Heitler-London approximation described here results in a binding energy of 3.2 eV and a proton-proton separation of 0.8 Å. With the complete Hamiltonian and improved calculation methods, good agreement with the experimental values of 4.74 eV for the binding energy and 0.74 Å for the separation of the nuclei is obtained.

Besides the strength of the covalent bond, its pronounced directional dependence is a characteristic feature. Since the shapes of the wave functions are largely given by the atomic orbitals, their maximum overlap, leading to the largest possible binding energy, is only possible in certain directions (except, of course for the spherical *s*-orbitals). As we shall see in Section 4.4, the charge density distribution can be determined very precisely by X-ray diffraction, by measuring the "atomic structure factors" of the atoms involved in the bonding. Some elements form covalently bonded crystals. As an example, Figure 2.9 shows the valence electron density in a germanium crystal. The germanium atoms are clearly not arranged linearly, but in a zigzag reflecting the

12 Walter Heinrich Heitler, *1904 Karlsruhe, †1981 Zurich
13 Fritz Wolfgang London, *1900 Breslau (now Wroclaw), †1959 Durham

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

directionality of the covalent bond. The maxima of the charge density are clearly visible and located *between* the atoms which is typical for covalent bonding.

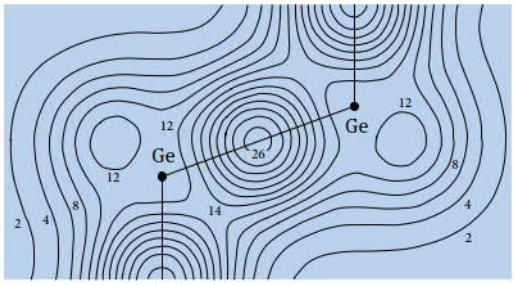


Fig. 2.9: Density distribution of valence electrons in a germanium crystal. The numbers labelling the lines of constant density indicate the electron density in elementary charges per primitive unit cell (see Section 3.3). The position of the nuclei is indicated by points. (After M.L. Cohen, Science 179, 1189 (1973).)

2.4.3 Types of Covalent Bonds

The covalent **tetrahedral bond** is a type of bond which is particularly important for solid state physics. It occurs, for instance, in solids of elements of the fourth main group, such as carbon, silicon or germanium. We discuss this special bond taking the example of diamond, but the arguments can also be applied to heavier elements in this group, since only the valence electrons in the outer shells contribute to the bonding and thus the configurations of the inner shells do not matter.

To create a covalent bond, we start with the free carbon atom which has the electron configuration $1s^2\,2s^2\,2p^2$. The formation of so-called sp^3-hybrid orbitals can be divided conceptually into two stages: First, there is a transition from the configuration $2s^2\,2p^2$ to the configuration $2s2p_x\,2p_y\,2p_z$, which is, however, energetically less favorable, since the p-states have a higher energy than the s-states. The second stage is the formation of suitable linear combinations to provide the optimal overlap of the wave functions with neighboring atoms. The initial energy loss is more than compensated for by the higher binding energy. Diamond, in which the carbon atoms are bound in this way, has a high binding energy of 7.36 eV/atom despite the small number of four binding partners. Silicon has 4.64eV/atom and germanium 3.87 eV/atom.

By linear superposition of the four different orbitals, four equivalent wave functions can be constructed which point from the center of a tetrahedron, centered on the atom, towards the four corners. These four new wave functions Ψ_i have the form

$$\Psi_i = \frac{1}{2}\left(\psi_s \pm \psi_{p_x} \pm \psi_{p_y} \pm \psi_{p_z}\right). \tag{2.26}$$

Here the sign combinations (+ + +), (+ − −), (− + −) and (− − +) are possible. If only positive signs are selected, the "club-shaped" p-wave functions become an orbital that

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

is oriented in the direction of the space diagonal of a cube.[14] As shown in Figures 2.10a and 2.10b, the additional superposition with the s-wave function inflates the club in one direction and causes it to shrink on the other end, since the p-wave function has the opposite sign there. The same applies to the three other hybrid orbitals. The tetrahedral form of the sp^3-hybrid orbitals shown in Figure 2.10c is then obtained.

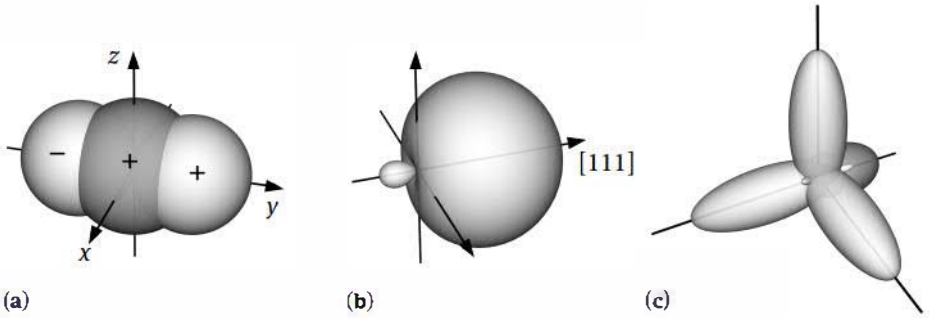


Fig. 2.10: Illustration of the formation of the sp^3-hybrid orbitals. **a)** The s-orbital is in the middle, the two "clubs" of the p_y-orbital can be seen left and right of it. **b)** Hybrid orbital in [111] direction. **c)** Orbitals of the tetrahedron-shaped sp^3-hybrid. In this case, for clarity, the "clubs" are shown somewhat slimmer than they actually are.

It is worth taking a closer look at the element carbon, since it exhibits a pronounced *allotropy*.[15] This is the term used to describe the ability of elements in the same state of aggregation to occur in different structural forms. Besides of the tetrahedral sp^3-orbitals of diamond, carbon can also form planar sp^2-hybrid orbitals based on the superposition of one $2s$- and two $2p$-wave functions. This results in three club-shaped orbitals creating a three-armed 120° stars. The fourth arm, the $2p_z$-orbital, sits perpendicular to the star plane. An interesting example is *graphene*, which has attracted a lot of attention in recent years owing to its unusual properties. It consists of only single layer of carbon atoms connected by sp^2-hybrid orbitals. Figure 2.11 shows the honeycomb structure of graphene. It has many properties that differ from those of "classical" solids.

So-called *carbon nanotubes* are closely related to graphene. They consist of one or more graphene layers rolled into tubes. These extremely thin-walled structures can have a lengths of up to a few centimeters. The atomic structure and the unusual properties of graphene and carbon nanotubes are discussed further in Chapter 2 und 7.

Graphite is yet another carbon modification which is also closely related to graphene. In graphite, graphene layers are stacked on top of each other as shown in

14 The two wave function "clubs" point in directions $\mathbf{r} = \hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}}$ or in the direction of $-\mathbf{r}$, where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ stand for the unit vectors of the Cartesian coordinate system.

15 Related to "allotropy" is the term "polymorphy", which refers to compounds.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

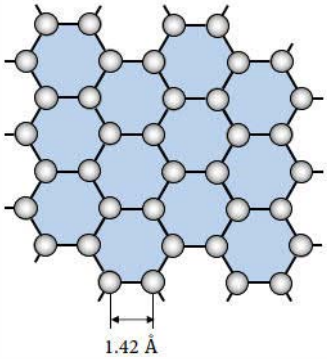


Fig. 2.11: Schematic view of a graphene layer. It consists of one layer of carbon atoms connected by sp^2-hybrid orbitals. The distance between adjacent atoms is 1.42 Å.

Figure 2.12, slightly staggered. The individual planes, termed basal planes, are primarily held together by Van der Waals forces. Additional bonding forces are provided by the π-electrons, which are delocalized and thus give rise to a high electrical conductivity. We can also see from the picture that there are alternating atomic layers which are laterally shifted against each other. While one half of the atoms has a neighboring atom in the underlying basal plane, the other half lacks such a neighbor.

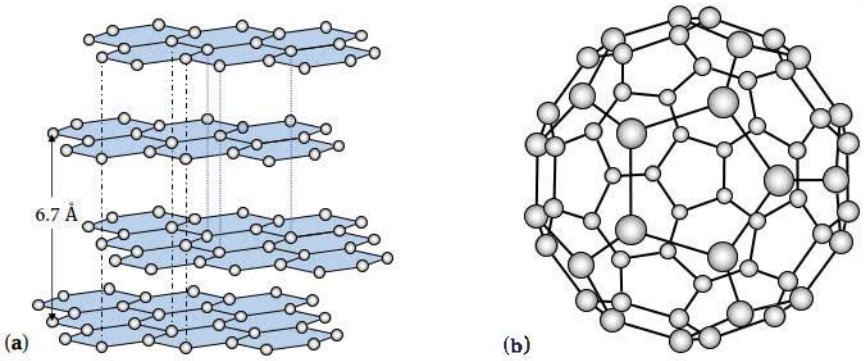


Fig. 2.12: **a)** Structure of graphite. The neighboring atomic layers are shifted laterally against each other. The atoms at positions indicated by dash-dotted lines have a neighbor in the neighboring layers. However, if one moves along the dotted lines, one only find atoms in every second layer. **b)** Structure of the fullerene molecule C_{60}.

The different interatomic distances reflect the different bond strengths: Within a single layer the atomic distance is 1.42 Å, but the distance between the planes is 3.35 Å. While the binding energy within the planes is very high at 4.3 eV, it is relatively low between the planes at 0.07 eV. The different bond strength cause an extremely strong anisotropy of the material properties. A direct consequence is the easy cleavage of graphite along the basal planes. It is also easy to understand that the electrical and thermal conductivity along the basal planes is very high, but relatively low perpendicular to the planes.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The elastic and magnetic properties as well as the hardness of graphite also depend strongly on the direction.

Another nice example of the consequence of sp^2-hybrid bonding is the molecule C_{60}, which was only discovered in 1985 and is known as *Buckminsterfullerene* or *Buckyball*. As Figure 2.12b shows, this molecule consists of 12 pentagons and 20 hexagons, contains 32 rings and has the shape of a football with a diameter of 7.1 Å. Similar to graphite, the π-electrons are delocalized and protrude from the surface of the sphere. C_{60}-molecules can form their own bonds and form three-dimensional crystals. At room temperature, C_{60}-crystals have a cubic structure with a distance of 10 Å from football to football. Shortly after the discovery of C_{60}, other molecules with similar structures were found. Besides C_{60}, there are C_{70}, C_{76}, ... C_{90} which are the best known representatives of this group.

Covalent bonds are also found in the elements of the fifth and sixth group of the periodic table, but in these cases, layer or chain structures are formed. Of course, there are also such bonds between different elements. Well-known examples are the cubic crystals boron nitride and gallium arsenide, which are constructed like diamond, each atom being surrounded by four atoms of the other kind. Each boron or gallium atom contributes three, each nitrogen or arsenic atom five electrons to the bond. Because of the different electronegativities of the elements involved, the bond between the atoms is not purely covalent, but also has an ionic contribution.

2.5 The Metallic Bond

Metals are characterized by the fact that their valence electrons are largely delocalized, i.e. almost uniformly smeared over the crystal. Since these electrons are able to move almost freely between the atomic cores, we speak of an **electron gas**. As a result the bonding between the atoms is non-directional and allows an optimal use of space. A quantitative treatment of the binding energy of metals is much more complex than in case of the other binding types. The main difficulty arises from the fact that the binding energy typically involves several contributions of comparable magnitude but different signs.

The situation is most simple in the case of the alkali metals, which come very close to the ideal of localized atomic cores and a homogeneous outer electron distribution. The smearing of the outer charges leads to a strong reduction in the kinetic energy of valence electrons and thus to a reduction in the total energy.

We will briefly discuss some of the challenges that occur when calculating the binding energy of metals. Even in the simplest case, namely alkali metals, a quantitative treatment of the binding is very complex and goes beyond the scope of this book. We first consider a very simple model in which the atomic cores are regarded as points whose charge is compensated by a spherical "electron cloud". The radius of this sphere, which is usually referred to as *Wigner-Seitz radius* r_s, is defined in such a way that

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

it contains just one valance electron in this case.[16,17] If N is the number of electrons and V the sample volume, then r_s is given by $V/N = 4\pi r_s^3/3$ and the charge density by $\varrho_\ell = -3e/4\pi r_s^3$. With this charge distribution we find for the electrostatic energy per alkali atom E_{Coul}/N the relation:

$$\frac{E_{Coul}}{N} = -\frac{e^2}{4\pi\varepsilon_0}\frac{9}{10\,r_s}\,. \tag{2.27}$$

Next we need to take into account the kinetic energy of the electrons, which, as mentioned above, is smaller than for isolated atoms, since in metals a much larger volume is available to each binding electron. However, the sign of this contribution is positive and thus causes repulsion. Here in anticipation, we use the simple expression for the free electron gas which we will derive in Section 8.2 and write:

$$\frac{E_{kin}}{N} = \frac{3}{5}\frac{\hbar^2}{2m_e}\left(\frac{9\pi}{4}\right)^{2/3}\frac{1}{r_s^2}\,, \tag{2.28}$$

where m_e is the electron mass. Beside these obvious contributions further terms appear as a consequence of the electron-electron interaction. The most important contribution is based on the exchange interaction, which we know in another form from the covalent bond and is based on the overlap of the wave functions of the delocalized electrons. Here we give the corresponding expression without going into the derivation:

$$\frac{E_{ex}}{N} = -\frac{3e^2}{16\pi^2\varepsilon_0}\left(\frac{9\pi}{4}\right)^{1/3}\frac{1}{r_s}\,. \tag{2.29}$$

The spin orientation also contributes to the binding energy because it favors a correlated electron motion. Since the corresponding term for alkali metals is relatively small, we neglect it and do not discuss it further here.

Thus, expressing r_s in units of the Bohr's radius $a_0 = 0.529$ Å and using the numerical values of the constants appearing, we then find for the total energy:

$$\frac{E_B}{N} = \left[-\frac{24.35}{(r_s/a_0)} + \frac{30.1}{(r_s/a_0)^2} - \frac{12.5}{(r_s/a_0)}\right]\ \text{eV/atom}. \tag{2.30}$$

Obviously, in this rough approximation, the binding energy depends only on r_s. Again using the fact that in equilibrium the first derivative of the binding energy in respect to the interatomic distance must disappear, we find the following relation for monovalent metals:

$$r_s \approx 1.6\,a_0\,. \tag{2.31}$$

Unfortunately, this result contradicts the actual situation: The alkali metals have rather different Wigner-Seitz radii, with values of r_s/a_0 falling between three and six.

16 Eugene Paul Wigner, *1902 Budapest, †1995 Princeton, Nobel Prize 1963
17 Frederick Seitz, *1911 San Francisco, †2008 New York City

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Evidently, we have oversimplified the problem. An important aspect is that because of the Pauli principle, the conduction electrons can hardly overlap with the relatively extended atomic cores, which we have assumed to be point-like in this simple model. To take this into account, we can use a relatively simple method here: we describe the effect of atomic cores on the valence electrons by using of a pseudopotential. A very simple potential of this kind, with which the principle idea can be illustrated, is shown in Figure 2.13. Up to a critical radius R_c, roughly corresponding to the ionic radius, the potential is not attractive. At R_c, it drops abruptly and then rises for $r > R_c$ as $\phi(r) = -e/4\pi\varepsilon_0 r$. A straightforward integration shows that the contribution of the Coulomb attraction (2.27) is thereby reduced by $E_{pseu} = 3Ne^2R_c^2/4\pi\varepsilon_0 r_s^3$. From this follows:

$$\frac{E_B}{N} = \left[-\frac{24.35}{(r_s/a_0)} + \frac{30.1}{(r_s/a_0)^2} - \frac{12.5}{(r_s/a_0)} + 41\frac{(R_c/a_0)^2}{(r_s/a_0)^3} \right] \text{ eV/atom} . \tag{2.32}$$

If we again use the fact that in equilibrium the first derivative of the binding energy with respect to the interatomic distance must disappear, then we can easily calculate the value of r_s/a_0, which in our simple model now depends on the critical radius R_c. With this still very strongly simplifying approach we find for r_s/a_0 the expression:

$$\frac{r_s}{a_0} \approx 0.82 + 1.82\frac{R_c}{a_0} + \ldots . \tag{2.33}$$

If, for example, we set the critical radius for sodium to the ionic radius of 0.97 Å, we obtain a value of $r_s/a_0 \approx 4.15$ in good agreement with the experimental value $r_s/a_0 = 3.99$. We obtain similarly good agreement for the other alkali metals. In the case that the ionic radius is known, the binding energy can be calculated with equation (2.32). It turns out, however, that the simple approximation used here is too rough for reliable predictions. While there is qualitative agreement in the case of the light alkali metals such as lithium and sodium, no stable metallic bond would be possible in the case of the heavier metals according to our simple analysis. However, we will not refine the treatment here,

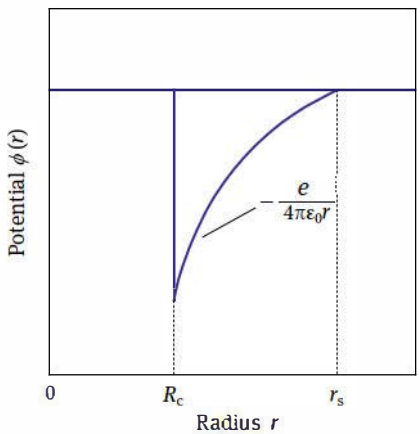


Fig. 2.13: Pseudopotential with critical radius R_c, which takes into account the fact that the free electrons overlap only slightly with the atomic trunk due to the Pauli principle. The critical radius R_c is roughly equal with the ionic radius.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

nor discuss the relatively complex calculations necessary for determining the binding energy of other metals.

We should simply note here, that both *s*- and *p*-electrons can contribute to the metallic bond. In alkali metals only *s*-electrons are present besides the core electrons and the model of the free electron gas (see Section 8.1) provides a good description of the electronic behavior. A purely metallic bond is characterized by the fact that the distance to the nearest neighbor is much greater than the ion diameter. In the transition metals, apart from the delocalized *s*-electrons, there are also strongly localized *d*-electrons. These form additional covalent bonds, determine the distance between the ions and make a significant contribution to the binding energy, but hardly contribute to the conductivity. The ionic radii and ion spacings of these metals are largely determined by the *d*-electrons.

Table 2.4 gives some examples. Here the ionic radii are listed together with the lattice spacing of alkali and transition metals. Finally binding energies for these metals are listed, which are considerably higher for the transition metals due to the covalent bonding contribution and the higher coordination number.

Tab. 2.4: Ionic radius, nearest neighbor separation and binding energy of some alkali and transition metals. (After Ch. Kittel, Introduction to Solid State Physics, Oldenbourg, 2013.)

	Li	Na	K	Cs	Fe	Co	Ni	Rh
Ionic radius (Å)	0.68	0.97	1.33	1.67	1.27	1.25	1.25	1.35
Nearest neighbor separation (Å)	3.02	3.66	4.52	5.24	2.48	2.50	2.49	2.69
Binding energy (eV/Atom)	1.63	1.11	0.93	0.80	4.28	4.39	4.44	5.75

2.6 The Hydrogen Bond

Hydrogen atoms often form a bond which cannot be assigned to the types of bonds described above. This so-called **hydrogen bond** exhibits several peculiarities which are responsible for the special properties of water and ice or the stability of the DNA double helix. At first glance, one would expect great similarities between alkali metals and hydrogen, since both have one *s*-electron in the outermost occupied shell. On this basis hydrogen should also be able to form ionic crystals. However, for the complete separation of electron and nucleus in hydrogen a large amount of energy is required, because the ionization energy of hydrogen atoms is 13.6 eV compared to only 5.1 eV for sodium. Ionic crystals with isolated H^+ ions do not occur because of this high ionization energy and the very small size of the proton.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Since hydrogen atoms have only one electron, they can form a real *covalent bond* with only one neighbor, as we know from many organic molecules. We speak of a hydrogen bridge when the hydrogen atom causes a bond between two atoms, whereby this is only possible when the hydrogen atom is covalently bonded to a strongly electronegative atom. The binding electron then largely, but not completely, passes over to the binding partner and the almost bare proton sits on the negatively charged ion. The positive proton attracts other negatively charged ions. Due to the small proton diameter of only 1 fm, this is only possible with one other partner due to spatial reasons, so that a bridge is formed between two electronegative atoms, e.g. between N, O or F. In most cases, the bond is asymmetric, i.e. of type A-H – B, with the proton H being at a greater distance from atom B than from atom A. In the case of water, the binding energy is 0.2 eV/bond. This means that such bonds can already open and close at room temperature, a property that is of crucial importance for the function of many biomolecules. With 1.6 eV/bond, the binding energy in hydrofluoric acid, i.e. in HF, is considerably higher.

Figure 2.14a shows schematically the hydrogen bridge between two water molecules. As mentioned above and indicated in the illustration, there are two equilibrium positions for the proton between the oxygen atoms. Depending on the charge state of the neighboring oxygen atom, which in turn depends on the orientation of the neighboring water molecules, the proton prefers one or the other position. The transition between the two positions usually takes place by thermally activated processes (cf. Section 5.1), but in some systems with hydrogen bridges, tunnel processes can also occur. The structure of ice can be seen in Figure 2.14b. The protons each link two oxygen atoms,

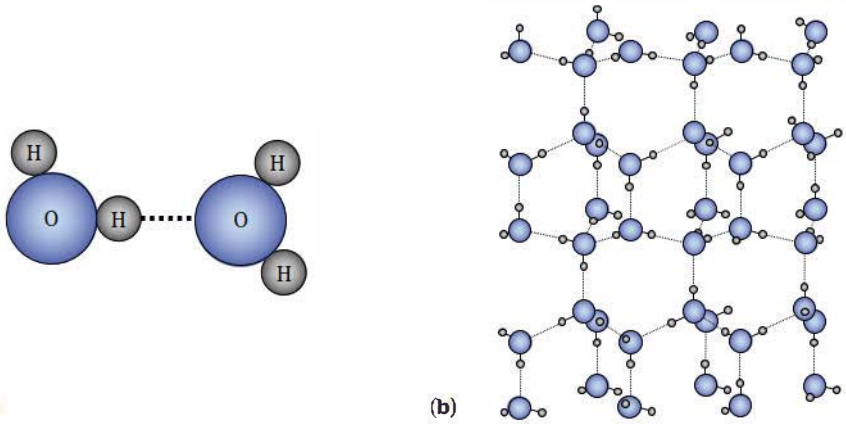


Fig. 2.14: Hydrogen bond. **a)** Schematic illustration of the hydrogen bond between two H_2O molecules. The spacing between the two oxygen atoms is about 2.7 Å, the length of the dotted bridge is about 1.7 Å. **b)** Arrangement of the water molecules in ice. The water molecules are not uniformly oriented even in very well-grown crystals. (After L. Pauling, The Nature of the Chemical Bond, Cornell University Press 1960.)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

but the order is not perfect even in well-grown ice crystals, because, owing to the relatively low binding energy, bonds are frequently broken thermally. The density anomaly of water at 4°C is a consequence of the dynamics of the hydrogen bonds.

Hydrogen bonds play an important role, not only in water, but also in many other inorganic compounds such as NH_3, HF or H_2S. The very high solubility of some oxygen, nitrogen and fluorine compounds in water is related to the formation of hydrogen bonds. This also applies to solutions of ammonia (NH_3) or methanol (H_3COH) in water. In general, hydrogen bonds are of fundamental importance for the function of many biologically relevant molecules.

2.7 Exercises and Problems

Note: *If the data required for the solution are not included in the problems, they can be found in one of the tables of this book.*

1. Types of Bonds. In solids, a distinction can be made between various types of bonds. However, these rarely occur in their pure form. Which bond types dominate in the solid phase of the following materials: argon, magnesium, graphite, diamond, crystalline quartz, quartz glass, polyethylene, GaAs, KBr and NH_3?

2. Solid Helium. Show that helium does not form a solid phase under normal pressure even at absolute zero. Estimate the binding and zero-point energy of a hypothetical helium crystal and compare the result with other noble gas crystals. Note that for a qualitative estimate the exact crystal structure is irrelevant. How can helium still be solidified?
Material parameters for helium: $\varepsilon = 0.86\,\text{eV}$, $\sigma = 2.56\,\text{Å}$.

3. Equilibrium Position. The interaction energy between two atoms can be expressed by the following equation:

$$U(r) = \frac{A}{r^n} - \frac{B}{r^m} \qquad \text{with } m < n\,.$$

(a) Determine the equilibrium position r_0 as a function of the parameters A, B, m und n.
(b) Discuss the variation with position of the force $F(r) = -\mathrm{d}U/\mathrm{d}r$ in the range $0 < r < \infty$. At what position for $r > r_0$ is the force maximum?

4. The Van der Waals Force Between Krypton Atoms. We investigate the Van der Waals interaction between two krypton atoms in more detail.
(a) Calculate the equilibrium separation between the two atoms.
(b) Calculate the energy reduction due to the attractive forces.
(c) How large is the energy increase due to the overlap of the wave functions?

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

(d) How large is the binding energy in equilibrium?
(e) Does the ratio calculated for the attracting and the repulsive part of the Lennard-Jones potential in equilibrium hold in all cases?

5. A Noble Gas Film. A monolayer of densely packed noble gas atoms sits on an unstructured substrate.
(a) Sketch the arrangement of the densely packed atoms in the monolayer.
(b) Determine the binding energy as a function of the interatomic separation.
(c) Calculate the equilibrium separation.
(d) Calculate the separation for xenon atoms.
(e) Compare the calculated value with that for a three-dimensional crystal.

6. Binding Energy. The binding energy of ionic crystals can be determined using the Born-Haber cycle. Use this method to calculate the binding energy of a $1\,cm^3$ KBr crystal. Assume the density to be $\varrho = 2.75\,g/cm^3$, and the following quantities (each referring to an atom or a molecule) ionization energy of potassium ($K \rightarrow K^+$) 4.34 eV, sublimation energy of potassium 0.90 eV, electron affinity of bromine ($Br \rightarrow Br^-$) 3.37 eV, dissociation energy of Br_2: 1.97 eV, and the reaction heat during the formation of KBr from metallic potassium and gaseous bromine: 4.05 eV.

7. Two-dimensional Ionic Lattice. Calculate the Madelung constant for a two-dimensional square ionic lattice by systematically enlarging the cell in question. What is the minimum size of this cell such that the deviation from the value of an infinitely large sample ($\alpha = 1.61554$) is less than 10^{-3}?

8. Ionic Crystals. The binding energy of ionic crystals is based on the electrostatic interaction between the ions.
(a) Determine the equilibrium distance R_0 of the ions for a crystal with binding energy $U_B = -695\,kJ/mol$ and the Madelung constant $\alpha = 1.748$.
(b) Calculate and discuss the parameter zA.
(c) The range of the repulsive potential can be determined more precisely with the help of a second quantity, for example from the measurement of the compression modulus $B = V\partial^2 U/\partial V^2$. Calculate the exponent of the repulsive potential if the compression modulus has the value $B = 1.75 \times 10^{10}\,N/m^2$.
(d) In this case, is the equilibrium separation greater or smaller than predicted by the Lennard-Jones potential?
(e) Which alkali halide crystal could it be?

9. Binding Energy of Lithium. Estimate the binding energy of lithium. Note that when comparing with the experimental value −1.63 eV/Atom, the ionization energy $E_{2s} = 5.39$ eV must be taken into account. Comment on the agreement with the experimental result.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

3 The Structure of Solids

After getting to know the different types of bonds, we turn to the spatial arrangement of atoms and molecules in solids in this chapter. Here, an important aspect is the distinction between crystals and amorphous solids. While crystals are characterized by their regular atomic structure, the structure of amorphous materials is irregular. However, before we examine the structure of the two different classes of solids more closely, we will briefly discuss their production, since it determines whether atoms form well-defined crystals or disordered amorphous solids. In this context, we will also discuss the peculiarities of the production of alloys. Closely linked to the question of structure is the further question of the type of order or disorder. We will also briefly address this point

3.1 The Production of Crystalline and Amorphous Solids

Except for organic materials, most solids are formed by cooling their melt. A typical, well-known example of this is the solidification of water into ice. In this chapter, we first deal with the question of under which circumstances ordered *crystals* are formed from the melt and under which disordered *amorphous solids* are produced.

3.1.1 Single Crystal Growth

In general, when a melt is cooled and no special precautions are taken, **polycrystalline solids** are usually formed. These consist of small crystallites which are connected to each other via largely disordered interfaces. Under certain conditions, **single crystals** can be obtained: these have a continuous regular atomic arrangement and accordingly the same direction-dependent physical properties over the entire volume. Single crystals are anisotropic, whereas polycrystalline materials have isotropic physical properties when the crystallites are randomly oriented and their macroscopic properties then reflect the mean value of the anisotropic crystal properties.

Single crystals are of great importance for both solid-state physics and for material science since they can be used to study the "pure" behavior of crystals. Furthermore, they also play an important role in technical applications. For example, very large and extraordinarily perfect silicon single crystals are produced in large numbers in order to manufacture semiconductor devices. Large silicon single crystals also have a wide field of applications in energy production by solar cells. Another example is piezoelectric single crystals that play an important role in the realization of precise clocks and in communication technology.

https://doi.org/10.1515/9783110666502-003

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Growing larger single crystals is usually a complex process and requires a great deal of experience. Of great importance is the high purity of the starting materials and to guarantee this, the technique of zone melting is often used. Here the starting material is placed in the form of a long (polycrystalline) rod in an elongated crucible and passed through a furnace designed in such a way that only a narrow slab ("zone") of the rod is melted. In this way, new material becomes liquid on the front side of the melting zone and solidifies on the back side. This acts as a cleaning process taking advantage of the fact that incorporating impurities into a crystal is usually energetically unfavorable. The concentration of impurities is therefore higher in the molten state than in the newly solidified crystal. In this way, any impurities are carried along in the melting zone and transported to the end of the rod or crucible.

The quality of the crystals is mainly determined by the growing conditions. In order to pull single crystals, crystallization must of course only start at one particular point in the melt, otherwise multiple starting points would lead to a polycrystalline sample. In many cases, a well-defined temperature profile is therefore established in the vessel and a crystallization nucleus is provided beforehand, i.e. a suitably oriented small seed crystal is introduced into the growth zone as a nucleus, which then grows by the addition of atoms or molecules. In order to produce the most perfect crystals, only small temperature differences in the growth zone can be tolerated, to allow the growth process to take place as close to thermal equilibrium as possible.

There are many methods of single crystal production, but most often crystals are drawn from solutions or melts. Two frequently used methods are shown schematically in Figure 3.1. Which process is most suitable for the production of a particular crystal depends on the specific requirements such as the size or the purity of the crystal and above all on the crystal properties themselves. When growing crystals from a solution, either the solvent is slowly evaporated at a constant temperature or the temperature is gradually lowered, keeping the solution in saturation. Very important from a technical point of view is drawing from the melt, which is discussed in more detail below. In recent years, epitaxial crystal growth methods have become increasingly important.

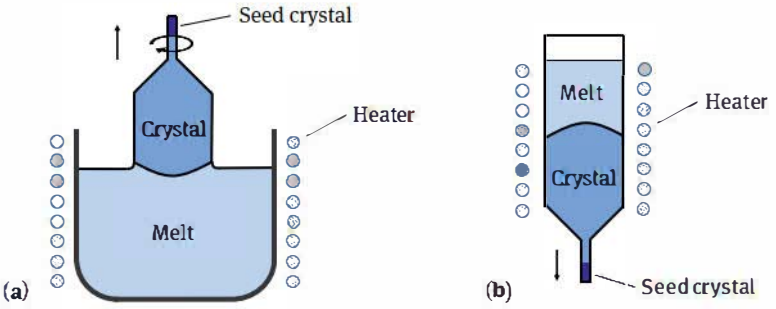


Fig. 3.1: Single-crystal growth methods. **a)** Czochralski-Kyropoulos technique and **b)** Bridgman-Stockbarger technique.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Here, crystalline layers of well-defined thickness are deposited on crystalline substrates from the gas phase or from a solution. The prerequisite for this is that the lattice spacing and structure of the substrate and deposited layer is essentially the same. There is a rich body of literature on the many different crystal growth methods.

As mentioned above, we will now briefly discuss the production of single crystals from the melt. Of great technical importance is the *Czochralski-Kyropoulos technique*,[1,2] which is used to produce very large, high-purity and low-defect silicon and germanium crystals. In this process, a seed crystal placed on a holder is slowly pulled up from the molten material (see Figure 3.1a). The temperature of the seed crystal is kept slightly below both the melting point and the temperature of the melt so that atoms attach to the seed crystal. In order to minimize temperature and pressure gradients, the holder supporting the growing single crystal is simultaneously slowly rotated. Typical drawing speeds are a few millimeters per hour.

In the *Bridgman-Stockbarger technique*,[3,4] the starting material is placed in a vertical cylindrical crucible with a conical lower end (see Figure 3.1b). In the first step, the starting material is melted in the upper part of the furnace. The crucible is then slowly lowered into the lower part of the furnace, where the temperature is a few Kelvin below the melting temperature. A single seed usually forms at the tip of the cone, from which a single crystal grows into the crucible.

In both the above processes there is a risk that the melt will be contaminated by the crucible, which is usually made of a high-melting-point, inert material such as graphite or platinum. Interestingly, crystals can be drawn without the need for a crucible by the zone melting method. The corresponding arrangement is shown schematically in Figure 3.2. A seed crystal is located at the lower end of a polycrystalline rod. The narrow

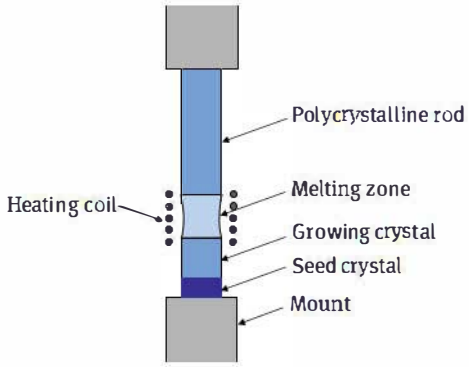


Fig. 3.2: Schematic representation of crystal growth using a crucible-free zone-melting technique. Starting with a seed, the single crystal grows from bottom to top.

1 Jan Czochralski, *1885 Exin, †1953 Posen
2 Spyro Kyropoulos, *1887 Makedonien, †1967 Alamogordo USA
3 Percy Williams Bridgman, *1882 Cambridge (USA), †1961 Randolph, Nobel Prize 1946
4 Donald C. Stockbarger, *1896 Walkerton (USA), †1952 Belmont

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

molten zone, located between the two solid ends of the sample, is held in place by the surface tension of the melt. The melting zone is slowly moved upwards along the rod, such that the single crystal grows from below. In this "floating zone technique", not only a single crystal is drawn, but also the starting material is simultaneously cleaned, as mentioned above.

3.1.2 Production of Alloys

Moving on from the growth of single crystals, in this section we ask the interesting question of what happens if we cool, not a pure material, but a melt consisting of a mixture of different materials. The final product will generally either be a *mixed crystal* or an *alloy*, if the melt consists of a mixture of molten metals. However, it should be pointed out that there is often no clear distinction between the two terms.

Most technically useful metallic materials fall into the class of alloys which usually consist of several *phases* forming microstructures in which a large number of elements occur. Phases, in this context, are volume sections that have the same composition and structure on a scale that is large compared to atomic dimensions. A **phase diagram** shows then which equilibrium states of a mixture occur as a function of temperature and relative concentration both in the liquid and in the solid state. Two simple diagrams of this kind are shown in Figure 3.4 and Figure 3.5 and are discussed below. In most cases, such diagrams are relatively complex. Particularly simple scenarios can be found in systems whose components can be mixed in more or less all proportions, such as Au-Ag, Au-Pd, Ni-Mn, Cu-Pt or Cu-Ni. These pairs of elements have the same crystal structure in their solid state and differ only slightly in their atomic dimensions. However, this unlimited solubility is rather an exception. For example, a maximum of only 0.2 % silver dissolves in aluminum, although the same arguments should also apply to these two elements as well.

Above all, we want to investigate the question of whether two substances can form *homogeneous* or *heterogeneous* mixtures and how the different phases are composed. Basically, two materials will dissolve in one another and form a homogeneous mixture if that has the smallest free enthalpy accessible to these components. A heterogeneous mixture occurs when the free enthalpy of the two coexisting phases is lower than that of the homogeneous mixture. In this case, the system breaks up into regions of different concentrations as a **miscibility gap** occurs.

The crucial quantity in the following thermodynamic consideration is the free enthalpy $G = U - TS + pV$. Here U denotes the internal energy, T the temperature, S the entropy, p the pressure and V the volume. However, since changes of pV do not play an important role in the following discussion, for simplicity, we will neglect this contribution and consider instead the free energy $F = U - TS$ of the system rather than the free enthalpy.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

We start our discussion with a simple thermodynamic consideration and ignore the spatial and temporal fluctuations of the system in question. We then apply the results to some basic examples. In order to keep things as simple as possible, we deal exclusively with binary alloys where the component metals have the same crystal structure.

In the following we consider a mixture of N_A atoms of substance A and N_B atoms of substance B, giving the total number of atoms, $N = N_A + N_B$. The composition of the system is expressed by the fraction of B-atoms $x = N_B/N$ and A-atoms $(1 - x) = N_A/N$. For further simplification, we assume that the interactions only occurs between atom pairs, where u_{AA}, u_{BB} and u_{AB} denote the potential energies of the respective bonds. These binding energies are negative in relation to the energy of isolated atoms. If the atoms are randomly distributed on crystal sites, the mean energy u_A per bond of an A-atom or u_B per bond of a B-atom is given by

$$u_A = (1 - x)u_{AA} + x\,u_{AB} \qquad \text{and} \qquad u_B = (1 - x)u_{AB} + x\,u_{BB} \,. \tag{3.1}$$

If each atom has z nearest neighbors, the mean energy u per atom is

$$u = \frac{z}{2}\left[(1 - x)u_A + x\,u_B\right] = \frac{z}{2}\left[(1 - x)^2 u_{AA} + 2x(1 - x)u_{AB} + x^2 u_{BB}\right] \,. \tag{3.2}$$

The factor 1/2 takes into account that two atoms are involved in each bond. The result can be expressed in the form

$$u(x) = \frac{z}{2}\left[(1 - x)u_{AA} + x u_{BB}\right] + u_M \,, \tag{3.3}$$

where the quantity u_M is the **energy of mixing**:

$$u_M = z\,x(1 - x)\left[u_{AB} - \frac{1}{2}(u_{AA} + u_{BB})\right] = z\,x(1 - x)\tilde{u}_M \,. \tag{3.4}$$

In this very simple model, the energy of mixing depends quadratically on the concentration x. If the mixed bond is stronger than the bond between the same components, then $\tilde{u}_M < 0$. Thus the free energy of the mixture is always smaller than that of the unmixed systems. In this case the components of the system are miscible in any ratio. For $\tilde{u}_M > 0$, however, the consequences are much more complicated.

We also have to take into account the **entropy of mixing** S, which arises from the number of possible arrangements of the atoms A and B:

$$S = k_B \ln \frac{N!}{N_A! N_B!} = k_B \ln \frac{N!}{[N(1 - x)]!(Nx)!} \,. \tag{3.5}$$

With the *Stirling*[5] *formula*, $\ln X! \approx X(\ln X - 1)$, which for large values of X is valid, we find

$$S = -Nk_B[(1 - x)\ln(1 - x) + x \ln x] \,. \tag{3.6}$$

5 James Stirling, *1692 Stirlingshire, †1770 Edinburgh

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Putting these expressions for the energy and the entropy of mixing back into our expression for the free energy $F = U - TS$, we find for the free energy of the mixture $F(x)$:

$$F(x) = Nu(x) - TS(x)$$
$$= N\left\{\frac{z}{2}\left[(1-x)u_{AA} + x\,u_{BB}\right] + u_M + k_B T\left[(1-x)\ln(1-x) + x\ln x\right]\right\}. \quad (3.7)$$

The properties of the mixture depend crucially on the ratio of the mixing energy to the thermal energy. If the thermal energy $k_B T$ is large enough, the systems always mix. If the (positive) mixing energy dominates $\tilde{u}_M$, the above-mentioned miscibility gap occurs. The crucial parameter is $p = z\tilde{u}_M/k_B T$. If $p < 2$, the free energy (3.7) has a minimum. If $p > 2$, then a maximum appears lying between two minima. This behavior is shown in Figure 3.3 for the case $\tilde{u}_M > 0$, where it was assumed that the bond between the A-atoms is somewhat stronger than that between the B-atoms. In the concentration range between the two minima the system is unstable and decays into two phases, which we will call α and β.

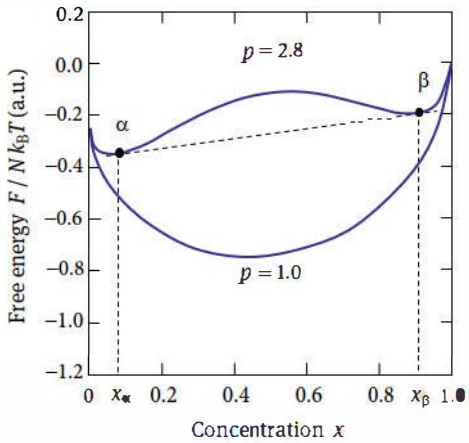


Fig. 3.3: The free energy of a binary mixture calculated with equation (3.7) for the parameters $p = 1.0$ and $p = 2.8$. A maximum occurs for $p > 2$. The mixture is unstable in a certain concentration range, the so-called miscibility gap. For $p < 2$ the free energy has only one minimum.

Let us first consider the free energy F of a mixture consisting of the two phases α and β. We express the free energy as

$$F(x) = N_\alpha f(x_\alpha) + N_\beta f(x_\beta)\,, \quad (3.8)$$

where N_α and N_β denote the number of atoms and $f(x_\alpha)$ and $f(x_\beta)$ the mean free energy per atom in the phases α and β, respectively. With $Nx = x_\alpha N_\alpha + x_\beta N_\beta$ and $N_\alpha + N_\beta = N$ equation (3.8) can be transformed as follows:

$$F(x) = N\left[\frac{x_\beta - x}{x_\beta - x_\alpha} f(x_\alpha) + \frac{x - x_\alpha}{x_\beta - x_\alpha} f(x_\beta)\right]. \quad (3.9)$$

The free energy of the mixture varies linearly with the concentration of the B-atoms and is therefore represented in the Fx-plane by a straight line.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The stability of the system is determined by the **chemical potential** $\mu = (\partial F/\partial N)_{T,V}$, which reflects the change in free energy with the number of particles. If the system decays into the phases α and β, in equilibrium the chemical potentials of the two phases

$$\mu_\alpha = \frac{\partial F_\alpha}{\partial N_B} = \frac{1}{N}\frac{\partial F_\alpha}{\partial x_B} \quad \text{and} \quad \mu_\beta = \frac{\partial F_\beta}{\partial N_B} = \frac{1}{N}\frac{\partial F_\beta}{\partial x_B} \tag{3.10}$$

must be equal. This means that at the points on the $F(x)$ curve corresponding to the two concentrations x_α and x_β the curve must have the same slope. This applies first of all to any pairs of points, but the lowest energy is obtained if, as shown in Figure 3.3, a common tangent can be placed on the curve.

Thus every mixture whose composition lies in the range between the two concentrations x_α and x_β will separate into two phases with precisely these compositions. In other words the miscibility gap mentioned above occurs.

The lower part of Figure 3.4 shows the phase diagram of the Cu-Ni system. Depending on the temperature and the composition, the system can exist as a homogeneous liquid (ℓ for liquidus), a homogeneous solid (s for solidus) or a two-component system (s + ℓ). The various areas are separated from each other by the *liquidus* and *solidus* curves. In the upper part of the figure the concentration dependence of the free energy of the two phases at a fixed temperature is shown schematically. The common tangent touches the curves at the concentrations x_ℓ and x_s. In equilibrium, for $x < x_\ell$, the system will take the form of a homogeneous liquid. Over the range $x_\ell < x < x_s$ the system consists of two phases, one solid with composition x_s and one liquid with the composition x_ℓ. For $x > x_s$ the equilibrium state is a homogeneous solid. Since the

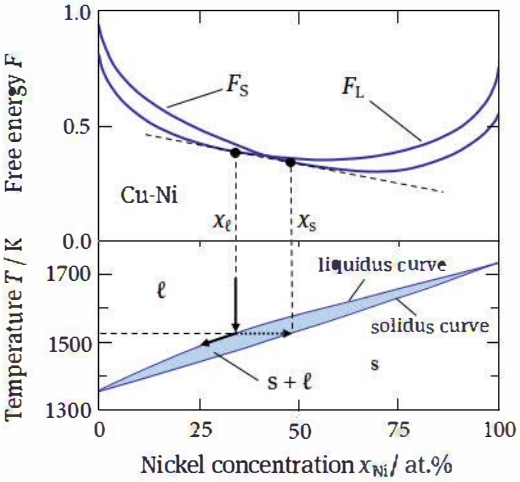


Fig. 3.4: A schematic representation of the free energy of liquid and solid for a binary mixture in equilibrium at $T = 1530$ K (upper part) and phase diagram of the copper-nickel system (lower part). (After F. Goodwin et al., *Springer Handbook of Condensed Matter and Material Data*, W. Martienssens, H. Warlimont, eds., Springer, 2005.)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

free energy of the solid decreases faster with temperature than that of the liquid,[6] the values of x_ℓ and x_s shift with temperature, as shown in the bottom part of Figure 3.4.

We can also see from this figure that binary melts do not solidify at a fixed temperature. If the temperature of the melt is lowered (downwards arrow), solidification begins when the liquidus curve is reached. The precipitating phase has a composition which corresponds to the concentration x_s (dotted arrow) and is therefore richer in nickel than the melt. As a result, copper accumulates in the liquid phase and lowers the solidification temperature. If the temperature is lowered further, the composition of the melt moves along the liquidus curve as indicated by the arrow following the liquidus curve, until the melting point of copper is reached. Since the course of the solidus curve also depends on temperature, the composition of the precipitating Cu-Ni alloy also changes continuously. This means that the resulting solid phase is not in equilibrium. Since atomic diffusion occurs relatively rapidly at temperatures just below the melting point of copper, homogenization of the sample can be achieved by annealing.

As a second somewhat more complicated example, we consider the behavior of Sn-Pb alloys, which in the past have frequently been used as soft solders. The corresponding phase diagram is shown in Figure 3.5. If a melt is cooled from high temperature, say with a composition 40at.% Sn / 60at.% Pb, it remains homogeneous until the phase separation curve is reached at about 543 K. Now the lead-rich α-phase containing a maximum of 29 at.% tin is precipitated.[7]

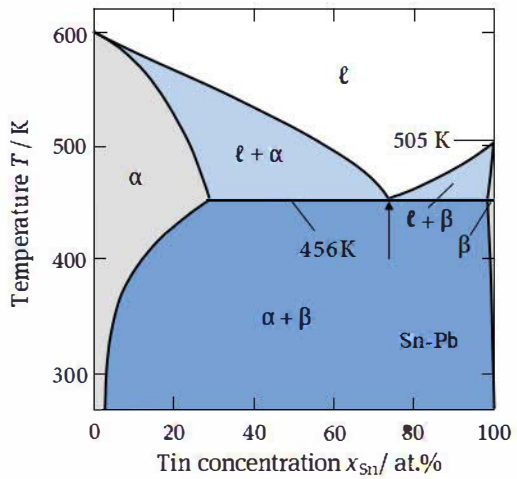


Fig. 3.5: Phase diagram of the Sn-Pb system. The eutectic point is highlighted by an arrow. (After F. Goodwin et al., *Springer Handbook of Condensed Matter and Material Data*, W. Martienssens, H. Warlimont, eds., Springer, 2005.)

6 The reason for this is that the vibrational spectra of the liquid and the solid are different, leading to different temperature dependences of the internal energy of the two phases.

7 In metallurgy, the phases are denoted by Greek letters. The meaning of the letters depends on the system under consideration.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

This causes tin to accumulate in the melt. As it continues to cool, the system runs along the phase separation line until the **eutectic point** is reached at 456 K with a content of 73.9 at.% tin (26.1 at.% lead) in the melt. If the temperature is lowered further, the melt fully solidifies with the formation of a mixture of the two phases α and β whose compositions will change further with decreasing temperature. However, the slow diffusion of the atoms in the solid usually ensure that the concentrations remain practically constant, except in the case of extremely slow cooling.

There are many binary systems in which the liquid phase remains at temperatures far lower than the lowest melting point of the constituents. This situation is often encountered with materials that have a miscibility gap in their solid but not in their liquid phase. A mixture with two liquidus curves is referred to as a **eutectic system**. The lowest temperature at which solidification occurs in such system, is called the *eutectic temperature* or *eutectic point*, and the associated composition the *eutectic composition*. For the Sn-Pb alloy, the melting points of tin and lead are 600 K and 505 K, respectively, but the eutectic point is 456 K.

The Sn-Pb soft solder traditionally used in the past has a composition that comes very close to being eutectic. Figure 3.6 shows an optical image of an eutectic Sn-Pb melt after solidification. Due to their different reflectivities, the alternating lamellae of the α- or β-phase can be clearly seen, which in this case they are essentially tin or lead.

Eutectics are of great interest not only from a scientific point of view, but also for practical applications. The system Au-Si is a case in point. While gold melts at 1336 K and silicon at 2177 K, the eutectic alloy with 69 % Au and 31 % Si melts at a mere 643 K. This property is of great importance in semiconductor technology because it allows gold wires to be welded to silicon components at relatively low temperatures.

Fig. 3.6: Optical image of a solidified Sn-Pb eutectic. (After Ch. Kittel, Introduction to Solid State Physics, Oldenbourg, 2013.)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

3.1.3 Glass Production

The term *amorphous solid* is not precisely defined. It implies that atoms are not found on regularly arranged sites as in crystals. However, the name does not indicate the degree of disorder. In addition to glasses, this class of materials also includes many thin-film systems produced by vapor deposition, plastics, and organic solids. We select the *glasses* from the multitude of amorphous materials because they are typical representatives of this substance class. The main difference between the production of crystals and the production of glasses is that rapid temporal and spatial temperature changes are avoided during crystal growth, whereas during glass production the melt is cooled relatively quickly in order to prevent the crystallization process.

As sketched in Figure 3.7, characteristic volume changes occur during melting and solidification, which are different for glasses and crystals. If a melt consisting of one component is cooled so slowly that the formation of ordered structures with atomic long-range order is possible, then crystallization occurs at the solidification temperature T_m. This is a first-order phase transition accompanied by a discontinuous volume change. In Section 5.4 we will briefly discuss the phenomenon of phase transitions.

If, on the other hand, a glass melt is cooled rapidly, it can be cooled below the crystallization temperature T_m without solidification occurring. Only considerably below T_m does the melt solidify into a glass. It is easy to imagine that atomic rearrangements constantly occur in a melt becoming slower and slower with decreasing temperature. However, since the breaking of atomic bonds in the melt and the orderly deposition of atoms on the crystal surface are prerequisites for crystal growth, this is suppressed in a supercooled melt. In the V-T-diagram, glass formation is indicated by a change of slope of the curve at the **glass transition temperature** T_g. Considerably below this temperature the arrangement of atoms is frozen within experimentally-accessible time

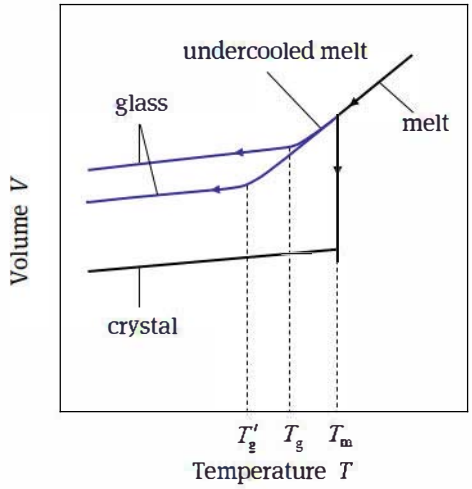


Fig. 3.7: Schematic illustration of the change of volume during solidification or crystallization of a melt. Crystallization occurs at a well-defined solidification temperature T_m. The glass transition takes place at T_g or T_g', depending on the cooling rate. Note that the volume changes are not drawn to scale.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

scales. Since the structure of the glasses hardly differs from that of their melts, they are often referred to as frozen liquids. As can be seen from the sketch, glasses have a larger volume and thus a lower density than their corresponding crystals.

The glass transition temperature T_g is typically about 2/3 of the solidification temperature of the corresponding crystal, i.e., $T_g \approx 2\,T_m/3$. For vitreous silica (i.e. quartz glass) consisting of pure SiO_2, T_g is 1350 K, whereas the melting point of crystalline SiO_2 is 1990 K. Depending on the composition, optical glasses or window glasses have T_g values of around 900 K, while in case of aqueous lithium chloride solution $LiCl{\cdot}7\,H_2O$, T_g of 137 K is far below the ice point. The cooling rates required for glass production are less than 10^{-6} K/s for good glass formers such as SiO_2, $LiCl{\cdot}7\,H_2O$ or for multi-component glasses, i.e. for substances that can easily be brought into the glass state. However, for most metallic glasses, such as CuZr, the cooling rates must be above 10^6 K/s to avoid crystallization, as these are poor glass formers. As indicated in Figure 3.7, T_g (within narrow limits) can be influenced by variation of the cooling rate. Lower cooling rates lead to a reduction of T_g and a slightly higher glass density. This changes the structure and consequently the physical properties to some extent. If a glass sample is tempered, i.e. if the sample is kept at a temperature just below T_g for a longer period of time, in many cases small crystalline areas will begin to form in the sample.

From the brief description of the processes at the glass transition and from the fact that the glass transition temperature and the resulting glass structure depend on the cooling rate, it can be concluded that glasses at room temperature are not in thermodynamic equilibrium but are instead in a metastable state[8]. In consequence, amorphous solids also exhibit a high residual entropy at absolute zero due to their structural disorder.

Since the processes taking place at the glass transition are not yet fully understood and their theoretical description is still incomplete, there is currently much intensive research in this field. Basically, two extreme ideas of the nature of the glass transition can be distinguished: either one treats the glass transition simply as a phase transition in the thermodynamic sense or one sees in it just a kinetic transition, in other words a process, in which the molecular movements are frozen with the rapid cooling. According to the current state of knowledge, the glass transition is probably a mixture of both, therefore new concepts and experiments are necessary to investigate them.

3.2 Order and Disorder

Since we shall be encountering the terms **order** and **disorder** again and again in the following chapters, we illustrate them by means of sketches of simple arrangements

8 Metastability is a weak form of stability outside the thermal equilibrium that can be destroyed over time.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

of atoms. In particular, we will discuss the structural difference between crystalline and amorphous solids. For a discussion of order and disorder we take a look at the Figures 3.8 –3.12, which show a range of actually occurring structures. For simplicity, we limit ourselves to two-dimensional representations: imagined as suitable cuts through a three-dimensional structure. We can thus understand Figures 3.8, 3.9a and 3.10a as cuts through well-ordered crystals, while the remaining figures show structures where different types of disorder prevail.

Packing hard spheres as densely as possible, we arrive at the arrangement in the plane as shown in Figure 3.8a. Since such a simple structure is typical of crystals with highly symmetrical structural units and non-directional bonding forces, they can be found in noble gas crystals and many metals. Often, however, the chemical bonds force a different arrangement, for example, grey arsenic consists of regular rings with six arsenic atoms as shown in Figure 3.8b. In real three-dimensional arsenic crystals such structures are stacked layer by layer.

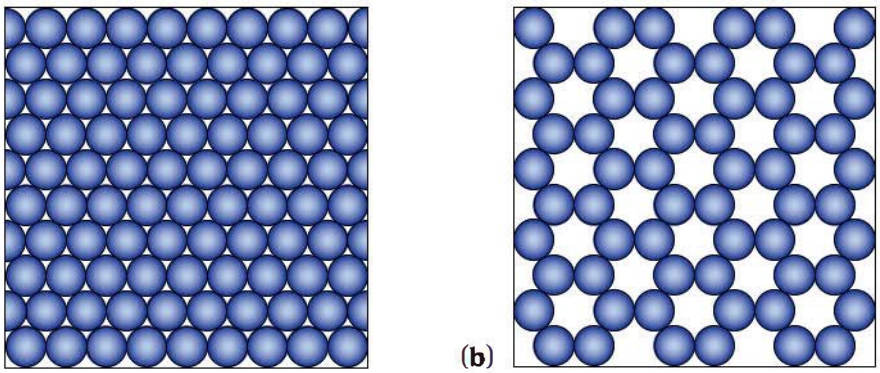


Fig. 3.8: Illustration of simple crystal structures. **a)** Crystal of high symmetry composed of identical atoms. **b)** Arrangement of arsenic atoms in arsenic crystals.

Figure 3.9a shows a similar ring structure, but the rings now contain two different kinds of atoms, A and B, which are divalent and trivalent, respectively. Here, too, one can imagine that crystals of the chemical composition A_2B_3 are formed when the layers shown are stacked on top of each other. We want to mention another possibility: the figure can also be seen as a cut through a crystal of the chemical composition AB_2, in which the small tetravalent A-atoms are connected by bivalent B-atoms. In this case, relatively rigid AB_4 tetrahedra would be the building blocks resulting in a three-dimensional cross-linking of the structure. Quartz crystals are an example of this kind having a structure that is similar to the one shown in Figure 3.9a, although the structure occurring in nature is still somewhat more complicated. However, we want to use this simple picture to illustrate the transition from crystal to glass.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

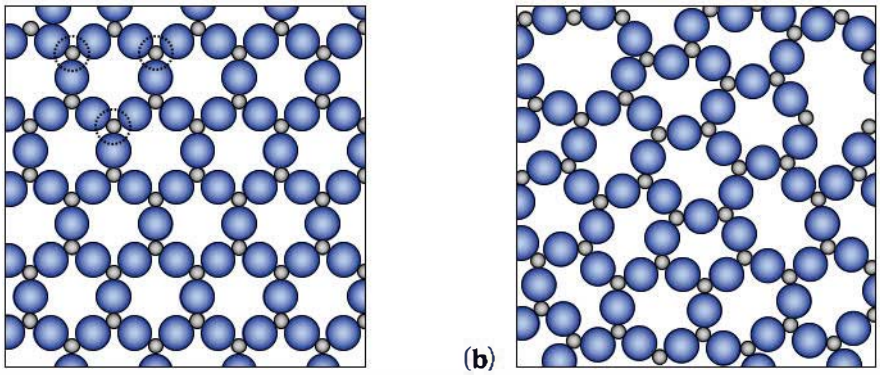


Fig. 3.9: Crystal and glass of the same chemical composition. **a)** A_2B_3 crystal. The large and small circles represent atoms of different valence, 2 and 3 respectively. In the upper left part, three B atoms of a hypothetical AB_2 crystal are indicated as dotted circles. These atoms reside above the drawing plane. **b)** A_2B_3 glass. Binding angles between atoms are subject to small variations. This results in irregularly shaped rings with no long-range order.

For this we turn to the amorphous structure in Figure 3.9b, which is composed of the same components as the crystalline structure of Figure 3.9a just discussed. The bivalent atoms are each connected to two adjacent trivalent atoms, thus ensuring a certain measure of **short-range order** simply arising from the local geometry favored by the chemical bonds. However, instead of regular identical rings, irregularly-shaped rings linking different numbers of atoms occur. The reason for this lies in the small deviations of the bond angles from the well-defined angles characteristic of crystals. These deviations lead to structural disorder and the loss of the **long-range order**, thus destroying structural correlations at large distances. The absence of this type of order is characteristic of amorphous solids. In the case of SiO_2, the change from the left to the right panel of the illustration would correspond to the transition from a (hypothetical) quartz crystal to vitreous silica (quartz glass), both being composed of similar structural units.In the following discussion we denote the amorphous modification of a material by an prefix to the chemical formula. For example, a-$Si●_2$ denotes vitreous silica and a-Si amorphous silicon.

Figure 3.1● shows two crystals with the composition AB. While that on the left has the two kinds of atoms arranged regularly, the structure on the right shows substitutional disorder with the same stoichiometry, i.e. with the same nominal composition, but with the atoms A and B statistically distributed on the existing sites. As we shall see in Section 5.4, solids can exist in both modifications between which reversible transitions occur. An example is the alloy CuZn, known as β brass, where at a critical temperature, an *order-disorder transition* takes place between the two structurally-different phases.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

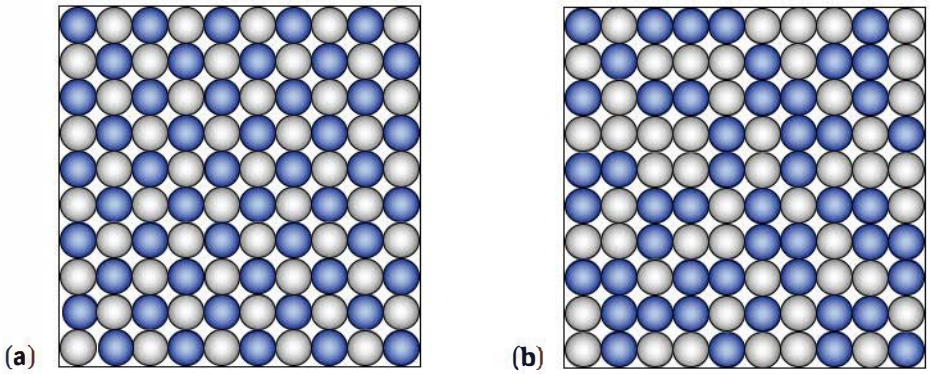


Fig. 3.10: Structure of an alloy AB. **a)** A regular crystal AB. **b)** A crystal with substitutional disorder in which both kinds of atoms are randomly distributed on the available sites.

In Figures 3.11a and 3.11b, the building blocks themselves are not fully spherically symmetric. The disorder here arises from the orientation of the building blocks, while their spatial arrangement remains completely regular. In Figure 3.11a, one type of atom carries a magnetic moment whose orientation is indicated by an arrow. In our two-dimensional example, the magnetic moment may point in four different directions. This kind of magnetic disorder can be found, for example, in paramagnetic salts when there is no external magnetic field. Of course, this is only a snapshot in time, since the orientations of the individual magnetic moments are constantly changing from thermal motion. In such systems, long-range order in the orientation of the moments can often occur below a critical temperature. This marks a phase transition from the paramagnetic to the ferromagnetic (or antiferromagnetic) state.

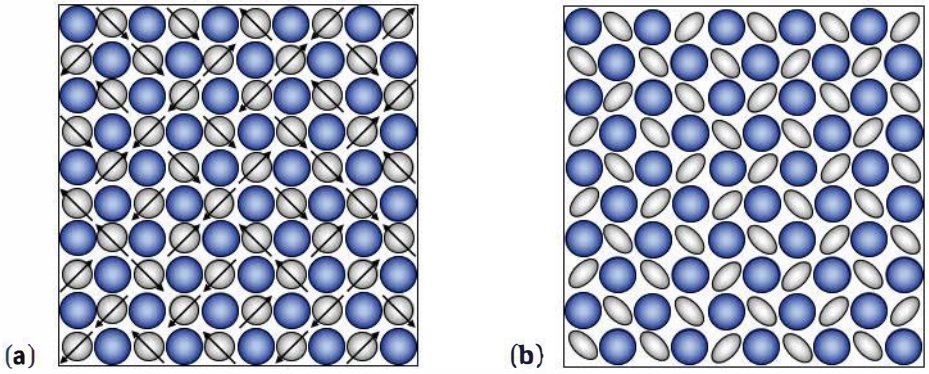


Fig. 3.11: Crystals with orientation disorder. **a)** Magnetic moments, indicated by small arrows, are statistically oriented. **b)** Ellipsoid-shaped molecules occupy existing equilibrium positions with no preferred orientation.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3.11b shows a crystal half of which consists of elliptical molecules or ions whose longitudinal axes point randomly in one of the two diagonal directions. An example is the ionic crystal CsCN, which is composed of spherical cesium ions and ellipsoid cyanide molecular ions. At higher temperatures, the CN ions are randomly oriented in the direction of the space diagonals of a cube and perform thermally-activated rotational jumps between the different equilibrium positions. On cooling, at a the critical temperature $T_c = 196\,K$ a phase transition takes place, leading to a change in the crystal structure, where all CN ions become aligned along the direction of one of the space diagonals.

Finally in this section, we look at *liquid crystals* which are generally made up of rod-like molecules. 3.12a shows a liquid crystal in the *nematic* phase. In these materials, the molecules are radomly distributed in space but with orientational order. Figure 3.12b shows a liquid crystal in the *smectic* phase. Here the molecules are aligned, as in the previous case, but are now additionally arranged in individual planes. However, crucially difference from a normal crystal, within the planes, the molecules are arranged randomly. Many modern display screens have a **Li**quid **C**rystal **D**isplay (LCD). In the above, we have seen that on one hand we have the ideal crystal with its strictly periodic structure and on the other the amorphous solid with its largely disordered structure. As can be seen in Figures 3.8 – 3.12, there is a broad spectrum of structures of solids in between, in which there is a pronounced, but imperfect order. In Chapter 4 we will see that there are a number of other defects that have a major influence on the physical properties of solids.

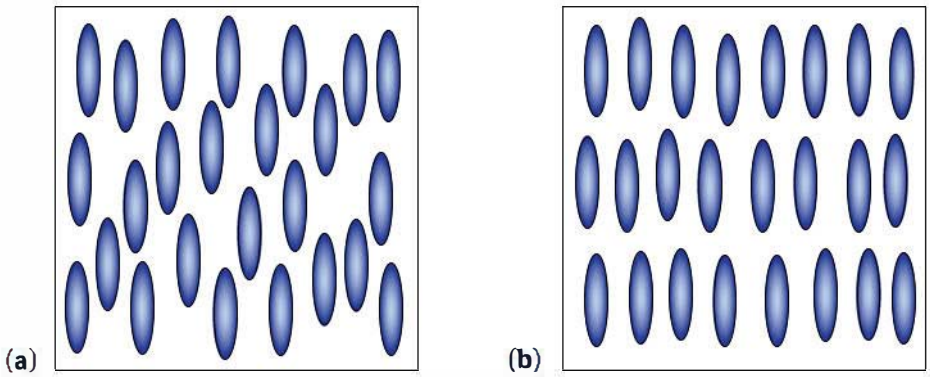


Fig. 3.12: Liquid crystals with oriented rod-shaped molecules. **a)** Nematic phase, **b)** smectic phase.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

3.3 The Structure of Crystals

In the following we first discuss the structure of crystals and then the structure of amorphous solids. The experimental aspects of structure determination will be discussed in Chapter 4.

3.3.1 Translational Symmetry and Crystal Systems

Since carefully-grown crystals stand out because of their regular shape, it is reasonable to assume that their building blocks are also regularly arranged. This thought was expressed as early as 1801 in a treatise from which the illustration of Figure 3.13 is taken. The illustration shows how the regular shapes of a crystal can be built up by the incremental addition of identical building blocks. In 1912 *M. von Laue*[9] made the suggestion that the regular structure of crystals might possibly be demonstrated by the use of X-rays, followed shortly by the experimental evidence provided by *W. Friedrich*[10] and *P. Knipping*[11].

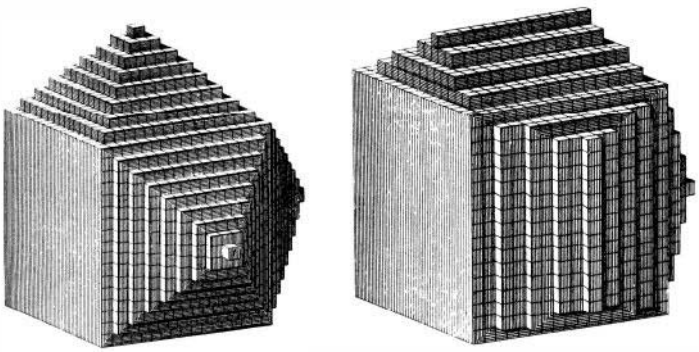


Fig. 3.13: Historic illustration for the construction of crystals from identical building blocks. (After R.-J. Haüy,[12] Traité de minéralogie, Paris, 1801.)

An ideal crystal consists of identical, identically-oriented groups of atoms arranged in a three-dimensional, infinitely extended and strictly periodic order. Each individual unit of the periodically-recurring structural groups is known as the **basis**. How many atoms form the basis depends on the substance. While in many metals the basis may consist

9 Max von Laue, *1879 Pfaffendorf (Koblenz), †1960 Berlin, Nobel Prize 1914
10 Walter Friedrich, *1883 Salbke (Magdeburg), †1968 Berlin
11 Paul Knipping, *1883 Neuwied, †1935 Darmstadt
12 René-Just Haüy, *1743 Saint-Just-en-Chaussée, †1822 Paris

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

of only one atom, in the case of complex protein crystals, the basis may comprise more than 10^4 atoms. If we assign a point in space to each structural unit of this kind, we can reduce the crystal structure to a lattice of points which enables us to describe the crystal structure mathematically in a simple way. This procedure is illustrated in Figure 3.14 for a two-dimensional structure. As indicated in the figure, the choice of basis is not unambiguous. However, it does not matter which position within the basis is chosen as the lattice point, because only a shift of the point lattice as a whole is related to this.

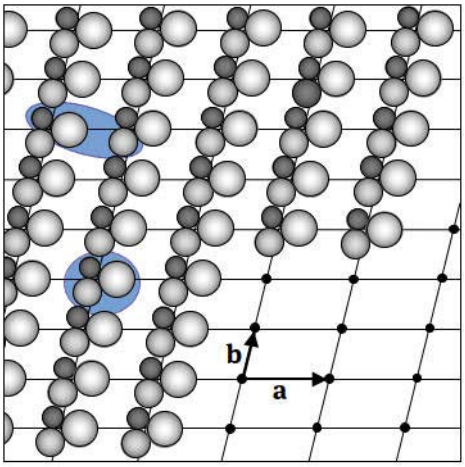


Fig. 3.14: Crystal structure and point lattice. The arrangement of the atoms of the basis (two possibilities are indicated by blue background ovals) and the choice of the origin of the basis with respect to the point lattice is of no importance. The vectors a und b define an oblique coordinate system, suitable for the description of the point lattice.

The symmetry of crystals can be described with the help of symmetry operations that transform the point lattice or crystal structure into itself. A distinction is made between the *translation symmetry* and the *point symmetries*. While in translational operations the crystal structure as a whole is shifted, in point-symmetry operations at least one point remains fixed in space.

To begin, we ignore the influence of the basis and look at the point lattices relating to different crystals, i.e. we assume a point-like basis. The most striking and important symmetry that we will often exploit is the **translational symmetry**. We choose any point in the crystal and look at its environment, which we denote by the symbol $\mathcal{U}$. A look at the two-dimensional crystal in Figure 3.14 shows that this environment repeats itself at regular intervals. We express this by noting that:

$$\mathcal{U}(\mathbf{r}) = \mathcal{U}(\mathbf{r} + \mathbf{R}) \,, \tag{3.11}$$

where $\mathbf{r}$ stands for an arbitrary position vector in real space. The translation operation taking us from one location to an equivalent one is described by the **translation vector** or **lattice vector R**, defined by

$$\mathbf{R} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \,. \tag{3.12}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The vectors **a**, **b** and **c** define a coordinate system which represents the symmetry of the point lattice. In the literature we can find various names for these vectors, for example, fundamental translation vectors or basis vectors. Here we use the term "bias vectors" which is customary in mathematics. In order to avoid confusion, it should be emphasized that the "bias vectors" do *not* describe the positions of the atoms within the basis, but simply define the point lattice. The components of **R** must be integer multiples of the basis vectors, and thus n_1, n_2 and n_3 are always integers. The lengths a, b and c of the basis vectors are called **lattice constants**.The parallelopiped spanned by the three basis vectors, is called the **elementary cell** or the **unit cell**. As Figure 3.15 shows, by putting such cells together space can be completely filled leaving no gaps or overlaps.

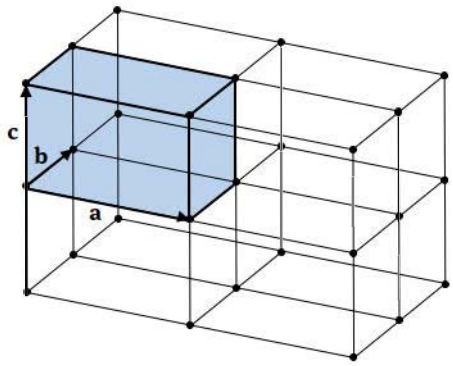


Fig. 3.15: Three-dimensional lattice. The vectors **a**, **b** and **c** define the unit cell, highlighted by light blue coloration and thicker edges.

However, the choice of the basis vectors is not unambiguous and several equivalent choices can be made as illustrated in Figure 3.16 for a two-dimensional lattice. The basis vectors drawn define unit cells of different shapes and sizes.

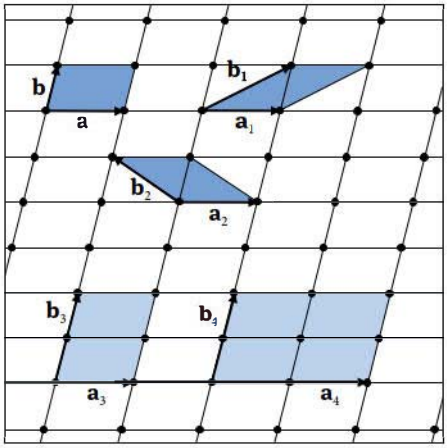


Fig. 3.16: A two-dimensional point lattice with different unit cells. The darker cells contain only one lattice point; they are therefore *primitive* unit cells. It should be noted that three of the four corners of the unit cells and thus also the corresponding lattice points are associated with neighboring cells. The two lighter, *non-primitive* cells contain two or four lattice points.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

A distinction is made between **primitive** and **non-primitive unit cells**. Primitive unit cells are the smallest possible containing only a single lattice point, usually chosen as the origin of the unit cell. They are characterized by the fact that *all* equivalent lattice points in space can be reached by the basis vectors via the relation (3.11). Non-primitive unit cells contain several lattice points and, starting from a position vector **r**, their basis vectors cannot reach all locations with the same environment using the relation (3.11).

Now we turn to **point symmetry operations**, which comprise rotation around an axis, mirroring and inversion. When such operations are performed, at least one point in space remains fixed. We first discuss **rotation** around an axis. Clearly the rotation of a point lattice by the angle 2π leads to congruence. Depending on the point lattice under consideration and a suitable choice of the axis of rotation, this can also be achieved by rotations by an angle of $2\pi/n$, where n is an integer number. However, this is only true when the value of n is 1, 2, 3, 4 or 6. In the two-dimensional case, it is easy to see that only regular triangles, quadrangles and hexagons allow a full coverage of the area without overlapping. Figure 3.17 shows that this does not apply to regular pentagons, heptagons or octagons. The same arguments also hold in three dimensions for filling the space with the corresponding regular polyhedra. Rotational symmetry of order n, also called n-fold rotational symmetry of a rotation axis, indicates how often congruence occurs during a rotation of the lattice within 2π. In the so-called *international notation*, which we will discuss in more detail below, this is simply expressed by the numbers 1, 2, 3, 4 or 6. We will meet concrete examples by discussing the symmetry of the basis and the point lattice of different crystals.

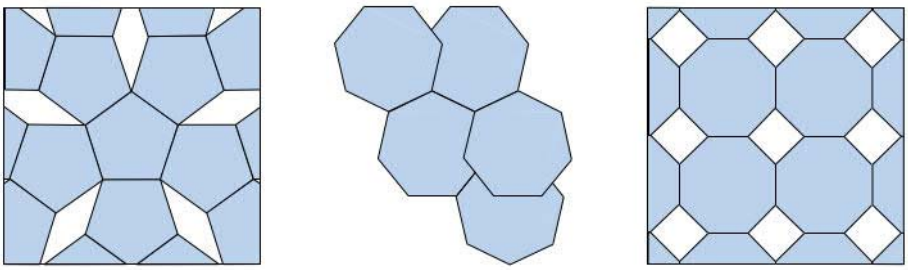


Fig. 3.17: Regular pentagons, heptagons or octagons cannot be arranged to fill two-dimensional space without overlapping. This means that symmetry axes with this order n cannot exist.

Another point symmetry operation is the **reflection** in a plane where not only an axis but the whole plane is fixed. Reflection symmetry can only occur in combination with rotational symmetry where the reflection plane either contains the axis of rotation or is perpendicular to it. The presence of such a reflection in a plane is indicated by the symbol m. The **inversion symmetry** (also called parity transformation) causes a transformation of the vector **r** into the vector $-\mathbf{r}$. This operation is expressed

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

by $x, y, z \rightarrow \bar{x}, \bar{y}, \bar{z}$ or by the symbol $\bar{1}$. Point lattices always obey inversion symmetry, but this can be lost in real crystals due to the missing symmetry of the basis.

In addition, there are two combined symmetry operations, often referred to as improper rotation: **rotary inversion** and **rotary reflection.** The rotary inversion is closely related to the rotation. It consists of a rotation by $2\pi/n$ with subsequent inversion. For its representation the symbols $\bar{1}$, $\bar{2}$, $\bar{3}$, $\bar{4}$ or $\bar{6}$ are used. At this point it should be mentioned that the point symmetry $\bar{2}$ corresponds to the reflection on a plane. In rotary reflection, the rotation is followed by a reflection on a plane perpendicular to the axis of rotation. The corresponding notation is $\frac{2}{m}$, $\frac{3}{m}$ and so forth.

Point lattices and their associated crystal structures can be classified according to their point symmetry operations. For a point lattice to be assigned to a certain **crystal system**, it must satisfy a *minimum* symmetry requirement. In this way, *seven* crystal systems can be distinguished, as listed in Table 3.1. In the column with the rotation order, the information in brackets indicates that for orthorhombic crystals two two-fold and for cubic crystals four three-fold axes of rotation are present. However, the classification made here based on the axes of rotation is still too restrictive in this form, because the simple rotation can also be replaced by the rotary inversion. This extension of the definition plays an important role especially when the crystals under consideration do not have a simple basis. For example, a crystal with a $\bar{6}$-axis is classified as hexagonal, although there is only a three-fold axis of rotation. The condition for the classification of a crystal as a cubic system is remarkable, because here the existence of four *three-fold* axes of rotation is required. These axes coincide with the space diagonals of the cube.

Tab. 3.1: The seven crystal systems and their basis vectors. The order of rotation refers to the rotational or rotary inversion axes (see text). The angle α is enclosed by the basis vectors **b** and **c**, and then, in cyclic order, β is enclosed by **a** and **c** and γ is enclosed by **a** and **b**.

Crystal system	Lattice constants	Angles	Order of rotation
triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$	1
monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90°, \beta \neq 90°$	2
orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90°$	2 (two)
tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$	4
hexagonal	$a = b \neq c$	$\alpha = \beta = 90°, \gamma = 120°$	6
trigonal (rhombic)	$a = b = c$	$\alpha = \beta = \gamma < 120° \neq 90°$	3
cubic	$a = b = c$	$\alpha = \beta = \gamma = 90°$	3 (four)

A crystal structure *can* also have *further* point symmetries which are not relevant for the classification into a crystal system. For example in the progression cubic → tetragonal → orthorhombic → monoclinic → triclinic, which goes from higher to lower symmetry

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

every preceding crystal system also has the symmetry elements of the following one. The same applies to the "side branches" hexagonal → orthorhombic, trigonal ⇢ hexagonal → monoclinic and cubic → trigonal. When specifying the crystal system for a certain crystal, the system with the highest number of matching symmetry elements is selected.

Many elements and simple compounds have cubic or hexagonal structure. This is intuitively understandable, since single atoms or simple molecules are often approximately spherical and can form highly symmetrical structures. However, if the basis is made up of less symmetrical molecules, then the crystal structure usually has less symmetry. Temperature often plays an important role in determining the symmetry of a crystal. Since higher temperatures can lead to thermal rotations of molecules the details of their shapes becomes less important. In consequence, as a rule of thumb, the high-temperature phases of solids often have a higher symmetry than the phases occuring at lower temperatures.

In many cases, the primitive unit cell does not fully express the symmetry of the point lattice. An example of this is the two-dimensional centered rectangular lattice shown in Figure 3.18. Starting from a primitive unit cell with the basis vectors $\mathbf{a}'$ and $\mathbf{b}'$, an oblique point lattice might be inferred. However, if instead a unit cell with two lattice points is used, a cell with higher symmetry can be defined: The angle between the new basis vectors **a** and **b** is then 90° and the additional symmetries of two different planes of reflection emerge.

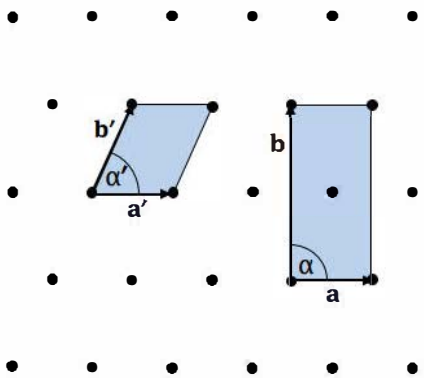


Fig. 3.18: Centered rectangular lattice. The rectangular, non-primitive unit cell contains two lattice points. The primitive unit cell does not reflect the full symmetry of the point lattice.

In order to exploit fully the symmetry, non-primitive unit cells are often used, with shapes chosen to encompass the highest possible number of point-symmetry elements. This leads to the 14 **Bravais lattices**,[13] shown in Figure 3.19 for three-dimensional point lattices. As can be seen in the figure, half of the Bravais lattices have a non-primitive unit cell. In the triclinic, trigonal and hexagonal crystal systems, the primitive unit

13 Auguste Bravais, *1811 Annonay, †1863 Le Chesnay

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

cell is identical to the corresponding Bravais lattice. In all other cases several Bravais lattices are assigned to the same crystal system.

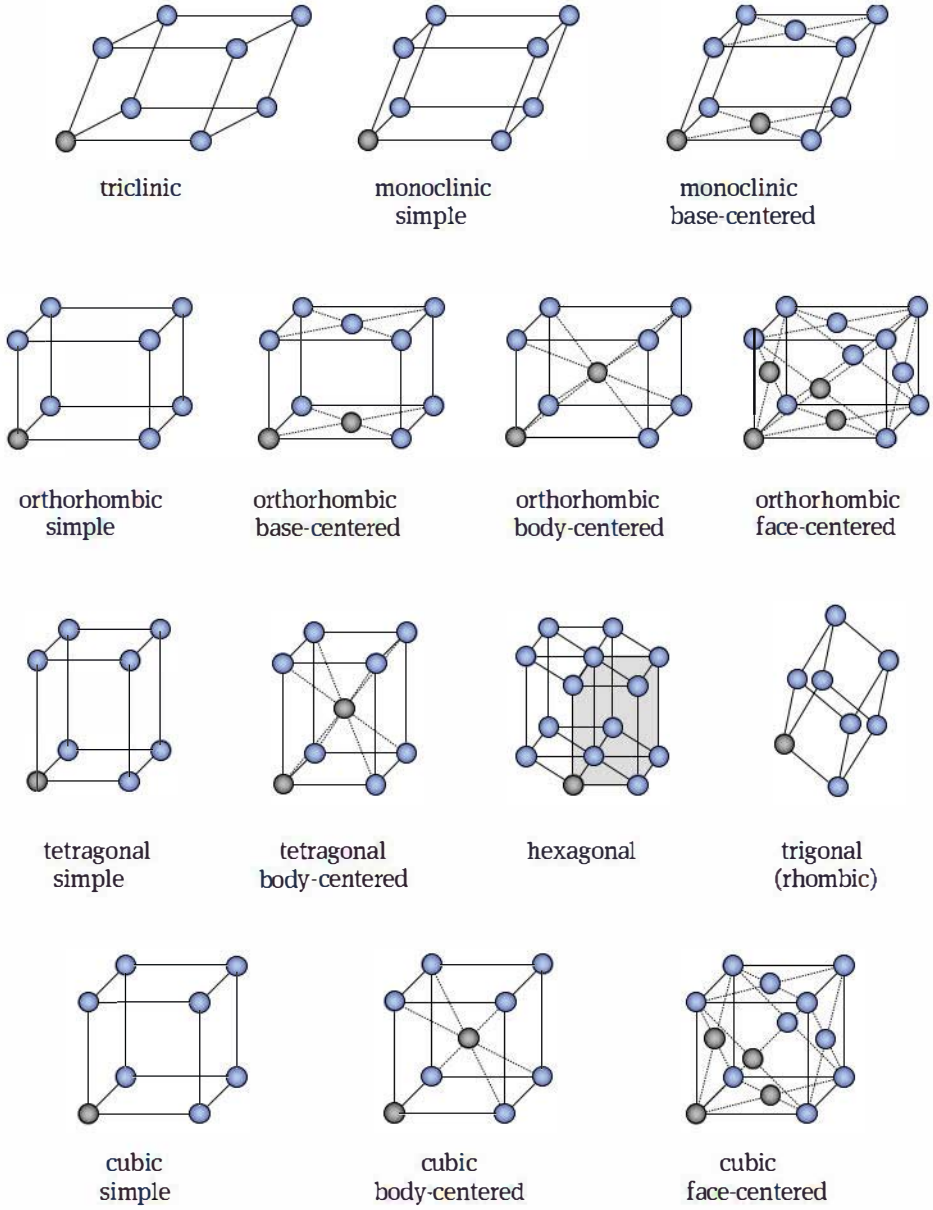


Fig. 3.19: Bravais lattices. The lattice points belonging to the respective unit cells are represented by dark-grey spheres. Lattice points assigned to neighboring unit cells are shown as light-blue spheres. The primitive unit cell highlighted in grey of the hexagonal Bravais lattice is a regular prism.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

3.3.2 Clusters and Quasicrystals

As we have seen in the previous section, five-fold symmetry is not compatible with the requirement for complete spatial coverage by identical structural units. Despite this, solids with this symmetry are indeed found in nature in the form of atomic **clusters**. In noble gases, for example, icosahedral clusters[14] consisting of twelve atoms on the surface and one atom in the center are preferred, since this atomic configuration with its five-fold symmetry leads to a particularly high binding energy. While periodically-constructed macroscopic crystals with five-fold symmetry are not possible, five-fold symmetry can be found in larger clusters containing up to 1000 atoms. In metal clusters, the bond is largely mediated by free, non-localized electrons. With this type of bond, a spherical shape of the cluster proves to be advantageous.

In 1984, D. Shechtman[15] surprisingly discovered an Al-Mn alloy that seemingly violated the oldest and most fundamental theorem in crystallography on the allowed orders of rotation. Later, further alloys were found which solidified into **quasicrystals** on the rapid cooling of their melts. Figure 3.20 shows a quasicrystal which clearly exhibits the "forbidden" shape. Despite their five-fold symmetry, quasicrystals proved to be ordered in structure studies. As we have seen, quasicrystals cannot be assembled like normal crystals from identical structural units. However, five-fold symmetry with complete space filling is possible if *two* kinds of rhombohedra are joined together in such a way that they form dodecahedron-shaped structural units. Dodecahedral quasicrystals have the highest symmetry of all known crystal lattices: They have six five-fold, ten three-fold and fifteen two-fold axes of rotation. Although such point lattices have a defined rotational symmetry, their lattice *lacks* the translational invariance typical of ordinary crystals despite their pronounced orientational order.

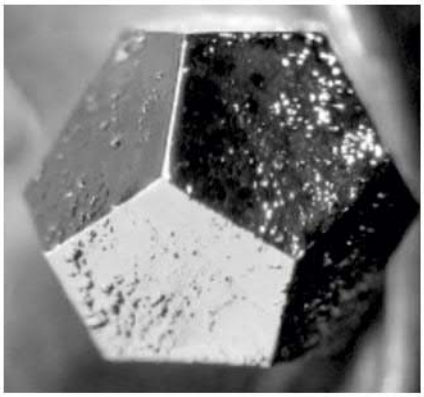


Fig. 3.20: A dodecahedral quasicrystal with the composition Ho-Mg-Zn. The quasicrystal, a few millimeters in size, clearly shows the dodecahedron shape. (After I.R. Fisher et al., Phil. Mag. B 77, 1601 (1998).)

14 An icosahedron consists of 20 slightly deformed tetrahedra, whose peaks are in the center.

15 Daniel Shechtman, *1941 Tel Aviv, Nobel Prize 2011

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In a quasicrystal, however, the atoms are not really arranged periodically but only "quasiperiodically". This means that quasicrystals have a locally regular structure, but are aperiodically arranged over large distances. This means that each cell is surrounded by a different environment of "unit cells". This fact can be illustrated very well in two dimensions. Figure 3.21 shows a so-called Penrose pattern[16,18], also called "Penrose tiling", which consists of two different rhombus shapes with the same edge length and the angles 36° and 72° respectively. Thus a plane can be filled without gaps. The light blue tint indicates that the pattern contains regular decagons of the same orientation. Each regular decagon consists of five small and five large rhombi. In addition, on closer analysis, more or less straight bands of rhombi can be identified, which are inclined by 72° to each other, but here we will not go further into this kind of translational symmetry. It can also be shown that the number of small rhombi divided by the number of large rhombi is given by the irrational number $(1 + \sqrt{5})/2$. Therefore, it follows that it is impossible to describe the Penrose pattern with a single unit cell of any size.

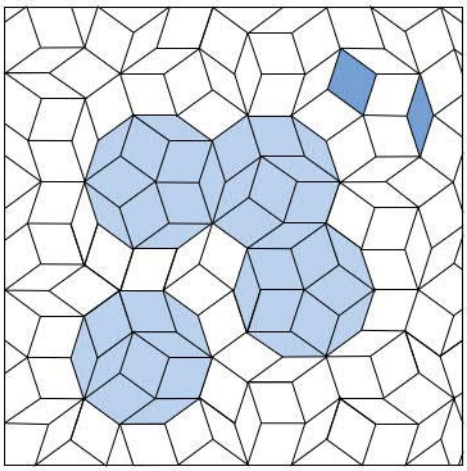


Fig. 3.21: Penrose pattern. The two types of underlying rhombi are highlighted in blue. Furthermore, four regular decagons are marked in light blue. Note that all decagons, some of which overlap, are oriented in the same way.

Quasicrystals are a modern and interesting field of solid state physics. Since their initial finding, various material systems have been discovered from which "crystals" of millimeter size can be produced with moderate cooling rates. These include Al-Li-Cu, Al-Cu-Fe or Zn-Mg-RE, where RE is a rare earths atom. There are also some quasicrystals composed of just two elements. This category includes $Cd_{5.7}Yb$, $Cd_{5.7}Ca$ and $Ta_{1.6}Te$, which have an icosahedral or dodecahedral structure.

16 This pattern was first described in 1974 by the mathematicians *R. Penrose* and independently by *R. Ammann*[17].

17 Robert Ammann, *1946 Boston, †1994 Billerica

18 Roger Penrose, *1931 Colchester, Nobel Prize 2020

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Owing to the lack of translational symmetry, the electrical conductivity of quasicrystals is rather low (cf. Chapter9) and decreases with decreasing temperature. Similarly, the thermal conductivity of quasicrystals is significantly lower than that of comparable crystalline materials. Quasicrystals are comparatively hard and brittle, are extremely resistant to corrosion and have low coefficients of friction and wetting. That said, the relationship between the macroscopic properties and their unusual structure is only partially understood at present.

3.3.3 Notation and the Influence of the Basis

In our discussion so far, we have largely ignored the influence of the basis on the crystal structure. As long as the basis consists only of a spherical atom, it does not influence the symmetry properties of the crystal. However, the situation changes fundamentally when the basis is more complicated. Figure 3.14, in which an asymmetric molecule was chosen as the basis, already showed this complication. Here, in contrast to the point lattice, the actual lattice with the basis, i.e. the crystal, does not obey inversion symmetry. In order to assign a crystal to a certain crystal system, the point lattice and basis must satisfy the minimum requirements for the symmetry elements. The problem associated with this is illustrated again in Figure 3.22, in which three cubic unit cells with hypothetical bases are represented.

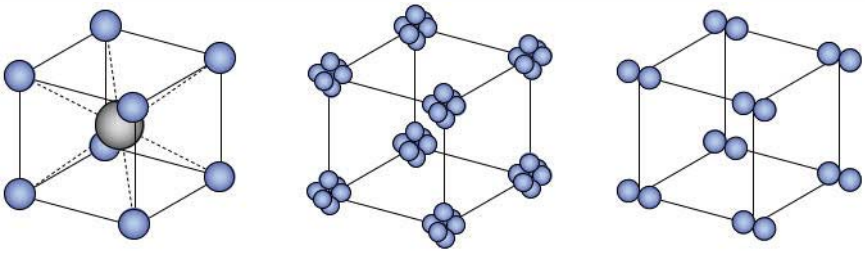


Fig. 3.22: Cubic unit cells with different bases. The two arrangements of basis atoms on the left are compatible with cubic symmetry. The arrangement on the right does not allow the classification into the cubic crystal system.

The addition of an atom at the center of the cube or having molecules with sufficiently high symmetry at the cube corners are compatible with the definition of the cubic crystal system, which requires four axes of rotation with three-fold symmetry. However, the two-atomic basis on the right is not compatible with this, since here the unit cell lacks the required axes of rotation with three-fold symmetry along the space diagonals.

The discussion on the influence of the shape of the basis leads us back to the question of symmetry operations. If we look for the symmetry elements of the cube in Figure 3.22a, we find beside the four axes of rotation with three-fold symmetry, which

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

are required for the classification into the cubic crystal system, three axes of rotation with four-fold symmetry and planes with reflection symmetry. Of course, there are further atomic arrangements with other symmetry elements, which also have four axes of rotation with three-fold symmetry and can thus act as the basis for cubic crystals. In a systematic search, five different arrangements can be found, signifying that the cubic crystal system can be subdivided into five **crystal classes**. We can illustrate this by taking a geometric object that has exactly the required symmetry elements. In Figure 3.23 the five possible classes are illustrated by cubes with different pattern on their surfaces. All five objects shown meet the requirements listed in Table 3.1 for inclusion in the cubic crystal system. The connection of such representations with the real structure of the basis is discussed in more detail below where the meaning of the notation is also explained. Corresponding considerations apply to the remaining six crystal systems. The set of symmetry elements compatible with the respective system are called crystallographic **point groups**.

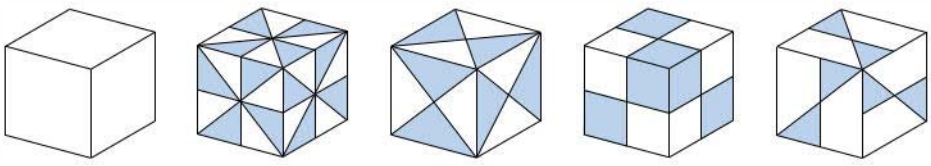


Fig. 3.23: Geometric objects with the symmetries of the corresponding cubic point groups. The first row below shows the corresponding international notation,[19] and the second the Schoenflies notation.

There are a total of 32 point groups or crystal classes that are distributed among the seven crystal systems. Their symmetry elements can be visualized with the help of *stereo-graphic projection*. This is mainly used in crystallography, but we will not go into this here, since it is not central to the discussion in this chapter.

Unfortunately, two different notations are used to denote the point groups. In the **international notation** (after *C. Hermann*[20] and *C.V. Mauguin*[21]), which we have used up to now without specific comment, rotary axes or rotary inversion axes and reflections are used for notation. This type of representation is mainly used to define the crystallography. The second one the **Schoenflies system** (named after *A.M. Schoenflies*[22]) is

19 There is also a shortened notation: Instead of $\frac{4}{m}\bar{3}\frac{2}{m}$ or $\frac{2}{m}\bar{3}$ one will then find the abbreviations $m\bar{3}m$ and $m\bar{3}$.

20 Carl Hermann, *1898 Lehe (Bremerhaven), †1961 Marburg

21 Charles-Victor Mauguin, *1878 Provins, †1958 Villejuif

22 Arthur Moritz Schoenflies, *1853 Landsberg an der Warthe, †1928 Frankfurt am Main

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

mostly used in group theory and spectroscopy. Unfortunately, there is no "one to one" correspondence of the different notations, since some symmetry operations are treated differently.

In Schoenflies notation, rotary inversion is replaced by rotary reflection, in which a reflection in a plane perpendicular to the axis of rotation follows after the rotation. In addition, the Schoenflies notation consists of a main symbol, which contains the order of the (vertical) axes of rotation. Here, C_n represents an n-fold vertical rotary axis, S_n is an n-fold rotary reflection axis and D_n stands for an n-fold vertical rotary axis, combined with a two-fold vertical rotary axis. Here "C" stands for cyclic, "S" (for "Spiegel", German for mirror) and "D" for dihedral. For mirror planes, the additional symbols "v", "h" and "d" are introduced as indices, depending on whether there are still n vertical (i.e. parallel to the axis of rotation) mirror planes, a horizontal (i.e. perpendicular to the axis of rotation) mirror plane or mirror planes, which contain the n-fold and halve the angles between the two-fold axes. Point groups that have more than one symmetry axis with $n > 2$ are abbreviated to T ("tetrahedron"), O ("octahedron") and I ("icosahedron"). As for the underlying regular polyhedra, these point groups have a large number of symmetry elements.

We will not go into the individual point groups and their names in more detail here, but will simply give an idea of the general procedure by means of two examples. Let us first consider the water molecule, shown in Fig. 3.24. It obviously has a two-fold axis of rotation and two mirror planes parallel to this axis. The denotation in the international notation is therefore $2mm$. In the Schoenflies notation the point group is called C_{2v}. Here C_2 stands for the axis of rotation, the index "v" indicates that there are also vertical mirror planes. As already mentioned, one can imagine a suitable geometric object that reflects the symmetry elements of the considered point group. A possible example in the present case is a cuboid with white and light-blue halves, as shown in Figure 3.24 on the right.

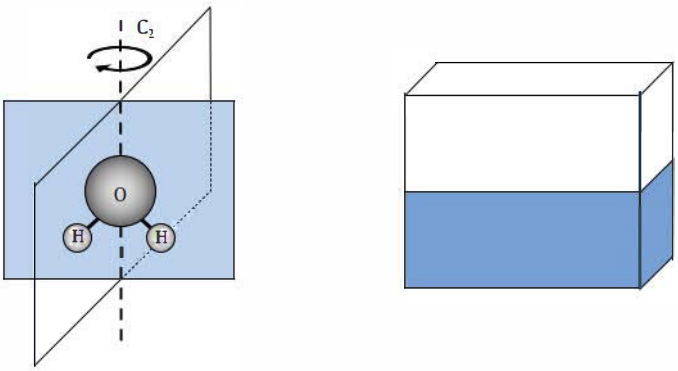


Fig. 3.24: Example for the point group $2mm$ or C_{2v}. The water molecule and the cuboid both have the symmetry elements of this point group.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In the second example we will deal with the point group of the molecule PF_3Cl_2, shown in Figure 3.25. It is clear that the molecule (and of course the corresponding geometric analog) has a six-fold rotary inversion axis and perpendicular to it a two-fold rotational axis. Furthermore, there is a vertical mirror plane. The corresponding name in the international notation is therefore $\bar{6}2m$. In the Schoenflies notation the corresponding point group is known as D_{3h}, the two-fold rotary axis being already included in the abbreviation D_n.

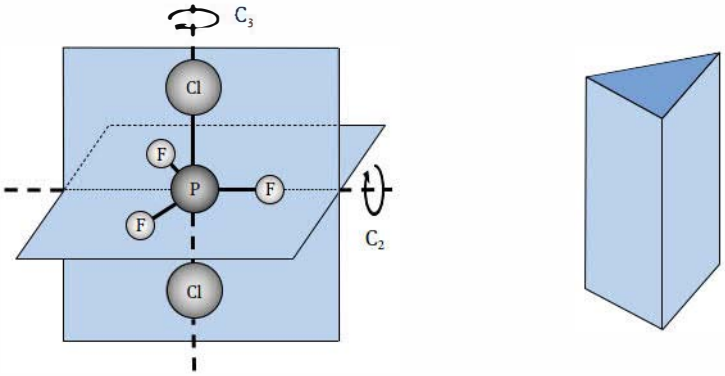


Fig. 3.25: Example for the point group $\bar{6}2m$ or D_{3h}. The triangular column reflects the symmetry elements of the phosphorus dichloride trifluoride molecule.

In addition to the symmetry operations already mentioned, further symmetry operations can be found by combining translation and point symmetry operations. These mixed symmetry elements include *screw axes* and *sliding mirror planes*, the meaning of which can be inferred from the names. If these additional symmetry operations are taken into account, the point groups can be further split and a total of 230 crystallographic **space groups** can be established. This is by no means a trivial topic, and we will not pursue it further as it is not important for the rest of the discussion.

Symmetry operations not only serve in the systematic classification of crystals, but also in the description of solid-state properties, where they often greatly simplify the mathematical treatment. For example, symmetry relations are of crucial importance in the theory of lattice vibrations and electronic states. They determine, among other things, the selection rules for optical transitions and play a crucial role in the question of the coupling of defects to their environment.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

3.3.4 Simple Crystal Lattices

In this section we will look at cubic and hexagonal crystals because they are easy to describe and very common in nature. We will be constantly referring to the number of **nearest neighbors**, often also referred to as the *coordination number*. In simple ionic crystals, this is the number of ions directly surrounding a selected ion, or the number of atoms directly bound to a central atom in the case of a more complex structure.

Cubic crystals. Referring back to figure 3.19 we see that there are three different cubic point lattices: the simple, the body-centered and the face-centered cubic lattice. They differ in the number and arrangement of lattice points in the unit cell and thus also in their **packing density** (also called the *packing ratio*). This is the name given to the fraction of the space filled by identical, touching spheres on the lattice points. The edge length of the cube-shaped unit cells is given by the lattice constant a.

The unit cell of the **simple cubic** lattice (abbreviation **sc** for "simple cubic") is shown again in Figure 3.26a. The reference atom at the coordinate origin $(0, 0, 0)$ is surrounded by six *nearest* neighbors at distance a and twelve *second nearest* neighbors at distances $a\sqrt{2}$. In this configuration, the packing density is 0.52, i.e. only about half of the space will be filled with hard spheres in this arrangement. Simple cubic crystal lattices with a monatomic basis only occur rarely in nature, e.g. α-Polonium or some metals under high pressure. However, this structure is often found in crystals with a polyatomic basis.

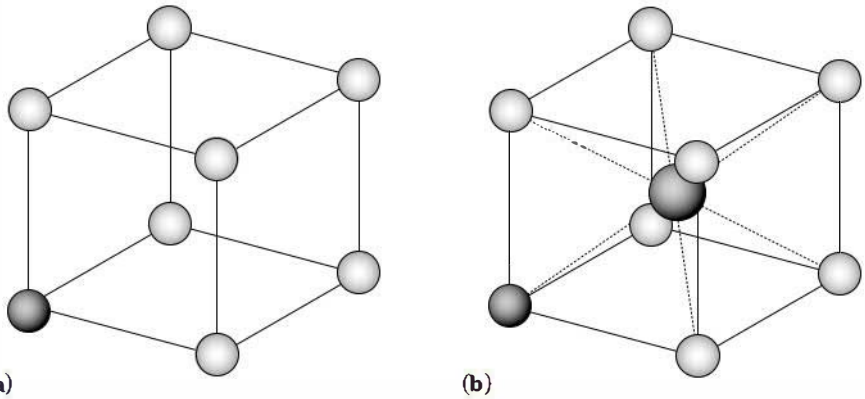


Fig. 3.26: The simple cubic lattice. **a)** A cubic lattice with a monatomic basis (dark colored). The lighter-colored atoms already belong to the neighboring unit cells. **b)** The cesium chloride structure. The unit cell contains the two dark-colored atoms or ions at the positions $(0, 0, 0)$ and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Here, too, the lighter-colored atoms are already components of the neighboring unit cells.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

A well-known example of a simple cubic lattice with a two-atomic basis is the **cesium chloride structure** shown in 3.26b. The two atoms or ions of the basis occupy the positions (0, 0, 0) and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ in the unit cell, which results in a much higher space filling than is possible with a monatomic basis. As one can see from the picture, the coordination number in this case is eight. This structure can be found in simple ionic crystals if the ions involved have comparable radii. Of course, this also applies to cesium chloride with the radii $r_{Cs^+} = 1.69$ Å and $r_{Cl^-} = 1.81$ Å.

Another example of a crystal with a simple cubic lattice is the mineral *perovskite* with the chemical formula $CaTiO_3$. In the unit cell the titanium ion is located at the origin, the calcium ion in the center of the cube and the three oxygen ions at positions $(\frac{1}{2}, 0, 0)$, $(0, \frac{1}{2}, 0)$ and $(0, 0, \frac{1}{2})$. Minerals with the perovskite structure are very common and are an essential part of the earth's crust. We will consider the perovskites $BaTiO_3$ and $SrTiO_3$ in Section 13.3 in discussing the properties of ferroelectrics,

Body-centered cubic lattices are very common in nature. Thus, the alkali metals, Ba, Cr, Eu, Fe, Mo, Nb, Ta, V and W and many compounds or alloys, exhibit this structure, for which the abbreviation **bcc** is used. As shown in Figure 3.27a, the unit cell contains two lattice points with coordinates (0,0, 0) and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. In a monatomic basis, the eight nearest neighbors are located at a distance of $a\sqrt{3}/2$ and the six second nearest neighbors are located at distance a. The packing density at 0.68 is substantially higher than that of the simple-cubic lattice mentioned above.

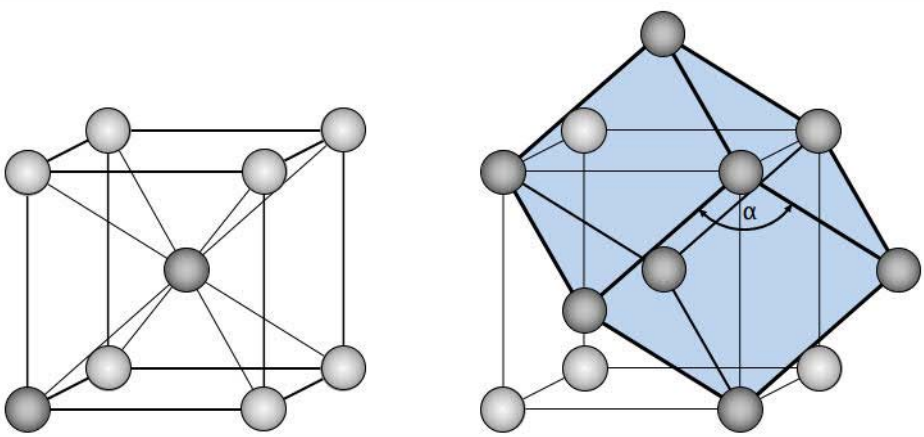


Fig. 3.27: Body-centered cubic lattice. **a)** Conventional unit cell with the two darker-colored lattice points. **b)** The primitive unit cell is highlighted in blue. The thin lines denote the non-primitive cubic unit cell. The angle α between the axes of the primitive unit cell is 109°28′.

Of course, instead of the cubic unit cell with two lattice points, one can also choose a primitive unit cell that contains only one lattice point according to the definition.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In this case, instead of the cube, a rhombohedron with low symmetry, edge length $a' = a\sqrt{3}/2$ and angle $\alpha = 109°\,28'$ represents the lattice. Figure 3.27b shows both the primitive and a non-primitive unit cell.

The **face-centered cubic** lattice is the third cubic lattice type, abbreviated **fcc**. Noble gas crystals, Ac, Al, Ag, Au, Ca, Cu, Ni, Pb, Pd, Rd, Sr, Th, Yb and many alloys or compounds have a face-centered cubic lattice. The cubic unit cell (cf. Fig. 3.28a) contains four lattice points with coordinates (0, 0,0), $(\frac{1}{2}, 0, \frac{1}{2})$, $(\frac{1}{2}, \frac{1}{2}, 0)$ and $(0, \frac{1}{2}, \frac{1}{2})$. The twelve nearest neighbors sit at a distance $a/\sqrt{2}$, the six second nearest at a distance a from a given atom. For equally sized hard spheres, 12 is the largest possible number of nearest neighbors. Face-centered cubic lattices therefore have the densest packing of spheres with a packing density of 0.74. Figure 3.28b shows the cubic and the primitive unit cell of this lattice type. As in the case of the body-centered cubic lattice, the primitive unit cell here is rhombohedral. It has edge length $a' = a/\sqrt{2}$ and the acute angle of the rhombi is 60°.

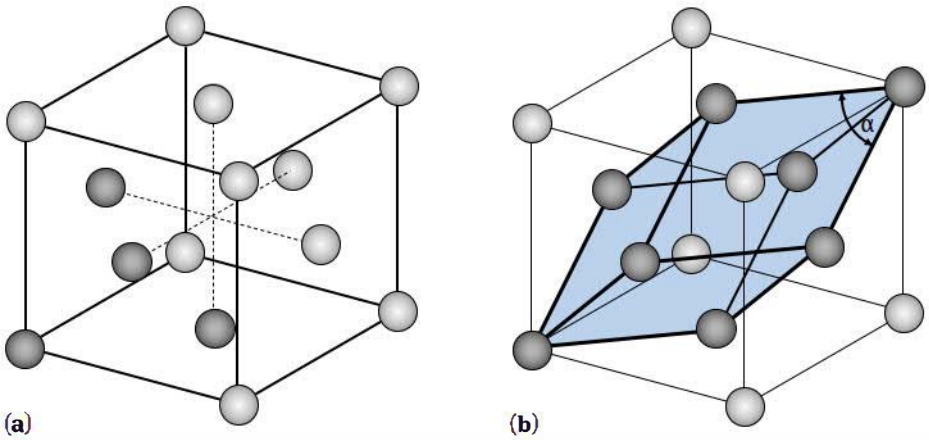


Fig. 3.28: Face-centered cubic lattice. **a)** Conventional cubic unit cell with four darker-colored lattice points. **b)** The primitive unit cell is highlighted in blue, with the non-primitive cubic cell indicated by the thin lines. The axes of the primitive unit cell are separated by $\alpha = 60°$.

A well-known example of a face-centered cubic lattice with a *two-atom* basis is the **sodium chloride structure** (cf. Figure 3.29), into which many alkali halides crystallize. The two different ions at the positions (0,0,0) and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ form the basis. We can therefore speak of two sublattices, one shifted from the other by half the space diagonal of the cube. This structure with the coordination number six occurs preferably when the radii of the ions involved differ strongly, as in the case of common salt with $r_{Na^+} = 0.99$ Å and $r_{Cl^-} = 1.81$ Å.

Other important examples of face-centered cubic lattices are *zinc blende* and *diamond*. The basis of zinc blende (ZnS) consists of a zinc atom and a sulphur atom having

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

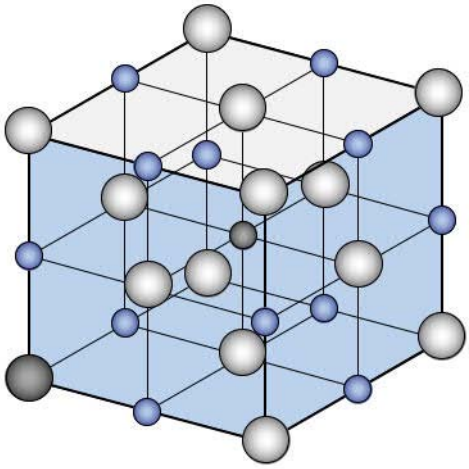


Fig. 3.29: The sodium chloride structure. The two dark-colored ions at the positions $(0, \bullet, \bullet)$ and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ form the basis.

the coordinates $(0, 0, 0)$ and $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Each atom is located in the center of a regular tetrahedron formed by atoms of the other kind as shown in Figure 3.30a. The coordination number is only four, the packing density, assuming atoms of the same size, only 0.34. Nevertheless, this crystal structure frequently occurs when the atoms are connected to four neighbors via covalent sp^3-bonds (cf. Section 2.4). Further examples showing the zinc blende structure are the compound semiconductors GaAs, GaP, CdS, InP and InSb.

Many elements such as carbon in the form of diamond, silicon, germanium or grey tin also have this structure. Although all positions are occupied by identical atoms, the primitive unit cell for these crystals contains two atoms as shown in Figure 3.30b. The diamond structure is a good example of the fact that, in the definition of the

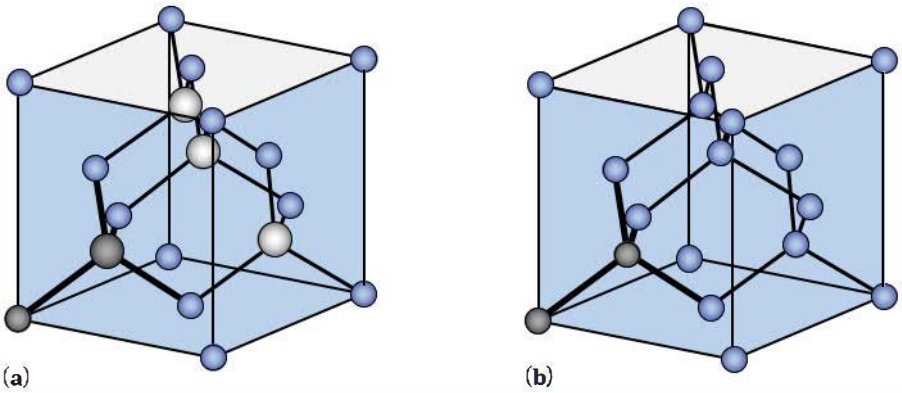


Fig. 3.30: Zinc blende and diamond structure. In both cases the atoms forming the basis are highlighted dark gray. **a)** Zinc blende. **b)** Diamond. Although the diamond lattice only consists of identical atoms, the basis is also bi-atomic.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

cubic crystal system, the three-fold, rather than the four-fold, rotary axes represent the decisive point symmetries.

At this point the question arises: Why are the structures of NaCl, CsCl and ZnS different? In order to answer this question, we first have to briefly discuss the term "*ionic radius*": In ionic crystals, the distance between two adjacent ions is determined by the sum of the radii r_A and r_B. Since the force between the ions under consideration is independent of direction, it is to be expected that the ions will be packed as densely as possible. However, there is one important limitation: The ion-ion contacts must be with ions of opposite charges, otherwise the binding energy is greatly reduced. This problem occurs when the ratio r_A/r_B becomes too large.

For the sodium chloride structure, the critical ratio of r_A/r_B is illustrated in Fig. 3.31. The arrangement of the ions in the plane parallel to the cube surfaces depicted in Figure 3.29 is shown. As can be seen from the figure, all the neighbors in this plane can just touch each other if the ratio r_A/r_B is given by

$$\frac{r_A}{r_B} = \frac{1}{\sqrt{2}-1} = 2.414\,. \tag{3.13}$$

If this value is exceeded, the oppositely charged ions can no longer touch each other. Crystals with higher values of r_A/r_B therefore have the less densely packed zinc blende structure. If, on the other hand

$$\frac{r_A}{r_B} < \frac{1}{\sqrt{3}-1} = 1.366\,, \tag{3.14}$$

the caesium chloride structure is typically preferred. However, it should be noted that this is only a rough guide.

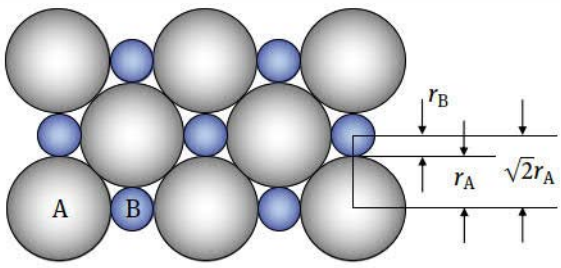


Fig. 3.31: Schematic illustration for calculating the critical ratio of ionic radii in ionic crystals. At the radius ratio depicted the transition between the sodium chloride and zinc blende structures takes place.

Hexagonal Close-packing of Identical Spheres. In addition to the face-centered cubic lattice just discussed, there is another arrangement of identical hard spheres, also leading to the highest possible packing density of 0.74. Figure 3.32 illustrates the two different ways of stacking the planes to create these structures. We take close-packed planes with hexagonal symmetry, which pack the spheres as closely as possible in a single plane. We place the first layer A. A second hexagonal layer B is placed

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

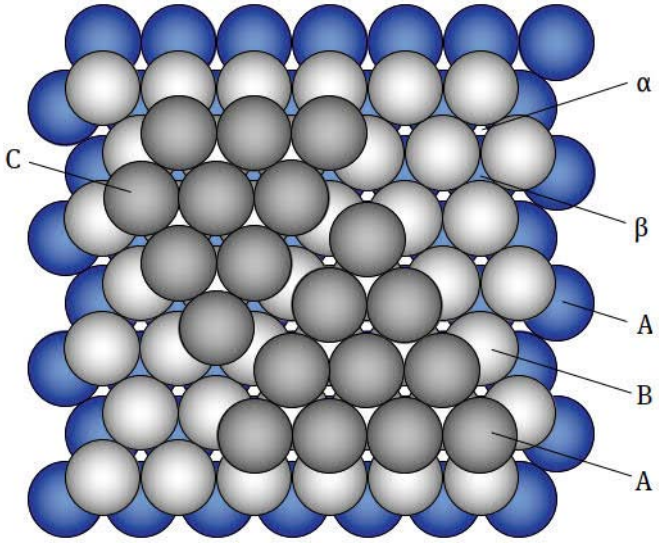


Fig. 3.32: Close-packing of identical spheres. The lowest layer A is blue, the middle layer B, light grey. The holes α in the second layer, sit above holes in the first layer, whereas the holes β are positioned above the centers of the spheres of the layer A. The atoms of the third layer (dark grey) can be either positioned over the holes α (shown left), or over the holes β (shown right). The stacking sequence ABC... (left) leads to the cubic structure, the sequence ABAB... (right) to the hexagonal structure.

on top with the spheres placed above the holes in the first layer, every second hole accommodating a sphere. There are now two possible arrangements for the third layer. As indicated in Figure 3.32, we can either place the third layer spheres over the holes α *or* over the holes β in the second layer, but the two alternatives are not equivalent regarding their position with respect to layer A. If we select the holes α, then the layer C is shifted with respect to both layer A and B. This stacking order ABC**ABC**ABC ... results in the already discussed face-centered cubic lattice. The planes represent a cut through the cubic crystal, which is defined e.g. by the lattice points (1, 0, 0), (0, 1, 0) and (0, 0, 1). As we will learn in the next chapter, these planes carry the notation (111).

If, on the other hand, on placing the third-layer spheres, we select the depressions β immediatedly above the spheres of layer A, then this layer is identical to layer A. In other words the third layer is the same as the first layer. This stacking sequence ABA**B**AB ... leads to the hexagonal close packed structure (hcp for "hexagonal close packed"). The following elements crystallize in the hexagonal structure: Be, Cd, Dy, Er, Gd, Hf, Ho, La, Lu, Mg, Nd, Os, Pm, Pr, Re, Ru, Sc, Tb, Tc, Tm, Ti, Y, Zn and Zr. The distance and number of the nearest and second nearest neighbors are identical with the ones of the face-centered cubic lattice. In consequence, the binding energies of the two structures hardly differ from each other.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 3.33 shows the unit cell of crystals with hexagonal close-packed structure. The blue prism with thickly drawn contours indicates the primitive unit cell with angle $\alpha = 120^0$ and the axes $a = b$, $c = a\sqrt{8/3} \approx 1.633\,a$. For real crystals, the ratio c/a usually deviates slightly from this ideal value. The atoms or molecules forming the basis are situated at $(0,0,0)$ and $(\frac{2}{3}, \frac{1}{3}, \frac{1}{2})$. The $\bar{6}$-fold rotational symmetry only becomes clearly visible through the joint representation of three primitive unit cells.

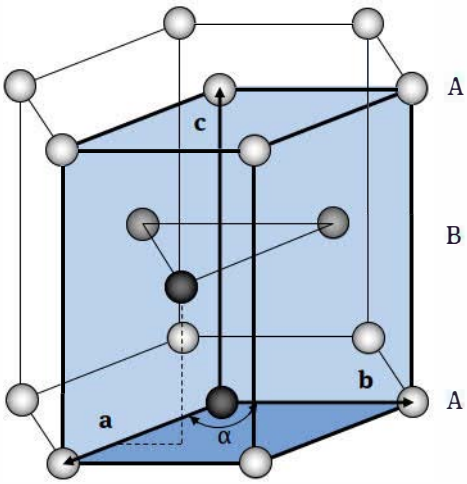


Fig. 3.33: Unit cell of the hexagonal closest-packed structure. The primitive unit cell (tinted blue) contains two lattice points (drawn in black). The symmetry of the structure becomes most obvious by joining three primitive unit cells. The layers denoted by A, B and A correspond to the layers in Figure 3.32, labeled in the same way.

3.3.5 The Wigner-Seitz Cell

The unit cells we have considered so far have taken the form of parallelepipeds with edges defined by the basis vectors. This choice proves to be convenient for representating lattice periodicity or for the Fourier analysis of the crystal structure as discussed in Section 4.3 below. However, if we are dealing with, for example, the distribution of electrons in a primitive unit cell, the computational effort can be considerably reduced if the lattice point coincides with the center of the unit cell. A primitive unit cell that meets this requirement is the **Wigner-Seitz cell**. This comprises the volume which is closer to the selected lattice point than to any other.

The construction principle is depicted in Figure 3.34a for the two-dimensional case. The selected lattice point is first connected by straight lines to the adjacent lattice points. Then at the midpoint the perpendicular is drawn. Clearly, the cell constructed in this way, not only has the symmetry of the lattice, but also gives complete coverage of the area. In two dimensions, unless the lattice is rectangular, the Wigner-Seitz cell is always a hexagon (regular or non-regular). In the case of three-dimensional lattices, a plane perpendicular to the connecting line is drawn at the center of the connecting line.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The Wigner-Seitz cell is then the polyhedron with the smallest volume that encloses the lattice point. The unit cell constructed in this way is primitive and has full lattice symmetry. To give a three-dimensional example, Figure 3.34b shows the Wigner-Seitz cell of the body-centered cubic lattice. The hexagonal surfaces are shared with the adjoining cells containing the nearest neighbors and the squares with those containing the second nearest neighbors.

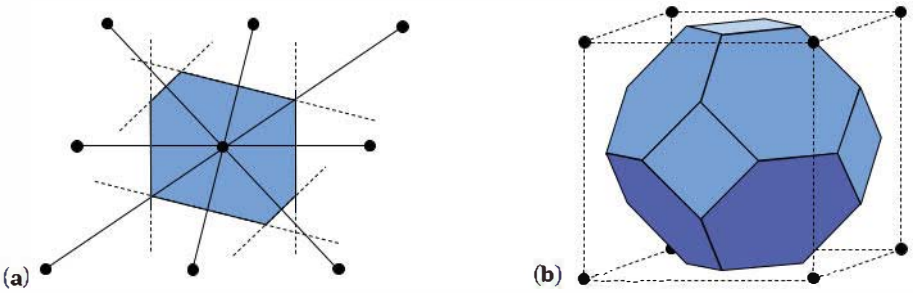


Fig. 3.34: The Wigner-Seitz cell. **a)** Illustration of the construction principle for a plane oblique lattice. Connecting lines are drawn from our central lattice point to each neighboring lattice points and at their midpoint perpendiculars are drawn. The enclosed area represents the Wigner-Seitz cell of this lattice. **b)** The Wigner-Seitz cell of the body-centered cubic lattice. The cubic unit cell is drawn with dashed lines, omitting the lattice point in the center of the cell.

3.3.6 Nanotubes

Before going on to talk about the surfaces of solids, let us briefly consider the structure of *carbon nanotubes*. Large graphene layers, briefly described in Section 2.4, are not mechanically stable. In addition, unsaturated bonds arise at the edges of the layers. In contrast, curved, closed structures such as fullerenes or nanotubes are mechanically and chemically much more stable. As mentioned in the previous chapter, nanotubes are tubular solids. While the diameters of the tubes can fall in the range of 1 – 50 nm, the lengths can exceed one centimeter. The tubes are closed at the ends by a hemispherical arrangement of carbon atoms, but we will not go into this detail here.

When a tube is formed from a graphene layer, the "honeycombs" of the graphene can be arranged in various ways with respect to the tube axes. Surprisingly, the orientation of the combs has a substantial influence on the electrical properties of nanotubes. Referring to Figure 3.35 we see the orientation of the two tube types showing both a particularly high symmetry.

Rolling a graphene layer such that the tube axis corresponds to the vertical direction of the layer shown in Figure 3.35, i.e. the lower end of the tube runs horizontally, the tube exhibits the so-called "zigzag structure". A sketch of such a tube is shown in

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

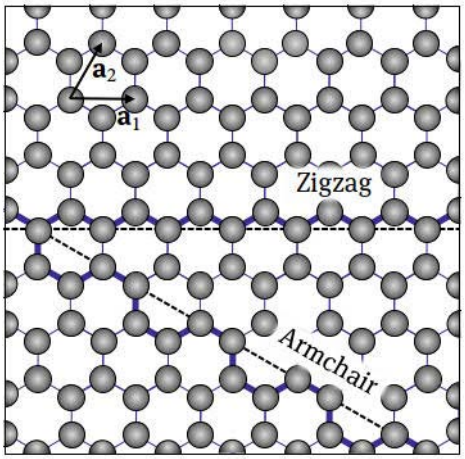


Fig. 3.35: The hexagonal graphene lattice. In addition to the basis vectors, the two directions leading to the armchair or zigzag structure are indicated by dashed lines.

Figure 3.36a. The second high-symmetry structure is the so-called "armchair structure". The axis and edge of this tube type are tilted from the zigzag structure by 30°. The resulting tube is shown in Figure 3.36b. In the first case, the axis of the tube runs parallel to one side of the hexagonal honeycomb, in the second case perpendicular to it. As already mentioned, the two tube types show very different electrical conduction properties: tubes with armchair structure show metallic properties, and tubes with zigzag structure are generally semiconducting.

In fact, the two structures described are special limiting cases. Usually there is no preferred angle between the cylinder axis and the hexagonal edges. In this case the hexagons run spirally around the tube axis forming a "chiral structure". Tubes of this type can have either semiconducting or metallic properties. The origin of the different behaviors is discussed in Section 8.5 below.

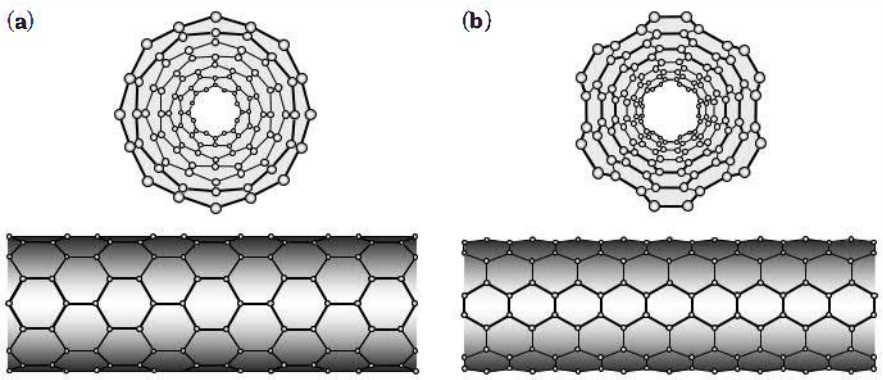


Fig. 3.36: Cross-section (top) and side view (bottom) of carbon nanotubes with **a)** zigzag or **b)** armchair structure.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

3.3.7 Surfaces of Solids

At this point, before we deal with the structure of disordered solids, we will briefly touch on the structure of crystalline surfaces. The "surface" is the name we give to the outermost atomic layers, about three layers deep, whose physical properties often differ significantly from those of the bulk. Ideally the surface is completely clean, hut under normal experimental conditions it is usually contaminated by impurity atoms, which may form thin layers or even be incorporated into the surface.

Even if the surface is free of impurities, the distance between the outermost atomic layers is generally changed. In particular, the distance *between* the uppermost layers is often significantly reduced, since the attractive forces are predominantly directed towards the bulk solid. On the other hand, in some cases the distance between the outer lattice planes may in fact increase. These changes in the structure of the surface, which are particularly pronounced in metals, are known as **surface relaxation**.

Almost invariably with non-metals, and sometimes with metals, a further change takes place, namely **surface reconstruction**. The abrupt discontinuity at the surface gives rise to unsaturated covalent or ionic bonds whose energy can be reduced by the formation of *superstructures*. Here the atoms arrange themselves in rows with alternately larger and smaller separations compared with those in the bulk volume, facilitating more bonds to neighboring atoms reducing the number of dangling bonds. Dangling bonds on the surface can often also be saturated by twisting the molecules or the direction of the bond, without needing the formation of superstructures.

When describing the structure of surfaces, instead of a lattice, it is usual to speak of a mesh. In this case, the unit cell is usually referred to as a *unit mesh*. There are *five Bravais meshes* in two dimensions: the oblique mesh, the rectangular mesh, the rectangular-centered mesh, the hexagonal mesh and the square mesh. Clearly, the situation in two dimensions is clearer than that in three: While there are 32 point and 230 space groups for three dimensions, there are only 10 point and 17 space groups in two dimensions.

The mesh of the unperturbed crystal, lying parallel to the surface, is taken as the reference mesh when describing the surface structure. The common notation describing surface meshes is illustrated in Figure 3.37. As in case of bulk samples, there are also two notations used in parallel, namely the matrix notation according to *P. L. Park*[23] and *H.H. Madden*[24] (1968) and the short notation of *E.A. Wood*[25] (1964), which we use here. In this figure, the somewhat larger atoms of the underlying solid body, the *substrate*, are shown covered by smaller, regularly arranged impurity atoms. In the figure are drawn three meshes, the reference mesh, with basis vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, and two surface

23 Robert Lee Park, *1931 Kansas City, †2020 Portland
24 Hannibal H. Madden, *1931, †2003 Port Angeles, USA
25 Elizabeth Armstrong Wood, *1912 New York City, †2006 Freehold, New Jersey

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

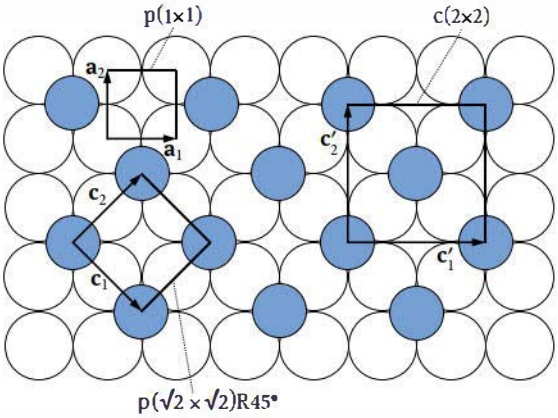


Fig. 3.37: The notation for surface meshes. The white circles represent the uppermost atomic layer of the substrate, on which the blue marked impurity atoms are arranged. The mesh p(1 × 1) is the reference mesh. The mesh $(\sqrt{2} \times \sqrt{2})$R45° is tilted by 45°, c(2 × 2) is a centered mesh.

meshes with basis vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ and $\mathbf{c}_1'$ and $\mathbf{c}_2'$. The mesh notation takes the form:

$$\left(\frac{c_1}{a_1} \times \frac{c_2}{a_2}\right) \mathrm{R}\alpha\,, \tag{3.15}$$

where Rα indicates the rotation of the mesh concerned with respect to the reference mesh. This latter is omitted if the two meshes are not tilted with respect to each other. The character "p" before the parenthesis indicates a primitive mesh, and "c" a centered mesh.

For the full characterization of impurity atoms on the surface, further information is required. The chemical formula of the substrate and the notation for the substrate surface are given, and at the end the chemical symbol of the impurity atoms. For example, the abbreviation Si(111)$(\sqrt{3} \times \sqrt{3})$R30°-Ag indicates that silver atoms are adsorbed on a (111)-surface of silicon, that the surface atom mesh is tilted by 30° from the substrate mesh and that the lattice constant is enlarged by a factor of $\sqrt{3}$. The notation (111) for the silicon surface will be discussed in detail in Section 4.3.

3.4 The Structure of Amorphous Solids

In an ideal crystal, the positions of all atoms are given exactly, since there is only one well-defined configuration. In other words, the coordinates of all the atoms can be given once the size and composition of a unit cell and the lattice structure are known. However, in amorphous solids the positions of the individual atoms is not easly characterized due to the lack of any periodicity. For an unambiguous indication of the atomic positions a suitable list would have to be established with an entry for *each* atom. Since this is not possible in practice, an amorphous structure can only be described with statistical methods thus in a much less detailed form than in crystals.

A simple, but important piece of information is that of the particle number density $n(\mathbf{r})$, which indicates how many particle centers are located within a volume ele-

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

ment dV at the position $\mathbf{r}$. The mean value $n_0 = \langle n(\mathbf{r}) \rangle = N/V$, where N stands for the number of atoms and V for the sample volume, represents the average particle density, which immediately allows our first conclusions about the structure. Compared to crystals, amorphous solids are less densely packed: average particle number densities and average mass densities are typically $1 - 10\,\%$ smaller than those in the corresponding crystals.

3.4.1 The Pair Correlation Function

Information about the local arrangement of the atoms can be obtained from the **pair distribution function** or **pair correlation function**. Although the definition does not distinguish between crystalline and amorphous solids, it only plays a significant role in understanding the structure of the latter. This function can be determined by diffraction experiments (see Section 4.5) or by simulations using the knowledge of the interaction potentials of the atoms.

If we assume for simplicity that the solid under consideration is composed of only *one* type of atom, then the pair correlation function reflects the probability that if an atom is already present at position $\mathbf{r}_1$ then a second atom will be located at $\mathbf{r}_2$. Formally, the pair distribution function $g(\mathbf{r}_1, \mathbf{r}_2)$ can be defined by the expectation value of the particle number densities $n(\mathbf{r}_1)$:

$$g(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{n_0^2} \langle n(\mathbf{r}_1)\, n(\mathbf{r}_2) \rangle \,. \tag{3.16}$$

The pair distribution function is normalized using the mean particle number density n_0 to yield the value one for large distances. It should also be noted that, by definition, since there is already a particle at $\mathbf{r}_1$, the probability of finding a second particle at this point must therefore vanish, since two atoms cannot be sited at the same location simultaneously.

Although, in an amorphous solid, the environment of each individual atom looks different, from a macroscopic point of view these materials are generally homogeneous and isotropic. Therefore, the pair distribution function will depend only on the distance $\mathbf{r} = (\mathbf{r}_2 - \mathbf{r}_1)$ between the two locations considered. Therefore, it reflects the particle number density found on average at a distance $r = |\mathbf{r}|$ from a chosen reference atom independent of direction, i.e.,

$$g(r) = \frac{n(r)}{n_0} \,. \tag{3.17}$$

To familiarize ourselves with this concept, we first take a look at the pair distribution function for a one-dimensional system. We randomly distribute N identical hard spheres with diameter d over a length L. The form of the distribution function depends crucially on the mean space between the spheres given by the parameter $\ell = (L - Nd)/N$.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

We arbitrarily pick out one sphere and determine the distances to the centers of the other spheres. If we average over many measurements, in other words, starting from different spheres, after normalization by the mean particle number density n_0 we obtain the pair distribution function, which is shown in Fig. 3.38 for various cases.

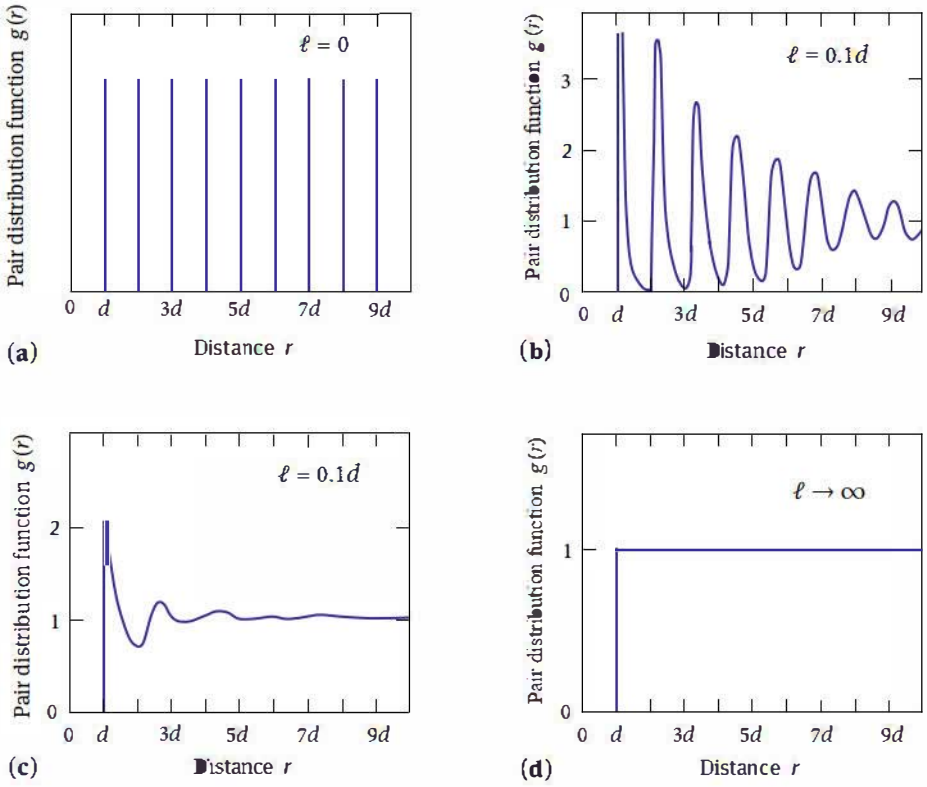


Fig. 3.38: The pair distribution function of a linear system with randomly distributed hard spheres. **a)** One-dimensional crystal ($\ell = 0$), **b)** liquid with $\ell = d/10$, **c)** liquid with $\ell = d/2$, **d)** diluted gas ($\ell \to \infty$). There is no vertical scale indicated for a) since the delta functions at integral values of d extent to infinity.

For the limiting cases $\ell \to 0$ and $\ell \to \infty$ the pair distribution function can be seen simply by inspection. In the first case, there is no space at all between the balls, which then lie tightly packed next to each other as in a one-dimensional crystal. The pair correlation function must be periodic and is non-zero only at the positions $x = md$ with $m = 1, 2 \ldots$ assuming the form of delta functions. In the second case, the mean distance between the spheres is large in relation to their diameter. We could therefore think of a dilute, one-dimensional gas. In this case, the probability of finding the center of a sphere in a given line element dr is relatively small and independent of the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

distance r to the selected sphere. In this case, the normalized pair distribution function is the same for all distances with the value 1. Of course, this value can only be found for $r \geq d$, since hard spheres can only approach this minimum distance. If we apply these arguments to the more general case of, say, a rigid random arrangement of the balls, i.e. that of an amorphous one-dimensional solid, or even of a liquid, the pair distribution function will lie somewhere between these two limiting cases. Depending on the value of parameter ℓ, more or less pronounced maxima occur, becoming increasingly blurred with increasing distance from the reference atom, due to averaging over many configurations (cf. Fig. 3.38b and Fig. 3.38c). The maxima at short distances indicate local order, which becomes more pronounced the closer the spheres are packed.

Figure 3.39 shows schematically the relationship between structure and particle density for a two-dimensional system. The sketch clearly shows that the nearest neighbors are responsible for the first maximum which is therefore to be found at $r \approx d$. Further analysis shows, as expected, that the area below the maximum is determined by the number of nearest neighbors. The next outward *coordination shell*, i.e. the second nearest neighbors, causes the second maximum and so on. As the distance increases, the maxima and minima of the particle number density and the pair correlation function become increasingly broadened due to the lack of positional order. At very large distances, the value n_0 or $g(r) = 1$ is finally reached. The same reasoning also applies to three-dimensional samples.

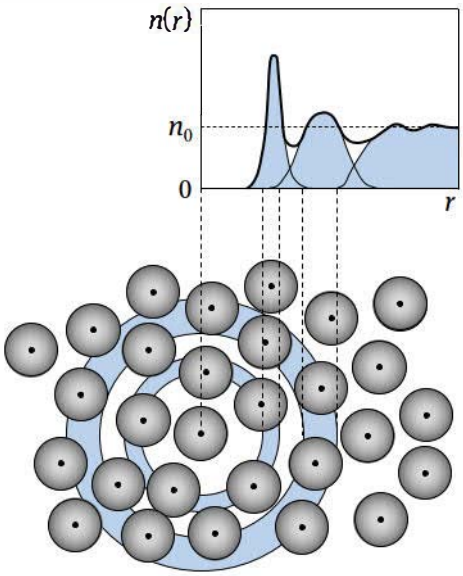


Fig. 3.39: An illustration of the relationship between structure and particle density in the two-dimensional case. As with the pair distribution function, the contribution of the starting atom is omitted.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

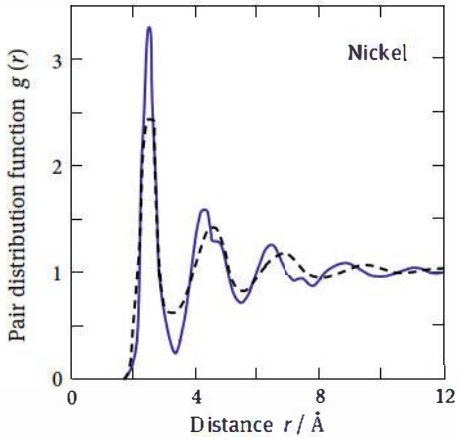


Fig. 3.40: Experimentally determined pair distribution function of amorphous (blue solid line) and liquid (black dashed line) nickel. (After Y. Waseda, *Structure of Non-Crystalline Materials*, McGraw-Hill, 1980.)

In Figure 3.40, the pair distribution function of an amorphous nickel film, determined by diffraction experiments (cf. Section 4.5), is compared with that of molten nickel. As can be clearly seen, the short-range order is much more pronounced in the amorphous phase than in the liquid. In both cases, the order decreases rapidly with the distance from the reference atom. The description of the structure of real amorphous solids is considerably simplified by the fact that only the distance r from the reference atom needs to be taken into account since no directional dependencies have to be considered. Of course, the pair distribution function obtained for the averaged local structure of an amorphous solid contains much less information than that used for crystalline solids.

3.5 Exercises and Problems

1. Ag-Cu Alloys. Figure 3.41 shows the phase diagram of the Ag-Cu system.

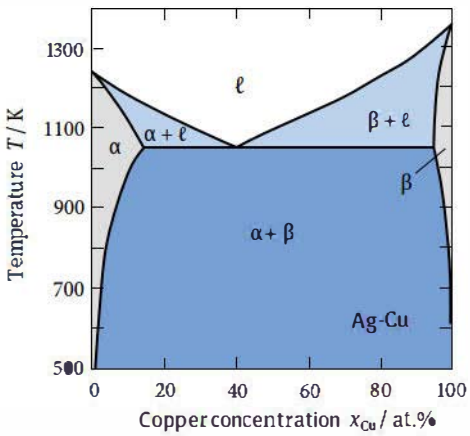


Fig. 3.41: Phase diagram of the Ag-Cu system. (After F. Goodwin et al., *Springer Handbook of Condensed Matter and Material Data*, W. Martienssens, H. Warlimont, eds., Springer, 2005.)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

(a) Read the copper concentration of the α-phase at 1100 K and at a temperature just below the eutectic temperature.
(b) Assuming a homogeneous melt of the Ag-Cu system containing 60 at.% copper, as the system cools between these two temperatures, name the phases that it passes through.
(c) What is the mixing ratio of the α- to β-phase for the specified alloy just below the eutectic temperature of 1050 K? How would this ratio change during further cooling?

2. Hypothetical Two-dimensional Lattice. Figure 3.42 shows a two-dimensional lattice. The spherical atoms have atomic radii $r_A = 2.0\,\text{Å}$ and $r_B = 0.8\,\text{Å}$, and atomic masses $m_A = 39u$ and $m_B = 12u$.
(a) Which point lattice describes completely the translation symmetry of the lattice? Specify the corresponding basis vectors.
(b) What would be the chemical formula and areal density (in g/cm^2) for this crystal?
(c) The atoms of a primitive unit cell form the basis of the crystal. Choose a basis with the highest possible symmetry and state the coordinates of the basis atoms.
(d) For which symmetries do the point lattice and basis match?
(e) Find a non-primitive unit cell with orthogonal basis vectors. Specify the basis vectors and the coordinates of the basis atoms.

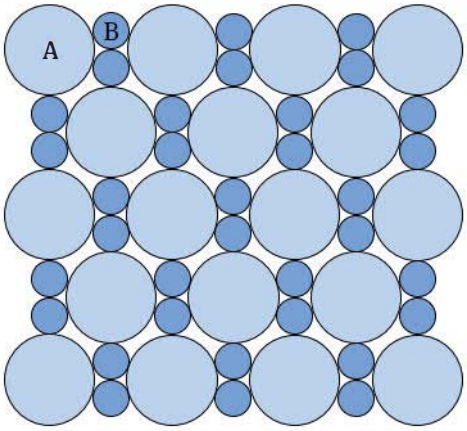


Fig. 3.42: A hypothetical two-dimensional crystal.

3. Allotropy of Iron. Many materials exist in several crystal structures. As mentioned in Section 2.4, this property is known as allotropy or polymorphism. Allotropic phase transitions usually occur during pressure or temperature changes. In iron, the transition from a body-centered cubic lattice (α-phase) to a face-centered cubic lattice (γ-phase) takes place at 1183K. Assuming that the distance between the nearest

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

neighbors does not change, calculate the relative volume change that occurs during this phase transition.

4. Cubic Crystals. The lattices of α-iron and copper are body-centered and face-centered cubic, respectively, with corresponding densities of $7.86\,g/cm^3$ (Fe) and $8.96\,g/cm^3$ (Cu). Take the atoms as hard spheres touching their nearest neighbors.
(a) Calculate the lattice constants of these metals in units of their atomic radii.
(b) What is the volume of the primitive unit cells in units of atomic radii?
(c) Calculate the lattice constants from the densities of the two metals.

5. The Phase Transition of Tin. Metallic tin (β-tin) transforms into a semiconducting phase (α-tin) below 286.4 K. This transformation is known as "tin pest". The densities of the two phases differ considerably, since $\varrho_\alpha = 5.77\,g/cm^3$ and $\varrho_\beta = 7.29\,g/cm^3$. While β-tin has a body-centered tetragonal lattice ($a = 5.83$ Å, $c/a = 0.545$), the lattice of α-tin is face-centered cubic ($a = 6.49$ Å). How many atoms make up the basis in each case?

6. Carbon Lattices. It is well known that carbon occurs in various modifications. Calculate the mass density of diamond, graphite and fullerene. Use the lattice constants $a = 3.57$ Å for diamond, $a = 2.46$ Å and $c = 6.71$ Å for hexagonal graphite and $a = 14.17$ Å for fullerene C_{60}-molecules packed to form a crystal lattice with a face-centered cubic unit cell as shown in Figure 3.43.

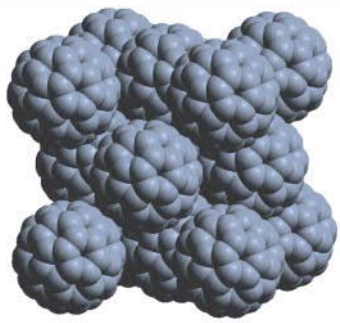


Fig. 3.43: Face-centered cubic crystal of C_{60}-molecules. (After B. Pevzner, www.godunov.com/bucky/fullerene.html.)

7. The Structure of Noble Gas Crystals. The noble gases form face-centered cubic crystals. Here we consider xenon and mistakenly assume that xenon crystals have a body-centered cubic lattice. With this assumption, calculate the nearest-neighbor distance. Determine the binding energy for this hypothetical crystal and compare it with the actual value.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

8. Hexagonal Lattice. Count the symmetry axes and symmetry planes of the hexagonal lattice. Use the non-primitive unit cell as shown in Figures 3.19 and 3.33.

9. Nanotubes. A 1 mm long carbon nanotube with armchair structure has a diameter of about 2.2 nm. How many atoms does it contain and what is the mass of the tube? Do a zigzag tube with the same dimensions give the same result? Why can the atoms that form the ends of the tubes be neglected in answering the question?

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

4 Structure Determination

In the investigation of the structure of solids, a number of aspects need to be considered. Of course, of particular importance is the structure itself, i.e. the structure of the lattice and the distribution of the atoms in the unit cells. However, in many cases, the question of deviations from the ideal structure also becomes significant. In this chapter, we will limit ourselves to the first aspect, the basic structure, the elucidation of which is provided by **diffraction** or **scattering experiments** . Since there is generally no clear distinction between the usage of the two terms in solid state physics, we will use both terms interchangeably in the following. For studying deviations from the ideal structure, i.e. the details of the local arrangement of atoms or defect structures, other methods are better suited. These include spectroscopic methods, the Mössbauer[1] effect and investigations based on electron or nuclear magnetic resonance. However, we will not discuss here these interesting and very successful measurement methods in detail.

After introductory remarks on the direct imaging of atomic structures and on the basics of scattering experiments, we derive in Section 3.2 an elementary theory of diffraction. We will see that the evaluation of scattering experiments on periodic structures is greatly simplified by introducing the concept of the *reciprocal lattice* which plays a very important role in solid state physics and will be widely used in the remaining chapters. Next, we discuss the structure determination of amorphous solids and finally we introduce the most commonly used experimental methods for structure determination.

4.1 General Remarks

The direct imaging of atoms allows direct access to the atomic structure of solids. This possibility became available in the latter part of the last century when atomic resolution of surface structures could be achieved by electron, scanning-tunneling or atomic-force microscopy. At the end of this chapter we will briefly discuss how these techniques work. As an example of direct imaging, Figure 4.1 shows a section through a thin film where silicon and germanium layers alternate. This high-resolution image was taken using transmission electron microscopy (TEM). The germanium atoms appear darker than the silicon atoms owing to their different imaging properties. In the section shown, 13 atomic layers of silicon alternate with two atomic layers of germanium. Such a regular arrangement of layers with different atomic compositions is known as a *superlattice*. The production of such layer structures has only become possible in recent years. We will discuss some of the interesting properties of these novel solid-state structures in Section 10.4.

1 Rudolf Ludwig Mößbauer, *1929 Munich, †2011 Grünwald (Munich)

https://doi.org/10.1515/9783110666502-004

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

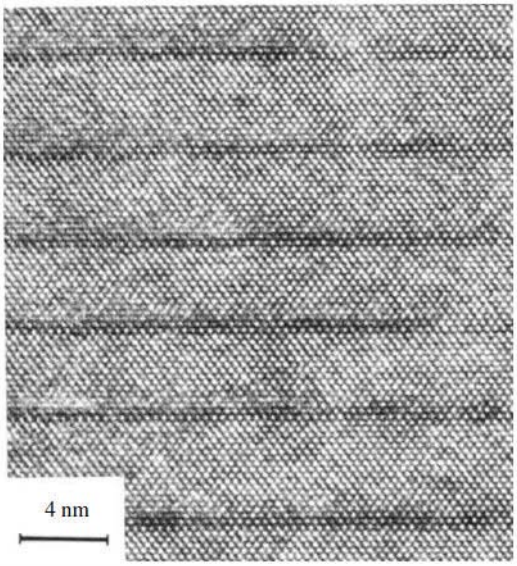


Fig. 4.1: A silicon-germanium superlattice. The image was taken with a high-resolution electron microscope in transmission. The individual atoms are clearly visible. Two atomic layers of germanium alternate with 13 atomic layers of silicon (appearing somewhat lighter in the image). (After E. Müller et al., Phys. Rev. Lett. **63**, 1819 (1989).)

The development of scanning tunneling and atomic force microscopy has given surface physics an enormous boost. Since these methods scan the profile of the sample surface, no information on the internal structure of the solid is obtained. As mentioned in the previous chapter, we often find that structures on the surface differ from those in the bulk part of the crystal, since different forces act on the atoms at the surface than act on those in the bulk. Surface imaging microscopes are therefore only of limited use for structural investigations but are of great importance for the study of surface properties. The high resolution, which can be achieved with atomic force microscopes in recent years, is demonstrated in Figure 4.2 showing the surface of a silicon crystal with atomic resolution.

Detailed knowledge on the internal structure of solids has been primarily obtained by diffraction experiments. The basic principle is simple: the sample is irradiated by

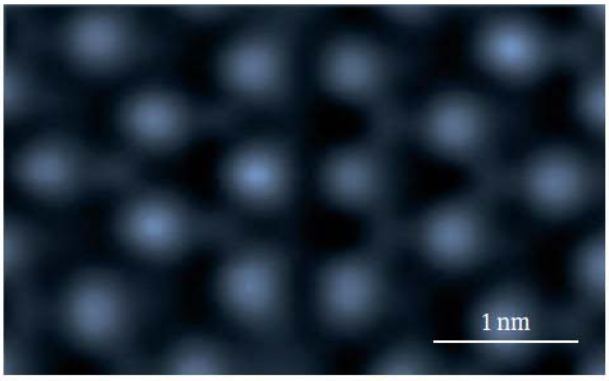


Fig. 4.2: Atomic force microscope image of the reconstructed silicon surface. On closer inspection, atoms at a deeper level can also be detected. (After M. Emmerich et al., Science **348**, 308 (2015).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

waves which are scattered by the atoms. From the resulting diffraction pattern, we can deduce the geometric arrangement of the atoms. X-rays are the most commonly used radiation for such diffraction experiments, but owing to their wave nature neutrons, electrons or light atoms can also be used. The strength of the interaction determines how far the particular radiation penetrates into the specimen and thus what fraction of the solid can be investigated. Excepting neutrons, which for non-magnetic atoms are scattered exclusively at the nuclei, the other probes interact with the electron shells. Owing to the strong Coulomb interaction between the incoming particles and the scattering atoms, electron or atomic beams are particularly suitable for the investigation of surfaces and thin films whereas X-rays and neutron beams are more suited to the investigation of massive solids.

Since we know from optics that diffraction is strongest when the light wavelength and the characteristic dimensions of the diffracted object are comparable, when determining the structure of solids, we need to ensure that the wavelength of the incident radiation is shorter than the atomic spacing, which is typically in the few Ångströms range. This means that, depending on the nature of the radiation used, particles with very different energies will satisfy this condition. To produce a wavelength of 1 Å, X-ray quanta need an energy of 12 keV, electrons 150 eV, neutrons 80 meV and helium atoms only 20 meV.

The observed scattered wave pattern is a superposition of the scattering contributions of all atoms. The atomic arrangement can thus only be reconstructed from diffraction experiments if the scattering is *coherent*, i.e. if the phases of the scattered waves emanating from the different scattering centers are correlated. In contrast, if the scattering is *incoherent*, the phases of the scattering waves are uncorrelated, and no conclusions can be drawn about the structure of the sample.

Scattering by solids can be either *elastic* and *inelastic* processes. If the electronic and vibrational states of the solid remain unchanged during the scattering process, the energy of the scattered radiation does not change. The scattering is then described as elastic. In inelastic scattering processes, however, energy is exchanged between the scattered particle and the scattering solid. Both processes play an important role in solid state investigations. Elastic scattering is used for structure determination and inelastic scattering for the investigation of excitation states.

4.2 Diffraction Experiments

In the following, we will develop a basic scattering theory, in which we consider the scattering process as a purely classical optical process. The quantum mechanical aspect appears only by the fact that the photons, neutrons, electrons and atoms are treated as waves. In deriving this, we neglect the influence of absorption, refraction, and changes in the intensity of the incident beam arising from the scattering process. Since in the derivation no specific properties of the interacting radiation are involved, the treatment

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

is applicable to all wave types. When describing the scattering of a particular radiation, the specific interaction mechanisms must, of course, be taken into account. Since, in addition to the classical method of structure determination with X-rays, structure determination with neutrons is also very important, we will discuss both experimental methods. We will keep in mind that neutron scattering provides information about the nuclear distribution, while the other scattering experiments provide information about the electron distribution.

4.2.1 Scattering Amplitude

We consider a sample with a distribution of scattering centers characterized by the **scattering density distribution** $\varrho(\mathbf{r})$. As mentioned before, in the case of X-rays, it is the electrons which act as thr scattering centers and in the case of neutron radiation it is primarily the nuclei, so that the appropriate scatter density reflects the electron or nuclear distribution respectively. If a plane wave with angular frequency ω_0, wave vector $\mathbf{k}_0$ and amplitude $A(t) = A_0 e^{-i(\omega_0 t - \mathbf{k}_0 \cdot \mathbf{r})}$ is incident on a scattering center, the scattering center becomes the starting point of a spherical wave. For the amplitude $A_c(t)$ of this scattered wave we write

$$A_c(t) = \frac{\tilde{A}}{R} e^{-i(\omega_0 t - k_0 R)} , \tag{4.1}$$

where R is the distance from the scattering center. The quantity $\tilde{A}$ reflects the scattering probability and is therefore specific both to the radiation used and to the nature of the scattering center.

The aim of scattering theory is to establish the relationship between the intensity of the scattered waves and the scattering density distribution $\varrho(\mathbf{r})$ and to derive the structure of the sample from this. The *phase difference* between the scattered waves arising from the different positions of the scattering centers and thus containing the structural information, is of crucial importance.

In an experiment, the radiation detector is typically far enough from the sample that we can treat the spherical scattered wave at the detector as approximately plane waves with wave vector $\mathbf{k}$. Referring to Figure 4.3 we see that the scattered wave generated by the depicted volume element, and the scattered wave from the origin O with position vector $\mathbf{r} = 0$, have a path difference of $\Delta s = (\mathbf{k} \cdot \mathbf{r}/k) - (\mathbf{k}_0 \cdot \mathbf{r}/k_0)$. Since the amplitude of the wave vector does not change during elastic scattering, we may set $k = k_0$, giving a phase difference $\Delta\varphi$ between the two beams of $\Delta\varphi = (\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{r}$. Thus we can express the contribution $dA_s(\mathbf{r}, t)$ of the volume element dV at location $\mathbf{r}$ to the amplitude of the scattered radiation at the detector at distance R_1 as:

$$dA_s(\mathbf{r}, t) = \varrho(\mathbf{r}) A_c \, dV = \frac{\tilde{A}}{R_1} \varrho(\mathbf{r}) e^{-i[\omega_0 t - kR_1 + (\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{r}]} \, dV , \tag{4.2}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

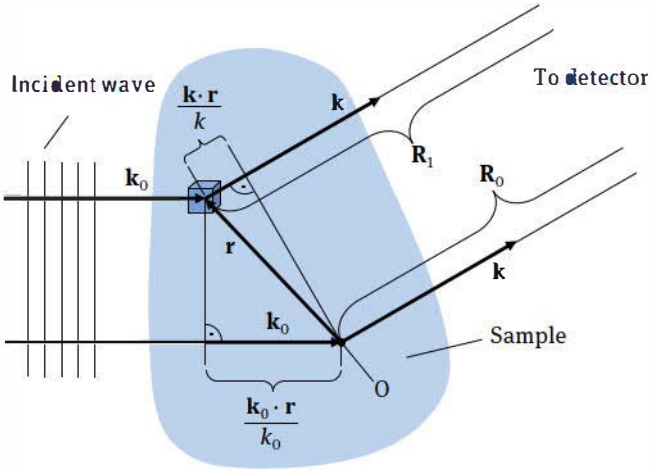


Fig. 4.3: To derive the scattering amplitude. A plane wave with the wave vector $\mathbf{k}_0$ is radiated on the sample. The wavefront of the scattered wave, which is propagating towards the far away detector, can also be seen as a plane wave characterized by the wave vector k. The vectors $\mathbf{R}_0$ and $\mathbf{R}_1$ connect the origin and the scatter volume, respectively, with the detector. The different quantities are explained in the text.

where we have neglected the possibility that the scattered wave can be repeatedly scattered in the sample, this simplification corresponding to the first Born approximation in quantum mechanical scattering theory. From the expression we obtain the total amplitude of the coherently scattered radiation at the detector by integration over the sample volume V_s. If the sample size is small compared with the detector distance, we can set $R_1 \approx R_0$ and obtain

$$A_s(t) = \frac{\widetilde{A}}{R_0} e^{-i(\omega_0 t - kR_0)} \int_{V_s} \varrho(\mathbf{r}) e^{-i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}} dV \,. \tag{4.3}$$

The quantities outside integral there are only quantities that do not depend on the position vector **r**, i.e. on the position of the scattering centers, but are determined only by the experimental setup. The information about the distribution $\varrho(\mathbf{r})$ of the scattering centers and thus about the structure of the sample under investigation is contained in the integral. This quantity

$$\mathcal{A}(\mathbf{K}) = \int_{V_s} \varrho(\mathbf{r}) e^{-i\mathbf{K}\cdot\mathbf{r}} dV \tag{4.4}$$

is called the **scattering amplitude**, where we make use of the abbreviation

$$\mathbf{K} = \mathbf{k} - \mathbf{k}_0 \tag{4.5}$$

for the **scattering vector**.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

It is clear from the form of equation (4.4) that it is the Fourier[2] transform of the scattered density distribution $\varrho(\mathbf{r})$. In principle, we can obtain the scattered density distribution $\varrho(\mathbf{r})$ (and thus the structure of the sample) from the measured scattering amplitude $\mathcal{A}(\mathbf{K})$ using the relationship

$$\varrho(\mathbf{r}) = \frac{1}{(2\pi)^3} \int \mathcal{A}(\mathbf{K})\, e^{i\mathbf{K}\cdot\mathbf{r}} d^3K \,. \tag{4.6}$$

Unfortunately, however, in diffraction experiments only the intensity of the scattered radiation can be measured, not the amplitude. This means that an important part of the information on the structure contained in the *phase* of the measurement signal is lost so that the transformation cannot be carried out directly. This leads to the *phase problem*, which we will discuss in Section. 4.4

Up to now, we have made no distinction between scattering by crystalline and by amorphous substances. However, analysis of the scattering data and the structural information so derived depends essentially on this. We will therefore deal with the two classes of material separately, first considering scattering by crystals and then dealing with scattering by amorphous samples. However, before going on to that we will first look at the Fourier description of periodic structures, which leads to the concept of the *reciprocal lattice*, since this considerably simplifies the mathematical treatment of the scattering on crystals.

4.3 The Fourier Description of Point Lattices

As discussed in Chapter 3, ideal crystals obey translational symmetry. Therefore the scattering density distribution $\varrho(\mathbf{r})$ is periodic and it is obvious to make use of this symmetry from the beginning when evaluating the scattering spectra. Therefore, we will expand the scattering density distribution into a Fourier series and investigate the individual Fourier components separately. As we will see, the Fourier expansion of crystal lattices leads to several interesting and important concepts, from which we will stress the *reciprocal lattice*, the *Brillouin zone*, and the *Miller indices*. After discussing these concepts we will return to our original problem, the determination of the crystal structure by diffraction experiments.

4.3.1 The Reciprocal Lattice

In the following we expand the three-dimensional periodic scattering density distribution $\varrho(\mathbf{r})$ into a Fourier series, using the fact that crystals have translational symmetry.

2 Jean-Baptiste-Joseph Fourier, *1768 Auxerre, †1830 Paris

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The Fourier expansion has the form

$$\varrho(\mathbf{r}) = \sum_{h,k,l} \varrho_{hkl}\, e^{i\mathbf{G}_{hkl}\cdot\mathbf{r}}\,, \tag{4.7}$$

where h, k, and l are independent integers. As usual, the Fourier coefficients ϱ_{hkl} are given by the equation

$$\varrho_{hkl} = \frac{1}{V_c}\int_{V_c} \varrho(\mathbf{r})\, e^{-i\mathbf{G}_{hkl}\cdot\mathbf{r}}\, dV\,. \tag{4.8}$$

The integration extends over the period of the function, which in our case means over the volume V_c of the primitive unit cell.

We write the vector $\mathbf{G}_{hkl}$ in the form:

$$\mathbf{G}_{hkl} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3 \tag{4.9}$$

where the vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$ define a new oblique coordinate system which we still need to determine. Since h, k and l may only take discrete values, each vector $\mathbf{G}_{hkl}$ represents a point on a new lattice, which is called the **reciprocal lattice**. The choice of the new coordinate system with basis vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$, and thus also the lattice they form, is subject to restrictions which result from the requirement of translational invariance of the corresponding lattice in real space.

The periodicity of the scattered density distribution $\varrho(\mathbf{r})$ can be expressed by the equation:

$$\varrho(\mathbf{r}) = \varrho(\mathbf{r} + \mathbf{R}) \tag{4.10}$$

where $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ (see equation (3.12)) is a lattice vector of the crystal. To simplify the notation in this chapter, we will use the terms $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ for the basis vectors rather than $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$.

While the Fourier coefficients ϱ_{hkl} are not position dependent, but the scattered density distribution $\varrho(\mathbf{r})$ according to (4.10) must have translational invariance, the position-dependent factor $\exp(i\mathbf{G}_{hkl}\cdot\mathbf{r})$ must also reflect this periodicity, i.e. the equation

$$\exp(i\mathbf{G}_{hkl}\cdot\mathbf{r}) = \exp[i\mathbf{G}_{hkl}\cdot(\mathbf{r}+\mathbf{R})] \tag{4.11}$$

must hold. This is only possible if $\exp(i\mathbf{G}_{hkl}\cdot\mathbf{R}) = 1$ or $\mathbf{G}_{hkl}\cdot\mathbf{R} = 2\pi m$, where m is an integer. If we expand the dot product, we see immediately that this condition is only fulfilled for arbitrary coefficients n_i of the lattice vector and for arbitrary h, k, l if the relation

$$\mathbf{b}_i\cdot\mathbf{a}_j = 2\pi\,\delta_{ij} \tag{4.12}$$

holds, where δ_{ij} is the Kronecker[3] delta. This means that the basis vectors of the reciprocal lattice are directly given by

$$\mathbf{b}_1 = \frac{2\pi}{V_c}(\mathbf{a}_2\times\mathbf{a}_3), \qquad \mathbf{b}_2 = \frac{2\pi}{V_c}(\mathbf{a}_3\times\mathbf{a}_1) \qquad \text{and} \qquad \mathbf{b}_3 = \frac{2\pi}{V_c}(\mathbf{a}_1\times\mathbf{a}_2)\,. \tag{4.13}$$

3 Leopold Kronecker, *1823 Liegnitz, †1891 Berlin

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The volume $V_c = (\mathbf{a}_1 \times \mathbf{a}_2) \cdot \mathbf{a}_3$ of the unit cell of the lattice in real space is related to the volume of the unit cell $(\mathbf{b}_1 \times \mathbf{b}_2) \cdot \mathbf{b}_3$ of the reciprocal lattice by the following relation:

$$(\mathbf{b}_1 \times \mathbf{b}_2) \cdot \mathbf{b}_3 = \frac{(2\pi)^3}{V_c} . \tag{4.14}$$

Equations (4.12) and (4.13) uniquely link both lattices. At this point, we should emphasize again that a *particular* reciprocal lattice vector G_{hkl} corresponds unambiguously to a *specific* Fourier coefficient ϱ_{hkl} of the scattering density distribution.

The vectors of the reciprocal lattice have dimensions of inverse length, i.e. they do not exist in the familiar **real space**, but rather in **k-space**, where **k** stands for the wave vector (which also has dimensions of inverse length). Since the momentum of a particle is given by $\hbar\mathbf{k}$, this space is also referred to as **momentum space**. In illustrations of the reciprocal lattice, we will usually label the axes with the symbol k, which we reserve for the wave vector or wave number.

To illustrate the relation between real and reciprocal lattice we will use some simple examples. For a *linear chain* with lattice constant a, the reciprocal lattice is also linear with lattice constant $2\pi/a$ according to (4.12). For an oblique *plane lattice*, the relation is illustrated in Fig. 4.4. The basis vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ of the reciprocal lattice lie perpendicular to the basis vectors $\mathbf{a}_2$ and $\mathbf{a}_1$ of the original lattice.

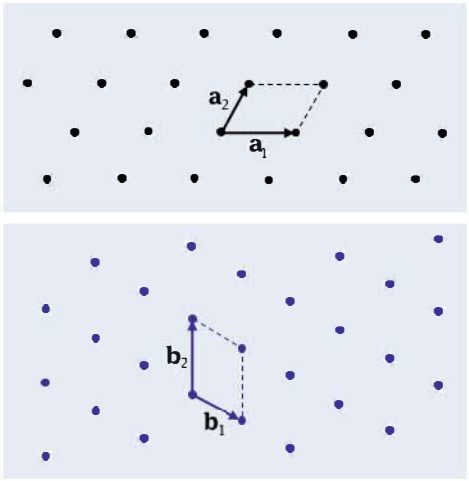


Fig. 4.4: Relationship between the real and reciprocal lattice. In the upper figure the original lattice is shown in black, in the lower figure the corresponding reciprocal lattice is shown in blue. The basis vector $\mathbf{b}_1$ of the reciprocal lattice is perpendicular to $\mathbf{a}_2$ and $\mathbf{b}_2$ is perpendicular to $\mathbf{a}_1$.

The unit cell of the *simple cubic lattice* in real and reciprocal space is shown in Figure 4.5. If $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are the unit vectors of the Cartesian coordinate system and a is the lattice constant, then $\mathbf{a}_1 = a\hat{\mathbf{x}}$, $\mathbf{a}_2 = a\hat{\mathbf{y}}$, and $\mathbf{a}_3 = a\hat{\mathbf{z}}$ define the real lattice. With $(\mathbf{a}_1 \times \mathbf{a}_2) \cdot \mathbf{a}_3 = a^3$ we find for the basis vectors of the reciprocal lattice: $\mathbf{b}_1 = 2\pi(\mathbf{a}_2 \times \mathbf{a}_3)/a^3 = 2\pi\hat{\mathbf{x}}/a$ with corresponding expressions for $\mathbf{b}_2$ and $\mathbf{b}_3$. The reciprocal lattice of the simple cubic lattice is thus also simple cubic. The lattice constant is $b = 2\pi/a$ and the volume of the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

primitive unit cell $(2\pi/a)^3$. It follows from the considerations above that in representations such as those in Fig. 4.5, two different spaces, real space and momentum space, are drawn one inside the other. Therefore caution must be taken in interpretation.

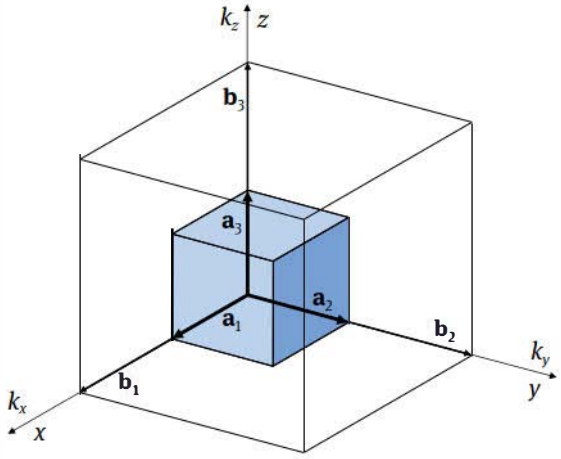


Fig. 4.5: Unit cell of simple cubic lattice in real (blue-tinted) and the reciprocal space with lattice constants a and $b = 2\pi/a$, respectively.

It is not quite so straightforward to establish this relationship with the non-simple cubic lattices. A somewhat longer calculation shows that the reciprocal lattice of the face-centered cubic lattice is the body-centered cubic lattice, and vice versa. We will come back to the structure of these lattices in the following section.

4.3.2 The Brillouin Zone

As for a lattice in real space, a Wigner-Seitz cell (see Section 3.3) for the reciprocal lattice can also be defined. This is called the **first Brillouin[4] zone** and, as we will see in the following chapters, it is of fundamental importance in solid state physics. The Brillouin zone plays an important role in lattice dynamics, in the description of the motion of conduction electrons, and in many other phenomena where lattice periodicity is of key relevance.

The construction principle of the Wigner-Seitz cell can be extended by including, not only the directly adjacent reciprocal lattice points, but also points somewhat further away leading to the construction of further cells. As in the construction of the original Wigner-Seitz cell, a vertical plane is setup in the middle of the connecting lines. This procedure leads to *Brillouin zones of higher order*, for which we now will introduce some simple examples.

4 Léon Nicolas Brillouin, *1889 Sévres, †1969 New York

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As mentioned above, the reciprocal lattice of a *linear chain* is itself also a linear chain but with lattice constant $2\pi/a$. The first Brillouin zone lies at the center of the reciprocal lattice, i.e. in the wave number range $-\pi/a < k \leq \pi/a$. The higher-order zones adjoin consecutively on the outside. In Figure 4.6 the first three Brillouin zones are indicated by differently colored lines. Note that the higher-order Brillouin zones each consist of two separate halves, which together have the same extension as the first Brillouin zone. The origin of the reciprocal lattice or the center of the first Brillouin zone is called **Γ-point**.

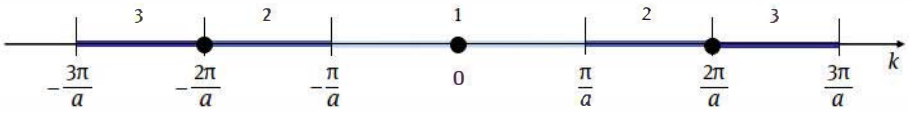


Fig. 4.6: The Brillouin zones of a linear chain. The series of Brillouin zones is indicated by differently colored lines. The perpendiculars at the center of the lines joining the origin to the various neighbors form the edges of the respective Brillouin zones. The Γ-point of the reciprocal lattice sits at the origin, $k = 0$.

Figure 4.7 shows the first three Brillouin zones of a 2-dimensional *rectangular* lattice. The first Brillouin zone is again a rectangle. However, with increasing zone order each zone is subdivided into more and more component parts, with the sum of the parts of each zone exactly equaling the area of the first. By moving the parts along reciprocal lattice vectors, they can each be joined again to form a rectangle with the area of the first Brillouin zone.

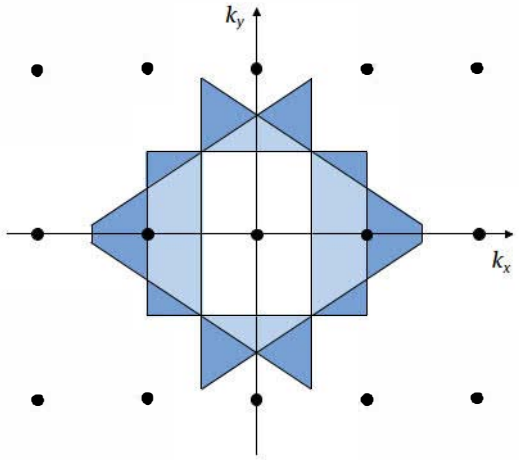


Fig. 4.7: The Brillouin zones of a rectangular lattice. The light blue area marks the second Brillouin zone, with darker blue marking the third.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As examples of the three-dimensional case, Figure 4.8 shows the first Brillouin zones of the face-centered cubic, the body-centered cubic and the hexagonal lattice. It is common to name points of high symmetry with abbreviations from group theory such as Γ, L, K, X, etc. As mentioned above, the origin of the first Brillouin zone is called *Γ-point*. As the polyhedra of the Wigner-Seitz cell fill real space, a periodic repetition of the polyhedra of the first Brillouin zone correspondingly fills reciprocal space without gaps. A comparison of Figure 4.8 and Figure 3.33b makes clear that the first Brillouin zone of the face-centered cubic lattice simply just corresponds to the Wigner-Seitz cell of the body-centered cubic lattice, and vice versa.

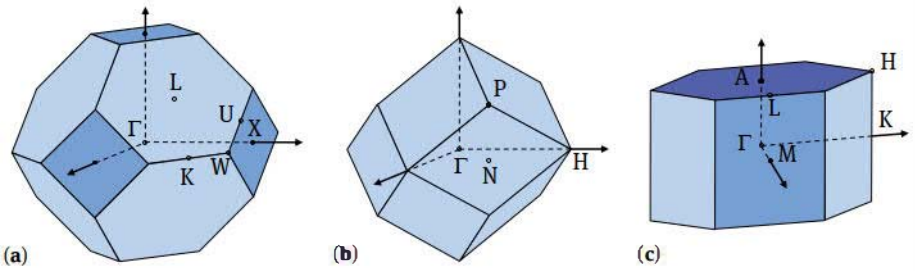


Fig. 4.8: First Brillouin zone of the face-centered cubic **(a)**, the body-centered cubic **(b)** and the hexagonal **(c)** lattice. Special points of high symmetry are marked by letters. (After N.W. Ashcroft, N.D. Mermin, *Festkörperphysik*, Oldenbourg, 2007.)

The three-dimensional Brillouin zones of higher order are also made up of an increasing number of components, which can be reassembled into a polyhedron with the same shape and size as the first Brillouin zone by translating them by the appropriate reciprocal lattice vectors. Figure 4.9 shows the first three Brillouin zones of the face-centered cubic lattice.

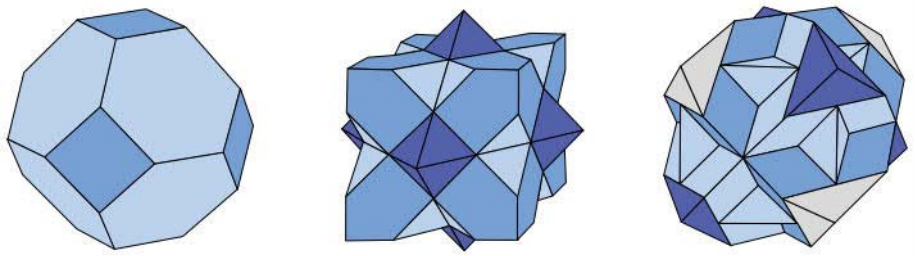


Fig. 4.9: The first three Brillouin zones of the face-centered cubic lattice. (After N.W. Ashcroft, N.D. Mermin, *Festkörperphysik*, Oldenbourg, 2007.)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

4.3.3 Miller Indices

A plane in the crystal containing a number of lattice points is called a **crystal** or **lattice plane**. Owing to the translation invariance of crystals, such a plane will have (infinitely) many equivalent planes all parallel to each other. For the characterization of lattice planes, it has also proved advantageous to use reciprocal lengths, as in the Fourier description of the lattice. This is the basis for indexing crystal planes using the **Miller**[5] **indices**. This procedure is demonstrated by the simple example of Figure 4.10.

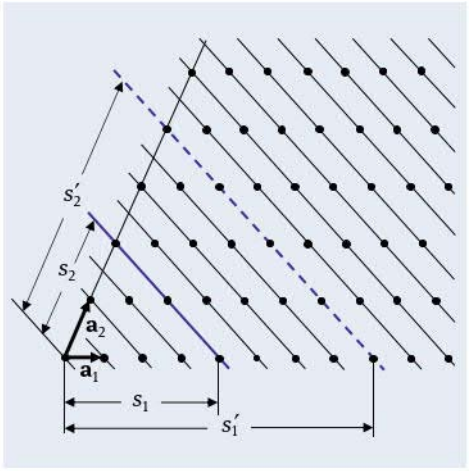


Fig. 4.10: Deriving the Miller indices. The diagram shows the intersection of one family of lattice planes with the plane of the diagram. These planes and the $\mathbf{a}_3$-axis run perpendicular to the diagram plane. The solid blue line highlights the plane with axes $s_1 = 4a_1$ and $s_2 = 2a_2$, and the blue-dashed line marks the plane with $s_1' = 8a_1$ and $s_2' = 4a_2$.

The figure shows a set of equivalent planes of a monoclinic lattice. We look first at the plane highlighted in blue. The intercepts of this plane with the axes s_i can be described in units of the basis vectors a_i. We thus find $\tilde{s}_1 = s_1/a_1 = 4$ and $\tilde{s}_2 = s_2/a_2 = 2$. In this monoclinic case, the plane runs perpendicularly to the drawing plane, being parallel to the $\mathbf{a}_3$-axis (which is clearly not shown), being also perpendicular to the drawing plane. Since then no intersection point occurs, the intercept is $\tilde{s}_3 = s_3/a_3 = \infty$. A whole family of such planes is indicated in the figure, for example, by the blue-dashed line with intercepts $s_1' = 8a_1$ and $s_2' = 4a_2$.

In the next step, we take the inverse of the intersection points to obtain:

$$h' = \frac{1}{\tilde{s}_1} = \frac{1}{4}, \qquad k' = \frac{1}{\tilde{s}_2} = \frac{1}{2} \qquad \text{and} \qquad l' = \frac{1}{\tilde{s}_3} = \frac{1}{\infty} = 0. \tag{4.15}$$

We multiply the numbers h', k', and l' by a factor m, chosen such that we obtain three integer numbers, normally as small as possible. These are the Miller indices. In the present case, with $m = 4$ (or $m = 8$ for the dashed plane) we get the Miller indices (hkl)

5 William Hallowes Miller, *1801 Llandovery Wales, †1880 Cambridge

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

of the plane drawn which turn out to be (120). The same procedure yields $(hkl) = (120)$ for the dashed plane as well. The abbreviation (hkl) thus stands for a set of parallel planes. We should note that in this convention there are no commas between Miller indices defining each plane. Here we should also note that the definition of the lattice planes in this way clearly depends on the choice of the basis vectors.

With the help of the Miller indices we can characterize not only the planes but also directions in the crystal. The $[hkl]$-*direction*, marked by square brackets, is the direction of the vector $[hkl] = h\mathbf{a}_1 + k\mathbf{a}_2 + l\mathbf{a}_3$. Here too, the smallest possible indices are used. With cubic crystals, the vector $[hkl]$ lies perpendicular to the (hkl)-lattice planes. Negative signs in the notation for planes and directions are indicated by overbars, e.g. $(\bar{1}00)$ or $[0\bar{1}0]$. If we do not want to pick out a particular set of lattice planes (hkl) or a special direction $[hkl]$, but rather the entirety of all symmetry-related *equivalent* lattice plane groups or crystal directions, then we write $\{hkl\}$ or $\langle hkl \rangle$. The various equivalent groups of lattice planes or directions can be transformed into each other by using rotational symmetries.

It is usually the low index planes which are of particular interest since these reflect the basic periodicities of the crystal lattices. In addition to the {100}-lattice planes, these are mainly the {110}- and {111}-planes. Figure 4.11 shows the position of some of these planes for cubic crystals. Occasionally we also find higher indexed planes that are not named according to the criterion of smallest possible numbers. These have a physical meaning if the basis is more complicated or if the unit cell under consideration is not primitive and contains several lattice points. Thus the (200)-plane is parallel to the (100)-plane, but intersects the axis at half the lattice spacing. In the case of body-centered cubic crystals, a lattice point lies at the cube center of the non-primitive unit cell (cf. Bravais lattice in Figure 3.26) in the (200)-plane.

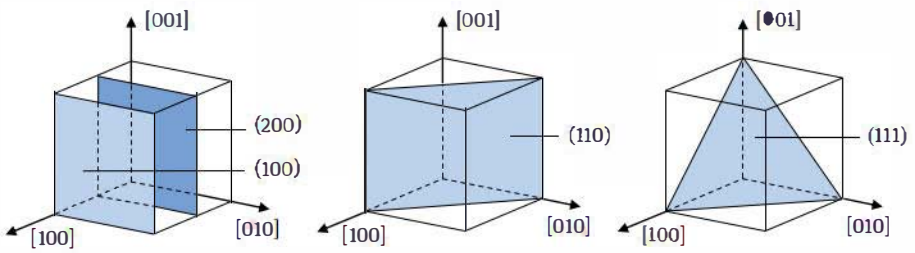


Fig. 4.11: Selected lattice planes of a cubic crystal. On the left the (100)- and (200)-planes are shown, and in the center and on the right the planes (110) and (111).

In addition, we should note that in surface physics, planes with large Miller indices often play an important role. If a crystal is cut at a small angle relative to one of its crystallographic axes, the surface created consists of a regular sequence of *steps*

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

of mono-atomic height separated by *terraces*. Such stepped or *vicinal* surfaces have particularly interesting physical properties, which we will not discuss here in detail.

When indexing the planes of hexagonal crystals, special care is required. If we use the Bravais lattice, the particular symmetry of this crystal system means that equivalent planes within the indexing scheme described are named differently. Thus, the planes (100) and $(1\bar{1}0)$ can be transformed into each other by a 60° rotation, which means they are equivalent, although we would not expect this from their indexing. This shortcoming can be avoided by introducing an additional index. The plans are denoted by $(hkil)$ and the additional index i is defined by $i = -(h + k)$. In our example we are thus dealing with planes $(10\bar{1}0)$ and $(1\bar{1}00)$ or more generally with $\{1\bar{1}00\}$-planes.

There is obviously a close relationship between the Miller indices and the reciprocal lattice. In order to find the connection, we choose, as sketched in Figure 4.12, an arbitrary lattice plane and define its spatial position by the three vectors **u**, **v** and **w**, which can also be expressed by the basis vectors of the respective lattice: $\mathbf{u} = \tilde{s}_1\mathbf{a}_1$, $\mathbf{v} = \tilde{s}_2\mathbf{a}_2$, and $\mathbf{w} = \tilde{s}_3\mathbf{a}_3$.

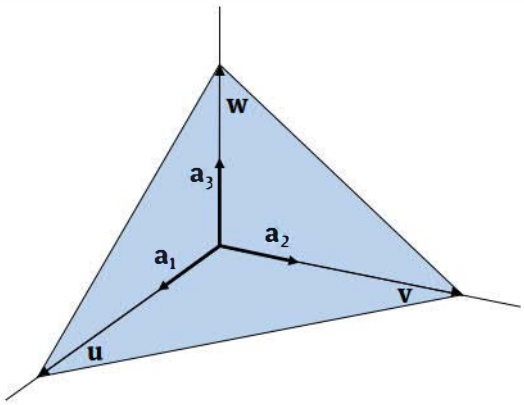


Fig. 4.12: To derive the relationship between the Miller indices and the reciprocal lattice: The vectors u, v and w, which can be expressed in units of the basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$, span a lattice plane of the crystal lattice.

Since the vectors $(\mathbf{u} - \mathbf{v})$ and $(\mathbf{w} - \mathbf{v})$ lie in the lattice plane, the cross product $(\mathbf{u} - \mathbf{v}) \times (\mathbf{w} - \mathbf{v})$ defines a vector $\hat{\mathbf{n}}$ perpendicular to the plane considered. Using the definition (4.13) we can express $\hat{\mathbf{n}}$ as:

$$\begin{aligned}\hat{\mathbf{n}} &= (\mathbf{u} - \mathbf{v}) \times (\mathbf{w} - \mathbf{v}) = (\mathbf{u} \times \mathbf{w}) - (\mathbf{v} \times \mathbf{w}) - (\mathbf{u} \times \mathbf{v}) \\ &= \tilde{s}_1\tilde{s}_3(\mathbf{a}_1 \times \mathbf{a}_3) - \tilde{s}_2\tilde{s}_3(\mathbf{a}_2 \times \mathbf{a}_3) - \tilde{s}_1\tilde{s}_2(\mathbf{a}_1 \times \mathbf{a}_2) \,. \end{aligned} \tag{4.16}$$

If we multiply $\hat{\mathbf{n}}$ by $-2\pi/\tilde{s}_1\tilde{s}_2\tilde{s}_3V_c$, we get

$$-\frac{2\pi}{V_c}\frac{\hat{\mathbf{n}}}{\tilde{s}_1\tilde{s}_2\tilde{s}_3} = -\frac{2\pi}{V_c}h'k'l'\hat{\mathbf{n}} = h'\mathbf{b}_1 + k'\mathbf{b}_2 + l'\mathbf{b}_3 = \frac{1}{p}\,\mathbf{G}_{hkl}\,, \tag{4.17}$$

where again we have used $h = mh'$ etc.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

This vector and the reciprocal lattice vector for the plane are parallel to each other, i.e. the reciprocal lattice vector $\mathbf{G}_{hkl}$ is perpendicular to the lattice planes (hkl). With elementary analytical geometry it can be further shown that the distance d_{hkl} between two neighboring lattice planes is given by

$$d_{hkl} = \frac{2\pi}{|\mathbf{G}_{hkl}|} \tag{4.18}$$

and is directly linked to the magnitude of the reciprocal lattice vector. Thus, the reciprocal lattice vector $\mathbf{G}_{hkl}$ represents the set of lattice planes (hkl): It is perpendicular to these planes, its magnitude is determined by the inverse plane distance. A reciprocal lattice vector is assigned to each set of lattice planes and vice versa. This interrelation becomes understandable if one considers that the reciprocal lattice vectors are determined according to equation (4.7) by the Fourier components of the scattering density distribution $\varrho(\mathbf{r})$.

4.4 Determination of the Crystal Structure

When discussing the scattering process we found that the amplitude of the scattered radiation is proportional to the scattering amplitude $\mathcal{A}(\mathbf{K})$. The observed scattering intensity $I(\mathbf{K})$ is therefore given by the relationship

$$I(\mathbf{K}) \propto |\mathcal{A}(\mathbf{K})|^2 = \left| \int_{V_s} \varrho(\mathbf{r})\, e^{-i\mathbf{K}\cdot\mathbf{r}} dV \right|^2 , \tag{4.19}$$

where the integration extends over the sample volume V_s.

Rather than the absolute value of the scattered radiation, it is its variation with the direction of observation that is crucial for the determination of the structure. This information is contained in the quantity $|\mathcal{A}(\mathbf{K})|^2$. However, for the sake of simplicity, we will nevertheless use the term "scattered intensity" which we will now calculate using the concept of the reciprocal lattice. We apply the expansion (4.7) of the scattering density distribution $\varrho(\mathbf{r})$ in terms of the basis vectors of the reciprocal lattice in (4.19) and obtain

$$|\mathcal{A}(\mathbf{K})|^2 = \left| \sum_{h,k,l} \varrho_{hkl} \int_{V_s} e^{i(\mathbf{G}-\mathbf{K})\cdot\mathbf{r}} dV \right|^2 . \tag{4.20}$$

In the further discussion we will usually omit the indices hkl in the reciprocal lattice vector and only give them when they are needed. Since the function $\exp[i(\mathbf{G} - \mathbf{K}) \cdot \mathbf{r}]$ oscillates, the contributions average out if the integration is taken over a scattering volume large in extent compared to the periods of the oscillations. From a physical point of view, this means that the scattering contributions of the individual atoms are smoothed out by interference, hut excluding certain special directions of observation

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

for which the **diffraction condition**

$$\mathbf{K} = \mathbf{G} \tag{4.21}$$

holds, and for which, the exponential function in (4.20) is unity. This reflects the fact that the superposition of the scattered waves is constructive due to the regular arrangement of the atoms. Formally this can be described by the relation

$$\int_{V_s} e^{i(\mathbf{G}-\mathbf{K})\cdot\mathbf{r}}\, dV \simeq \begin{cases} V_s & (\mathbf{K} = \mathbf{G}) \ , \\ 0 & (\mathbf{K} \neq \mathbf{G}) \ . \end{cases} \tag{4.22}$$

In the limiting case of an infinitely large sample volume, the Fourier representation of the δ function in the three spatial directions is shown on the left-hand side. Radiation scattering only occurs if the condition of (4.21) is strictly fulfilled. In the case of a finite sample volume or a finite penetration depth of the radiation, this scattering condition is "softened". We will discuss the important consequences of this in more detail in a moment. If we wish to observe a "diffraction peak", i.e. scattered radiation, the crystal orientation must be adjusted, for a given scattering geometry, until the scattering condition is fulfilled.

If we now calculate the scattering intensity with the help of (4.20), we find that the summation can be omitted, because each reciprocal lattice vector $\mathbf{G}_{hkl}$ just represents *one* Fourier coefficient ϱ_{hkl} of the scattering density distribution. Thus, we obtain for the scattered intensity the simple expression

$$|\mathcal{A}(\mathbf{K} = \mathbf{G})|^2 = |\varrho_{hkl}|^2 \, V_s^2 \ . \tag{4.23}$$

The quadratic dependence of the scattering intensity on the sample volume is surprising at first glance since we would expect the intensity of the scattered radiation to increase proportionally to the number of scattering atoms. The reason for this surprising prediction is that we have oversimplified the derivation of this result. In general, for a finite volume V_s the diffraction condition $\mathbf{K} = \mathbf{G}$ becomes less strict since with decreasing sample volume the diffraction peaks become broadened. Depending on the angle, the scattering intensity of a diffraction peak passes through a maximum whose *height* is given by (4.23), but whose width decreases with the sample volume. A more detailed consideration shows that the width of the peak is proportional to V_s^{-1}. From this, it follows that the *integral intensity* of the diffraction peak is linear in increasing scattering volume, i.e. is indeed proportional to the number of scattering centers as expected.

Fundamentally, the scattering volume is limited by the finite penetration depth of the incident radiation. With X-ray scattering, depending on the scattering strength of the sample under investigation, a proportion of $10^{-5} - 10^{-3}$ of the incident intensity will be scattered at each lattice plane. This leads to a strong decrease in intensity of the incident beam within a few micrometers of the surface, thus limiting the thickness

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

of the irradiated layer and thus also the scattering volume. The influence of a finite sample volume may also be visible with polycrystalline samples. If the sample consists of very small crystallites, a peak broadening may be observed under suitable conditions, allowing conclusions to be drawn about the size of the crystallites. However, the finite width of the diffraction peaks may also depend on factors other than the sample size. For example, the spectrum and the divergence of the incident radiation can already give rise to an apparent broadening of the peaks.

In advanced theories of scattering, other effects such as absorption or refractive index must also be taken into account in addition to the effective scattering volume. Of great importance is the fact that the diffracted wave will itself be further scattered by the lattice planes partly in the direction of the incident beam. These effects are incorporated in the *dynamic theory*, mainly developed by *P.P. Ewald*[6]. Of course, this treatment is much more informative, but also more complicated than the relatively simple discussion presented here.

4.4.1 Ewald Sphere and Bragg Condition

If X-rays with wave vector $\mathbf{k}_0$ are incident on a randomly oriented crystal, there is usually no diffraction peak, since the diffraction condition $\mathbf{K} = \mathbf{G}$ is not satisfied for the chosen observation direction, in other words, the beam will pass through the crystal with no particular interaction. If the direction of the incident beam is given, the crystal must be orientated appropriately and the observation of any output beam must also be made in a suitable direction. The required orientation of the crystal in relation to the incident radiation can be pictured geometrically by the construction of the Ewald sphere, as shown in Figure 4.13. Here the reciprocal lattice is shown and the vector $\mathbf{k}_0$ is drawn such that its endpoint lies at the origin of the reciprocal lattice. Then around the starting point of $\mathbf{k}_0$ we draw a circle (or a sphere in the three-dimensional case) with radius k_0 to mark the set of all scattering vectors $\mathbf{K} = (\mathbf{k} - \mathbf{k}_0)$ for elastic scattering with $|\mathbf{k}_0| = |\mathbf{k}|$. The scattering condition (4.21) is fulfilled if the Ewald sphere passes through a point on the reciprocal lattice, in which case, there will be a diffracted beam in the $\mathbf{k}$-direction with scattering vector $\mathbf{K} = \mathbf{G}$.

So far we have assumed an infinitely-extended perfect crystal lattice with infinitely sharp reciprocal lattice points. However, if we are dealing with crystals with finite dimensions, the reciprocal lattice points become "smeared". The smaller the diffracting volume, the less sharp the lattice points and the broader the observed diffraction peaks become. This also applies when the incident radiation of Figure 4.13 has only a limited penetration depth. The finite frequency width and the divergence of the incident radiation can also lead to a softening of the diffraction condition. This effect could be

6 Paul Peter Ewald, *1888 Berlin, †1985 Ithaca

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

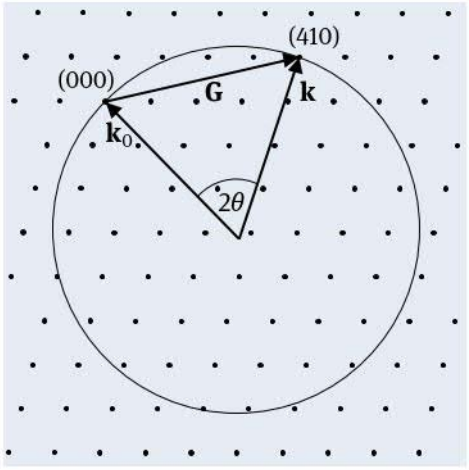


Fig. 4.13: A two-dimensional representation of the Ewald sphere. For the given orientations of the crystal and the incident ray $\mathbf{k}_0$ there will be a reflected ray $\mathbf{k}$ by diffraction at the (410)-lattice planes.

included in the construction of the Ewald sphere by giving a finite thickness to the perimeter line (2D) or perimeter shell (3D).

We want to look at our results from a different angle. By the spatial arrangement of the radiation source, the radiation detector and the orientation of the sample, the experimenter determines $\mathbf{K}$ in the reciprocal lattice. If a peak is observed, the diffraction condition also determines the reciprocal lattice vector $\mathbf{G}$ with the number triplet (hkl). This in turn is connected with the Fourier component of the scattering density distribution and thus with the associated array of lattice planes. This means that the observed scattering occurs at a periodic variation of the scatter density whose period is given by $d_{hkl} = 2\pi/|\mathbf{G}_{hkl}|$.

The relation between the scattering vector and the period of the density variation provides a clear interpretation of the diffraction experiment (see Figure 4.14). If θ is the angle between the direction of the incident wave and the planes of the scattering density variation, then because of $k_0 = k$ for the magnitude of the scattering vector: $K = |\mathbf{k} - \mathbf{k}_0| = 2k_0 \sin\theta = (4\pi \sin\theta)/\lambda$. If this relationship and equation (4.18) are inserted into the scattering condition (4.21), the known **Bragg condition**[7] is obtained

$$2\, d_{hkl}\, \sin\theta = \lambda\ . \tag{4.24}$$

The angle θ, for which a diffraction peak is observed, is often called the glancing angle. The straightforward derivation gives $2d \sin\theta = n\lambda$ where n is an integer. The appearance of the factor n has a simple explanation: The periodic modulation of the scattering density distribution caused by the lattice planes considered does not correspond to a pure sinusoidal modulation. We also have to take into account higher harmonics with periods d_{hkl}/n, which also cause constructive interference of the scattered radiation.

7 William Henry Bragg, *1862 Wigton, †1942 London, Nobel Prize 1915

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

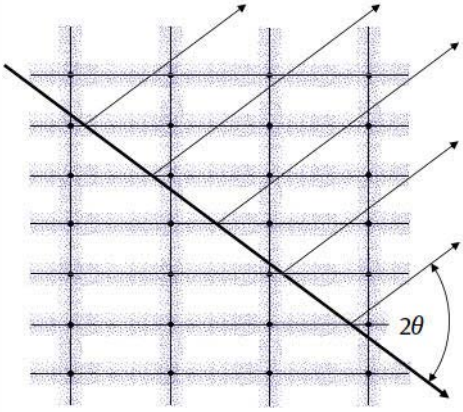


Fig. 4.14: Illustration of the Bragg reflection at scattering density variations with tetragonal symmetry. The figure shows the incident beam being reflected by the density fluctuations of the horizontal lattice planes. The reflected partial rays overlap constructively if the Bragg condition is fulfilled. The position of the atomic nuclei is indicated by points. The partial beams reflected by the vertical planes are not shown.

Since the reciprocal lattice also includes lattice points which can be represented by an integer multiple $n\mathbf{G}$ of the original vector, then $\mathbf{K} = n\mathbf{G}$ is also a condition for the constructive superposition of the scattered waves.

4.4.2 The Structure Factor

The arguments above enable us to predict in scattering experiments which reflections *are allowed* by the lattice periodicity of crystals, but not the *magnitude* of these peaks. The *intensity* of the peaks is determined following (4.23) by the Fourier coefficients ϱ_{hkl} of the scattering density distribution and thus contains information about the structure of the basis. If the basis is not monatomic, interferences between the contributions of the various atoms in the unit cell lead to a variation of the peak intensity. When calculating the scattering intensity, we should clearly consider the contributions of the individual atoms separately and then to sum them up. As shown schematically in Figure 4.15, we divide up the position vector $\mathbf{r}$ by defining $\mathbf{r} = \mathbf{R}_m + \mathbf{r}_\alpha + \mathbf{r}'$. The vector $\mathbf{R}_m$ simply determines the position of the unit cell under consideration and is unimportant to the following discussion.

The arrangement of the atoms α within of the unit cell is described by $\mathbf{r}_\alpha$. The scattering properties of the individual atoms are determined by the distribution of the scattering density in the atom concerned. Here we have to integrate over the position vector $\mathbf{r}'$, showing the position of the scattering volume dV' in the atom. We use the expression for $\mathbf{r}$ and instead of (4.8) above, we obtain the expression:

$$\varrho_{hkl} = \frac{1}{V_c}\int_{V_c} \varrho(\mathbf{r})\, e^{-i\mathbf{G}\cdot\mathbf{r}} dV = \frac{1}{V_c}\sum_\alpha e^{-i\mathbf{G}\cdot\mathbf{r}_\alpha}\int_{V_\alpha} \varrho_\alpha(\mathbf{r}')\, e^{-i\mathbf{G}\cdot\mathbf{r}'} dV' = \frac{1}{V_c}\sum_\alpha f_\alpha(\mathbf{G})\, e^{-i\mathbf{G}\cdot\mathbf{r}_\alpha}. \quad (4.25)$$

The integration of the density distribution of the scattering centers is made over the volume V_α of the individual atoms, the summation extending over all the atoms of the basis. In the final term, we introduce the **atomic structure factor** $f(\mathbf{G})$ as an abbrevi-

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

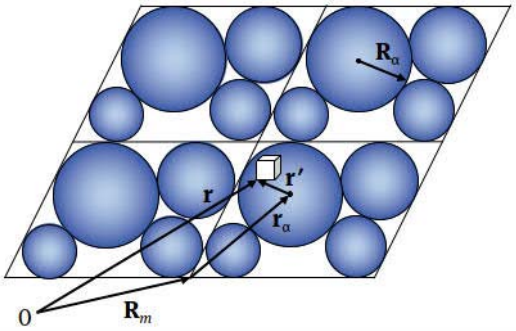


Fig. 4.15: Splitting of the position vector when deriving the structural and atomic structure factor. The atoms are considered as spherical objects with the radius R_α. The position of the unit cell is defined by $\mathbf{R}_m$, that of the atomic centers in the cell by $\mathbf{r}_\alpha$ and the position of the scattering volume dV' within the atom is defined by $\mathbf{r}'$.

ation, describing the specific properties of the scattering center under consideration, which in turn depend on the incident radiation. We first turn to the evaluation of the sum and then discuss the meaning of the atomic structure factor itself.

The summation is simplified if the position of the atoms in the unit cell is expressed in terms of the basis vectors of the lattice in real space. For the position vector of the atoms we therefore write $\mathbf{r}_\alpha = u_\alpha \mathbf{a}_1 + v_\alpha \mathbf{a}_2 + w_\alpha \mathbf{a}_3$, where the components u_α, v_α and w_α are determined by the arrangement of the atoms in the basis. If we calculate the dot product in (4.25), we obtain for the **structure amplitude** or the **structure factor** $\mathcal{S}_{hkl} = \varrho_{hkl} V_c$ the important result

$$\mathcal{S}_{hkl} = \sum_\alpha f_\alpha(\mathbf{G})\, e^{-2\pi i(hu_\alpha + kv_\alpha + lw_\alpha)} . \tag{4.26}$$

As simple examples of the calculation of the intensity of the scattered radiation we consider the three cubic lattices. If the basis of the simple cubic lattice is monatomic, then there is no summation and if we take the atomic center as the unit cell origin, we obtain $\mathcal{S}_{hkl} = f(\mathbf{G})$. In this case the angular dependence of the scattering intensity is determined only by the atomic structure factor, i.e. by the scattering properties of the atoms. If we had chosen a different coordinate for the position of the atom in the unit cell, we would have introduced a phase factor but which would disappear when the intensity is calculated.

If we now take a simple cubic lattice but with a diatomic basis, for example simple cubic cesium chloride lattice with the basis Cs^+ and Cl^-, we find by inserting in (4.26):

$$\mathcal{S}_{hkl} = \begin{cases} f_{Cs} + f_{Cl} & h+k+l \quad \text{even}, \\ f_{Cs} - f_{Cl} & h+k+l \quad \text{odd}. \end{cases} \tag{4.27}$$

Crystals with this structure cause diffraction peaks with very different intensities depending on whether the contributions of the two ion types are either added or subtracted. Depending on the cross sum of the indices, strong and weak peaks occur.

For body-centered cubic lattices with a monatomic basis, we find

$$\mathcal{S}_{hkl} = f\left[1 + e^{-i\pi(h+k+l)}\right] = \begin{cases} 2f & h+k+l \quad \text{even}, \\ 0 & h+k+l \quad \text{odd}. \end{cases} \tag{4.28}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

This result indicates that lattice planes where the sum of the Miller indices is odd do not give rise to reflections. Surprisingly, therefore, the diffraction peaks of the (100)-planes should be missing. As we can see in Figure 4.16, there is a simple, obvious explanation for this. There is a path difference of 2π between the rays scattered from adjacent (100)-planes if the scattering condition (4.21) is fulfilled, so that the individual rays overlap constructively as expected. However, the contribution of these planes is compensated by that of the intermediate (200)-planes, since their scattering contributions, as indicated in Figure 4.16, have a phase difference π at the given scattering angle with respect to that of the (100)-planes. The (200)-reflection, occurring at a larger angle, is not deleted, since here, as can be readily seen, the scattering contributions of the (100)- and (200)-planes constructively reinforce.

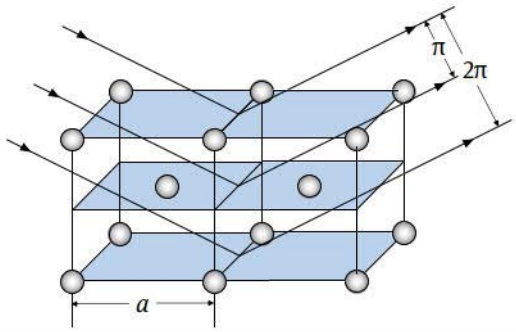


Fig. 4.16: The missing (100)-reflection in body-centered cubic lattices. The scattering contributions of adjacent planes cancel out by destructive interference.

At this point we should mention an interesting effect that occurs during the structure determination of *cesium iodide* with X-rays. As we will see later, for X-rays the atomic structure factor is proportional to the number of electrons in the scattering atoms. Since cesium and iodine ions have the same number of electrons, they thus have the same scattering strength, so that according to (4.27) those reflections with an odd sum of Miller indices disappear and only reflections with even values of $(h + k + l)$ occur. Therefore, in this experiment CsI behaves as a body-centered cubic crystal with a simple basis.

For the face-centered cubic lattice with a simple basis, we find

$$S_{hkl} = f\left[1 + e^{-i\pi(k+l)} + e^{-i\pi(h+l)} + e^{-i\pi(h+k)}\right] = \begin{cases} 4f & \text{all indicies are even or odd,} \\ 0 & \text{otherwise.} \end{cases} \tag{4.29}$$

So diffraction peaks only occur if *all* indices are either even or odd.

We now explain the complete disappearance of certain reflections in more detail. The transition from the lattice in real space to the reciprocal lattice is made following transformation rules for the basis vectors. In the derivation of the reciprocal lattice we pointed out that the result of the transformation depends on the choice of the unit cell in real space. We will illustrate this in the case of the two-dimensional square

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

lattices in Figure 4.17. In the upper left corner, a primitive unit cell is drawn with its corresponding reciprocal lattice below. However, on the right, for the same square lattice, we choose a unit cell which, in analogy to the body-centered cubic lattice, has a lattice point in the center. The application of the same transformation rule as for this primitive unit cell now leads to the reciprocal lattice shown below, with lattice-points now correspondingly closer. For clarity, all quantities on the right-hand side are distinguished by a prime suffix.

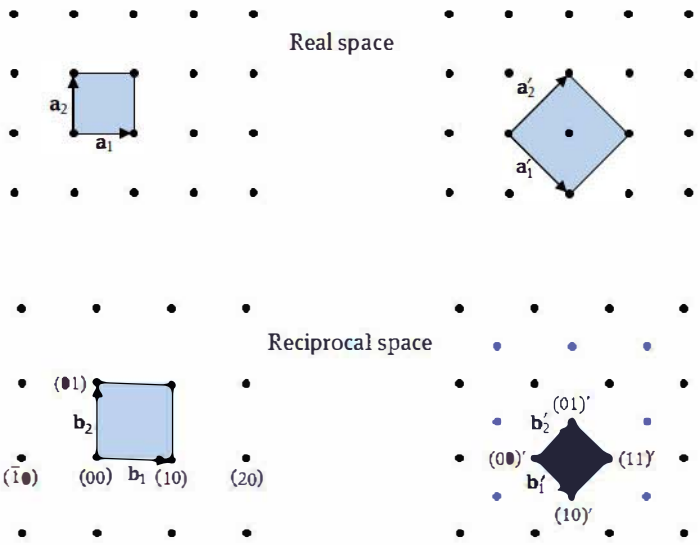


Fig. 4.17: The dependence of the reciprocal lattice on the choice of the unit cell in real space. On the left, a primitive unit cell is assumed and the reciprocal lattice is shown below. On the right, a "centered" square lattice was chosen as the unit cell. In the reciprocal lattice shown below additional blue lattice points appear. The indexing is clearly different in the two cases.

We note that the points on the-right-hand side with values (10) and (01), for example, and more generally all lattice points with odd values of $(h+k)$ have no correspondence on the left-hand side of the picture. The reciprocal lattice points with even values of $(h + k)$, on the other hand, do have correspondences in the reciprocal lattice of the left-hand picture. For example, the $(11)'$ lattice point on the right corresponds to the (10)-lattice point on the left. Whenever we assume a non-primitive unit cell, i.e. choose a "to large" unit cell, then additional reciprocal lattice points appear. Since their corresponding structural factor disappears, they cannot be observed in experiments. The absence of the (100)-reflection for the body-centered cubic lattice discussed above has the same cause. This simple observation shows that the choice of the unit cell has a considerable influence on the assignment (and notation) of the diffraction peaks, but clearly the physical result of any measurement will not be influenced by the choice.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

4.4.3 Atomic Structure Factor

In discussing the scattering intensity, we have so far used the *scattering strength* of individual atoms. Now let us look at the **atomic structure factor**

$$f_\alpha(\mathbf{K}) = \int_{V_\alpha} \varrho_\alpha(\mathbf{r})\, \mathrm{e}^{-\mathrm{i}\mathbf{K}\cdot\mathbf{r}}\, \mathrm{d}V \tag{4.30}$$

which is defined by (4.25). We have rewritten the corresponding expression here, but simply replacing the reciprocal lattice vector **G** by the scattering vector **K**, since the scattering strength of individual atoms does not depend on the crystal structure. For clarity, we use for the position vector the denotation **r** instead of $\mathbf{r}'$.

Before we discuss the importance of the atomic structure factor in X-ray diffraction, let us take a look at neutron scattering, since the conditions there are particularly simple. Except for the weak interaction with the magnetic moment of the electrons, which is only important for ordered magnetic structures, neutrons are scattered by nuclei which are small compared to the neutron wavelength, which in diffraction experiments is comparable to the lattice spacing. This means that $\mathbf{K}\cdot\mathbf{r} \ll 1$, so that the exponential factor in (4.30) is close to one. The integration thus leads to a constant direction-independent atomic structure factor $f(\mathbf{K}) = -b$, whose value reflects the specific scattering properties of the nucleus under consideration. Usually this constant is called **scattering length** b. The sign is fixed by convention, a negative sign indicating that the interaction between neutron and nucleus is attractive, which holds in most cases. This does not play a role in our considerations, since the scattering intensity is determined by $|b|^2$. Both the sign and the absolute value of the scattering length vary from element to element and from isotope to isotope.

The situation is more complex for the diffraction of X-rays. Since the extension of the scattering object, i.e. the electron cloud of the atoms, is comparable to the wavelength of the incident radiation, we have to consider the interference between the contributions of all the volume elements of the electron distribution. Equation (4.30) therefore leads to a very different result from that of neutron scattering. In the case of X-ray diffraction, the atomic structure factor $f(\mathbf{K})$ is called the **atomic scattering factor** or **atomic shape factor**. In the following discussion we assume for simplicity that the charge associated with each atom is spherically symmetrical and that any charge density in the space between the atoms can be neglected. The scattering factor then depends only on $|\mathbf{K}|$, so that in spherical coordinates for the atomic scattering factor $f_\alpha(K)$ we can use the expression:

$$f_\alpha(K) = \int \varrho_\alpha(r)\, \mathrm{e}^{-\mathrm{i}\mathbf{K}\cdot\mathbf{r}}\, \mathrm{d}V = \int_0^{R_\alpha} \mathrm{d}r \int_0^{\pi} \mathrm{d}\vartheta \int_0^{2\pi} \mathrm{d}\varphi\, \varrho_\alpha(r)\, r^2 \sin\vartheta\, \mathrm{e}^{-\mathrm{i}Kr\cos\vartheta}\,. \tag{4.31}$$

Here φ stands for the azimuth and ϑ for the polar angle between **K** and **r**. The integration over the distance r takes place up to the "surface" R_α of the atom. Spherical symmetry

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

simplifies the integration over φ and ϑ allowing it to be performed analytically, yielding:

$$f_\alpha(K) = \int_0^{R_\alpha} 4\pi r^2 \varrho_\alpha(r) \frac{\sin Kr}{Kr} \, \mathrm{d}r \, . \tag{4.32}$$

For the complete calculation of the atomic scattering factor, the specific charge density distribution $\varrho(r)$ of the atom must be known. Only in the simplest cases can the integration over the radius vector be carried out analytically, such as in the case of the free hydrogen atom. Here the scattered density distribution $\varrho(r)$ is determined by the square of the known wave function $\psi_0(r) = (\pi a_0^3)^{-1/2} \exp(-r/a_0)$.[8] Thus the scattering density is given by:

$$\varrho(r) = |\psi_0(r)|^2 = \frac{1}{\pi a_0^3} \mathrm{e}^{-2r/a_0} \, , \tag{4.33}$$

with a_0 the Bohr radius. If we insert $\varrho(r)$ in (4.32) and take the integration limit to infinity $R_\alpha \to \infty$, we find:

$$f_\mathrm{H}(K) = \frac{1}{\left[1 + \left(\frac{1}{2} a_0 K\right)^2\right]^2} \, . \tag{4.34}$$

This function, and the result of a numerical calculation for free aluminium atoms, is shown in Figure 4.18. As expected, the radiation from extended atoms depends strongly on the direction. For aluminum, experimental values are also shown, as determined with the help of (4.25) from scattering experiments on crystals.

We now take a closer look at the limiting case $Kr \to 0$. Since here $(\sin Kr)/Kr \to 1$, equation (4.32) simplifies to an integration over the full distribution of the scattering density, which naturally gives the total number Z of electrons in the scattering atoms or ions. Therefore we find:

$$f(Kr \to 0) \approx \int_0^{R_\alpha} 4\pi r^2 \varrho(r) \mathrm{d}r = Z \, . \tag{4.35}$$

This limiting case is realized in processes scattering in the forward direction, since here the scattering vector goes to zero $K \to 0$. As the scattering intensity is proportional to f^2, it is proportional to Z^2 for small scattering angles and is determined exclusively by the number of electrons. The same limiting case is found when the scattering atom is small compared with the wavelength of the incident radiation, since in this case $R_\alpha \ll \lambda$ and therefore $KR_\alpha \ll 1$. This means that, in this case, the atoms simply act as point-like scattering objects, and thus we get a result similar to that for neutron scattering: The atomic scattering factor is independent of the scattering vector and

8 To ensure that the atomic scattering factor is dimensionless the charge of the electron in the expression for $f(\mathbf{K})$ is usually omitted.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

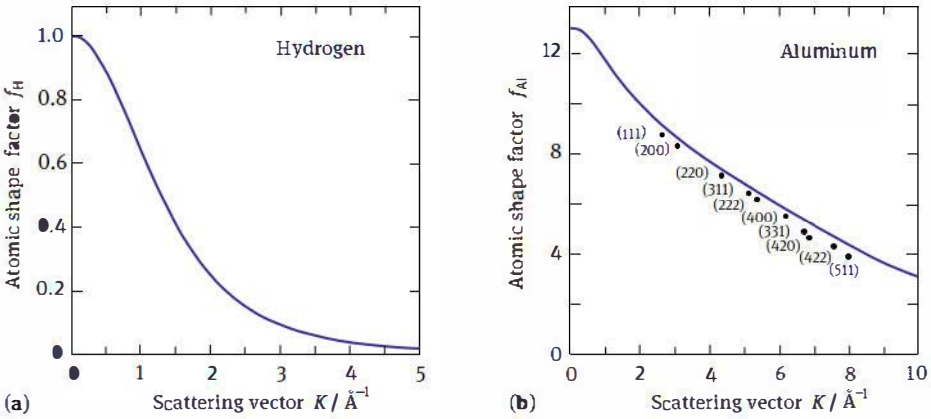


Fig. 4.18: Atomic scattering factors as a function of the scattering vector $K = (4\pi \sin\theta)/\lambda$. **a)** The atomic scattering factor for free hydrogen atoms according to (4.34). **b)** The numerically calculated atomic scattering factor for aluminum. The points are experimental values derived from measurements of the scattering intensity on crystals in various directions. (After B.W. Batterman et al., Phys. Rev. **122**, 68 (1961).)

reflects only the scattering intensity of the atom. In X-ray scattering experiments, the atomic form factor thus indicates the scattering amplitude observed due to a deviation of the charge distribution of the atom from a point-like form. The finite extension of the atoms causes a reduction of the scattering intensity, except for forward scattering, due to the interference of the scattering contributions.

4.4.4 Surfaces and Thin Layers

Now we will briefly discuss the determination of the structure of surfaces or thin layers. For this purpose, we first have to understand what the reciprocal lattice of a two-dimensional structure looks like. Let us call the basis vectors of the surface lattice $\mathbf{c}_1$ or $\mathbf{c}_2$, those of the reciprocal lattice $\mathbf{c}_1^*$ or $\mathbf{c}_2^*$ and using the definition (4.12) of the reciprocal lattice, we find:

$$\mathbf{c}_1 \cdot \mathbf{c}_2^* = \mathbf{c}_2 \cdot \mathbf{c}_1^* = 0 \qquad \text{and} \qquad \mathbf{c}_1 \cdot \mathbf{c}_1^* = \mathbf{c}_2 \cdot \mathbf{c}_2^* = 2\pi \,. \tag{4.36}$$

The reciprocal lattice points of a two-dimensional lattice can be represented in three-dimensional space as infinitely long rods perpendicular to the surface. We can imagine that the two-dimensional lattice is created from a three-dimensional lattice in the limiting case that the length of the basis vector $\mathbf{a}_3$, perpendicular to the surface, goes to infinity. In this case, the reciprocal lattice points move closer and closer together along the $\mathbf{a}_3$ direction forming bars at the limiting case, as shown in Figure 4.19a. The notation of the rods by the indices hk of the reciprocal lattice vector $\mathbf{g} = h\mathbf{c}_1^* + k\mathbf{c}_2^*$ is only partially shown in the illustration for clarity.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

On this basis, we can use the Ewald sphere to determine which possible diffraction peaks occur when the Ewald sphere intersects with one of the rods of the reciprocal lattice. Figure 4.19b shows this situation for the peaks lying in the plane of the figure. Of course, the Ewald sphere also cuts rods in front of and behind the drawing plane, so that the number of possible reflections increases. Here it should be noted that, regardless of the direction of the incident radiation or of the orientation of the sample, the diffraction condition can always be fulfilled, in contrast to diffraction by three-dimensional lattices. We will talk about the experimental realization in the last section of this chapter.

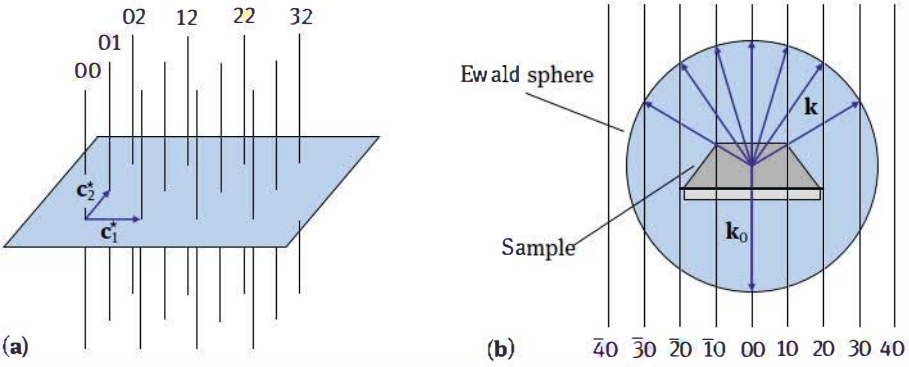


Fig. 4.19: a) The rods of the reciprocal space of a periodic solid surface. The reciprocal lattice basis vectors $\mathbf{c}_1^*$ and $\mathbf{c}_2^*$ are also shown. **b)** The Ewald sphere for vertical incidence of the probe radiation. The points of intersection of the Ewald sphere with the rods define the wave vector $\mathbf{k}$ of the diffraction peaks. For the sake of clarity, only reflections lying in the plane of the drawing are shown.

4.4.5 The Phase Problem and the Diffraction Peak Width

As already mentioned in Section 4.2, the scattering amplitude $\mathcal{A}(\mathbf{K})$ cannot be determined directly, since radiation detectors are not able to register the phase of the scattered radiation. Instead of $\mathcal{A}(\mathbf{K})$ only the intensity, i.e. $|\mathcal{A}(\mathbf{K})|^2$, can be measured. Owing to the missing phase information, the Fourier transformation of $|\mathcal{A}(\mathbf{K})|^2$ according to equation (4.6) cannot really be carried out, so that the structure cannot be directly derived from the measured scattered intensity. This can be clarified by the simple example of *Friedels rule*[9]: Assuming that any absorption can be neglected, the scattering density distribution $\varrho(\mathbf{r})$ is a real quantity, so that according to (4.4) we can write $\mathcal{A}(\mathbf{K}) = \mathcal{A}^*(-\mathbf{K})$. The intensity is therefore $|\mathcal{A}(\mathbf{K})|^2 = |\mathcal{A}^*(-\mathbf{K})|^2 = |\mathcal{A}(-\mathbf{K})|^2$, i.e.,

9 Jacques Friedel, *1921 Paris, †2014 Paris

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

that independently of the actual structure, the image of the diffraction peaks always shows inversion symmetry. It cannot be deduced from the experiment whether "front" and "back" of a crystal differ. This restriction is hardly important when determining the structure of crystals with a simple basis. In such cases experience or a comparison with known crystals will help.

However, when investigating complex unit cells, especially proteins, the lack of phase information is very disadvantageous. To address the phase problem there are essentially three approaches: the Patterson method, the direct method or the use of anomalous dispersion. We will not go in detail into these relatively complex methods here, but only briefly describe the basic procedure.

In the **Patterson method**[10], a Fourier transformation of the *scattering intensity* into real space is performed. The result is a map of maxima whose positions are determined by the position vectors, each of which connects two atoms of the base. In the Patterson map, the height of each maximum is proportional to the product of the electron density of the two atoms involved. Since the number of maxima increases quadratically with the number of atoms in the base, the method in its original form is only applicable to crystals with a relatively small base.

The phase problem can be significantly reduced for very large molecules by the incorporation of heavy atoms, by *isomorphic substitution*. Since these atoms scatter particularly strongly owing to their large number of electrons, they thus provide reference points in the Patterson map. Since there are only a few heavy atoms, their positions can be determined. With these reference points, further analysis of the structure of large molecules is then possible.

The **direct method** exploits statistical relationships between sets of structural factors to derive possible values for each phase. Starting from a few reflections whose phases are known (starting reflections), the unknown phases of the other reflections can be estimated with a certain probability. The probability that the structure thus determined is the correct one increases with the number of reflections available for the solution. If only a few are measured, the corresponding program cannot find the correct solution or can find it only with great difficulty. With this mathematically very complex procedure the structure of the basis can be decoded, even if it contains a few hundred atoms.

If measurements are carried out at two different wavelengths, where one wavelength must be close to the resonant absorption of one type of atom, **anomalous dispersion** is used. In this case, the scatter density distribution and thus the scatter amplitude is no longer real but complex. The result provides an additional independent data set with which the missing phase information can be found.

In neutron scattering, the fact that different isotopes of an element have different scattering properties is often used. Therefore, measurements on samples with two

10 Arthur Lindo Patterson, *1902 Nelson/Neuseeland, †1966 Philadelphia

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

different isotope components yield two independent data sets providing different structural information. A very important example is provided by the replacement of hydrogen by deuterium.

In the final part of this section we refer briefly to the width of the diffraction peaks. As already mentioned, the dispersion of the incident rays and their divergence, contribute to the width, as do the finite penetration depth of the rays and the disorder in the crystals. In this context it is an interesting question whether the width of the peaks also depends on the temperature. We are tempted to initially answer this question with a yes, since the thermal movement changes the position of the atoms and thus also the reciprocal lattice in time. On average, however, the periodicity of the lattice remains the same. Experimentally, we find that the thermal motion does not influence the width of the diffraction peaks, but reduces its *intensity*. We will discuss this observation ("The Debye-Waller factor") in Section 6.3.

4.5 Scattering Experiments on Amorphous Solids

Owing to their irregularity, the structure of amorphous solids is much more difficult to access, both theoretically and experimentally, than that of crystals. In crystals, the incident radiation is only diffracted constructively by the lattice planes in very specific directions. Since there are no regular lattice planes in amorphous materials, scattering occurs in all directions. However, the scattering intensity *is* modulated by the near-order of the atoms and depends on the scattering angle. In the following we look for a relation between the scattering intensity and the pair distribution function, which describes the structural composition of amorphous materials according to (3.17). The procedure is similar to that of the derivation of the structure factor of crystals. We assume spherical atoms whose scattering strength is described by the atomic structure factor. Thus the scattering amplitude $\mathcal{A}(\mathbf{K})$ caused by an atom m with the position vector $\mathbf{r}_m$, can be written in the form $\mathcal{A}_m(\mathbf{K}) = f_m \exp(\mathrm{i}\mathbf{K}\cdot\mathbf{r}_m)$. This scattering contribution interferes with that of all other atoms n at the locations $\mathbf{r}_n$. The scatter intensity of the whole sample, expressed by $|\mathcal{A}(\mathbf{K})|^2$, is obtained by summing the scattering amplitudes $\mathcal{A}_m(\mathbf{K})$ of all atoms and calculating the square value:

$$|\mathcal{A}(\mathbf{K})|^2 = \sum_m f_m(\mathbf{K})\, \mathrm{e}^{\mathrm{i}\mathbf{K}\cdot\mathbf{r}_m} \sum_n f_n^*(\mathbf{K})\, \mathrm{e}^{-\mathrm{i}\mathbf{K}\cdot\mathbf{r}_n} = \sum_{m,n} f_m(\mathbf{K})\, f_n^*(\mathbf{K})\, \mathrm{e}^{-\mathrm{i}\mathbf{K}\cdot\mathbf{r}_{mn}} . \tag{4.37}$$

For the occurrence of interferences in the scattered radiation, it is not the absolute but the relative coordinates of the atoms which are crucial. This quantity appears in the last part of the above expression as the vector $\mathbf{r}_{mn} = (\mathbf{r}_n - \mathbf{r}_m)$. Equation (4.37) applies to arbitrarily structured samples. For crystals, the summation over the vectors $\mathbf{r}_{mn}$ can be carried out with the help of the reciprocal lattice, since there the translation invariance of the lattice can be used. In the case of amorphous solids or liquids, this summation cannot be performed in its entirety.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

To simplify matters, in the following discussion we assume that the substance concerned consists of only one type of atom. (Extension to multi-component systems is conceptually possible without much effort.) For a *single* type of atom, equation (4.37) simplifies to:

$$|\mathcal{A}(\mathbf{K})|^2 = \sum_m f^2(\mathbf{K}) + \sum_m \sum_{n \neq m} f^2(\mathbf{K})\, e^{-i\mathbf{K}\cdot\mathbf{r}_{mn}}, \tag{4.38}$$

where the contribution of the atoms m is written separately. In evaluating this expression we go from the sum to an integration via the particle number density $n_m(\mathbf{r})$. Since each atom in amorphous solids has on average the same environment, $n_m(\mathbf{r})$ depends neither on the direction nor on the choice of atom. We therefore replace the sum $\sum_n$ by the corresponding integral and obtain:

$$|\mathcal{A}(K)|^2 = \sum_m f^2(K) + \sum_m f^2(K) \int_{V_s} \mathrm{d}r \int_0^\pi \mathrm{d}\vartheta\, n(r)\, 2\pi r^2 \sin\vartheta\, e^{-iKr\cos\vartheta}, \tag{4.39}$$

where ϑ represents the polar angle between $\mathbf{K}$ and $\mathbf{r}$ and we integrate over the entire sample volume V_S. We have already met this integral in equation (4.31) when calculating the atomic scattering factor, but there instead of the particle number density $n(r)$ we had the atomic scattering density distribution $\varrho(r)$. Taking the result from (4.31) we get:

$$|\mathcal{A}(K)|^2 = \sum_m f^2(K) + \sum_m f^2(K) \int_{V_s} 4\pi r^2 n(r)\, \frac{\sin Kr}{Kr}\, \mathrm{d}r\,. \tag{4.40}$$

As summing over the index m yields the number N of atoms present, equation (4.40) simplifies to:

$$|\mathcal{A}(K)|^2 = N f^2(K) \left[1 + \int 4\pi r^2 n(r)\, \frac{\sin Kr}{Kr}\, \mathrm{d}r\right]. \tag{4.41}$$

The quantity $n(r)$ is identical to the particle density function as used in the definition of the pair correlation function $g(r)$ in Section 3.4. As we saw there, in amorphous solids the two quantities are directly linked via $g(r) = n(r)/n_0$, where n_0 is the mean particle number density.

Since the integral extends over the entire sample volume, it therefore contains information not only about the average environment of the reference atom, but also about the whole sample. The contributions of the neighbors, the subject of interest here, can be separated from those of all the other atoms by a simple device. We simply subtract the mean particle number density n_0 from $n(r)$ and add it separately. After suitable summing of the terms and division by $N f^2(K)$ we find:

$$\frac{|\mathcal{A}(K)|^2}{N f^2(K)} = 1 + \underbrace{\int 4\pi r^2 [n(r) - n_0]\, \frac{\sin Kr}{Kr}\, \mathrm{d}r}_{\text{Local structure}} + \underbrace{\int 4\pi r^2 n_0 \frac{\sin Kr}{Kr}\, \mathrm{d}r}_{\text{Distant atomes}}. \tag{4.42}$$

The first term, the quantity unity, describes the scattering contribution of the reference atoms and contains no interference effects. The factor $[n(r) - n_0]$ in the second term

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

disappears after a few atomic distances, since $n(r)$ rapidly approaches the value n_0 (see Section 3.4). Therefore, only neighboring atoms contribute to this scattering fraction. The interference term thus reflects the local structure around the (arbitrarily chosen) reference atom. The main contribution to the third term arises from atoms at a large distance from the reference atom, because of the factor r^2. This contribution is determined by the macroscopic properties of the sample. The factor $(\sin Kr)/Kr$ means that this scattering contribution only occurs in practice in the forward direction, i.e. at $K \approx 0$. Even at relatively small values of the scatter vector, the integral oscillates so rapidly that the integral disappears. In most scattering experiments the intensity in the immediate vicinity of the transmitted beam is deliberately blocked by an aperture and not further resolved. Since this scattering contribution, known as **small-angle scattering**, is insignificant for the determination of the local order, we will leave it out of consideration in the subsequent discussion.[11]

A closer look at equations (4.37) and (4.42) shows that $|\mathcal{A}(K)|^2$ is not directly proportional to the Fourier transform of the pair correlation function. The reason for this is that the reference atoms, for which the interferences between different scattering contributions do not play a role, also contribute to the scattered signal. They are described in (4.42) by the first term, with simple value unity, or by Nf^2. In fact, $|\mathcal{A}(K)|^2/Nf^2(K)$ is the Fourier transform of the autocorrelation function of the scatter density distribution. Since $|\mathcal{A}(K)|^2$ is a measured quantity and the atomic structure factor $f(K)$ is known from other measurements, the autocorrelation function and thus the pair correlation function can also be obtained directly from the scatter data with the help of a Fourier transformation.

We will now briefly discuss the evaluation of the experimental data. As already mentioned, in the simple case of an amorphous solid composed of only one kind of atom, the autocorrelation function, and thus the pair correlation function $g(r)$, can be obtained by the Fourier transformation of $|\mathcal{A}(K)|^2$. Very often, however, rather than $g(r)$, it is the radial density distribution function RDF that is given, and which is linked to the particle density via the relationship $\mathrm{RDF} = 4\pi r^2 n(r)$. The RDF indicates the average number of atomic centers located on a thin spherical shell with radius r around the reference atom. At large distances from the reference atom, since $n(r)$ approaches n_0, the value of the RDF increases with the square of the radius of the atomic shell.

Figure 4.20 shows a simple example of the RDF of amorphous germanium and also, for comparison, a corresponding curve for polycrystalline germanium, which, due to the random orientation of its small crystallites, also has a structure independent of direction on average. The first maximum occurs in both cases at about 2.4 Å and has the same appearance. This means that both polycrystalline and amorphous germanium have the same distance to, and the same number of, nearest neighbors. Thus the order of

11 Of course there are also applications for the investigation of larger structures. Special experimental techniques have been developed for necessary small angle investigations.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the nearest neighbors is nearly as high in the amorphous phase as in crystalline material. The broadening of the second maximum in the case of the amorphous material indicates variations in the angle of bonding, which leads to a distribution in the distance between the next nearest neighbors. In the amorphous material further coordination shells become less and less pronounced. At large distances the radial density distribution function gradually loses all structure. These measurements confirm the ideas about the structure of amorphous solids which we developed in Section 3.4.

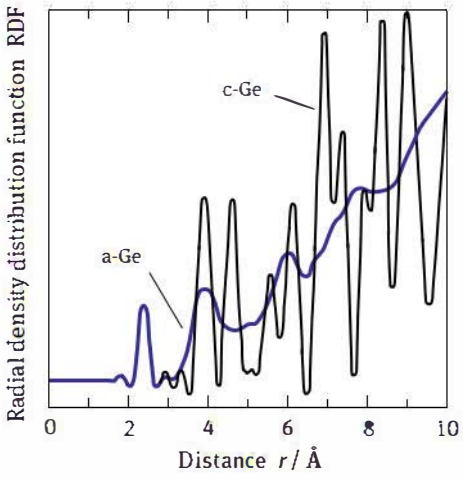


Fig. 4.20: Radial density distribution function of amorphous (blue) and crystalline (black) germanium. (After R.J. Temkin et al., Adv. Phys. **22**, 581 (1973).)

It should be noted that in scattering experiments an averaging over all existing atomic arrangements always takes place. Furthermore, the pair correlation function or the RDF basically contains a one-dimensional representation of a three-dimensional structure. However, the determination of this apparently simple function already causes considerable difficulties. Theoretically, the Fourier transformation requires knowledge of the exact dependence of $|A(K)|^2$, but experimentally, the scattering intensity cannot be determined for all values of the scattering vector K. In addition, measurements at very small and very large scattering vectors prove to be difficult and have large uncertainties. The unavoidable restriction of the measuring range and the ever-present measuring errors can give rise to additional structures in the autocorrelation function during the Fourier transformation which are difficult to distinguish from "real" structural features.

A further fundamental difficulty arises in the investigation of multicomponent systems. In the case of a two-component substances with two atomic types A and B, such as vitreous silica a-SiO_2, one function is not sufficient to describe the structure, because there are three partial pair correlation functions: In addition to the two functions which reflect the correlation between the A and B atoms, there is also the mixed pair distribution function describing the distribution of one type of atom with respect to the other. For the calculation of these three functions, three independent sets of scattering

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

data are required. Although measurements using neutrons and X-rays provide two independent sets of data, investigations using electron or atomic beam diffraction provide no additional information because the scattering is from the electron shell like in case of X-rays. In exceptional cases, the different scattering properties of dissimilar isotopes allow for two independent measurements with neutrons, like for example, measurements on the metallic glass $Ni_{76}P_{24}$ with different concentrations of $^{58}_{28}Ni$ or $^{60}_{28}Ni$. Since the scattering lengths of the two isotopes clearly differ, in this particular case it was actually possible to create a complete set of partial radial density distribution functions.

Figure 4.21 shows X-ray and neutron scattering measurements on vitreous silica. In both cases the scattering intensity is plotted in arbitrary units as a function of the scattering vector K. For X-ray scattering, the scattering intensity is proportional to $(Ze)^2$, whereas for neutron scattering it depends on the scattering length b, $I(K) \propto |b|^2$. It is noticeable that the intensity of the X-ray radiation decreases rapidly with scattering vector. The reason for this is the angle dependence of the atomic structure factors, which was shown in Figure 4.18. As already mentioned, in neutron scattering the corresponding factor, the scattering length, does not depend on the scattering angle, so that structures can still be resolved even with scattering vectors $K > 40\,\text{Å}^{-1}$. It is also noticeable that the periodic variations caused by the pair distribution function are very different in intensity. This is due to the fact that silicon and oxygen atoms behave differently with respect to X-rays and neutrons. In experiments with X-rays, the

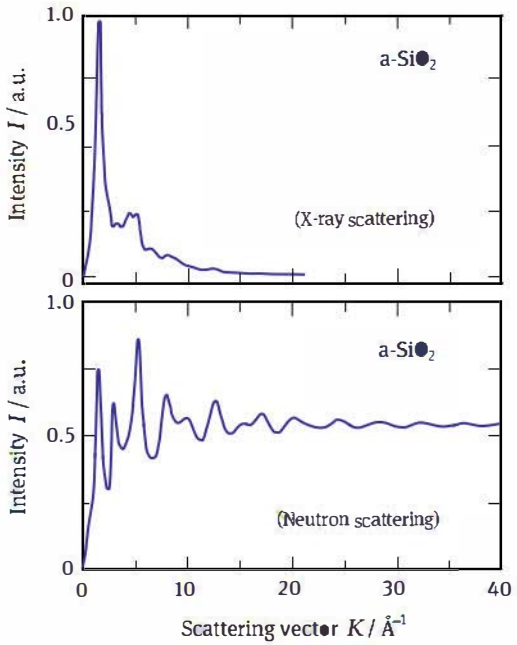


Fig. 4.21: X-ray and neutron scattering by vitrous silica. The scattering intensity I is shown in arbitrary units as a function of the scattering vector K. (After R.L. Mozzi, B.E. Warren, J. Appl. Cryst., **2**, 164 (1969) (upper picture) and A.C. Wright, J. Non-Cryst. Solids **179**, 84 (1994) (lower picture).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

scattering at the silicon atoms predominates, since they scatter about three times more than oxygen atoms. For neutrons, however, $|b|^2$ is determining. Therefore the neutron scattering signal of the oxygen atoms is about twice as large as that of the silicon atoms.

How do we arrive at the structure of the sample from the scattering data $|\mathcal{A}(K)|^2$? Even with single-component systems this requires considerable numerical effort. In general, we proceed in several steps. First, we assume a probable model of the structure. Its pair correlation function or autocorrelation function is determined and the expected scatter intensity is calculated with the help of the Fourier transformation. A comparison with the measured curve will probably show deviations. The model is then improved, a new comparison made, and the same procedure followed until agreement between the measured and the calculated curve is satisfactory. Additionally, data from other structure-sensitive experiments are usually used. Of course, even with the best agreement, there is still no proof that the model actually corresponds to the real structure, because pair correlation functions are often quite ambiguous. Under certain circumstances, very different structural models can lead to very similar correlation functions. Currently, *Monte Carlo methods* and *molecular dynamics simulations* can also play an important role in the evaluation of scattering data. The large numerical effort required can now be accomplished with modern computers within reasonable times.

4.6 Experimental Methods

At the beginning of this chapter we incides that structural investigations can be made with various probes. For determining the structure of extended samples, X-ray and neutron diffraction are the most suitable methods. Measurements can be made with a continuous or a monochromatic radiation spectrum depending on the technique used. However, since more detailed investigations are generally carried out with monochromatic radiation, we sketch a scheme for such a scattering experiment in Figure 4.22. In order to pick out a narrow wavelength band from the broad spectrum of the source,

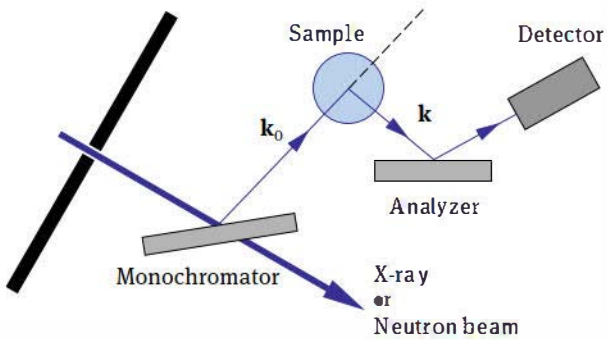


Fig. 4.22: Schematic setup for a scattering experiment with monochromatic radiation, using X-rays or neutrons as probes.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Bragg reflection from a single crystal is often used. Here the polychromatic X-ray or neutron beam is incident on a *monochromator crystal* where, as discussed earlier, most of the incident beam passes through this crystal almost without diffraction, but that component with the "requisite" wavelength is reflected at the Bragg angle. This part of the beam with wave vector $\mathbf{k}_0$ is then used for the scattering experiment.

Reflected from the monochromator the beam impinges on the sample where it is diffracted. The scattered radiation with wave vector $\mathbf{k}$ is either detected directly, or falls onto a second single crystal (known as the *analyzer crystal*) where it undergoes a further Bragg reflection. In the case of inelastic scattering, discussed in detail in Section 6.3, the analyzer crystal is used to measure the energy loss of the scattered radiation. For investigations of elastic scattering (where there is no change of wavelength) an analyzer crystal is not strictly necessary, but can be used to suppress any inelastic scattering background.

Due to the relatively low technical effort required, X-rays from X-ray tubes are mostly used for structure determination, where, depending on the measuring method, the continuous bremsstrahl spectrum or the characteristic X-rays is used. For some time now, synchrotrons specially designed for structural investigations have become available. Synchrotron radiation is characterized by its wide wavelength range, its high intensity with small beam divergence, and complete polarization. For a long period, X-rays were exclusively detected photographically. Nowadays however, semiconductor detectors, scintillation counters, gas discharge counters or cryogenic micro-calorimeters are used.

As we have seen, the intensity of the X-ray reflections is proportional to the square of the atomic structure factor. Since this factor increases with the number of electrons, the scattering increases quadratically with the nuclear charge number of the scattering atoms. As a result, light elements are hardly observable, especially in the presence of heavy elements, and require particularly sensitive methods for detecting the scattering intensities. Hitherto, hydrogen atoms could not be detected at all in many cases. Recently, however, modern detectors and synchrotron radiation sources have been used for this purpose and have brought significant progress, such that today hydrogen compounds, which of course play a major role in polymer physics and in biological investigations, can now be studied in X-ray scattering experiments.

In neutron scattering, conditions for detecting light atoms are much more favorable, since the absolute value of the scattering length, while varying from nucleus to nucleus, is of the same order of magnitude for all elements. However, we need to take into account that most nuclei scatter also to some extend incoherently, thus giving rise to a scattering background and are therefore not equally suitable for structural investigations. The incoherent scattering cross section is particularly large with hydrogen. Nontheless, until recently the position of hydrogen atoms in solids has been mainly investigated with neutrons, making use of the weak coherent scattering contribution. A significant improvement of the signal can be achieved by replacing the light hydrogen by deuterium, where the ratio between coherent and incoherent scattering cross section is substantially better.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

A further difficulty in X-ray diffraction is differentiating between elements that are adjacent in the periodic table, since their atomic scattering factors are similar. An example was shown in Section 4.4: In caesium iodide, the two ion types are hardly distinguishable in X-ray diffraction experiments. The diffraction pattern of CsI is therefore similar to that of a body-centered cubic "xenon crystal" with a monatomic base.[12] With neutrons, on the other hand, atoms of neighboring elements can usually be readily distinguished, since the nuclei have different scattering properties. Another important field of application of neutron scattering is the investigation of magnetic structures (see Section 12.4), which is based on the interaction of neutrons with the magnetic moments of the electron shell.

Suitable neutron beams can be obtained either from nuclear reactors designed to produce a particularly high neutron flux, or from spallation sources, in which a beam of high-energy protons or electrons is fired from an accelerator at a heavy metal target. Since in both cases the neutrons produced have too high an energy (i.e. too short a wavelength) for structural investigations, they must first be slowed down in a moderator. There they reach thermal velocities and thus an appropriate wavelength for diffraction experiments on solids. As described above, with the help of Bragg reflection, a narrow velocity range can be extracted from the wide (thermal) velocity distribution, i.e. a "monochromatic" neutron beam is generated. The scattered neutrons are detected with suitable counter tubes, for example, with BF_3 proportional counters, in which ^{10}B nuclei are converted by neutron capture into ^{7}Li and ^{4}He nuclei, which ionise the gas in the counter. More recently ^{3}He detectors are becoming more common, where a ^{3}He nucleus captures a neutron, then splitting into a ^{3}H nucleus and a proton which can be detected. With the increasing development of measurement techniques, structural investigations can be carried out on ever smaller samples. For neutron investigations, samples with a volume of about 1 mm^3 are typically required. For measurements with X-rays the samples can be considerably smaller.

4.6.1 Measurement Procedures

The diffraction condition (4.21) involves two quantities which can be varied in the experiment, wavelength and crystal orientation. The latter determines the scattering angle for a given direction of the incident wave. If we want to find reflections, at least one of the two quantities must be changed during the measurement. Therefore, most experiments can be assigned to one of the three measuring methods discussed in the following, although the measuring techniques and the respective equipment are completely different depending on the nature of the probe.

12 Real xenon crystals have face-centered cubic structure.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The Rotary Crystal Method. In the rotary crystal method, which was introduced by *M. de Broglie*,[13] monochromatic radiation and single crystals are used. In the simplest case, the crystal is rotated around one axis, as shown schematically in Figure 4.23 for X-ray diffraction. During rotation, the reciprocal lattice rotates with the crystal. A reflection always occurs when a group of planes of the rotating lattice fulfills the Bragg condition. Since this can occur repeatedly for certain reciprocal lattice points as we rotate the crystal completely around 360°, the angle of rotation can usually be restricted. Reflections originating from lattice planes parallel to the vertical axis of rotation are not defracted in the horizontal plane. If the reflecting lattice planes are inclined with respect to the axis of rotation, the resulting reflected beams have a vertical component. In modern equipment, the reflections are not, as shown schematically in figure, recorded on film, but with X-ray detectors as discussed above.

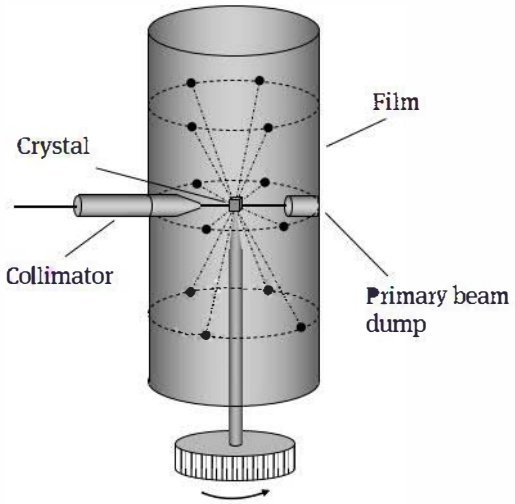


Fig. 4.23: The rotary crystal method. Schematic representation of the experimental setup and the diffration peaks observed on the film.

There are various techniques that aim to assign reflections as clearly as possible. These include the *Weissenberg-Böhm-goniometer*[14,15] and the *precession camera*, in which the film is moved synchronously with the crystal rotation, preventing overlapping of the reflections. In all cases the evaluation of the measured data is considerably simplified if the axis of rotation of the examined crystal coincides with an axis of high crystal symmetry. To achieve this, the crystal is usually oriented beforehand using the Laue method, as discussed below.

13 Louis-César-Victor-Maurice, de Broglie, *1875 Paris, †1960 Neuilly-sur-Seine
14 Karl Weissenberg, *1893 Vienna, †1976 The Hague
15 Johann Böhm, *1895 Budweis, †1952 Prague

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The Powder Method. The powder method, also called *Debye-Scherrer*[16,17] *method*, is particularly suitable for the precise determination of lattice constants. In fact, most available lattice data have been obtained with this method. In this method, finely powdered crystalline or fine-grained polycrystalline samples are irradiated with monochromatic waves. As shown schematically in Figure 4.24a for the case of X-ray diffraction, the beam, collimated by an aperture, falls on the sample. Crystallites whose lattice planes are by chance positioned to satisfy a particular Bragg condition give rise to reflections at an angle of 2θ to the primary beam. Owing to the random orientation of the crystallites, the diffracted rays lie on a conical shell concentric with the primary beam as indicated in the figure. In the simplest case, the scattered radiation is detected in the form of *Debye-Scherrer rings* on X-ray film. A Debye-Scherrer image of Cu_3Au, an alloy which is discussed in Section 5.4, is shown in Figure 4.24b.

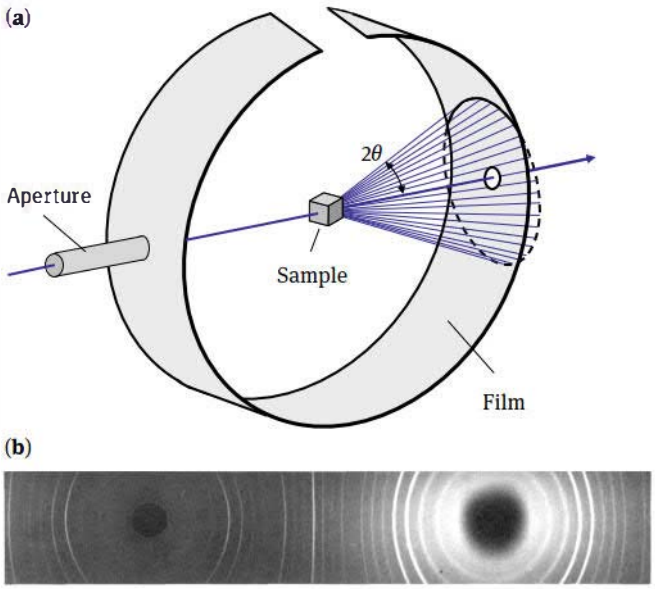


Fig. 4.24: **a)** Schematic representation of the Debye-Scherrer method. Monochromatic X-rays strike the sample at the center. The conically scattered X-rays are collected by a film strip. **b)** Powder image of an ordered Cu_3Au alloy. The diffraction rings were recorded on film as in the sketch. (After Ch. Kittel, *Introduction to Solid Physics*, Oldenbourg, 2013).

Figure 4.25a shows the scattering intensity of crystalline quartz as a function of the scattering angle. The data was recorded with a powder diffractometer, where the intensity of the reflections was recorded with a counter. Of course, the Debye-Scherrer

16 Peter Debye, *1884 Maastricht, †1966 Ithaca, Nobel Prize 1936
17 Paul Scherrer, *1890 St. Gallen, †1969 Zurich

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

method can also be used for neutron scattering. The result of such a measurement on a $YBa_2Cu_3O_7$ powder sample is shown in Figure 4.25b. This material is one of the high temperature superconductors, which we will discuss in detail in Section 11.5.

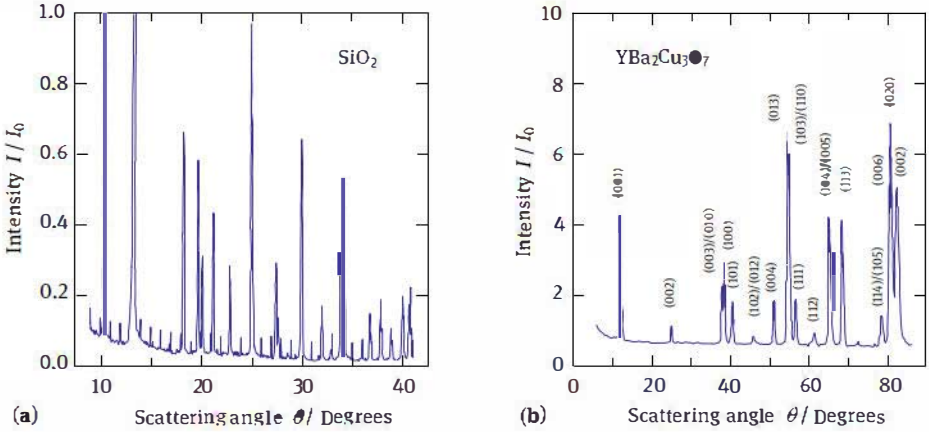


Fig. 4.25: Debye-Scherrer images. **a)** Powder diffractometer image from a quartz crystal sample. (Author unknown.) **b)** Neutron diffractogram of a $YBa_2Cu_3O_7$ powder sample. The designation of the diffration peaks refers to the orthorhombic unit cell. (According to G. Schatz, A. Weidinger, Nuclear Solid State Physics, Teubner, 1992.)

The angular dependence of the scattering intensity in powder images is very similar to that which occurs in diffraction experiments on amorphous materials. In powder scattering the beam averages all crystal orientations, in amorphous solids the beam averages all local atomic arrangements. Thus we can see that an accumulation of randomly oriented crystallites causes similar scattering to that from amorphous solids. Nevertheless, there is a significant difference between these two classes of substances. Since the crystallites show not only short-range, but also long-range order, the observed rings are not broadened but sharp.

The Laue Procedure. In the Laue process, the single crystal is irradiated with a continuous spectrum, so that many crystal planes simultaneously fulfill equation (4.21) and cause reflections. If on irradiating along an axis of symmetry of the crystal, a diffraction image with the symmetry elements of that axis is produced. Therefore, for a crystal with an unknown structure, a photograph is first taken in a random direction to determine the main axes of symmetry. Then we radiate along these axes and thus determine the crystal symmetry from the Laue images. Laue images are also used for orientating known crystals. Figure 4.26 shows as an example the Laue image of a lectin crystal taken with a synchrotron X-ray source. This is a complex protein molecule that occurs in peas.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

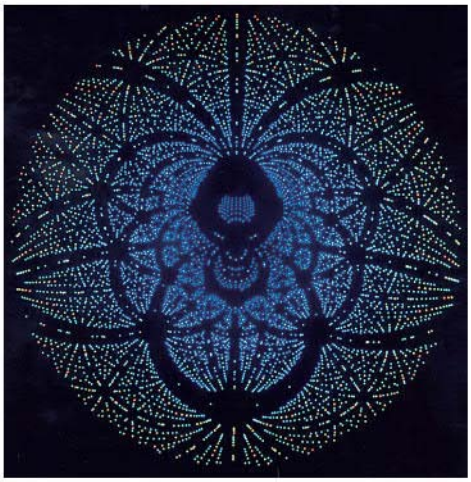


Fig. 4.26: Laue image of a lectin crystal. The two-fold symmetry of the crystal is clearly visible. (The image was taken from a report of the CCLRS Synchrotron Radiation Source, Daresbury Laboratories, United Kingdom).

4.6.2 Measurements on Surfaces and Thin Films

Since electron or atomic beams interact particularly strongly with solid samples, only very thin layers can be investigated in transmission experiments. For this purpose electrons are accelerated to about 50–100 keV before striking the sample, usually in the form of a foil. Due to their short wavelength, the scattering only takes place into a small opening angle of about 3–5°. After passing through the foil the electrons are detected. As an example for such a measurement, Figure 4.27 shows the scattering at an iron layer about 9 nm thick, where the electrons were recorded on a photographic plate.

The iron film was produced by evaporation at 4.2 K, i.e. at liquid-helium temperature, and then irradiated. When warmed to room temperature, the initially amorphous film anneals into a polycrystalline layer of crystallites with no preferential orienta-

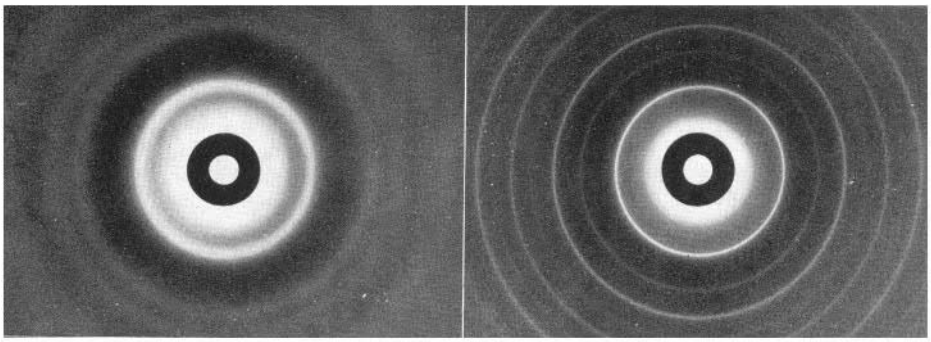


Fig. 4.27: Electron diffraction images of a 9 nm thick iron film in the amorphous and the crystalline state. (After T. Ichikawa, phys. stat. sol. (a) **19**, 707 (1973).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

tion. Due to this phase transition, the smeared rings become as sharp as those from Debye-Scherrer photographs of crystalline powders. The pair correlation function of amorphous nickel shown in Figure 3.39 was obtained from similar data.

The diffraction of low-energy electrons with energies between 10 eV and 1000 eV is particularly suitable for the investigation of surfaces or near-surface layers. It is a standard method in surface physics known as *LEED* (**L**ow **E**nergy **E**lectron **D**iffraction). In this case, only electrons that are backscattered within the first atomic layers of the sample are detected. Since the scattered electrons are very sensitive to surface contaminations, the experiments are usually performed in ultra-high vacuum ($p < 10^{-8}$ Pa). A typical experimental arrangement is shown schematically in Figure 4.28a. The electrons are accelerated by a voltage U before striking the sample. The reflected electrons pass through two grids whose potential is adjusted so that only those electrons that have lost minimal energy on contact with the sample can pass. They are then accelerated and made visible on a hemispherical fluorescent screen.

As an example, Figure 4.28b shows the diffraction image of a (111)-surface of face-centered cubic platinum. Here we should point out an important difference between LEED and X-ray scattering measurements. The strong scattering and absorption of electrons means that the backscattered electrons come from a very thin layer on the surface, so that the sample volume can be regarded as a two-dimensional structure. The third direction in space, i.e. the direction perpendicular to the surface, is practically irrelevant. Therefore, we are working under the conditions we have seen in Section 4.4 for the scattering of electrons at surfaces. Since the diffraction condition in two-dimensional samples has a much less restrictive effect, even with monoenergetic electrons multiple reflections may occur, arising from different reciprocal lattice vectors. LEED-images with monoenergetic electrons superficially show a certain similarity to the Laue images discussed above.

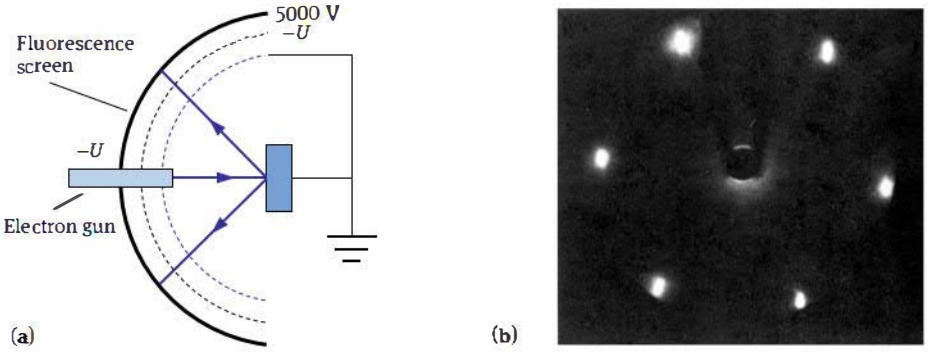


Fig. 4.28: a) Schematic arrangement for observing LEED reflections. **b)** The diffraction image of a Pt(111) crystal surface taken at an electron energy of 51 eV. (After G.A. Somorjai, *Chemistry in two dimensions: Surfaces*, Cornell University, 1981).

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Direct Imaging of the Surface. In Figure 4.1, at the beginning of this chapter, we saw that by using a **transmission electron microscope** (TEM) thin layers can be imaged with atomic resolution. The electrons penetrate the sample and are then detected on a detector plate. The resolution limit d is given by the Rayleigh criterion[18]

$$d = \frac{\lambda}{2A_N} \approx \frac{0.6\text{nm}}{A_N\sqrt{U}}\,, \tag{4.43}$$

where A_N stands for the numerical aperture. It is important to note that the wavelength of the electrons depends on the acceleration voltage U (as $\lambda \propto U^{-1/2}$). Consequently, it should be possible to achieve sub-atomic resolution with voltages around 100 kV. Despite the theoretical limit not having been reached so far, owing to lens aberrations, modern devices still achieve a resolution of about $d \approx 0.05$ Å.

The direct imaging of atoms on a surface is possible by means of **scanning probe microscopy**, where the resolution limit is in the Ångström range. The best-known instrument in this category, the Scanning Tunneling Microscope (STM), invented by *G. Binnig*[19] and *H. Rohrer*[20] is shown schematically in Figure 4.29a. In this device, a sharp metal tip, from which ideally a single atom "protrudes", is moved over the conductive sample surface at a distance of about one nanometer. The position of the tip can be controlled with picometer accuracy by piezoelectric rods. A bias voltage is applied to the sample relative to the tip and the current between tip and sample is measured. This is the tunnelling current which depends exponentially on the distance between the tip and the sample. If the STM is operated in the so-called feedback mode, as the tip moves over the surface the current is maintained constant by adjusting the tip distance. In this way, the tip follows the topography of the surface, which is recorded.

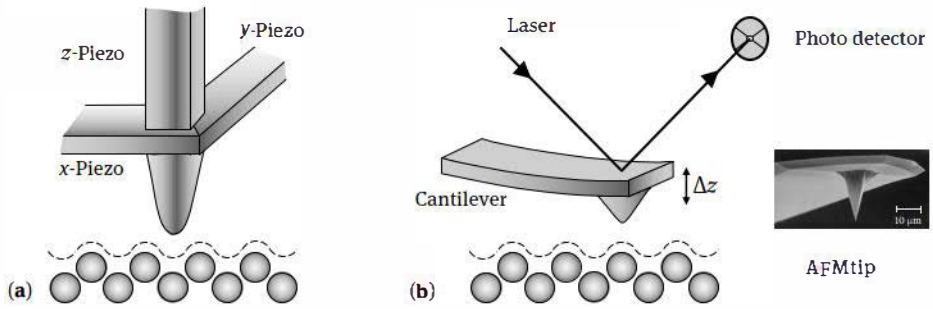


Fig. 4.29: **a)** Scheme of a scanning tunneling microscope. In feedback mode the piezo elements move the tip over the sample in such a way that the current remains constant. **b)** Schematic of a scanning force microscope. The deflection of the cantilever is monitored by the position of the laser beam.

18 John William Strutt, Lord Rayleigh, *1842 Maldon, †1919 Witham, Nobel Prize 1904
19 Gerd Binnig, *1947 Frankfurt am Main, Nobel Prize 1988
20 Heinrich Rohrer, *1933 Buchs, Sankt Gallen, †2013 Wollerau, Nobel Prize 1988

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The resolution limit is about 0.1 nm in the lateral and about 0.01 nm in the vertical direction.

The similar **A**tomic **F**orce **M**icroscope (AFM) was developed shortly after the STM. This device measures the force between tip and sample and can be used on both conductive and insulating samples. As shown in Figure 4.29b, a sharp tip is attached to the end of a millimeter-sized carrier, usually called a cantilever. The force exerted by the sample on the tip causes the cantilever to bend, which can be measured by using the back of the cantilever as a reflector for a laser beam, which is then detected by a photo detector. This kind of microscope can achieve a resolution of about 0.3 nm in both horizontal and vertical directions. We have already seen the result of such a measurement in Figure 4.2, in which the surface of silicon is imaged by an AFM. Several different operating modes have been developed for AFMs, but we will not pursue the details here.

4.7 Exercises and Problems

1. Miller Indices. "Nanowires" are often prepared on crystal surfaces that are slightly inclined with respect to a plane with low indices.

(a) Calculate the angle between the (1 20 0) and the (010)-plane of a cubic crystal.
(b) What is the distance between the (120 0) planes for copper (face-centered cubic) with a lattice constant $a = 3.61$ Å?
(c) Calculate the minimum photon energy required to produce a first order X-ray diffraction peak?

2. The Ewald Sphere. Thallium iodide (density $\varrho = 7.29\,\mathrm{g/cm^3}$) exhibits a cesium chloride structure above 441 K.

(a) In a diffraction experiment, X-rays are incident on the sample along the [010]-direction. The scattered radiation is detected in the $[3\bar{1}0]$-direction. At which wavelength and at which angle is the reflection observed?
(b) Are there any other reflections under the given conditions?
(c) Can the $[1\bar{9}0]$ reflection be observed?

3. Bragg Reflection. On a copper powder at room temperature ($T = 293$ K) a reflection at the Bragg angle $\theta = 47.75°$ is observed. What is the coefficient of thermal expansion when at $T = 1293$ K the angle is 46.60°?

4. Two-dimensional Hexagonal Lattice. Specify the two lattice vectors of the two-dimensional hexagonal lattice in Cartesian coordinates and display the Wigner-Seitz cell graphically. Determine the reciprocal lattice vectors and draw the corresponding lattice. Construct the first Brillouin zones and calculate their areas.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

5. Structural Factor of Perovskite. As mentioned in Section 3.3 perovskite ($CaTiO_3$) has a simple cubic lattice.

(a) Calculate the structure factor of perovskite.
(b) What shape do $\mathcal{S}_{100}$, $\mathcal{S}_{200}$ and $\mathcal{S}_{110}$ have?
(c) Which of the structural factors is the largest, which the smallest? To answer this question exactly, the atomic scattering factors would have to be known. For a simple estimation we can assume that the atomic form factors increase proportionally to the nuclear charge.

6. Body-centered Cubic Lattice. There is a clear connection between the lattice in real and reciprocal space. Here we take a closer look at the body-centered cubic lattice.

(a) Specify three vectors that represent the primitive unit cell in real space.
(b) Calculate the vectors of the primitive unit cell of the reciprocal lattice.
(c) Show that the reciprocal lattice has cubic symmetry.

7. Powder Diffraction. The powder of a material with cubic symmetry is measured with X-ray diffraction at wavelength 1.54 Å. In forward direction rings are observed at the angles 28°, 32°, 46°, 54° and 57°, as well as in backward direction at 168°, 160°, 146°, 133° and 130°.

(a) Calculate the value of $\sin^2\theta$ for the different rings and compare the values with those expected for the different cubic lattices. Which cubic lattice is it?
(b) Assume that the smallest observed ring in the forward direction is caused by the reflection with the smallest Miller indices and calculate the lattice constant.
(c) To determine the lattice constant, use the ring with the smallest diameter in the reverse direction.

8. Laue Method. A polonium single crystal (α-Po) is examined with the Laue method. Estimate the maximum number of diffraction peaks when the X-ray tube is operated at 40 kV. α-Po has a simple cubic lattice and has the atomic mass 209.98u and the density 9196 kg/m^3.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

5 Structural Defects

In previous chapters we have always been dealing with ideal crystals or ideal amorphous solids. However, many physical properties of real materials arise from defects, that is, deviations of the material from the ideal. Already their finite size can influence their properties significantly. Furthermore, we have to remember that the perfect arrangement of the approximately 10^{23} atoms in a macroscopic crystal will in any case be degraded by the simple action of entropy which means the appearance of defects in real crystals is inevitable.

In this chapter, we first discuss the most important types of defects, both local and extended and examine in detail many of their properties in the context of ordered crystalline solids. Then we move on to discuss the rather more difficult subject of defects in amorphous materials.

Sadly for a book of this scope, these are major topics in themselves and we have to omit for the most part the experimental methods used for investigating defects in solids and also the material-science aspects of defects which are of outstanding importance in practical applications.

Defects in real crystals are local or extended deviations from the perfect periodic arrangement of the lattice. Besides point defects, typically of atomic size, macroscopically extended defects such as line and planar defects, also occur in crystals. Naturally, the definition of defects in amorphous materials is much more difficult. While we find point defects similar to those in crystals, and also new types of defects such as unsaturated or broken bonds, normally absent in crystals, amorphous solids show no line or planar defects analogous to those in crystals.

5.1 Point Defects

Point defects exert a considerable influence on the properties of solids. They are caused by atoms being removed from, added to, or exchanged within a crystal. Accordingly, we can distinguish three types of point defect: *vacancies*, *interstitials* and *impurities*. As shown in Figure 5.1, in the vacancy case an atom is missing, while interstitials are formed when an extra added atom is inserted *between* regular lattice sites. Impurities can occupy both regular sites, and also interstitial sites. We also speak of point defects, even if several atoms are involved, provided that the irregularity of the lattice is limited to a few atomic spacings. Point defects are particularly important when solids have relatively few extended defects. In that case, point defects determine the atomic diffusion, limit the electrical conductivity of metals, limit or enhance the electrical conductivity of semiconductors and are responsible for the electrical properties of ionic crystals at high temperatures.

https://doi.org/10.1515/9783110666502-005

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

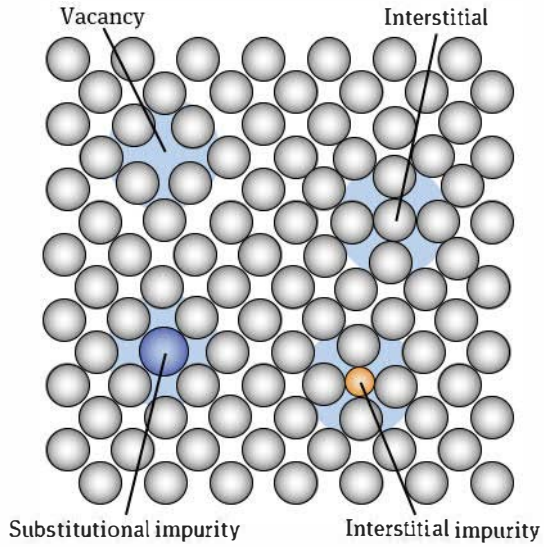


Fig. 5.1: Schematic representation of point defects in crystals. The location of the point defects is highlighted by light-blue filled circles in the background. The slight distortions of the lattice caused by the defects are also hinted at.

5.1.1 Vacancies

The simplest conceivable defect is a missing atom. This point defect is called a **vacancy** or **Schottky defect.**[1] It occurs mainly in simple solids and plays a major role in the properties of metals and ionic crystals. Scanning tunneling microscopy can be used to make vacancies on surfaces visible. Figure 5.2 shows the surfaces of a platinum and a silicon crystal with atomic resolution. In the left image, the regularly arranged

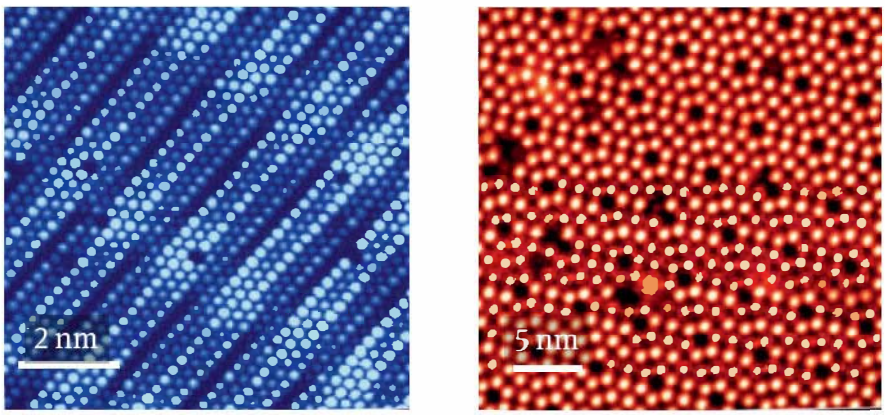


Fig. 5.2: Vacancies on the surface of platinum **(a)** and silicon **(b)**. Both pictures were taken with a scanning tunneling microscope. (After G. Ritz et al. Phys. Rev. B **56**, 10518 (1997).)

1 Walter Schottky, *1886 Zurich, †1976 Forchheim

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

platinum atoms with several vacancies in the middle part of the picture are clearly visible in the central part of the figure. The right-hand image shows vacancies on a silicon surface. Both surfaces show reconstructions (see Section 3.3.7).

Density of Vacancies. We can calculate the number of vacancies found in crystals in thermal equilibrium. The crucial quantity is the free enthalpy $G = U - TS + pV = F + pV$, where U is the internal energy, S the entropy, p the pressure, V the sample volume and F the free energy. Assuming constant pressure and temperature, the change of the free enthalpy is: $\mathrm{d}G = \mathrm{d}U - T\mathrm{d}S + p\,\mathrm{d}V$. Provided the number of vacancies is not large enough to noticeably change the volume of the sample, the $p\mathrm{d}V$ contribution to the free enthalpy can be neglected and it is sufficient to consider the free energy F alone.

The vacancy contribution to the internal energy of the sample, ΔU_L, can be given immediately. If E_L is the energy required to generate a vacancy, then for N_L vacancies we get $\Delta U_\mathrm{L} = N_\mathrm{L} E_\mathrm{L}$. We also have to take into account that vacancies also contribute to the entropy, and in two ways. First, each vacancy causes a change in the thermal vibrations in its immediate vicinity. In order to take this into account, we attribute a vibration entropy S_L to each vacancy.[2] Secondly, the configuration entropy also needs to be taken into account. This contribution is based on the multitude of different possible arrangements of the N_L vacancies on the $(N + N_\mathrm{L})$ lattice sites. This is a similar effect to that which played an important role in the mixture entropy of alloys we saw in Section 3.1. This contribution is proportional to the logarithm of the number of possible arrangements and has a strong influence on the number of vacancies. Thus, using the arguments of (3.5), we find for the contribution ΔF_L of the vacancies to the free energy:

$$\Delta F_\mathrm{L} = N_\mathrm{L} E_\mathrm{L} - T\Delta S = N_\mathrm{L} E_\mathrm{L} - N_\mathrm{L} T S_\mathrm{L} - k_\mathrm{B} T \ln\left[\frac{(N + N_\mathrm{L})!}{N! N_\mathrm{L}!}\right] . \tag{5.1}$$

For thermodynamic equilibrium, the free energy must be at a minimum. Thus we need the density of vacancies for which $\partial F/\partial N_\mathrm{L} = 0$. Starting from this requirement, assuming $N_\mathrm{L} \ll N$ and using the Stirling formula $\ln X! \simeq X(\ln X - 1)$, which is valid for large values of X, the following relationship is obtained:

$$N_\mathrm{L} = N\, \mathrm{e}^{S_\mathrm{L}/k_\mathrm{B}}\, \mathrm{e}^{-E_\mathrm{L}/k_\mathrm{B}T} . \tag{5.2}$$

We would therefore expect that in thermal equilibrium the number of vacancies should increase exponentially with temperature. The crucial quantity here is the energy E_L required to form a vacancy.

In principle, a vacancy is created when an atom is brought from inside the bulk of the crystal to the surface, where it is bound more weakly than in the bulk. Thus E_L

2 In some textbooks this contribution to entropy is not considered in the derivation of the vacancy density, because it does not influence the temperature dependence of the vacancy density. S_L is basically derived from equation (7.3). Since entropy is given by $S = -(\partial F/\partial T)_V$, a change in the frequency spectrum due to a vacancy causes an additional entropy to occur.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

and the binding energy should be comparable, for which we therefore expect values around 1 eV/atom. The vibrational entropy depends on the crystal structure and typical values of S_L/k_B are between 0.5 and 5. If we assume $E_L \approx 1$ eV and $S_L/k_B \approx 3$, we would calculate a vacancy concentration of $N_L/N \approx 2 \times 10^{-4}$, for 1000K, but only $N_L/N \approx 3 \times 10^{-16}$ for room temperature.

Now let us discuss the experimental confirmation of the exponential increase of the vacancy concentration and the experimental value of E_L. Despite the above simplification $p\mathrm{d}V \approx 0$, the generation of vacancies is associated with a measurable increase in volume at higher temperatures. When analyzing measurements, we must take into account that, besides the effect of vacancies, normal thermal expansion (cf. Section 7.1) also contributes to the volume change. The two contributions can be separated, if in addition to the change in volume of the sample, the variation of the lattice constant a is determined very accurately by means of X-ray diffraction. If the contribution $\Delta V/V$ due to the lattice expansion is subtracted from the measured macroscopic volume change $(3\Delta a/a)$, we obtain the vacancy contribution to the volume change.[3] An observable effect can only be expected at relatively high vacancy concentrations, in other words, near the melting point. Since the melting temperature is mainly determined by the value of the binding energy, a measurable concentration of vacancies already occurs at room temperature for low-melting-temperature materials. Figure 5.3 shows the result of such a measurement for sodium, which has a melting point 371 K not far above room temperature and can therefore be studied in an easily accessible temperature range.

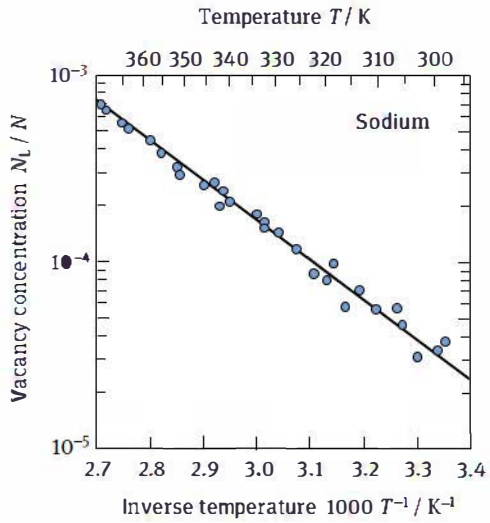


Fig. 5.3: Temperature dependence of the vacancy concentration N_L/N of sodium, measured via the volume change. By plotting $\log(N_L/N)$ as a function of the inverse temperature, the exponential dependence on temperature becomes immediately visible. (After R. Feder, H.P. Charbnau, Phys. Rev. **149**, 464 (1966).)

3 The simple relation $\Delta V/V = 3\Delta a/a$ is actually only valid for isotropic materials.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

From the measured data, we can deduce a value of $E_L = 0.42\,\text{eV}$ for the formation energy and $S_L/k_B = 5.8$ for the entropy factor. Since the binding energy of sodium is 1.11 eV/Atom, the result makes clear that the formation and binding energies are not equivalent, even if the orders of magnitude are comparable. As seen in Figure 5.1, we also need to take into account, among other things, that the lattice near the vacancy created adjusts to the new conditions. This so-called relaxation of the lattice results in lowering the overall energy and thus causes a noticeable reduction of E_L.

For materials with a higher melting point, the binding energy U_B/N and thus also the formation energy E_L is higher. Thus, from equation (5.2), we would find a considerably lower vacancy concentration at room temperature than for sodium. For a comparison of the binding and formation energy, Table 5.1 shows values for a number of materials.

Tab. 5.1: Binding energy U_B/N, formation energy E_L for vacancies and activation energy E_D for volume diffusion (see Section 5.1.5) for different materials. In order to ensure the comparability of the numerical values, the ionic crystals do not have U_B/N or E_L per ion pair but the value per ion. For ionic crystals, E_D refers to the positive ions. (Data from different sources.)

	Al	Cu	Zn	Au	LiF	LiCl	NaCl	KCl
U_B/N (eV)	3.39	3.49	1.35	3.81	5.36	4.38	4.07	3.68
E_L (eV)	0.75	1.18	0.42	0.94	1.34	1.06	1.01	1.15
E_D (eV)	0.56	0.88	0.40	0.78	0.65	0.41	0.86	0.89

In ionic crystals there is an additional complication arising from the charge on the ions. In order to avoid a high Coulomb energy, charge neutrality must be maintained when a vacancy is generated. A calculation very similar to the above, but with the constraint that the sample remains uncharged, leads to the following prediction for the number of vacancies in diatomic ionic crystals:

$$N_L^+ = N_L^- = \sqrt{N^+N^-}\, e^{(S_L^+ + S_L^-)/2k_B}\, e^{-(E_L^+ + E_L^-)/2k_BT} = \sqrt{N^+N^-}\, e^{S_P/2k_B}\, e^{-E_P/2k_BT}, \qquad (5.3)$$

where $S_P = (S_L^+ + S_L^-)$ is the entropy and $E_P = (E_L^+ + E_L^-)$ the formation energy of the ion pairs. The charge neutrality requirement means that the probability of defect formation is determined by the sum of the individual formation energies. The numerical values in Table 5.1 for the formation energies E_L of ionic crystals therefore do not refer to individual ions, but reflect *half* the formation energy per ion pair, i.e. $E_P/2$.

An interesting consequence of the constraint for **charge neutrality** is charge compensation. This effect occurs when impurity ions are introduced whose charge differs from that of the host ions. For example, if NaCl is doped with $CaCl_2$, the density of the crystal is surprisingly *reduced* when the heavier calcium atoms are incorporated. We can understand this if we consider that for each Ca^{2+} ion incorporated, in order

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

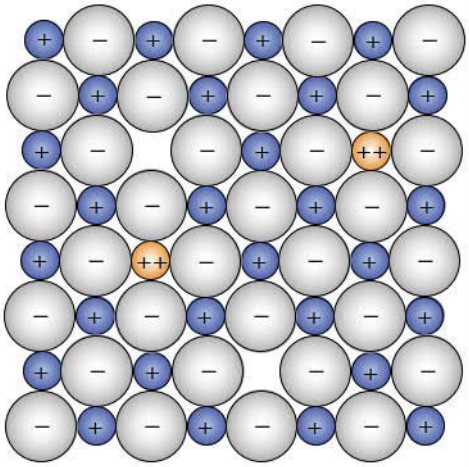


Fig. 5.4: Ca^{2+} ions in NaCl crystals. The incorporation of calcium ions gives rise to Na^{+} vacancies and thus a reduction of the mass density.

to maintain charge neutrality an additional Na^{+} vacancy must be created. This effect is illustrated in Figure 5.4. In this case, the vacancy density is no longer determined by the temperature, but by the concentration of added calcium ions. Doping can thus cause a large number of vacancies already at room temperature, which, as we will see, also play an important role in the electrical conductance.

5.1.2 Color Centers

The study of **color centers** in alkali halides was historically the starting point for the investigation of point defects and played an important role in the development of solid state physics. Pure alkali halide crystals are transparent in the visible spectral range. They become colored by contamination or by irradiation with X-ray, gamma or high-energy particle radiation. As an example, Figure 5.5 shows the optical absorption spectrum of an irradiated KCl crystal. The absorption bands that occur can be assigned to various types of point defect. Here we will discuss in more detail the simple **F-center**, which causes the strongest absorption band (labeled F in the figure). There are also more complicated centers involving several vacancies or ions, denoted by the abbreviations F_A, M, N, R, V_K, ... Similar spectra are also found for the other alkali halides, but with the absorption bands at different wavelengths.

Color centers are clearly demonstrated when an NaCl crystal is heated in sodium vapor and then quenched. After quenching (rapid cooling) the originally clear crystal shows a yellow-brown coloring. During heating, sodium atoms from the vapor phase accumulate on the surface and change the stoichiometry. The structure is stabilized by the formation of chlorine vacancies, which take up one electron from the sodium atoms and diffuse into the bulk of the crystal. These are the F-centers. The magnetic moment of these electrons can easily be detected in electron spin resonance (ESR)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

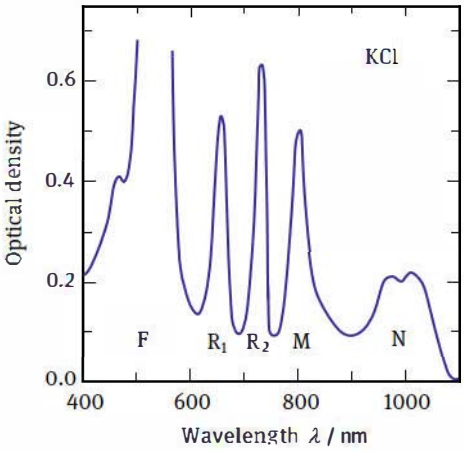


Fig. 5.5: Optical absorption spectrum of KCl after irradiation with X-rays. The various bands arise from different defect types, labeled F, R_1, R_2, M and N. (After R.H. Silsbee, Phys. Rev. **138**, A180 (1965).)

measurements: While no ESR signal can be observed in a defect-free crystal, a strong measurement signal immediately appears if any F-centers are present.

The electron of the F-center is not located at the center of the volume vacated by the chlorine ion. It is energetically more favorable for the electron to sit near one of the six neighboring positive metal ions. However, since this is a quantum system, the electron sits in a superposition of all the sites near a neighboring metal ion, as illustrated in Figure 5.6. As with ordinary atoms, transitions between the eigenstates of the F-centers can be excited by optical irradiation, with the frequency of the absorption lines being a characteristic of the alkali halide crystal in question. Since the electron motion is strongly coupled to the vibrations of the neighboring ions, the spectral lines are not narrow, but relatively wide similar to the optical absorption bands of molecules. The crystals, which are transparent when pure, are colored because color-center absorption bands are usually fall in the visible spectrum range. As already mentioned, further bands arising from more complex defects can occur. For example, the M-center consists of two immediately adjacent chlorine vacancies, each of which contains one electron

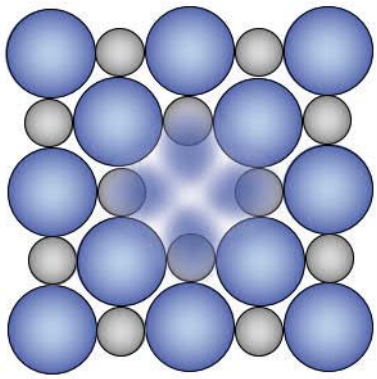


Fig. 5.6: Schematic representation of the charge distribution in an F-center.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

for charge compensation. Then there is the *R*-center which consists of three *F*-centers lying next to each other.

As seen in Figure 5.7a, the absorption bands of the various alkali halides caused by *F*-centers occur at different wavelengths. At first sight we might think that the band shifts to longer wavelengths with increasing cation mass. This observation seems to indicate that, contrary to our simple idea of the structure of the *F*-centers, cations play an important role. However, a look at Figure 5.7b makes it immediately clear that it is not the mass but the size of the ions that matters and thus the lattice constant is the crucial quantity. The wavelength at maximum λ_{max} is related to the distance R_0 of the nearest neighbors via the simple relationsbip $\lambda_{max} = AR_0^2$. For alkali halides with the NaCI structure, the constant of proportionality has the value $A = 6 \times 10^{12}\,\mathrm{m}^{-1}$. This relationship is easily explained: in the simplest approximation, the electron of the *F*-center moves in a potential well, with size determined by the distance to the nearest neighbors. We know from quantum mechanics that the spacing of the energy levels of a particle in a potential well is inversely proportional to the square of the well dimension, hence the dependence on R_0^2. We should note that with just this simple idea, not only the relative shift of the absorption band but also its position can be predicted with surprising accuracy.

When an *F*-center is excited to a higher level by absorption, it can then relax back to its ground state by luminescence. Here it is interesting that there is a large shift of the luminescence band compared to the excitation wavelength. For example, while the *F*-centers of KCl absorb at 560nm, luminescence radiation is observed at

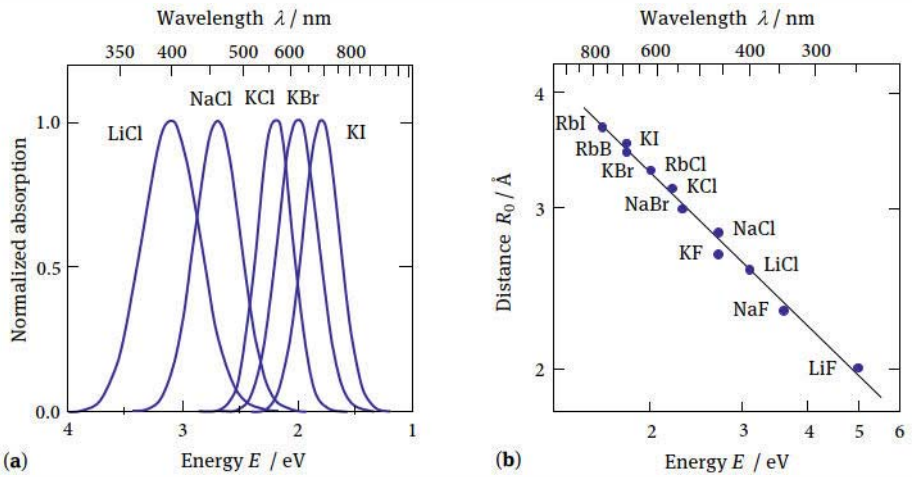


Fig. 5.7: Optical absorption by *F*-centers. **a)** Absorption bands of the *F*-centers in different alkali halides, **b)** Photon energy or wavelength in the maximum of the absorption bands as a function of the distance R_0 of the nearest neighbors plotted on a logarithmic scale. (After G. Miessner, H. Pick, Z. f. Physics **134**, 604 (1953).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

about 1000 nm. The reason for this shift is the fact that over the time the F-center is excited the surrounding lattice adjusts to these new conditions.

When an F-center is excited or ionized, this takes place from a lattice configuration which has already adjusted to the electron configuration, i.e. from a relatively low energy ground state. The de-excitation or recombination of the electron, on the other hand, takes place in a non-relaxed vacancy corresponding to a higher ground state energy. The energy of the emitted photon is therefore smaller than that which was initially absorbed.

Crystalline materials containing color centers also have important technical applications. They are used as the active medium in tunable infrared lasers. Here, the shift of the luminescence band compared to the absorption band noted above is exploited and a simple inversion of the levels involved in the emission can be achieved. However, this application does not make use of the properties of the simple F-centers, but rather those of more complicated color centers. For example, in the first color-center laser, F_A-centers in KCl were excited to produce stimulated emission, the F_A-center being a vacancy in which an immediately adjacent K^+- is replaced by an Na^+-ion.

5.1.3 Interstitials

In addition to vacancies, point defects also include **interstitials** which, as the name suggests, occupy the spaces between regular lattice atoms. Clearly, the insertion of an additional atom in the environment causes a very strong distortion of the lattice. As shown in Figure 5.1, the lattice adapts by the displacement of neighboring atoms. The energy E_I required to create interstitials goes primarily into the distortion of the surroundings and therefore depends strongly on the size of the inserted atoms. Interstitials therefore occur mainly in materials with an open structure, i.e. in solids with low packing densities. Interstitials and vacancies are, to a certain extent, in competition. In ionic crystals, the energy of creation for the two defect types is comparable. In alkali halides the Schottky defects predominate, whereas in silver halides the interstitials prevail. In densely packed metals practically no interstitials are found and vacancies are the dominant point defects.

Interstitials can be formed when atoms on regular lattice sites are displaced into adjacent interstices of the lattice leaving a vacancy. If the interstitial and vacancy formed are in close proximity, this is called a **Frenkel defect**.[4] This type of defect leaves the stoichiometry of the sample untouched. Because of their local charge neutrality, Frenkel defects in ionic crystals often occur rather than simple interstitials. Their equilibrium concentration can be derived in the same way as for Schottky defect pairs in ionic crystals leading to similar results.

4 Yakov Il'ich Frenkel, *1894 Rostov-on-Don, †1952 St. Petersburg

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Although interstitials are seldom found in metals in thermal equilibrium, they can be produced in large numbers by bombardment by high-energy particles, since atoms are knocked out of their lattice sites when the particles are decelerated. This is known as *radiation damage*, which plays an important role in reactor technology.

5.1.4 Impurities

Real solids inevitably contain impurities. Since these often have a strong influence on the properties of solids, they are often specifically introduced in the context of technical applications. Well-known examples are the refinement of metals and the doping of semiconductors. Depending on size, chemical properties, and temperature, impurities occupy various places in the crystal lattice. As shown in Figure 5.1, **substitutional** impurities occupy a regular site whereas **interstitial** impurities take up positions between the regular lattice sites. There is often a strong interaction between the impurities and the inherent defects of the crystal. For example, large impurities in metals usually surround themselves with vacancies, allowing the reduction of the high elastic stresses in the surroundings.

The structure of point defects is usually even more complicated for crystals with a poly-atomic basis. For example, for crystals with a diatomic basis, interchanging neighboring chemically-different atoms effectively results in two regular lattice sites being occupied by "impurities", although no foreign atoms have been added. This type of defect is of great importance for compound semiconductors because it causes unwanted doping. These so-called "antisite defects" occur particularly frequently when the atoms involved have comparable diameters, as is the case with the semiconductor indium antimonide, for example.

5.1.5 Atomic Transport

Individual atoms or ions can move through a solid under external influences such as an electric field or driven by a concentration gradient. The prerequisite for this to happen is almost always the existence of structural defects. Single atoms generally move via jumping in and out of vacancies or via interstitial positions, whereas extended defects such as dislocations (Section 5.2) are required for the motion of larger atomic structures.

Diffusion of Vacancies. In principle, the direct exchange of positions between two adjacent atoms is the simplest transport mechanism. However, the high elastic stresses caused by this process makes it unlikely and thus this transport mechanism is normally of no great significance. In contrast, diffusion via vacancies or interstitials is of extraordinary importance for atomic transport in solids. The elementary steps in vacancy migration are illustrated on the left-hand side of Figure 5.8. On the right-hand

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

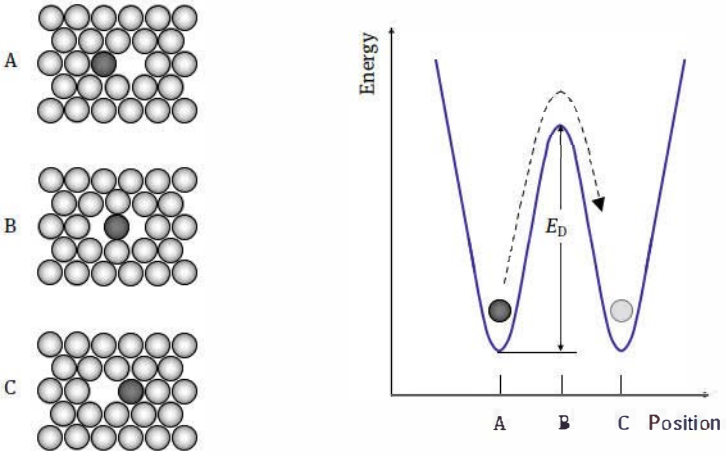


Fig. 5.8: Schematic representation of vacancy propagation. The dark-shaded atom jumps into the vacancy, moving to the right and propelling the vacancy in the opposite direction. In doing this, the atom moves in the potential sketched on the right. The potential barrier, which for a three-dimensional lattice is a saddle point in the potential landscape, determines the activation energy E_D.

side, the potential through which the atom moves when the neighboring lattice site is unoccupied is shown.[5] As indicated, an atom jumps from a regular lattice site on to the neighboring vacancy, causing this vacancy to move in the opposite direction by one lattice unit. With no applied field, the vacancy can move in any direction. With decreasing temperature the jump rates and thus the distances covered per unit time become smaller and smaller.

Now we consider the diffusion of vacancies quantitatively. For an atom next to the vacancy to surmount the potential barrier into the free space, sufficient thermal energy must be available. The probability that this will actually happen is determined by the Boltzmann factor[6] $\exp(-E_D/k_BT)$, which includes both the **activation energy** E_D and the temperature. If the atom vibrates with the **attempt frequency** ν_0 in its potential well, the jump frequency ν, i.e. the number of *successful* attempts per time unit, is given by the expression

$$\nu = \nu_0 \, e^{-E_D/k_BT} \, . \tag{5.4}$$

The attempt frequency is determined by the frequency of the lattice vibrations and is therefore of the order $\nu_0 \approx 10^{13}\,s^{-1}$ (cf. Chapter 6). Processes in which a potential threshold is overcome with the help of thermal energy, are described as **thermally activated**. This fundamental jump mechanism also plays an important role in many other areas of physics.

5 We should be plotting the free enthalpy here, but under the given conditions the simplified treatment leads to the correct result.

6 Ludwig Eduard Boltzmann, *1844 Vienna, †1906 Duino (Trieste)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

According to *Fick's law*[7] the diffusion current $\mathbf{j}$, i.e. the number of particles crossing unit area per unit time, is given by the **diffusion constant** D (also known as the *diffusion coefficient*) and the gradient in the number density $n_L = N_L/V$ of the vacancies:

$$\mathbf{j} = -D\,\mathrm{grad}\, n_L \,. \tag{5.5}$$

The diffusion constant is determined by the frequency and range of the jumps.[8] For the diffusion of vacancies, the jump range is given by the distance a of the nearest neighbors and the jump rate by (5.4), thus we can write:

$$D = \alpha a^2 \nu = \alpha a^2 \nu_0 \, e^{-E_D/k_B T} = D_0 \, e^{-E_D/k_B T} \,. \tag{5.6}$$

The numerical factor α is 1/6 for cubic crystals. The activation energy E_D is usually comparable to the energy E_L required to form vacancies, and therefore also lies in the energy range of 1 eV. If we use the values for copper (E_D = 0.88 eV), we find for room temperature the diffusion constant to be $D \approx 10^{-18}$ cm^2/s. Since the **mean diffusion length** L, also called *diffusion path*, for three-dimensional samples is given by $L = \sqrt{6Dt}$, the vacancies only move about 2 nm per hour at room temperature.

Let us briefly look again at the vacancy concentration at room temperature. Given the energy E_L = 1.18 eV from Table 5.1 for the generation of a vacancy in copper, from equation (5.2) we would expect the vacancy concentration to be vanishingly small. In reality, however, there are quite substantial numbers of vacancies in copper, as well as in other metals. This is because in the solid phase the density of vacancies decreases during cooling mainly by the process of migration to the surface and thus depends on the time needed to reach the surface.[9] When the sample is cooled quickly, the distances that can be covered are so small that the vacancies remain trapped inside the sample and a thermal equilibrium concentration cannot be established.

Interstitial Diffusion. The simplest case of diffusion occurs when interstitial impurities migrate in the lattice. If the substitutional incorporation of atoms is not constrained by covalent bonding and the impurity diameter is smaller than that of the host atom, impurities often occupy interstitial sites. Since, as shown in Figure 5.9, the impurities have to "squeeze through" between the host atoms at the regular sites when jumping from one interstitial site to another, they move in a potential landscape that largely resembles that of vacancy diffusion.

Once interstitials with their comparatively high formation energy E_I have been generated, they are able to move relatively easy through the sample. Typical activation

7 Adolf Eugen Fick, *1829 Kassel, †1901 Blankenberge

8 Although the diffusion constant in crystals is directional and thus a tensor, we treat it as a scalar quantity here for the sake of simplicity.

9 As we will see in Section 5.2, there is another mechanism for reducing the concentration of vacancies: vacancies can also disappear at "edge dislocations" inside the crystal, but even here they must first diffuse to the dislocations.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

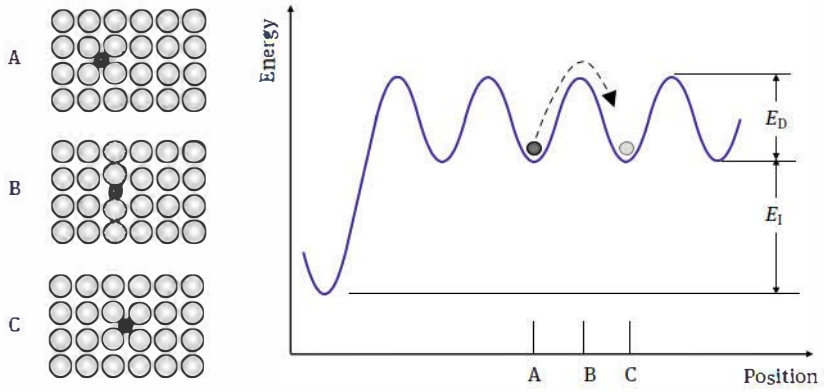


Fig. 5.9: Schematic representation of the motion of an interstitial. On the left-hand side the motion of an atom is shown and on the right-hand side the potential energy of the atom is sketched, where E_I is the energy required to produce an interstitial and E_D is the height of the potential barrier that must be overcome during each diffusion jump.

energies E_D for interstitial diffusion are usually smaller than those for vacancy diffusion. Equation (5.4), for thermally activated diffusion also applies again. The diffusion constant is given by equation (5.6), predicting an exponential increase with temperature. The data in Figure 5.10 for the diffusion of interstitial nitrogen atoms in iron indeed show the expected behavior. From the measurement, the numerical values $E_D = 0.85$ eV and $D_0 = 0.05$ cm^2/s can be deduced. This means that at 1100 K, but far from the melting point $T_m \approx 1800$ K, a nitrogen atom jumps to a neighboring site about once per nanosecond, whereas at room temperature it can hardly move at all.

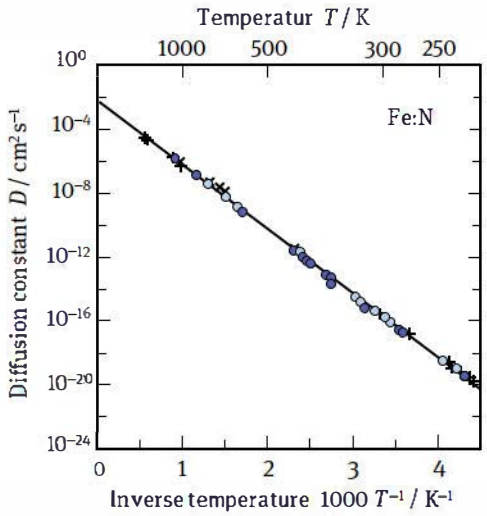


Fig. 5.10: The temperature dependence of the diffusion coefficient of nitrogen in iron. The data were determined by various measurement methods and provided by a number of authors. It is remarkable that over the temperature interval studied the diffusion coefficient varies by about 16 orders of magnitude! (After A.E. Lord, Jr., D.N. Beskers, Acta Met. **14**, 1659 (1966).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Vacancy diffusion. The term "vacancy diffusion" is a source of considerable confusion, because this does *not* mean the diffusion of vacancies, hut the diffusion of atoms on regular lattice sites enabled by the presence of vacancies. Depending on whether the diffusing atoms are (substitutional) impurities or themselves atoms of the host material we can speak of **impurity diffusion** or **self-diffusion**, respectively. The diffusion constant can be measured with radioactive tracers where the spatial or temporal change of a known initial distribution of radioactive atoms is followed.

Vacancy diffusion on a regular lattice relies on two factors, the jump probability itself, and the probability of finding a neighboring vacancy. The vacancy concentration is given by equation (5.2), and the jump probability by equation (5.4). Thus for the diffusion we find:

$$D = D_0 \, e^{S_L/k_B} \, e^{-(E_L+E_D)/k_B T} \, . \tag{5.7}$$

The fact that the concentration of vacancies is also considered here makes it clear why the diffusion coefficient for vacancy diffusion is considerably lower than that for interstitial diffusion.

The strong temperature dependence of the diffusion process is exploited in the doping process of semiconductors. Here, impurities such as phosphorus or arsenic are often allowed to diffuse from the surface into the semiconductor, where the "doping profile" can be controlled by time and temperature. After cooling to room temperature, the diffused impurities become essentially immobile.

Charge Transport in Ionic Crystals. The *electrical conductivity* in ionic crystals provides an excellent example of the application of the concepts discussed in this section. Here electrical charge transport takes place not by conduction electrons as in metals, but via the diffusion of vacancies or interstitials.

The electrical conductivity σ_{el} is defined by the general expression:

$$\sigma_{el} = n_q q \mu \, , \tag{5.8}$$

where n_q is the number density of the charge carriers, q is their charge and μ their mobility.[10] For diffusing ions, the mobility and diffusion coefficient are connected through the **Einstein-Smoluchowski relation**[11,12]

$$\mu k_B T = qD \, . \tag{5.9}$$

In most alkali halide crystals, the current is transported via vacancy diffusion. However, in the silver halides, interstitial diffusion occurs because not only is the activation

10 The term "mobility" is discussed in more detail in Section 8.2 in the context of the electrical conductivity of metals.

11 Albert Einstein, *1879 Ulm, †1955 Princeton, Nobel Prize 1921

12 Marian von Smoluchowski, *1872 Vorder-Brühl, †1917 Karkowo

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

energy for diffusion relatively small in these materials, but also the generation of interstitials requires relatively little energy. Since the diffusion constants of different ions vary considerably, it is usually sufficient to take into acount only the contribution of the ions with the highest mobility. In the case of sodium chloride, for example, the mobility of the small sodium ions is much higher than that of the large chlorine ions, while in silver halides, it is the silver ions that primarily move.

Inserting the Einstein-Smoluchowski relation into (5.8), along with equation (5.7) we obtain:

$$\sigma_{el} = \frac{n_q q^2 D}{k_B T} = \frac{n q^2 D_0}{k_B T} e^{S_P/2k_B} e^{-E_P/2k_B T} e^{-E_D/k_B T}, \qquad (5.10)$$

where n is the number density of the ion pairs, S_P and E_P the entropy and formation energy of the vacancy pairs and E_D the activation energy of the ions with the higher mobility. The exponential functions give rise to an extraordinarily rapid temperature dependence for the conductivity. Ionic crystals are therefore good insulators at room temperature, but good conductors at high temperatures.

The involvement of vacancies in charge transport is particularly evident in the case of *charge compensation*. In the case already mentioned of doping NaCl with calcium (see Figure 5.4), a large number of vacancies are already present at room temperature and available for current transport. Although the current is proportional to the number density n_{Ca} of the calcium ions, the charge is primarily transported by sodium ions which are deposited at the cathode. The calcium ions themselves contribute very little to the current. At lower temperatures, the increase in conductivity with temperature is initially determined by the activation energy E_D for the diffusion of the sodium ions, i.e. $\sigma_{el} \propto n_{Ca} \exp(-E_D^{Na}/k_B T)$. At higher temperatures the number of thermally generated vacancies dominates and the conductivity is described by (5.10).

As an example, Figure 5.11 shows the electrical conductivity of sodium chloride on a logarithmic scale as a function of inverse temperature. This way of plotting the data

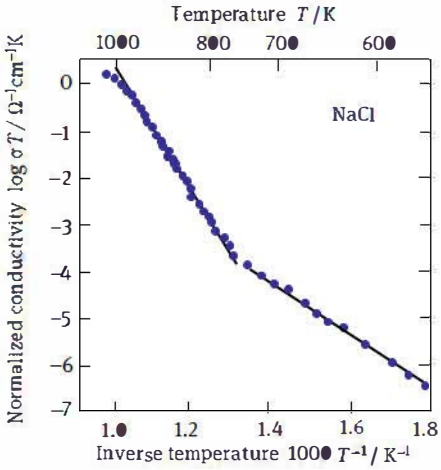


Fig. 5.11: The electrical conductivity of NaCl. The logarithm of the conductivity is plotted as a function of inverse temperature. The two clearly distinct temperatures regions can be attributted to different mechanisms of vacancy generation. (After W. Lehfeldt, Z. Phys. **85**, 717 (1933).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

illustrates clearly that there are two temperature ranges with different temperature dependences of the conductivity. The rapid increase in the conductivity at higher temperatures arises from thermally generated vacancies. The deviation from this behavior at the highest temperatures is due to the generation of vacancy pairs. The weaker increase at the lower temperatures is driven by vacancies generated by the impurities in the sample.

5.2 Extended Defects

While, as we have seen, point defects can strongly influence the electrical and optical properties of crystals, they have hardly any effect on the *mechanical properties*, because these are mainly influenced by extended defects. Extended defects include *dislocations* which can be thought of as a line of lattice imperfections. In addition to the one-dimensional dislocations, polycrystalline materials can also contain two-dimensional defects, so-called *grain boundaries*, marking the border separating differently-oriented crystallites. We briefly discuss both types of defects here.

5.2.1 Mechanical Strength

If a thin rod of length L is pulled, it first stretches according to *Hooke's law*.[13] The well-known relation applies

$$\sigma = E \frac{\delta L}{L} , \tag{5.11}$$

where σ is the mechanical stress, E the modulus of elasticity and $(\delta L/L)$ the fractional change of length or strain. If the stress is further increased, the behavior of the rod depends to a large extent on the material under investigation. In the case of *brittle* materials, which includes glasses, ionic crystals and ceramics, the rod will crack with no significant changes in the sample being detectable beforehand. In this case, the breakage already occurs at strains $(\delta L/L) < 0.01$. Brittle fractures are always associated with the existence or formation of cracks that occur either on the surface or inside the sample. We will not deal with the very important but complex phenomenon of crack formation here.

As shown in Figure 5.12a, bars made of *ductile* materials behave differently. Up to point A, Hooke's law is followed. Between A and B, the relationship between stress and strain becomes non-linear. When the stress is relieved, the curve is followed in reverse, i.e. the deformation is reversible. While brittle materials break at stresses above point B, ductile materials can be stretched further. *Plastic deformation* sets in, the material

13 Robert Hooke, *1635 Freshwater, †1703 London

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

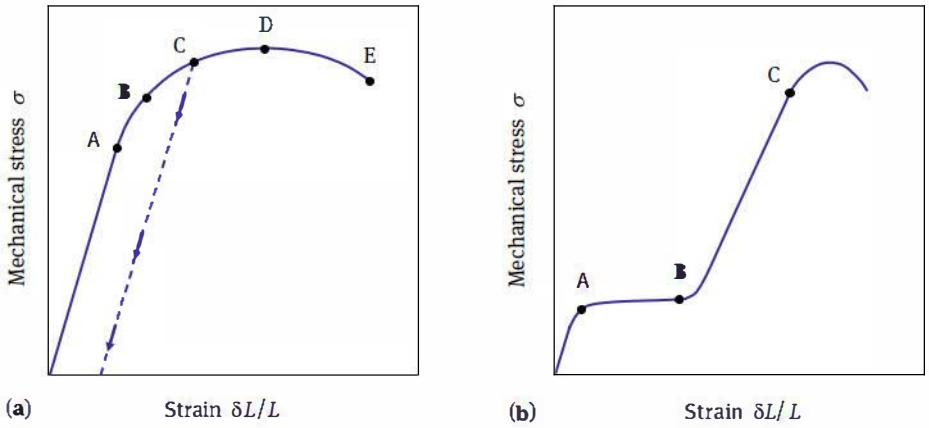


Fig. 5.12: Typical tensile-strain diagrams of polycrystalline and monocrystalline specimens. **a)** The strain of a ductile polycrystalline workpiece under mechanical tensile stress. The significance of points A – E is explained in the text. If the sample is strained to point C, then on decreasing the stress, the dashed curve is traversed. **b)** The strain of a single crystal under tensile stress. A typical characteristic is the occurrence of plastic deformation between points A and B, where the sample extends at almost constant stress.

"flows" and the resulting deformation becomes irreversible. If the stress is now reduced, for example starting from point C, the dashed curve is followed in the direction of the arrows. At the end when the strain returns to zero, a permanent deformation of the rod remains. When a new stress is applied, the elongation of the bar is reversible almost up to point C: the bar has been "cold hardened". At the maximum of the curve, point D, the maximum tensile strength is reached. At even higher values of strain, the bar deforms locally and finally ruptures at point E. Since the irreversible processes also involve the development of the deformation over time, such tensile strain diagrams are usually measured at constant deformation rates $d(\delta L/L)/dt$ of about $10^{-3}\ s^{-1}$.

The tensile-strain diagram of single crystals, e.g. of an aluminum single crystal, is shown in Figure 5.12b. Since crystals have anisotropic mechanical properties, the orientation of the sample with respect to the direction of pull plays an important role. In general, irreversible processes, such as plastic deformation, commence in single crystals at much lower stresses than in polycrystalline materials. As can be seen in Figure 5.12b, the linear region is already left at relatively low tensile stress. The crystal then continues to stretch at almost constant tensile stress until point B is reached. Between points B and C the plastic flow ceases and the crystal becomes more firm again. Above point C it starts to soften once more. In the following, our main focus is on the plastic deformation between points A and B, where parts of the crystal slip over each other as a whole. As can be seen in the schematic drawing 5.13, slightly shifted bands are formed, which surprisingly exhibit the same perfection as the original single crystal.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

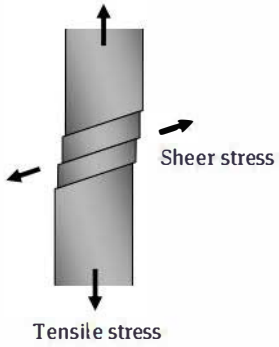


Fig. 5.13: Plastic deformation of single crystals. Depending on the relative crystal orientation, a tension applied to a crystal during a tensile test can give rise to a shear stress parallel to particular crystal planes, leading to macroscopic sections of the crystal slipping bodily along these directions.

Critical Shear Stress. Before discussing the slip mechanism, let us first estimate the *critical shear stress*, above which we would expect the lattice to become unstable and plastic deformation to occur. We will find that the simple estimate based on ideal crystal structure leads to completely wrong predictions in the case of the plastic deformation of single crystals.

For this purpose, referring to Figure 5.14a, we consider a crystal, attached at the base, subjected to a shear stress. On the atomic level (see right-hand side of the picture) the shear stress causes a deflection δx of the lattice planes relative to each other. According to Hooke's law $\sigma = Ge$, there is a linear relationship between the mechanical shear stress σ and the resulting distortion $e = \delta x/d$, where G is the shear modulus. As the stress is increased, the situation will eventually arise where the atoms of adjacent lattice planes come to lie on top of each other. If the system is released at this point, the shifted lattice planes, depending on their exact position, can either return to their original position or jump into a new stable position, advanced by one lattice constant from the original.

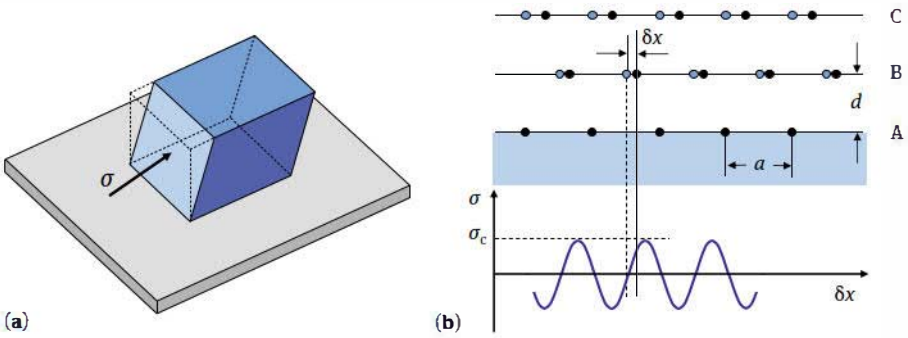


Fig. 5.14: To estimate the critical shear stress. **a)** Basic arrangement. **b)** Variation of the required shear stress σ as a function of the displacement δx of the lattice plane B. The equilibrium of the lattice planes is indicated by light blue circles.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In the following we approximate the effective shear stress by the sine function

$$\sigma = \frac{Ga}{2\pi d} \sin\left(\frac{2\pi\delta x}{a}\right) . \tag{5.12}$$

The pre-factor $Ga/2\pi d$ results from the requirement that for small deflections Hooke's law remains valid. The critical shear stress σ_c is reached at $\sin(2\pi\delta x/a) = 1$ because then the force is sufficient the lattice planes to slip over one another. With the simplification $a \approx d$, the critical shear stress should therefore be given by the relationship

$$\sigma_C = \frac{Ga}{2\pi d} \approx \frac{G}{2\pi} . \tag{5.13}$$

A comparison between these estimated values and those actually measured on aluminum samples (cf. Table 5.2) shows that the measured critical shear stresses are much smaller than those calculated. Contrary to expectations, the difference is particularly large for single crystals and much smaller for polycrystalline materials. The calculated critical shear stress is in best agreement with that of the Al-alloy duralumin, which is used extensively in technical applications and is mixed with large quantities of other materials. This suggests that for single crystals the mechanism of plastic deformation is not consistent with the simple mutual slippage of lattice planes as assumed in the critical shear stress estimation, but that other mechanisms are responsible. In fact, the flow process in crystalline solids is dictated by the movement of dislocations, which we will now examine in more detail.

Tab. 5.2: Shear modulus and critical shear stress of different aluminum samples.

	$G/2\pi \approx \sigma_c^{calc}\left(\frac{N}{m^2}\right)$	$\sigma_c^{exp}\left(\frac{N}{m^2}\right)$	$\sigma_c^{calc}/\sigma_c^{exp}$
Single crystal	4×10^9	4×10^5	10 000
Polycrystal	$\simeq 4 \times 10^9$	2.6×10^7	150
Duralumin	$\simeq 4 \times 10^9$	3.6×10^8	10

5.2.2 Dislocations

We now briefly describe the microscopic structure of the two basic types of dislocation. To illustrate a **edge dislocation**, we imagine cutting in a crystal and inserting an additional partial lattice plane into the section. The edge of this plane is called the **dislocation line**. In its vicinity, the lattice is strongly distorted over a range of several atomic distances and only at larger distances is the crystal stress free. In the simplest case the dislocation line extends from one surface of the crystal to the opposite surface. Figure 5.15a shows a 3-dimensional picture of such an edge dislocation with the added

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

layer in blue. The dislocation can be characterized by the procedure illustrated in Figure 5.15h which shows a plane section of the crystal with the added layer again in blue and the dislocation line perpendicular to the drawing plane.

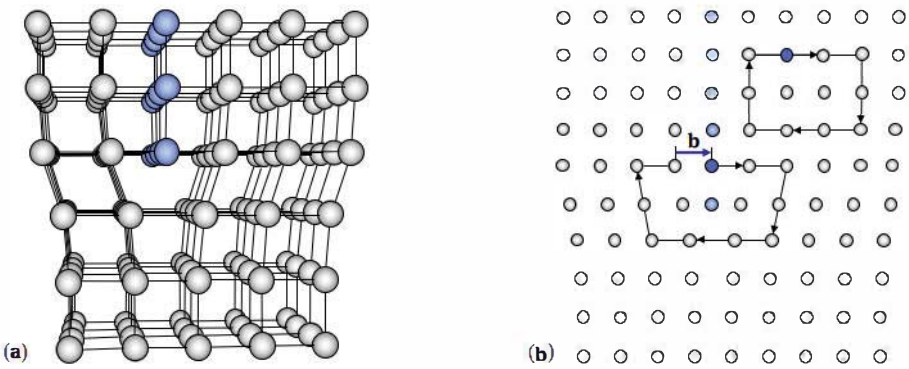


Fig. 5.15: Edge dislocation. **a)** 3D-representation of an edge dislocation in a cubic crystal with the added layer colored blue. **b)** To define the Burgers vector: Again the blue atoms mark the inserted lattice plane. Two clockwise paths are drawn starting from a dark-blue atom in each case, one in the bulk crystal and one enclosing the dislocation line. The gap which opens up when the dislocation line is included is the Burgers vector b.

First, a closed path is followed in the undisturbed part of the crystal by going from lattice site to lattice site. If the same path is followed but now enclosing the dislocation line, a gap in the path opens up. This missing step is called the **Burgers vector b**. If the sequence of steps is taken in both cases in the same sense of circulation the direction of the Burgers vector is independent of the path around the dislocation line. As can be seen in the figure, in the case of an edge dislocation the Burgers vector[14] is *perpendicular* to the dislocation line. The Burgers vector represents the magnitude and direction of the lattice distortion resulting from a dislocation in a crystal lattice and together with the dislocation line provides a complete description of any type of dislocation.

A **screw dislocation** can be visualized, as shown in Figure 5.16a, by cutting a crystal part way along a plane and sliding one half across the other by a single lattice vector. The cut edge marks the dislocation line. The Burgers vector can be determined in the same way as for edge dislocations, but it now points *in the direction* of the dislocation line. If one follows a path on a lattice plane perpendicular to the dislocation line going around it, one finds that this path is helical as for a screw, hence the name.
For some time now, it is possible to image dislocations with scanning tunneling or atomic force microscopes. As an example, Figure 5.16b shows a screw dislocation on

14 Johannes Martinus Burgers, *1895 Arnhem, †1981 Washington D.C.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

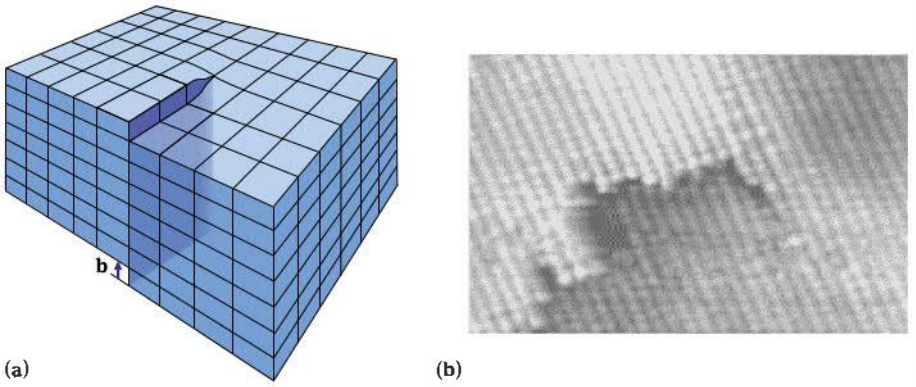


Fig. 5.16: Screw dislocation. **a)** Schematic representation of a screw dislocation for a tetragonal lattice. The dislocation line and the Burgers vector **b** are parallel. **b)** Screw dislocation on the surface of a $Pt_{25}Ni_{75}$ crystal. The atomic resolution image was obtained with a scanning tunnelling microscope. (Courtesy of M. Schmid, P. Varga, Inst. General Physics, Vienna University of Technology).

the surface of a $Pt_{25}Ni_{75}$ crystal. The similarity with the schematic representation is unmistakable. If we draw a path around the dislocation line as shown in Figure 5.15b, we find the Burgers vector perpendicular to the drawing plane.

We have introduced edge and screw dislocation as different types of dislocation. Basically, however, these are only dislocations with a particular orientation of the dislocation line with respect to the Burgers vector. The enclosed angle is 90° or 0° in these special cases, respectively. It can be shown that the Burgers vector along a dislocation line of any shape retains its direction with respect to the crystal. This means that on the path along a dislocation line, the character of the dislocation determined by the angle between the dislocation line and Burgers vector can change, as illustrated in Figure 5.17.

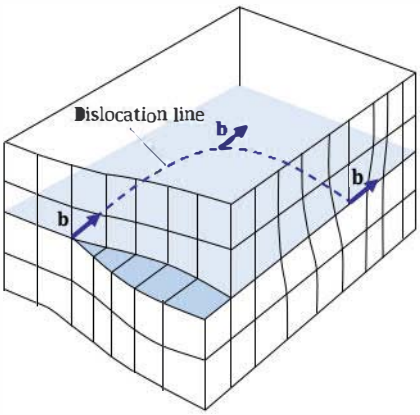


Fig. 5.17: Mixed dislocation. Dislocation line starting as screw dislocation on one side of the crystal and ending as edge dislocation on another side of the crystal. The Burgers vector has the same orientation along the dislocation line.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In the simplest case, the dislocation line traverses the whole crystal, starting at one surface and ending at another. However, for topological reasons, dislocation lines cannot terminate inside the crystal. Therefore, if the dislocation exists entirely within the crystal it must take the form of a loop. Such *dislocation loops* are usually self-contained and run completely inside the crystal. The simplest example is an intercalated lattice plane, the extent of which is small enough that nowhere it reaches the crystal boundary. The dislocation line runs along the boundary of this plane, forming a closed dislocation loop. Along the path of such a dislocation line, the dislocation repeatedly changes its character.

If we look again at Figures 5.15 and 5.16, we will see that an edge dislocation, but not a screw dislocation, can provide a source or sink for vacancies and interstitials. For example, if an atom diffuses from the bottom end of the inserted plane in Figure 5.15 into the crystal, an interstitial is formed and at the same time a small piece of the dislocation line moves up creating a so-called jug. Many of these processes are needed to move up the entire dislocation line. This process is called climb. The energy required for the formation of interstitials is much lower in this type of formation than in the undisturbed lattice because the lattice is already distorted in the vicinity of the dislocation. Conversely, if an interstitial diffuses to the step dislocation, it is "destroyed" there, because the described process now takes place in the opposite direction. In the same way, vacancies can appear or disappear. This makes it clear that our discussion in Section 5.1 was too simple when we mentioned that in real crystals the density of vacancies does not correspond to the thermal equilibrium value, because during the cooling phase the vacancies usually cannot reach the surface. Since vacancies are destroyed not only at the surface but also at edge dislocations, their diffusion path and thus their actual concentration is somewhat reduced.

It would be useful to know the equilibrium number of dislocations. To address this we need to know the contribution of the dislocations to the internal energy. Thus we need to know what energy is stored in the elastic distortion of the lattice. We will not make this calculation here but note that the energy required per atom in a dislocation line is comparable to the energy required to form interstitials. In contrast, the configurational entropy, which plays an important role for vacancies or interstitials, is negligible. This becomes clear when we realize that a dislocation is a continuous line of point defects. The requirement for a continuous line considerably reduces the number of conceivable arrangements of point defects and thus the configuration entropy of dislocations. The contribution of the entropy to free energy can therefore be neglected to a good approximation. The free energy, which is known to be a minimum at thermal equilibrium, is therefore at its minimum value when *no* dislocations are present. That said, in practice, typically about 10^8 dislocations/cm^2 are observed, where the *dislocation density* is the number of dislocation lines passing through unit area in the crystal. The dislocations are formed during the production of the samples and represent a frozen, metastable state. In "good" single crystals, about 10^5 dislocations/cm^2 can be found, in very carefully grown crystals the density

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

can decrease to below 10 dislocations/cm^2, whereas values of around 10^{12} cm^{-2} are found for cold formed metals.

The Visualization of Dislocations. There are a number of methods for making dislocations visible. A simple method is based on the chemical etching of sample surfaces. This takes advantage of the fact that the etching speed increases in the vicinity of a dislocation because the atoms in the strongly distorted structure are more weakly bound. If a dislocation ends at a surface, during etching a small pit is formed at this point which can be easily seen with a microscope. This relatively straightforward technique is often used to determine the dislocation density.

With transparent crystals, dislocations in the interior of the crystal can also be made visible, exploiting the particular ease with which interstitials can move along dislocations. Thus, silver or copper ions are first diffused into the sample and by suitable thermal treatment, precipitates are formed. These precipitates can be observed with a microscope. Dislocations can also be imaged in thin films using transmission electron microscopy. The image contrast here arises from local density variations in the vicinity of the dislocation lines.

As shown schematically in Figure 5.18a, dislocations influence crystal growth. Since the bond between atoms is much stronger at the step of a screw dislocation than on the free surface, crystal growth occurs preferentially at the step. The result is spiral crystal growth as shown in Figure 5.18b. In this example, an image by an atomic force microscope shows a graphite surface on which the growth spiral is clearly visible.

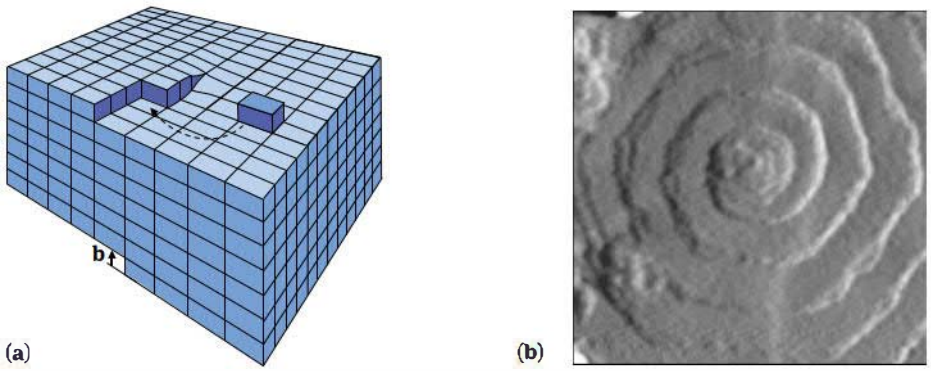


Fig. 5.18: **a)** Schematic representation of crystal growth at a screw dislocation, leading to a spiral step front. **b)** Growth spiral on a graphite surface, imaged with an atomic force microscope. (After J.A. Rakovan, J. Jaszczak, American Mineralogist **87**, 17 (2002).)

Plastic Deformation. We now revisit the observation that the measured critical shear stress for the plastic deformation of single crystals is much smaller than estimates

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

based on equation (5.13). The reason for this is illustrated in Figure 5.19. Under the influence of shear stress, dislocations are formed which can migrate through the crystal. Seen from the outside, the upper half of the crystal slides over the lower half. As the individual lattice planes move in a stepwise manner with just one plane moving at a time, the critical stress is only exceeded locally. The actual critical shear stress σ_c is much lower with this mechanism than the estimated value where we assumed the simultaneous movement of all lattice planes. Depending on the acting forces and the type of existing or generated dislocations, the shearing motion can be vertical or parallel to the dislocation line.

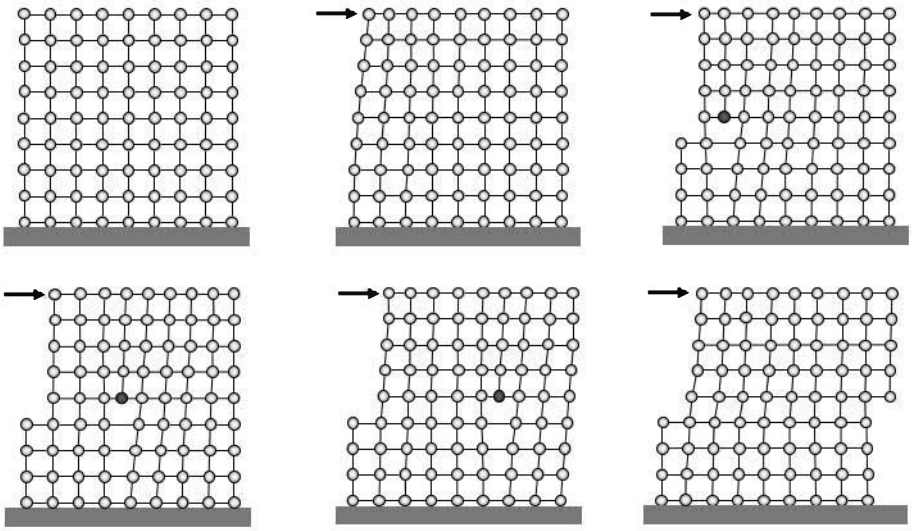


Fig. 5.19: The formation and propagation of an edge dislocation under shear stress. The dislocation line, highlighted by the dark-shaded atom, moves through the crystal held at the base. The arrow indicates the direction of the applied shear stress.

Whiskers consitute an interesting special case. Whiskers are fine hairlike crystals containing a single screw dislocation running along the axis and can grow spontaneously in a number of situations for example during precipitation from a supersaturated solution. Since, when the whisker is bent, the dislocation is not subject to shear stress parallel to the Burgers vector, the stress cannot cause slipping. As a result critical shear stresses have been observed in, for example, tin whiskers a thousand times greater than those observed in bulk tin samples. Tin whiskers can cause serious problems in microcircuits by bridging small gaps between conductors, leading to short circuits as depicted in Figure 5.20.

Since plastic deformation is mediated by the movement of dislocations, the actual strength of a material is essentially governed by those mechanisms which block the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

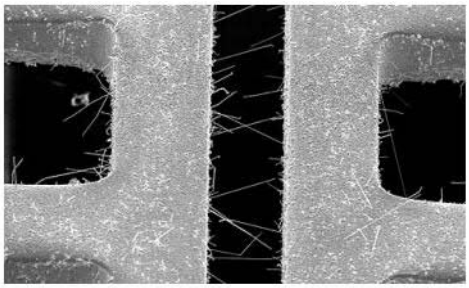


Fig. 5.20: Bridging of leads by tin whiskers. (After P. Goradia et al., Conference Paper: Indian Surface Finishing Conference Mumbai 2014.)

movement of dislocations. Since the diffusion of atoms invariably weakens the pinning of dislocations, the strength of materials decreases with increasing temperature.

The understanding of pinning mechanisms is of utmost importance in materials science. Basically, the propagation of dislocations is hindered by lattice defects, since overcoming them requires additional energy. Therefore, an increase in the strength of a material can be achieved by the deliberate generation of defects. An important technique for changing the mechanical properties in this way is therefore by the incorporation of impurities. For example, interstitial carbon atoms in iron, oxygen atoms in silicon or substitutional zinc atoms in copper increase the strength and the yield point of the host lattice. This form of **hardening** is called *solid solution strengthening*. Other processes are precipitation or particle hardening and dispersion hardening. Precipitates are small particles of a second phase which can be formed by the agglomeration of impurities and whose shape can be changed by thermal treatment. Dispersion particles are particles of a second phase, which were already present in the molten state and become incorporated in the host crystal. For example, graphite precipitates play an extremely important role in the mechanical properties of cast iron.

An example of the anchoring of dislocations can be seen in the electron-microscope image of Figure 5.21. Here it can be clearly seen that dislocation propagation under the influence of shear is hindered by defects, seen from the resulting kinks at the dislocation pinning sites.

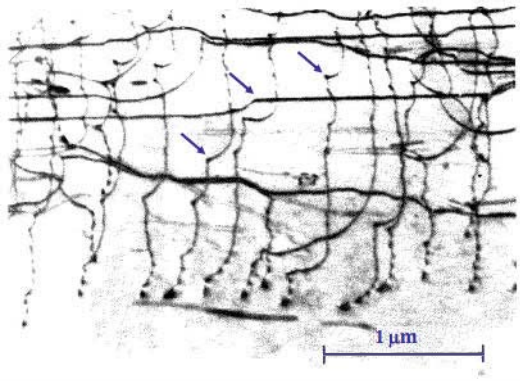


Fig. 5.21: Electron microscope image of dislocations in a magnesium oxide crystal after application of mechanical stress. Buckling occurs at the pinning points along the dislocation lines. Three of the many kinks are marked by arrows. (After B.K. Kardashev et al., phys. stat. sol. (a) 91, 79 (1985).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

A further means of strengthening materials is the formation of dislocation networks. As the interpenetration of dislocations costs energy, the movement of the dislocation lines is severely hindered as the number increases. This effect occurs during plastic deformation, when strain hardening occurs due to the generation of additional dislocations. Furthermore, since real solids are mostly polycrystalline, the boundaries between the crystallites, the *grain boundaries*, which we will discuss later, provide substantial obstacles to the dislocation movement. The hardening arising from this effect is known as *fine grain hardening*.

Finally, we address a quantitative problem here. In Figure 5.19, we show one section of the crystal moving relative to the other by just one lattice constant, while a single dislocation crosses the crystal. Thus *macroscopic* displacement is only possible if new dislocations are constantly being generated during the deformation. One means of generating dislocations is based on the **Frank-Read source**,[15,16] which is briefly explained here. Its mode of operation is shown in Figure 5.22. We imagine a starting dislocation loop V of which only the straight section 1 is drawn, running between anchor points A and B. The remaining part (from B to A) is not shown as it runs below the plane of the figure. The straight section of the dislocation bulges between the anchor points under the influence of a force acting from the bottom of the figure and passes through the stages 1 to 7. At "snapshot" 5 the bulges touch and recombine with the creation of a new dislocation loop V′. The original dislocation V returns to the initial state 1 and the process can be repeated indefinitely.

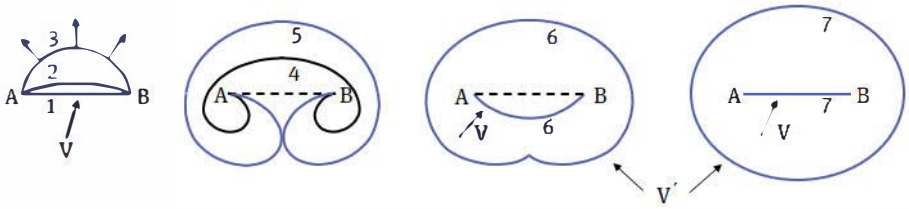


Fig. 5.22: Frank-Read mechanism for generating dislocations. Under the influence of shear stress the dislocation V goes through different stages 1 to 7. At the end of the cycle the original dislocation V is back to its initial state and in addition the dislocation V′ is created. States 2 and 4 are shown in black.

Figure 5.23 shows a Frank-Read source in silicon. The dislocations are decorated with copper precipitates and thus made visible. Since silicon is opaque in the visible wavelength range, the image was taken in infrared light. In addition to the newly formed dislocations, the part of the original dislocation running inside the crystal is also visible.

15 Frederick Charles Frank, *1911 Durban, †1998 Bristol
16 William Thornton Read, AT&T Bell Laboratories

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

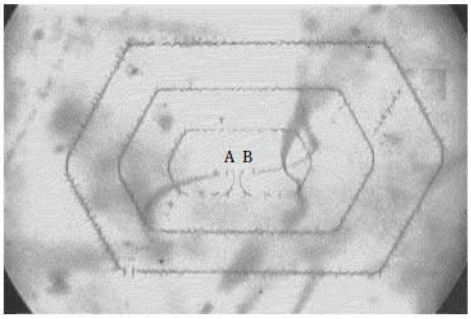


Fig. 5.23: Frank-Read source in silicon. The dislocations are decorated with copper precipitates. A and B mark the dislocation anchor points as shown in Figure 5.22. The image was produced under infrared illumination, as silicon is opaque in the visible. (After W.C. Dash, Dislocations and Mechanical Properties of Crystals (Wiley, New York, 1957).)

5.2.3 Grain Boundaries

As described in Section 3.1, special care has to be taken when growing single crystals. Most solids turn out to be polycrystalline, because without special precautions, crystal growth begins at various points during cooling of the melt, namely where there are already sufficiently large nuclei. This results in solids composed of many small crystallites, as seen in Figure 5.24.

Fig. 5.24: Optical image of a polycrystalline copper sample. The individual crystallites with an average size of about 3 mm are clearly visible.

Two neighboring crystallites or grains are differently oriented in polycrystalline materials. The interface region between them is called a grain boundary. These are two-dimensional defects which have a strong influence on the properties of polycrystalline samples, especially on the mechanical and electrical properties. During annealing, the grain boundaries heal to some extent, as the greatly increased diffusion rates in the vicinity of the grain boundaries facilitate the rearrangement of atoms into energetically more favorable arrangements, with the growth of large crystallites at the expense of the small ones.

The properties of grain boundaries depend very much on the growth conditions and the orientation of the adjacent crystallites. If the orientations of the neighboring crystallites are very different, the grain boundaries can be regarded as a disordered "internal planar surface" region comprising an accumulation of point defects and

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

dislocations. Under mechanical stress the atoms in these grain boundaries can move relatively easily, with the result that diffusion along grain boundaries is increased.

If the alignment of identical lattice planes at two different points in good single crystals is determined with high accuracy, it is often found that these can be slightly tilted with respect to each other by an angle $< 1^{\circ}$. The grain boundary between such *mosaic blocks* is called **a small angle grain boundary**. As shown schematically in Figure 5.25a, such boundaries comprise a series of edge dislocations, periodically repeated at relatively large intervals along the grain boundary. Under suitable loading, such grain boundaries can move as a whole through the crystal perpendicular to the dislocation lines. If crystals with small angle grain boundaries are etched , the dislocations become visible as small etch pits, as discussed above. The tilt angle θ between the mosaic blocks can thus be determined from the distance d between the etch pits. The structure of cleanly formed grain boundaries can be studied with high resolution electron microscopy, as shown in Figure 5.25b, illustrating a grain boundary in aluminum.

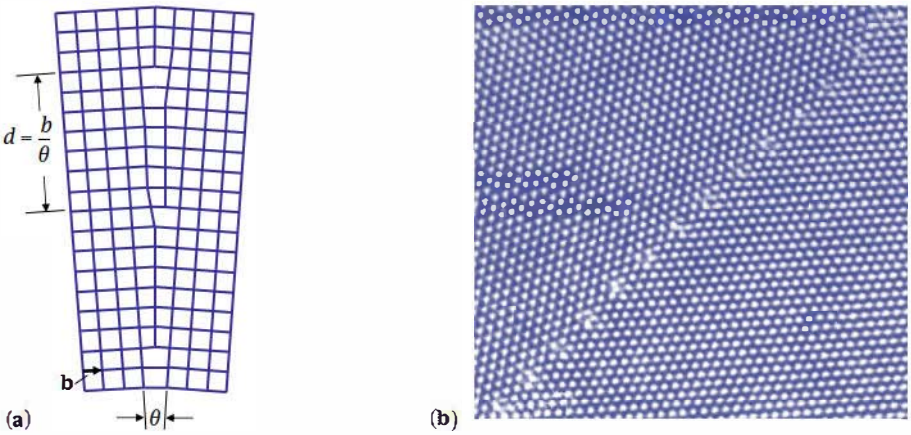


Fig. 5.25: **a)** Schematic representation of a small angle grain boundary. The tilt angle determines the distance between the dislocation lines. **b)** High-resolution electron microscope image of a grain boundary in aluminum. (Photograph by EM Group, Cambridge University.)

Finally we introduce one further extended defect, where we find so-called **stacking faults**, which occur in particular with close-packed crystal structures. In a face-centered cubic crystal, the successive (111) planes are arranged according to the pattern ...ABC ABC ABC ABC... (cf. Section 3.3). However successive layers can be placed in alternative positions, which may result in a stacking order ... ABC ABC AB $\bar{C}$ BA CBA CBA ... . Plane C, marked with a "hat", indicates a layer about which the structure of the crystal is mirrored. This is called *twin formation*. Another possibility would be the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

arrangement ... ABC ABC AB Å BC ABC ABC ... . At plane Å, the regular face-centered cubic sequence is interrupted by a narrow slab of hexagonal close-packed structure.

5.3 Defects in Amorphous Solids

Since the atomic structure of amorphous materials lacks translational symmetry, it is not possible to define point defects such as we did for crystals in Figure 5.1. Terms like vacancies or interstitials lose their precise meaning. In amorphous materials, the deviations from a periodic atomic arrangement give rise to the appearance of "voids" of various sizes, in some ways similar to the vacancies of the crystals. Their number depends on the cooling rate during glass production or other manufacturing processes of amorphous solids. As vacancies in crystals, the voids of amorphous solids have a major impact on the diffusion of impurities and influence the viscosity of glasses in the vicinity of the glass transition.

The **dangling bond** is a typical feature of amorphous substances and which is not found in crystalline materials. In glasses used for technical applications, such defects are mainly caused by metallic ions or other impurities that break the bonds. However, they can also occur in pure amorphous materials and have been investigated in detail. *Unpaired electrons* can easily be detected by electron spin resonance methods, because they cause an easily recognizable signal owing to the magnetic moment of their spins. Remarkably, the number of such detected defects depends strongly on the coordination number of the amorphous solid. For example, in the tetrahedrally bonded semiconductors a-Si or a-Ge about $10^{19} - 10^{20}$ dangling bonds per cm^3 are found. On the other hand, in chalcogenide glasses such as a-As_2S_3 or in amorphous selenium the number of unpaired electrons is below the detection limit. The reason for the concentration of unsaturated bonds differing by so many orders of magnitude was not understood for a long time.

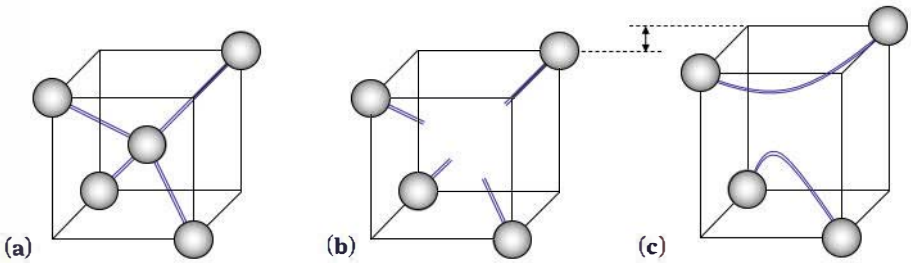


Fig. 5.26: Vacancy in crystalline silicon. **a)** Atomic configuration in a perfect Si crystal. **b)** Immediately after the removal of an atom, four bonds are broken. **c)** The arrangement of atoms and bonds after the reconstruction is complete (not to scale).

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

We will take the dangling bonds in silicon as an example. We first look at c-Si, from which an atom is removed. As shown in Figure 5.26 this creates four open bonds oriented towards the missing atom. This state is not stable. The four unpaired electrons form new bonds so that no unpaired electrons remain. At the same time, the arrangement of the atoms changes, as indicated in the picture by the vertical extension of the atomic positions.

As shown schematically in Figure 5.27a, the saturation of all covalent bonds in amorphous silicon cannot always be fulfilled. The electron shown lacking a partner, cannot form a bond and therefore remains unpaired. If layers of pure amorphous silicon are produced, e.g. by vapor deposition, the occurrence of a very large number of dangling bonds cannot be avoided. This is due to the fact that the angle of the rigid sp^3-hybrid bond of silicon cannot be greatly changed. Since changing the angles and distances are a prerequisite for a random arrangement of the atoms in a network, a large number of bonds must remain broken. The existence of dangling bonds is therefore a characteristic of fourfold coordinated amorphous solids. As we will see in Section 10.3, open bonds largely determine the electrical properties of amorphous semiconductors: They cause such high conductivity that doping a-Si, as is common in crystalline semiconductors, has no noticeable influence on the electrical properties. Layers of pure amorphous silicon are therefore unsuitable for applications in semiconductor technology.

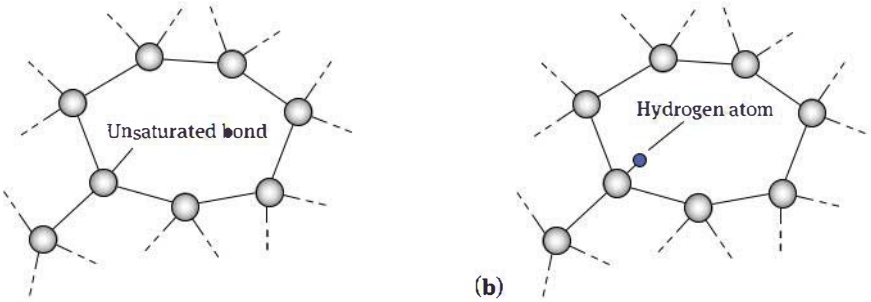


Fig. 5.27: **a)** Schematic two-dimensional illustration of a dangling bond in amorphous silicon. **b)** Hydrogen passivates the dangling bond.

Nonetheless, solar cells are produced from this material. In order to improve fundamentally the electrical properties for technical applications, large quantities of hydrogen are incorporated during the production of the layers. Due to its monovalent nature, hydrogen causes saturation of the free bonds but without itself causing any new unsaturated bonds. The corresponding defect structure is shown in Figure 5.27b.

As already mentioned, no unpaired electrons can be detected in the chalcogenide glasses, although structural defects do occur. To discuss defects in this class of materi-

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

als, we take amorphous selenium as a particularly simple example, as it behaves in the same way as chalcogenide glasses, but has a simpler structure. The electrons of the outermost shell in selenium are in the s^2p^4-configuration. In a-Se, chains appear in addition to ring-shaped structures, as shown in Figure 5.28a. The bond between the atoms of the chains is formed by p-electrons, but only two of the four p-electrons contribute to the bond. Each atom therefore still has a non-bonding p-electron pair.

At the ends of the (variable length) chains, there must be unpaired electrons. We refer to such an end atom as a $C_1^\bullet$ defect to indicate that it is a neutral chalcogen atom that is only singly coordinated. The middle picture shows two chain ends, i.e. two $C_1^\bullet$ defects and one continuous chain. The first two chains react with each other: By transfer of an electron they form one triply-coordinated and positively charged C_3^+ defect and one negatively charged, singly-coordinated C_1^- defect. Both defects now only have electron *pairs*, i.e. they are without unpaired electron spin and thus cannot be detected in electron spin resonance experiments.

It takes additional energy to generate the C_1^--defect, since the excess electron is repelled by the electrons already present. On the other hand, the formation of the third bond in a C_3^+-defect is associated with an energy gain. Overall, the reaction $2C_1^\bullet \dashrightarrow C_3^+ + C_1^-$ is exothermic, so that this type of defect structure is a typical feature of chalcogenide glasses. This combination of these two defect states is called *valence alternating pair* (VAP). Similar defects are also found in other amorphous solids, but their defect structure is usually more complicated. This is especially true for amorphous compounds.

An additional difficulty in describing theoretically such defect structures is the relaxation of the environment during the creation or annihilation of these defects, as the potential energy is lowered by local structural rearrangements. This relaxation

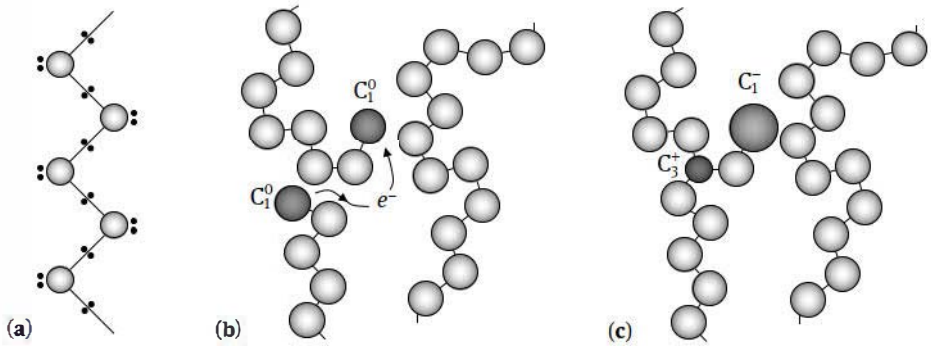


Fig. 5.28: The formation of a valence-alternating pair in amorphous selenium. a) Part of a selenium chain with binding and non-binding electrons located between or directly on the atomic cores. b) Schematic diagram of three chains consisting of selenium atoms. Two chains end in the picture with a $C_1^\bullet$ defect each, indicated by dark shading. c) Valence-alternating $C_1^- - C_3^+$ pair. The chain depicted on the right is not involved in the formation of the charged defect pair.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

is more pronounced in amorphous solids than in crystals because the network has a greater variability in the spatial arrangement of the atoms. However, such structural changes occur relatively slowly, since a larger number of atoms are involved.

Two-dimensional defects, such as grain boundaries, do not occur in amorphous materials. This results in an interesting and extremely important characteristic of glasses: they are in most cases transparent. Although some materials that show no absorption in the visible spectral range and are thus optically transparent as single crystals, are milky or white as *polycrystalline materials*. This is due to the light scattering that occurs on the interfaces between the differently oriented crystallites. The absence of grain boundaries is therefore of crucial importance for the optical properties of glasses and thus for numerous of their applications in daily life.

5.4 The Order-Disorder Transition

We conclude this chapter with a brief discussion of order-disorder transitions. As already shown in Figure 3.9, the atoms of substitutional alloys are located on regular lattice sites, but it is a question of statistical probability which type of atom occupies which site. Such a disordered state usually occurs when the melt is rapidly cooled. A well-known example of this is the copper-gold alloy, which can exist in any composition ratio and in which the disordered state can be "frozen". As we will see in Chapter 9, the electrical resistance at low temperatures, the residual resistance, is a measure of the deviations from the ideal crystal, since the electrons are scattered by the impurities while moving through the lattice. Therefore, "wrongly" arranged atoms increase the residual resistance. Figure 5.29a shows the *residual resistance* of copper-gold alloys as a function of the gold concentration. As expected, one finds the maximum resistance at maximum disorder if the alloy consists of equal parts of gold and copper. For completeness we should note that the resistance follows the Nordheim rule,[17] which states that the residual resistance is proportional to $x(1 - x)$, where x stands for the concentration of one of the two components. The agreement between the experimental data and this simple description is impressive. However, if the alloys are cooled sufficiently slowly below 680 K, ordered phases are formed at certain composition ratios. This reduces the degree of disorder and the residual resistance drops. As can be seen in Figure 5.29b, at the correct values $x = 0.25$ and $x = 0.5$, the intermetallic compounds Cu_3Au or CuAu are formed with a resistance much lower than that of disordered alloys of the same composition. The formation of the order can be quantified very well by means of X-ray diffraction. In Figure 4.24b we have shown an example of a Debye-Scherrer image of the ordered Cu_3Au phase.

17 Lothar Wolfgang Nordheim, *1899 Munich, †1985 La Jolla

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

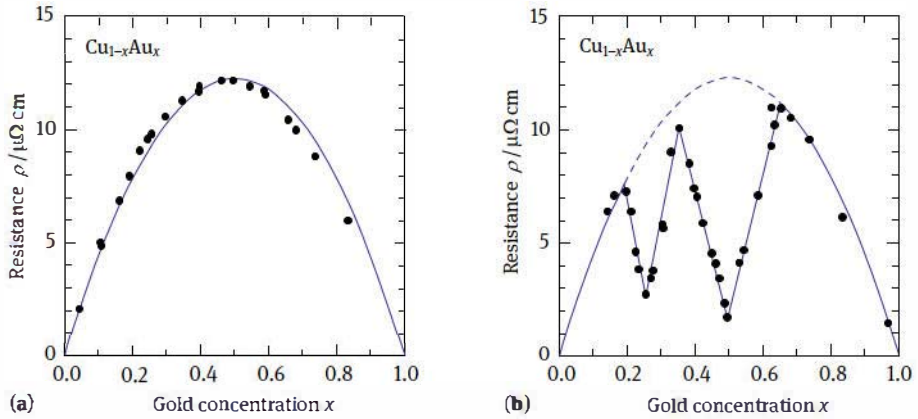


Fig. 5.29: The residual resistivity of copper-gold alloys with the resistivity of the pure metals subtracted. The parabolas show the prediction of the Nordheim rule. **a)** The resistivity of disordered alloys. **b)** The resistivity of tempered samples. There are clear minima at the composition of the intermetallic compounds CuAu and Cu_3Au. (After C.H. Johanson, J.O. Linde, Ann. Phys. **25**, 1 (1936).)

The question now arises how the structure changes during the transition from the disordered to the ordered phase. As an instructive example, we take the behavior of the alloy CuZn, consisting of equal parts of copper and zinc and known as "β-brass". At room temperature, this alloy is well-ordered, based on a simple-cubic lattice with a diatomic basis. The corners of the cube are occupied by copper atoms and the centers by zinc atoms, giving the cesium chloride structure. If the sample is heated, a transition to a completely disordered phase occurs at a critical temperature of about 735 K, the exact value depending on the precise composition. Above this temperature the lattice sites are statistically occupied with the alloy having a body-centered cubic lattice. This process is reversible and the order is restored on cooling. In general, the transition from the ordered to the disordered phase can be of first- or second-order, depending on the system. If the phase transition is first order, latent heat, the heat of transformation, is released during the transition. In the second-order transition, a maximum is observed in the specific heat, but no latent heat is released. In the case of the above-mentioned copper-gold alloys, for example, there is a first-order order-disorder transition. In the case of the CuZn alloy, there is a second-order transition, the specific heat for which is shown in Figure 5.30. The temperature dependence of heat capacity through the transition shows the typical λ-shape.

To characterize the degree of order, we introduce a parameter s as a measure of the long-range order. If the sample consists of N atoms and there are R atoms in the "right" and F atoms in the "wrong" places, then

$$s = \frac{2R - N}{N} = \frac{N - 2F}{N} \,. \tag{5.14}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

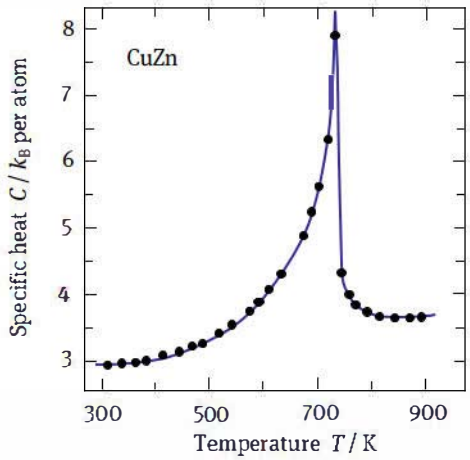


Fig. 5.30: Temperature dependence of the specific heat of β brass. The λ-shaped maximum is a characteristic feature of a phase transition of second order. (After F.C. Nix, W. Shockley, Rev. Mod. Phys. **10**, 1 (1938).)

In the case of ideal order, all N atoms are in the right places, i.e. $R = N$ and $s = 1$. In case of complete disorder, half of the atoms are in the wrong places, i.e. $s = 0$. The schematic temperature dependence of this parameter is shown in Figure 5.31 for the two phase transitions of different order. While in the case of the second-order phase transition, the long-range order steadily approaches zero as the critical temperature T_c is approached, a discontinuous jump occurs at the first-order transition. The different temperature dependences of the two types of *order parameter* also make it clear why only the first-order transition is associated with a latent heat.

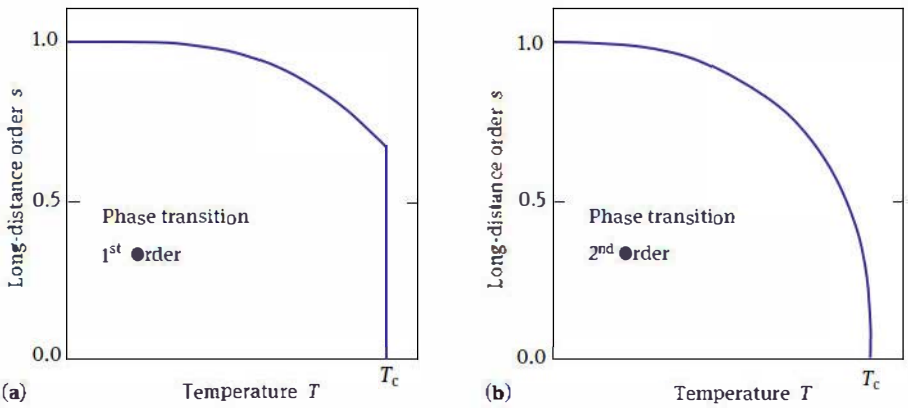


Fig. 5.31: Schematic temperature dependence of the order parameter. **a)** Phase transition of first order. At T_c the order parameter makes a jump. **b)** Phase transition of second order. The order parameter changes continuously.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

With the help of X-ray diffraction the degree of order can be monitored. Looking at the case of the CuZn alloy in more detail, as noted above, the ordered phase of β-brass takes the form of a simple cubic lattice with a diatomic basis. As with caesium chloride CsCl (Section 3.3), the structure factor is given by $S_{hkl} = f_{Cu} + f_{Zn} \exp[-i\pi(h + k + l)]$. All reflections occur, although with different intensities, as predicted by equation (4.27). In the disordered phase the scattering intensity is determined by the averaged structure factor. With the average atomic shape factor $\langle f \rangle = (f_{Cu} + f_{Zn})/2$, for the structure factor we can write: $\langle S_{hkl} \rangle = \langle f \rangle \{1 + \exp[-i\pi(h + k + l)]\}$. On average, the alloy behaves as a body-centered cubic crystal and according to reference (4.28) the reflections with an odd sum $(h + k + l)$ are not expected. The additional lines observed in Debye-Scherrer images in the ordered phase are called *superstructure lines*.

5.5 Exercises and Problems

1. Diffusion of Vacancies in Gold. During the growth of single crystals, vacancies become incorporated. To reduce their number, crystals are often annealed just below the melting temperature ($T_m = 1337$ K). The lattice constant of the face-centered cubic lattice is $a = 4.08$ Å.

(a) Specify the basic jump length.

(b) Estimate the time required at room temperature and at 90% of the melting temperature for a vacancy to diffuse to the surface from a depth of 1 cm. The activation energy for the jump process is $E_D = 0.78$ eV.

2. Vacancy Diffusion. The diffusion of sodium ions in NaBr (density $\varrho = 3.20$ g/cm^3) was investigated at 753 K and 923 K using the tracer method. For the diffusion coefficient D the values of 3.22×10^{-15} m^2/s and 3.62×10^{-13} m^2/s were found, respectively.

(a) How large is the sum of the formation and activation energies of the vacancies?

(b) Calculate the prefactor $D_0 \exp(S_L/k_B)$.

(c) Calculate the electrical conductivity of NaBr at 823 K.

3. Vacancies and Thermal Expansion. In thermal equilibrium, the concentration of vacancies can be easily calculated. For copper the relevant parameters are $S_L = 1.5\, k_B$ and $E_L = 1.18$ eV.

(a) Calculate the vacancy concentration at room temperature and at 90% of the melting temperature $T_m = 1358$ K.

(b) At these temperatures, how large is the volume expansion $\Delta V/V$ caused by the vacancies?

(c) Calculate the coefficient of linear thermal expansion caused by the vacancies at these temperatures.

(d) Compare the result with the literature value ($\alpha = 1.65 \times 10^{-5}$) and with the result of problem 3 in Chapter 4.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

4. Defects in Iron Oxide. The mineral FeO, also called wiistite, has a sodium chloride structure (lattice constant $a = 4.31$ Å) and occurs only in non-stoichiometric compositions owing to missing iron atoms. What is the concentration of these vacancies if the sample under consideration has a density of $\varrho = 5730\ \mathrm{kg/m^3}$?

5. Color Centers. There are two simple models to describe the optical properties of F-centers, namely the "hydrogen model" and the "square-well potential model". Examine the properties of the three alkali halides LiF, NaCl and RbI within the framework of these models.

(a) In the hydrogen model, it is assumed that the captured electron moves in the field of a positive point charge e. The surrounding ions are taken into account by the dielectric constant $\varepsilon_r = n'^2$, where n' is the refractive index. Calculate the energy difference between the ground state and the first excited state.

(b) In the square well potential model, it is assumed that the electron is trapped in a cubic infinitely-high potential well with side $L = 2R_0$. Here R_0 is the distance to the nearest neighbors. Calculate the energy difference between the ground state and the first excited state.

(c) Compare your results with Figure 5.7b.

The alkali halides LiF, NaCl and RbI have the same structure, and their lattice constants have the following values: 4.02, 5.64 and 7.34 Å. The refractive index is given by $n' = 1.41$, 1.56 and 1.65, respectively.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

6 Lattice Dynamics

Most solid state properties can be attributed either to the motion of atoms around their equilibrium positions or to the motion of almost free electrons. In the former case, we speak of lattice dynamic and in the latter of electromagnetic properties. This clear distinction is possible because electrons move much faster than the heavier atomic cores due to their low mass. If atoms are displaced from their equilibrium positions in solids, a new distribution of electrons "instantaneously" appears in response, with an increase in the total energy. If the atoms return to their initial position, the energy is fully recovered. The electrons are not excited during this process but remain in their ground states independent of the coordinates of the atomic cores. This allows us to treat the two subsystems separately in what is called the *adiabatic approximation* or *Born-Oppenheimer approximation*[1].

We will discuss the dynamics of the lattice in two stages. First, in this chapter, we consider the vibrational states that occur in crystals and amorphous solids. We will assume initially that the displacement of atoms from their rest positions is associated with a parabolic variation in the potential energy of the solid. This approximation allows, for example, for the description of the specific heat, which is an important thermodynamic quantity. Then in the next chapter we will discuss properties where the weak anharmonicity of the lattice potential plays a crucial role, such as the thermal conductivity and thermal expansion.

6.1 Elastic Properties

Before we discuss elastic vibrations at the atomic level, we will first deal with solid state vibrations using the concept of an **elastic continuum**, which does not take into account the atomic structure. Of course, this simplification is only viable if the processes under consideration occur on length scales large compared to atomic separations.

We start with the deformation of solids under the influence of external forces. We will assume initially that the deformations are so small that there is a linear relationship between the force and the deformation. As we have seen in Section 5.2, considerable deviations from linear behavior occur at higher mechanical stresses, but in this chapter we will not go into the details of the non-linear properties.

1 Julius Robert Oppenheimer, *1904 New York, †1967 Princeton

https://doi.org/10.1515/9783110666502-006

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

6.1.1 Stress and Deformation

While parts of the following may seem complicated because of the large number of indices used, in fact the underlying concept of elastic deformation is quite simple. Referring to Figure 6.1, illustrating the various stress components, we start by imagining a small cube of the sample material whose edges are parallel to the orthogonal coordinate system. Three orthogonal forces are applied to each side of the elemental cube, with units of force per unit area. This is the (mechanical) **stress**.

The total stress on the material is described by the Cauchy stress tensor[2] $[\boldsymbol{\sigma}]$:

$$[\boldsymbol{\sigma}] = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix}, \tag{6.1}$$

where the first index of the tensor components σ_{ij} indicates the direction of the force, and the second index indicates the area to which the force is applied. In order to determine the stress components, we consider the small cube referred to above. In Figure 6.1 the components are shown following the sign convention that a stress is positive if it acts in the positive direction of the coordinate axes (tension) and negative if it acts opposite direction (compression). We should also note that, in equilibrium, for each force there must always be a corresponding counterforce to satisfy Newton's third law. This requirement has an important additional consequence for the non-diagonal stress components: In order that there is no resulting angular momentum on the specimen, in each case a second stress component must act to fulfill the condition $\sigma_{ij} = \sigma_{ji}$. This

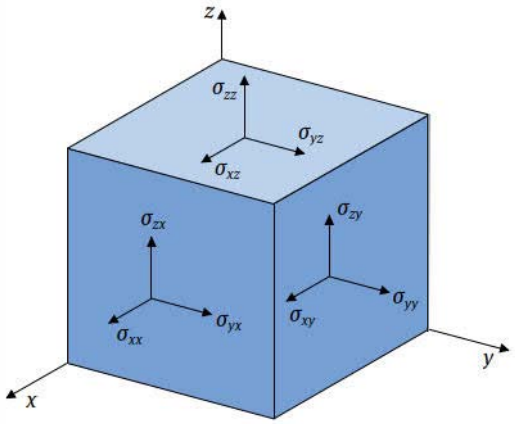


Fig. 6.1: Illustration of the stress components. For each surface, we can define the three stress components acting on it pointing along each of the three coordinate axes. For each stress component, σ_{ij}, the first index i indicates the direction of the force, and the second, j, the surface to which it is applied.

2 Baron Augustin Louis Cauchy, *1789 Paris, †1857 Sceaux

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

not only ensures that the stress tensor is symmetrical, but also reduces the number of independent stress components from 9 to 6.[3]

The **deformation** or strain is described by means of the strain tensor [**e**]:

$$[\mathbf{e}] = \begin{pmatrix} e_{xx} & e_{xy} & e_{xz} \\ e_{yx} & e_{yy} & e_{yz} \\ e_{zx} & e_{zy} & e_{zz} \end{pmatrix} . \tag{6.2}$$

The significance can be presented as follows: we consider the two adjacent points $P(\mathbf{r})$ und $Q(\mathbf{r} + \Delta\mathbf{r})$ on a sample subject to an external force. The force causes a different displacement at these two points, so that after the deformation we find P and Q at positions $(\mathbf{r} + \mathbf{u})$ and $(\mathbf{r} + \Delta\mathbf{r} + \mathbf{u} + \Delta\mathbf{u})$, respectively. While $\mathbf{u}$ describes the purely translational part of the displacement, $\Delta\mathbf{u}$ characterizes the deformation. If the relative displacement $\Delta\mathbf{u}$ is small compared to $\Delta\mathbf{r}$, we can then linearize and substitute the displacement components with $i = x, y, z$ by

$$\Delta u_i = \frac{\partial u_i}{\partial x}\Delta x + \frac{\partial u_i}{\partial y}\Delta y + \frac{\partial u_i}{\partial z}\Delta z . \tag{6.3}$$

The dimensionless displacement gradients are used to define the components e_{ij} of a strain tensor in the following form:

$$e_{ii} = \frac{\partial u_i}{\partial i} \qquad \text{and} \qquad e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial j} + \frac{\partial u_j}{\partial i}\right) . \tag{6.4}$$

The symmetrical components thus defined satisfy the condition $e_{ij} = e_{ji}$, which ensures that a rigid-body rotation of the sample is not interpreted as deformation.[4]

To familiarize ourselves with the concepts, we look at the special cases, of linear stress, uniform compression and shear. In Figure 6.2a a force in the y-direction acts on a cube with edge length L. The tensile stress σ_{yy} stretches the cube by the amount Δy. This stress causes the strain component e_{yy}, which is given by $e_{yy} = \partial u_y/\partial y = \Delta y/L$. The dashed arrow shows the stress that must act in the opposite direction to keep the sample in equilibrium. For simplicity, the transverse contraction that occurs at the same time is not shown.

Figure 6.2b shows the deformation of a cube caused by uniform compression. All sides of the cube are subjected to uniaxial stress, causing a relative volume reduction $\Delta V/V \approx -(e_{xx} + e_{yy} + e_{zz})$.

Finally, Figure 6.2c shows the shearing of a cube by the stress components σ_{yz} and σ_{zy}.[5] The two additional stress components ensuring the equilibrium of forces

3 In some textbooks, the indices are defined in reverse, with the first index indicating thearea on which the force acts and the second indicating the direction. However, since the stress tensor is symmetrical, the order of the indices is not important in the definition.

4 In addition, it should be noted that in some books the factor $\frac{1}{2}$ is omitted.

5 Here we consider *pure shear*, where the stress components are directed in such a way that no torque occurs. As we can easily see, just the stress component σ_{yz} together with the corresponding stress

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

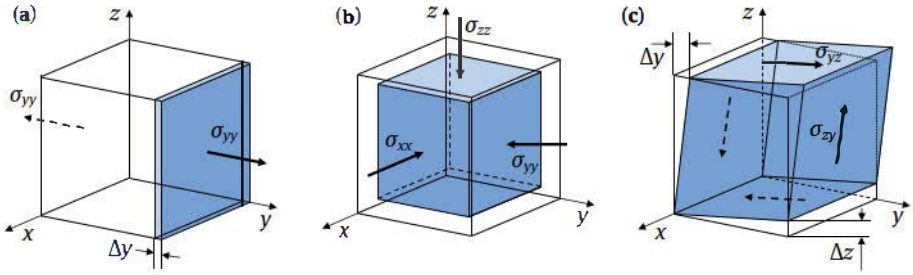


Fig. 6.2: Illustration of some fundamental aspects of the theory of elasticity. **a)** Stretching of a cube in the y-direction. A tension σ_{yy} causes elongation in the y-direction by an amount Δy. **b)** Uniform compression under pressure from all sides. For the sake of clarity, the forces acting on the bottom or rear sides are not shown. **c)** Pure shear deformation. In a) and c) the counterforces required for mechanical equilibrium are shown as dashed arrows.

are shown as dashed arrows. For the two equally large strain components, we have $e_{zy} = \partial u_z/\partial y = \Delta z/L$ and $e_{yz} = \partial u_y/\partial z = \Delta y/L$. The sample volume remains constant during shear.

6.1.2 Elastic Constants

Hooke's law, i.e. the linear relationship between the stress and the strain, has the general form:

$$\sigma_{ij} = \sum_{kl} c_{ijkl} e_{kl} \,. \tag{6.5}$$

The coefficients c_{ijkl} are called the elastic constants or the stiffness constants and constitute the components of the **elasticity tensor**. The elasticity tensor is a 4th order tensor with 81 components. Due to the symmetry of σ_{ij} and e_{kl}, the following equalities apply: $c_{ijkl} = c_{jikl} = c_{ijlk}$. This reduces the number of independent components to 36. The further relation $c_{ijkl} = c_{klij}$ follows from the quadratic dependence of the elastic energy on the deformation, giving finally 21 independent components.

The symmetry relations allow us to introduce a shortened notation, the so-called *Voigt[6] notation*, which combines two indices as follows. These are $xx \to 1$, $yy \to 2$, $zz \to 3$, $yz \to 4$, $zx \to 5$, and $xy \to 6$. This frequently used notation can also be used for the stress and strain tensor. However, there is a small "complication" with the

component in the opposite direction on the opposite side of the cube, which prevents translation, already cause shearing, but the torque does not disappear in this case. This type of shear is called *simple shear*. In the discussion of critical shear stress in Section 5.2, we have used this type of shear as a basis.

6 Woldemar Voigt, *1850 Leipzig, †1919 Göttingen

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

strain tensor: $e_{xx} \dashrightarrow e_1$, $e_{yy} \dashrightarrow e_2$, $e_{zz} \dashrightarrow e_3$, $2e_{yz} \dashrightarrow e_4$, $2e_{zx} \rightarrow e_5$ and $2e_{xy} \dashrightarrow e_6$. This modification follows from the expression for the energy density.

In many cases it is more convenient to use the inverse tensor instead of the elasticity tensor. The coefficients s_{ijkl} also known as the *compliance constants* are defined (in analogy with (6.5)) as:

$$e_{ij} = \sum_{kl} s_{ijkl}\sigma_{kl} \,. \tag{6.6}$$

The actual number of independent components depends on the crystal system. While all 21 elastic moduli are required to describe the elastic properties of triclinic crystals, only three are needed for the higher symmetry cubic crystals. The components of the elasticity tensor can be deduced from symmetry. For cubic crystals the following arguments apply: the three cubic axes of the crystal are equivalent. Therefore, the diagonal components for uniaxial strain and shear must be the same, giving $c_{11} = c_{22} = c_{33}$ and $c_{44} = c_{55} = c_{66}$. The same argument can also be applied to shear forces, i.e. $c_{12} = c_{13}$, etc. Since no normal stresses occur during shear, then $c_{14} = 0$, etc. Finally, shear along one direction does not produce shear perpendicular to that direction, so $c_{45} = 0$, and so on. Thus, the elasticity tensor for cubic crystals must take the following form:

$$[c] = \begin{pmatrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{12} & 0 & 0 & 0 \\ c_{12} & c_{12} & c_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{44} \end{pmatrix} . \tag{6.7}$$

In the case of cubic crystals, the elastic constants and compliance coefficients are related as follows:

$$c_{11} - c_{12} = \frac{1}{s_{11} - s_{12}}, \qquad c_{11} + 2c_{12} = \frac{1}{s_{11} + 2s_{12}} \qquad \text{and} \qquad c_{44} = \frac{1}{s_{44}} \,. \tag{6.8}$$

The elastic constants and the density of some cubic crystals are listed in Table 6.1.

For **isotropic materials** such as amorphous solids or polycrystalline substances with randomly oriented crystallites, the elasticity tensor looks like (6.7) but in addition the relationship $c_{11} = (c_{12} + 2c_{44})$ applies. In these materials there are therefore only two independent constants, namely $\lambda \equiv c_{12}$ and $\mu \equiv c_{44}$, which are usually called **Lamé constants**[7]. They are related to the frequently used Young's modulus[8] or elastic modulus E, the lateral strain coefficent ν (Poisson's ratio[9]), the bulk modulus B and the shear modulus G as follows

7 Gabriel Lamé, *1795 Tours, †1870 Paris
8 Thomas Young, *1773 Milverton, †1829 London
9 Siméon Denis Poisson, *1781 Pithiviers, †1840 Paris

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Tab. 6.1: Density and elastic constants of some cubic elements. (After W. Martienssen, *Springer Handbook of Condensed Matter and Material Data*, W. Martienssen, H. Warlimont, eds., Springer, 2005.)

Crystal	ϱ (g/cm^3)	c_{11} (10^{11} Pa)	c_{44} (10^{11} Pa)	c_{12} (10^{11} Pa)
Na	0.97	0.076	0.043	0.063
K	0.86	0.037	0.019	0.032
Cu	8.96	1.69	0.753	1.22
Ag	10.50	1.22	0.455	0.92
Au	19.30	1.91	0.422	1.62
Al	2.70	1.08	0.283	0.62
Cr	7.19	3.48	1.00	0.67
Fe	7.86	2.30	1.17	1.35
Ni	8.90	2.47	1.22	1.53
Si	2.33	1.66	0.796	0.639
Ge	5.32	1.29	0.671	0.483

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu}\,, \qquad \nu = \frac{\lambda}{2(\lambda + \mu)}\,, \tag{6.9}$$

and

$$B = \frac{3\lambda + 2\mu}{3}\,, \qquad \mu = G\,. \tag{6.10}$$

Table 6.2 shows the modulus of elasticity and Poisson's ratio of some technically important polycrystalline materials. Since the exact numerical values depend strongly on the manufacturing conditions, only rough values are given.

Tab. 6.2: Elastic constants of polycrystalline, isotropic materials. Al, Pb, Cu, Zn and Sn are cast samples.

	Young's modulus E / GPa	Poisson's ratio ν
Aluminum	68	0.3
Lead	55	0.5
Glass	55	0.16
Copper	76	0.4
Steel	200	0.3
Tungsten	400	0,3
Zinc	76	0.3
Tin	27	0.3

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

6.1.3 Sound Propagation

As in gases and liquids, elastic disturbances propagate in solids in the form of elastic waves. However, in addition to *longitudinal* sound waves, in solids *transverse* sound waves can also propagate owing to the shear stiffness of solids. Denoting the displacement of a selected small volume element from its equilibrium position by **u** and the wave vector of the sound wave by **q**, then for longitudinal waves $\mathbf{u} \parallel \mathbf{q}$ and for transverse waves $\mathbf{u} \perp \mathbf{q}$.

The properties of sound waves follow from the wave equation in the theory of elasticity. In the general case this equation is complex and we will not derive it here. Instead we derive the equation for the special case of a longitudinal wave in an isotropic medium. We will then generalize this equation. To start we take a sample subject to a spatially varying mechanical stress σ_{xx} which causes a displacement u_x of the small volume dV (see Figure 6.3) whose density is ϱ. The net force acting on the volume element is given by

$$dF_x = [\sigma_{xx}(x + dx) - \sigma_{xx}(x)]\,dy\,dz = \frac{\partial \sigma_{xx}}{\partial x} dx\,dy\,dz\,. \tag{6.11}$$

With Newton's second law[10] the equation of motion is given by:

$$\varrho \frac{\partial^2 u_x}{\partial t^2} = \frac{\partial \sigma_{xx}}{\partial x}\,. \tag{6.12}$$

If we also take into account that $\sigma_{xx} = c_{11} e_{xx} = c_{11}(\partial u_x/\partial x)$ (according to 6.5) we obtain the wave equation

$$\varrho \frac{\partial^2 u_x}{\partial t^2} = c_{11} \frac{\partial^2 u_x}{\partial x^2}\,. \tag{6.13}$$

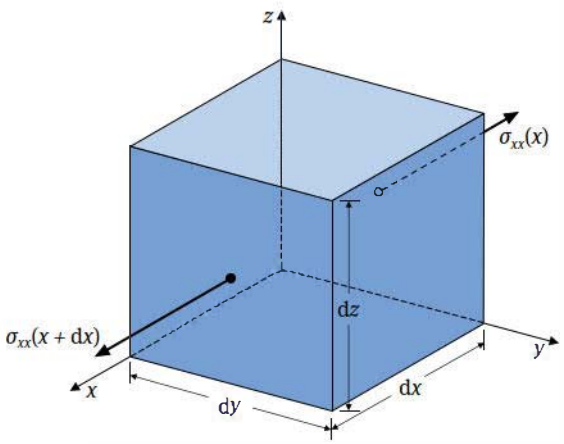


Fig. 6.3: Volume element of an isotropic solid which is subjected to a spatially varying stress in the x-direction.

10 Isaac Newton, *1642 Woolsthorpe-by-Colsterworth, †1727 Kensington

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

For anisotropic crystals, in addition to the stress component σ_{xx}, all the other components acting on the volume element must also be taken into account. If we treat the other two components of displacement u_y and u_z in the same way as with u_x, instead of (6.12), we obtain:

$$\varrho \frac{\partial^2 u_i}{\partial t^2} = \sum_j \frac{\partial \sigma_{ij}}{\partial j} = \sum_{jkl} c_{ijkl} \frac{\partial^2 u_l}{\partial j \partial k} \qquad (i,j,k,l = x,y,z) \,. \tag{6.14}$$

As a simple but instructive example, let us take a closer look at sound propagation in cubic crystals. We align our Cartesian coordinate system in such a way that the axes coincide with the ⟨100⟩-directions. Taking into account the elasticity tensor (6.7), the wave equation then takes the form:

$$\varrho \frac{\partial^2 u_x}{\partial t^2} = c_{11} \frac{\partial^2 u_x}{\partial x^2} + c_{44} \left(\frac{\partial^2 u_x}{\partial y^2} + \frac{\partial^2 u_x}{\partial z^2} \right) + (c_{12} + c_{44}) \left(\frac{\partial^2 u_y}{\partial x \partial y} + \frac{\partial^2 u_z}{\partial x \partial z} \right) . \tag{6.15}$$

The equivalent equations for the displacements u_y and u_z are obtained simply by the cyclical exchange of the indices x, y, z.

We do not want to discuss the relatively complicated general solution, but some special cases. First, we consider a longitudinal wave propagating in x-direction. In this case, we can set $\partial u_x/\partial y = \partial u_x/\partial z = 0$ and $u_y = u_z = 0$ and find

$$\varrho \frac{\partial^2 u_x}{\partial t^2} = c_{11} \frac{\partial^2 u_x}{\partial x^2} \,. \tag{6.16}$$

This equation is identical to (6.13), the wave equation for the propagation of longitudinal waves in an isotropic medium. As a solution we choose the plane wave $u_x = U_x \exp[-\mathrm{i}(\omega t - qx)]$, where U_x is the amplitude and q the wave number. Inserting this result into $\varrho\,\omega^2 = c_{11} q^2$ we obtain the simple **dispersion relation**[11]

$$\omega = \sqrt{\frac{c_{11}}{\varrho}}\, q = v_\ell\, q \,. \tag{6.17}$$

There is a *linear* relationship between frequency and wave number. The proportionality constant v_ℓ is the longitudinal velocity of sound, which is independent of frequency and wave number.[12] A linear relationship is also found with all other wave types as we will discuss later. This means that in elastic continua, other than for very large displacements, sound propagation is dispersion-free. For the velocity of longitudinal waves in the [100]-direction in cubic crystals, the following applies:

$$v_\ell = \frac{\omega}{q} = \sqrt{\frac{c_{11}}{\varrho}} \,. \tag{6.18}$$

11 By dispersion relation we mean the relationship between the frequency of the wave and its wave vector or wave number.

12 We do not distinguish here between the phase velocity $v = \omega/q$ and the group velocity $v_g = \partial\omega/\partial q$, since for linear dispersion there is no difference between the two quantities.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

We now investigate the propagation of a transverse wave in the x-direction (again with x parallel to the [100]-direction) with particle displacement in the y-direction. Assuming $u_x = u_z = 0$ and $u_y = U_y \exp[-i(\omega t - qx)]$ we find for the wave equation and the resulting sound velocity the relations:

$$\varrho \frac{\partial^2 u_y}{\partial t^2} = c_{44} \frac{\partial^2 u_y}{\partial x^2} \qquad \text{and} \qquad v_t = \frac{\omega}{q} = \sqrt{\frac{c_{44}}{\varrho}} \,. \tag{6.19}$$

Because we have cubic symmetry here, with propagation along the [100]-direction similar expressions are also obtained when the displacement is in the z-direction. Both shear waves propagate with the same velocity v_t. However, this does not apply if the sound propagation is in an arbitrary crystal direction.

Shear waves propagating in the xy-plane with displacements in z-direction are a special case. With $u_z = U_z \exp[-i(\omega t - q_x x - q_y y)]$, from the wave equation (6.15) we find

$$\varrho\, \omega^2 = c_{44}(q_x^2 + q_y^2) = c_{44}\, q^2 \,. \tag{6.20}$$

With this polarization the velocity of sound is independent of the direction of propagation within the xy-plane.

By means of sound velocity measurements in the [100]-direction, the two elastic constants c_{11} and c_{44} can be determined in an elegant way. The third constant c_{12} is obtained by measuring the sound velocity of a transverse wave of suitable polarization in [110]-direction. To show this, we consider waves whose wave vectors and displacements lie in the xy-plane. Taking for example $u_x = U_x \exp[-i(\omega t - q_x x - q_y y)]$ and $u_y = U_y \exp[-i(\omega t - q_x x - q_y y)]$ along with (6.15) results in the two equations:

$$\begin{aligned} \varrho\, \omega^2 u_x &= \left[c_{11} q_x^2 + c_{44} q_y^2\right] u_x + (c_{12} + c_{44})\, q_x q_y u_y \,, \\ \varrho\, \omega^2 u_y &= \left[c_{11} q_y^2 + c_{44} q_x^2\right] u_y + (c_{12} + c_{44})\, q_x q_y u_x \,. \end{aligned} \tag{6.21}$$

The solution in [110]-direction takes a particularly simple form, for which we write $q_x = q_y = q/\sqrt{2}$. Using the condition that the constraints equation of the system disappears, we obtain the two solutions

$$\varrho\, \omega^2 = \frac{1}{2}(c_{11} + c_{12} + 2c_{44}) q^2 \qquad \text{and} \qquad \varrho\, \omega^2 = \frac{1}{2}(c_{11} - c_{12}) q^2 \,. \tag{6.22}$$

By inserting into (6.21) we find that in the case of the first solution, the particle displacements satisfy the relationship $u_x = u_y$. This means that $\mathbf{u} \parallel \mathbf{q}$, so we are dealing with a longitudinal wave. For the second solution, $u_x = -u_y$. The displacement occurs in the $[1\bar{1}0]$-direction, so that $\mathbf{u} \perp \mathbf{q}$. We therefore have a transverse wave, which is particularly suitable for determining the elastic constant c_{12}.

We briefly summarize our results on sound propagation in the xy-plane in Figure 6.4, which shows the sound velocities of **vitreous silica** (a-SiO_2) and gal-

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

lium arsenide (GaAs).[13] Let us first look at the curves for vitreous silica. The velocity of sound waves does not depend on the direction of propagation and as with all glasses, a-SiO_2 shows isotropic elastic behavior. Since the transverse waves of different polarization are degenerate, only two circles appear in Figure 6.4a. The velocity for longitudinal waves is given by (6.18) and that for the transverse waves by (6.19). The longitudinal sound velocity of 5973 m/s is about 50% higher than the transverse sound velocity, 3766 m/s.

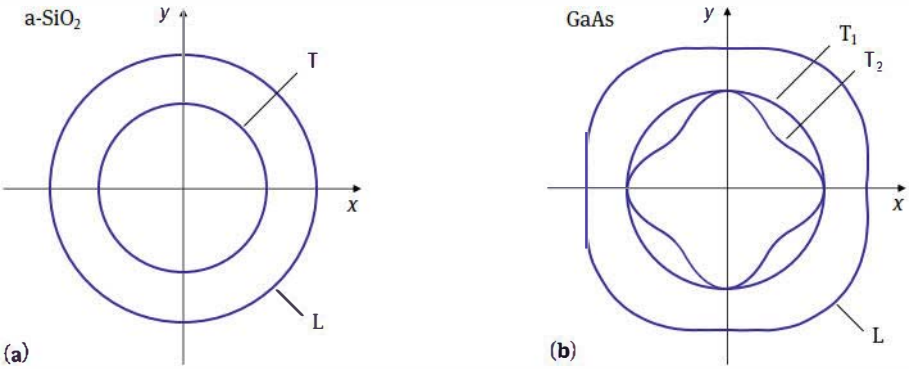


Fig. 6.4: Directional dependence of the velocity of sound. **a)** Wave propagation in vitreous silica is isotropic. The two differently-polarized transverse waves (T) are degenerate. **b)** Sound velocity in gallium arsenide when propagating in the xy-plane. The transverse waves T_1 and T_2 are polarized perpendicular and parallel to the drawing plane, respectively. In the ⟨100⟩-directions they are degenerate, in the other directions the velocities differ significantly.

Figure 6.4b shows the angular dependence of the sound velocity in gallium arsenide in the xy-plane. As expected, GaAs is elastically anisotropic, although the crystal structure is cubic. Except for the ⟨100⟩-directions, the velocities of the three differently polarized waves differ significantly. The two transverse waves are degenerate only along the major axes. In the ⟨100⟩-directions the longitudinal waves propagate at 4730 m/s, whereas the transverse waves propagate at 3340 m/s.

It is important that the polarization of the sound waves in any direction of propagation is generally not exactly parallel or perpendicular to the wave vector **q**. This means that the particle motion is not exactly longitudinal or transverse. Therefore, one speaks of **quasi-longitudinal** or **quasi-transverse waves**, depending on which type of wave is predominant. In cubic crystals sound waves with *pure polarization* only occur in ⟨100⟩-, ⟨110⟩- and ⟨111⟩-directions. There the analysis of the measured data is considerably simplified.

13 In the technical literature, it is not the velocity but the inverse velocity, the so-called "slowness", that is generally plotted.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Very simple experimental setups can be used for generating plane sound waves in bulk samples, as shown in the schematic diagram of Figure 6.5. Two piezoelectric transducers are simply attached to the ends of a sample and a very thin layer of a viscous liquid is used to establish acoustic contact between the transducers and the sample. A high-frequency pulse of, for example, 1 GHz and a duration of 1 μs is applied to the transmitting sound transducer which is thus briefly excited into oscillation.[14] This emits a sound pulse into the sample, which, in the case of samples with plane-parallel ends, runs repeatedly back and forth and is registered at the receiving end with each reflection. For samples of 1 cm length, the time interval between two echoes is typically about 5 μs. If the sample length is known, the sound velocity can be determined from the time sequence of the echoes. Due to the *ultrasonic attenuation*, the observed signal amplitude decreases with the running distance x or with the running time t. The relationship $I = I_0 \exp(-x/\ell) = I_0 \exp(-vt/\ell)$ is found for the sound intensity I, where ℓ is the **mean free path length** and I_0 the intensity at time $t = 0$. Various mechanisms contributing to the attenuation of sound waves are discussed in the next chapter.

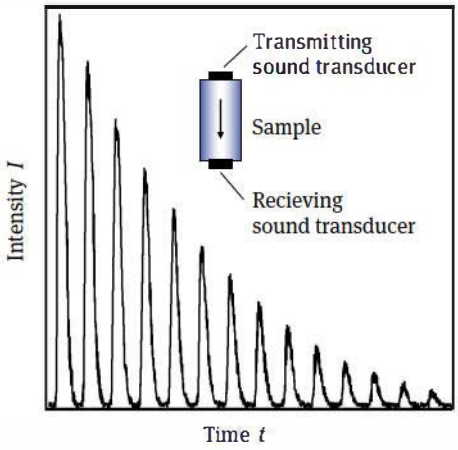


Fig. 6.5: Ultrasonic echoes in a glass sample. The oscilloscope image shows only the envelope of the sound pulses and not the high-frequency 1 GHz oscillations. The measurement setup is shown schematically at upper right.

By suitable selection of the orientation of the piezoelectric transducers, longitudinal or transverse waves can be generated and detected and thus the two constants c_{11} and c_{44} can be separately determined for isotropic materials. For anisotropic crystals, sound propagation in several different directions and with different polarization must be measured if all the elastic constants are to be determined using this technique.

14 In most experiments only *one* transducer is used. In these cases the transducer is used both as transmitter and receiver

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

6.2 Lattice Vibrations

We now move on from the simple limiting case of the elastic continuum and take a closer look at the dynamics of real crystals and deal with their vibrational excitations. As mentioned above, we again use the adiabatic approximation and disregard anharmonic effects. We start with very simple special cases and then generalize the concept. An important result will be that we will find considerable deviations from the behavior of continuum sound waves, when the wavelengths become comparable to the lattice constant. Furthermore, we find that reciprocal lattice vectors also play a very important role in the lattice dynamics.

We will study the dynamics of crystals with monatomic and diatomic bases separately, since these two types of crystals behave qualitatively differently. Extending the treatment to crystals with larger bases is conceptually simple, hut requires more extensive calculation. To make the mathematics as simple as possible, we begin with a major simplification: we consider a one-dimensional *linear chain*. One might think that any results derived in this way would have only a very limited applicability to three-dimensional crystals, but for reasons explained below, the results in fact closely reflect the 3-D behavior.

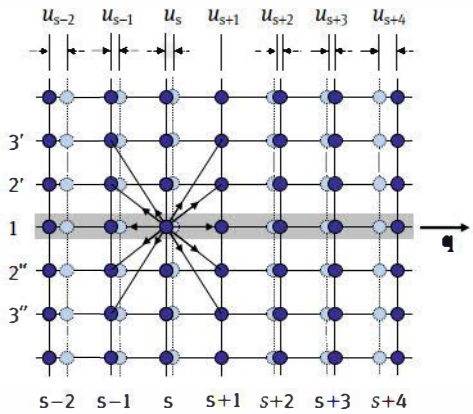


Fig. 6.6: Propagation of longitudinal waves in an orthorhombic crystal. The displacement of the atoms and the propagation of the wave is in the [100]-direction. The momentary position of the atoms is shown by dark blue points, their position at rest by light blue points. The arrows symbolize the forces acting on the selected atom. The considered linear chain is marked by a grey bar.

If we consider the propagation of a plane wave along a crystal direction with high symmetry, all forces that do not act in the direction of the displacement compensate each other due to symmetry. This is shown in Figure 6.6, which illustrates a section along the basis (i.e. along a (●01)-plane) of an orthorhombic crystal with a monatomic basis. If a longitudinal wave propagates in the [1●0]-direction, the atoms are only affected by forces along this direction. For example, in the displacement of the lattice planes shown, the atoms 1, $2'$, $2''$, ... of plane $(s-1)$ exert forces on the reference atom in plane (s). The resulting force, however, points exclusively in the direction of the displacement, since the perpendicular forces compensate each other.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Corresponding arguments also apply to the propagation of transverse waves in directions of high symmetry, because only forces perpendicular to the wave vector act. Since the distance between the lattice planes does not change in the direction of propagation, the forces in the longitudinal direction compensate each other. For cubic crystals, the previous arguments are also valid for the ⟨110⟩- and ⟨111⟩-directions.

In general, during wave propagation in directions of high symmetry, the displacement of the atoms is either purely longitudinal or purely transverse, allowing the mathematical treatment to be reduced to a one-dimensional problem. It is sufficient to pick out a chain, because all atoms of a lattice plane perform the same motion. In the equation of motion *effective forces* enter, since each atom is not only affected by the atoms along the chain, but also by the other neighboring atoms. The forces perpendicular to the displacement compensate each other as described above.

However, for wave propagation in an arbitrary direction, simple symmetry arguments cannot be used in general. In these cases, the forces acting on atoms do not necessarily act in the direction of propagation or perpendicular to it. The atoms then describe trajectories that can no longer be approximated by the motion of a linear chain. Such waves have a *mixed polarization* and are called *quasi-longitudinal* or *quasi-transverse*, depending on which displacement predominates, as in the case of sound waves.

6.2.1 Lattice with a Monatomic Basis

We will now set up the equation of motion for a particular atom in plane s (see Figure 6.6). For this purpose, we first consider the forces acting on a single lattice plane. In the harmonic approximation, a small displacement of an adjacent plane $(s + n)$ causes a force proportional to the displacement u_{s+n} to act on the plane s. If the plane s is also displaced, the resulting force F_s is proportional to the difference $(u_{s+n} - u_s)$. When calculating F_s, we must therefore add the contribution from all planes and obtain

$$F_s = \sum_n \mathcal{C}_n(u_{s+n} - u_s) \,, \tag{6.23}$$

summing over all integers n. The **force constants** $\mathcal{C}_n$ reflect the strength of the interaction between the planes s and $(s + n)$ and are therefore different for longitudinal and transverse displacements.

In constructing the equation of motion, we make use of the fact that the problem is one-dimensional and can be reduced to that for a *linear chain*. Picking out the chain indicated by the grey bar in Figure 6.6 we can write the equation of motion for the reference atom in plane s:

$$M\frac{\mathrm{d}^2 u_s}{\mathrm{d}t^2} = \sum_n C_n(u_{s+n} - u_s) \,. \tag{6.24}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

where M is the atomic mass and C_n is the effective force constant. The latter takes into account the interaction of the *reference atom* with all the atoms in the plane $(s + n)$.

Guessing a solution for the displacement of the atoms, we choose a propagating plane wave of the form:

$$u_{s+n} = U e^{-i[\omega t - q(s+n)a]}, \tag{6.25}$$

where a is the equilibrium separation of the lattice planes. Inserting this expression into (6.24), we obtain:

$$\omega^2 M = \sum_n C_n (1 - e^{iqna}). \tag{6.26}$$

Since, for reasons of symmetry, $C_{-n} = C_n$, we can limit the summation to positive values of n and obtain:

$$\omega^2 = \frac{1}{M} \sum_{n=1}^{\infty} C_n \left(2 - e^{iqna} - e^{-iqna}\right) = \frac{2}{M} \sum_{n=1}^{\infty} C_n [1 - \cos(qna)]. \tag{6.27}$$

In most solids, the interaction between the atoms falls off so rapidly that the summation can be truncated after the nearest or next nearest neighbor without causing any major error. In the first case we only need the first force constant C_1 and we find:

$$\omega^2 = \frac{2C_1}{M}(1 - \cos qa) = \frac{4C_1}{M} \sin^2\left(\frac{qa}{2}\right) \quad \text{and} \quad \omega = 2\sqrt{\frac{C_1}{M}} \left|\sin\left(\frac{qa}{2}\right)\right|. \tag{6.28}$$

If we also take into account the interaction with the next nearest neighbor with the inclusion of the C_2 force constant term, we obtain:

$$\omega^2 = \frac{4C_1}{M}\left[\sin^2\left(\frac{qa}{2}\right) + \frac{C_2}{C_1}\sin^2(qa)\right]. \tag{6.29}$$

The dispersion curve for the case where we only consider the interactions with the nearest neighbors is shown in Figure 6.7. If further force constants are also included, quantitative changes occur, but the qualitative behavior is hardly affected. How good the respective approximation is depends on the system under consideration. While, in

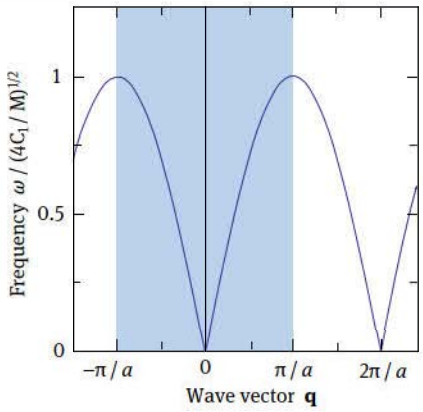


Fig. 6.7: The dispersion curve for a linear chain with only the interactions between nearest neighbors taken into account. The first Brillouin zone is highlighted in blue.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the case of molecular and noble gas crystals, the interactions drops off rapidly with distance and neighbors beyond nearest neighbors hardly contribute at all, interactions with more distant neighbors is of importance in metals.

To understand lattice vibrations better, let us first familiarize ourselves with a number of particular aspects. First, we look at the phases of neighboring atoms. With the help of (6.25) we find:

$$\frac{u_{s+1}}{u_s} = \frac{U e^{-i\omega t}\, e^{iq(s+1)a}}{U e^{-i\omega t}\, e^{iqsa}} = e^{iqa} \,. \tag{6.30}$$

Since the phase difference between neighboring atoms can only be between 0 and 2π, the wave vector can be limited to the values

$$-\pi < qa \leq \pi \qquad \text{or} \qquad -\frac{\pi}{a} < q \leq \frac{\pi}{a} \,. \tag{6.31}$$

This is just the range of the *first Brillouin zone* of a linear chain or of a primitive lattice in the direction of the main lattice axes. If a wave vector $\mathbf{q}'$ lies outside the first Brillouin zone, it can be moved back into it simply by adding a reciprocal lattice vector $2\pi p/a$, where p is a positive or negative integer:

$$q' = q + \frac{2\pi p}{a} \,. \tag{6.32}$$

This procedure is called **reduction to the first Brillouin zone**. The validity of this procedure is easily shown by inserting (6.32) into (6.30). For the phase difference between neighboring atoms, the relationship holds

$$\frac{u_{s+1}}{u_s} = e^{iq'a} = e^{iqa} e^{\pm 2\pi i p} = e^{iqa} \,. \tag{6.33}$$

Equation (6.32) reflects the fact that the wave vector of an oscillation of coupled mass points reflects the wave amplitude only at the actual position of the points. In other words, from a physical point of view, it is not reasonable to speak of a wave *between* the

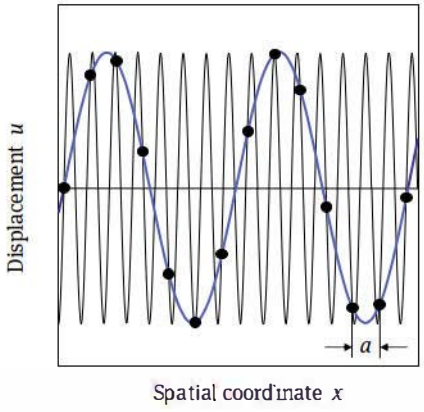


Fig. 6.8: The periodic displacement of a linear chain of atoms. The instantaneous positions of the atoms (dots) can be described formally either by a wave with $\lambda < 2a$ (black) or a wave with $\lambda > 2a$ (blue). The wave numbers of the two waves differ by $2\pi/a$, i.e. by a reciprocal lattice vector.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

mass points. This is illustrated in Figure 6.8, showing a given configuration of displaced atoms which can be represented both by a wave with $\lambda < 2a$ and with $\lambda > 2a$.

In this context, it is surprising that the addition of the reciprocal lattice vector apparently changes the propagation direction of the wave: a wave, which runs in the direction of the positive x-axis and whose wave vector lies in the second Brillouin zone, has a wave-vector which points in the opposite direction after the addition of the reciprocal lattice vector $G = -2\pi/a$. This means that the *phase velocity* $v = \omega/q$ changes its sign. As can readily be seen from Figure 6.9, the *group velocity* $v_g = \partial\omega/\partial q$ and thus the energy transport does not change. In fact, the two descriptions of atomic displacements are equivalent. This result will be of great importance in Chapter 7 when discussing thermal conductivity.

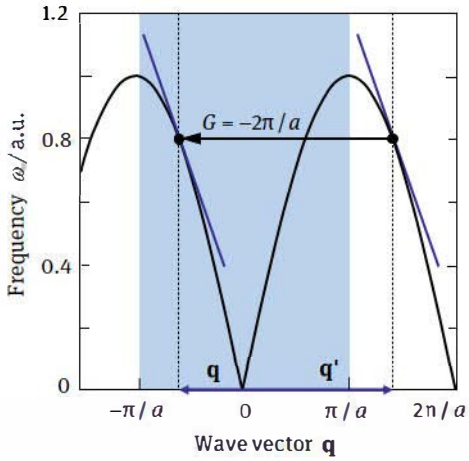


Fig. 6.9: Reduction to the first Brillouin zone. By adding the reciprocal lattice vector $G = -2\pi/a$, the wave vector $\mathbf{q}'$ is replaced by vector $\mathbf{q}$. However, the *slope* of the dispersion curves for the specified wave vectors is not changed by this operation and therefore the group velocity is not affected.

It is worthwhile taking a closer look at the lattice vibrations in the limiting cases of very long and very short wavelengths. In the **long-wavelength** limiting case, i.e for $q \to 0$, with an expansion of the cosine term of (6.27) only the first two terms are of importance and the equation simplifies to:

$$\omega^2 = \frac{2}{M}\sum_{n>0} C_n \left[1 - \cos(qna)\right] \approx \frac{q^2 a^2}{M}\sum_{n>0} n^2 C_n \,. \tag{6.34}$$

Clearly in this case there is a linear relationship between ω and q, as we have already seen in the form of equation (6.17) for sound waves. The resulting phase and group velocities follow directly from the definitions $v = \omega/q$ and $v_g = \partial\omega/\partial q$. This gives us a microscopic expression for the elastic constants c_{11} and c_{44}. Applying the result to a simple-cubic crystal with density $\varrho = M/a^3$, and comparing (6.34) with (6.18) and (6.19) leads to the result:

$$c_{11} = \sum_{n>0} \frac{n^2}{a} C_n^{\ell} \qquad \text{and} \qquad c_{44} = \sum_{n>0} \frac{n^2}{a} C_n^{t} \,, \tag{6.35}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

with the indices "ℓ" or "t" added to distinguish between longitudinal and transverse displacements. At this point it should be emphasized again that all the arguments here also apply to transverse waves. However, with a few exceptions, the values of the force constant C_n^t and thus the restoring forces for shear motions are always smaller than C_n^ℓ. Therefore the general relationship $v_\ell > v_t$ applies.

In the limiting **short-wavelength** case i.e. for lattice vibrations with wave vectors at the *boundary of the Brillouin zone*, the phase shift between the displacement of adjacent lattice planes is particularly interesting. From (6.30) it follows directly that for $q \to \pm\pi/a$, $u_{s+1}/u_s = \exp(\pm i\pi) = -1$. In other words, neighboring lattice planes thus oscillate with the opposite phase. This means that waves with the wave vector $\pm\pi/a$, i.e. with wavelength $\lambda = 2a$, take the form of standing waves. We can think of this in another way, as with X-rays, elastic waves are subject to Bragg reflection. This follows from the scattering theory discussed in Section 4.2, since this is independent of the nature of the incident waves. We use the lattice constant a in (4.24) as the period of density variation and obtain $2a \sin\theta = \lambda$. If the waves propagate perpendicularly to the lattice planes, then $\theta = 90^\circ$ and we expect a Bragg reflection for wavelength $\lambda = 2a$. The Incoming and reflected wave are thus superimposed on each another, forming a standing wave.

The occurrence of standing waves at the boundary of the Brillouin zone is also reflected in the behavior of the group velocity $v_g = \partial\omega/\partial q$. By deriving (6.28), we find that the group velocity for $q \to \pi/a$ approaches zero independently of the force constant C_n. This remark applies exclusively to directions with high lattice symmetry, which we have implicitly assumed in our treatment. In the general case, while the normal component of the group velocity disappears, the dispersion curve at the boundary of the Brillouin zone may well have a finite slope.

Although the discussion here has been in the context of the propagation of lattice waves in directions of high symmetry, the conclusions are also qualitatively valid for other directions. The reduction to the first Brillouin zone is possible in all cases and the dispersion curves in arbitrary direction are similar. Some examples of this will be discussed in Section 6.3 below.

6.2.2 Lattice with a Polyatomic Basis

If the primitive unit cell contains several atoms, then further effects occur. To explain the basic principles, it is again sufficient to study the particularly simple case of lattice vibrations propagating along directions of high symmetry, with the results also applying qualitatively propagation in arbitrary directions.

We consider a lattice with a diatomic basis consisting of an atom A and an atom B, and a lattice constant of a, where the atoms are arranged as shown in Figure 6.10. As in the case of the monatomic lattice, we can imagine that it is a (001)-plane of an orthorhombic crystal. For an approximate description of the behavior of three-

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

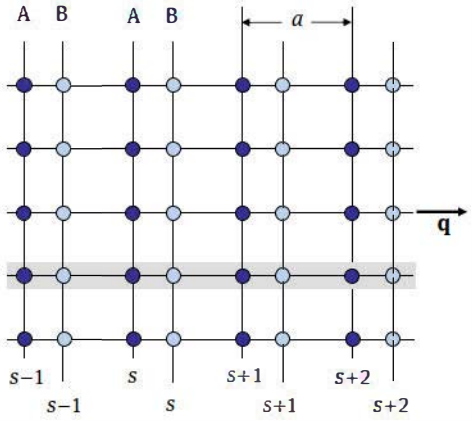


Fig. 6.10: For derivating the vibrations for a lattice with a diatomic basis. We take each plane to contain only atoms of one kind, with dark blue points representing atoms A, the light blue points atoms B. The linear chain under consideration is marked by the gray bar.

dimensional crystals we again consider the linear chain. In Figure 6.10 the dark blue circles denote atoms A and the light blue circles atoms B. Their positions in the linear chain are labelled by the lattice-plane indices. As in the case of monatomic-basis crystals, forces perpendicular to the propagation direction of a longitudinal wave compensate each other. The linear-chain approximation proves to be good even if the atoms in adjacent planes are offset from each other.

We assume that the forces decrease so rapidly with distance that only interactions with the immediately adjacent atoms needs to be considered and that the force constants are the same for the next adjacent pairs of planes. The equations of motion for atoms A with displacement u and mass M_1 and for atoms B with displacement v and mass M_2 are analogous to (6.24) and can be easily found with the help of Figure 6.10. This results in:

$$M_1 \frac{d^2 u_s}{dt^2} = C(v_s + v_{s-1} - 2u_s) ,$$

$$M_2 \frac{d^2 v_s}{dt^2} = C(u_s + u_{s+1} - 2v_s) . \tag{6.36}$$

The simplifications assumed above describe well the vibrations of ionic crystals for example. If, on the other hand, the crystal consists of diatomic *molecules*, then this treatment is not suitable.

Taking again as a solution a plane wave, but with different amplitudes U and V for the two types of atom:

$$u_s = U e^{-i(\omega t - qsa)} \quad \text{and} \quad v_s = V e^{-i(\omega t - qsa)} , \tag{6.37}$$

and inserting (6.37) into (6.36) we find:

$$(2C - \omega^2 M_1)U - C(1 + e^{-iqa})V = 0 ,$$

$$-C(1 + e^{iqa})U + (2C - \omega^2 M_2)V = 0 . \tag{6.38}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The condition for the existence of non-trivial solutions results in

$$\omega_{a,o}^2 = C\left(\frac{1}{M_1} + \frac{1}{M_2}\right) \mp C\sqrt{\left(\frac{1}{M_1} + \frac{1}{M_2}\right)^2 - \frac{4}{M_1 M_2}\cdot \sin^2\left(\frac{qa}{2}\right)}, \qquad (6.39)$$

where the indices "a" and "o" indicate the negative and positive sign representing *acoustic* and *optical*, respectively. The reason for this notation is explained below. The shape of the dispersion curves is shown in Figure 6.11. Since there are two separate curves, we speak of two **branches**. Thus the lattice vibrations of a linear chain with a monatomic or a diatomic basis already differ in the number of branches.

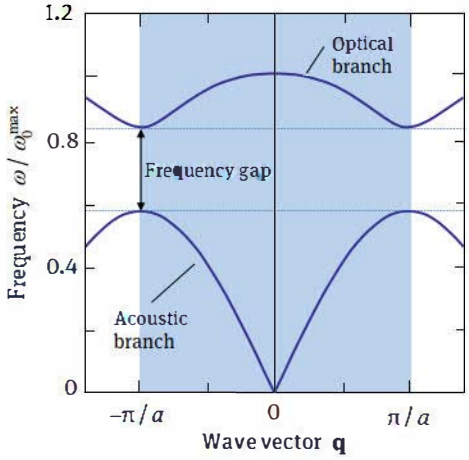


Fig. 6.11: The dispersion curves for lattice vibrations in crystals with a diatomic basis. The optical branch lies above the acoustic branch, separated by a *forbidden* band.

We note that the **acoustic branch** with the minus sign in (6.39) corresponds to the behavior of a chain with a monatomic basis, discussed above. This can easily be seen for the special case $M = M_1 = M_2$, where the basis vector is twice as large for a two-atomic basis and the lattice spacing a is replaced by $2a$ in the above equation. Since we have already discussed the acoustic vibrations in detail, we do not need to discuss them further here.

Now we examine the **optical branch**, i.e. where the square root term in (6.39) carries the positive sign. For low wave vectors, $q \to 0$ we obtain the dispersion relation

$$\omega_o^2 = \frac{2C}{\mu}, \qquad (6.40)$$

where μ, the *reduced mass*, is given by $\mu^{-1} = (M_1^{-1} + M_2^{-1})$. As can also be seen in Figure 6.11, near the Γ-point, i.e. at $q \approx 0$, the frequency of the optical branch is almost independent of the wave vector. The fact that acoustic and optical lattice vibrations can have two completely different frequencies at the same wave vector arises from the different phase relation of the two types of atom.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

We get more insight into this by looking at the relative phases of the vibrating atoms. From equation (6.38) we obtain for $q \to 0$:

$$\frac{U}{V} \approx \frac{2C}{2C - \omega^2 M_1} \,. \tag{6.41}$$

For the acoustic branch $\omega \approx 0$ and thus $U \approx V$. The two kinds of atoms oscillate with (almost) the same amplitude and the same phase. In contrast, we find for the optical branch $U/V = -M_2/M_1$, because here we have $\omega_o^2 = 2C/\mu$. This means that all atoms with mass M_1 have the opposite phase to those with mass M_2. This type of displacement can be most easily represented for transverse lattice vibrations, although it is of course also valid for longitudinal waves. Figure 6.12 illustrates the displacements of the atoms for the two branches in the case of (relatively) long wavelengths.

Obviously, long-wavelength acoustic lattice waves, i.e. for $q \to 0$, are identical to normal sound waves. However, in the case of optical lattice vibrations, an oscillating electric dipole moment is associated with the antiphase displacement of the sublattices if the oscillating atoms carry opposite charges. This means that such vibration modes can couple to electromagnetic waves which largely determine the optical properties in the infrared range (cf. Section 13.3). Crystals with such properties are said to be **infrared active**. Well-known examples are ionic crystals. We find a different behavior when the basis consists of two identical atoms, as in the case of diamond or silicon. Here, the counter oscillations of the sublattices against each other do not give rise to a dipole moment. Therefore, no infrared absorption is observed. As we will see in the following Section 6.3, such vibrations can be detected by Raman scattering.

At the boundary of the Brillouin zone, i.e. for $q \to \pi/a$, assuming $M_1 < M_2$, we find for the dispersion relation, the expressions $\omega_o^2 = 2C/M_1$ and $\omega_a^2 = 2C/M_2$. If we put this result into one of the two equations (6.38), we get the amazing result $V/U = 0$ and $U/V = 0$. In other words, depending on the branch under consideration, either the sublattice of the heavy or the light atoms sits at rest while the other sublattice oscillates. For lattice vibrations with wave vectors near the boundary of the Brillouin zone, the

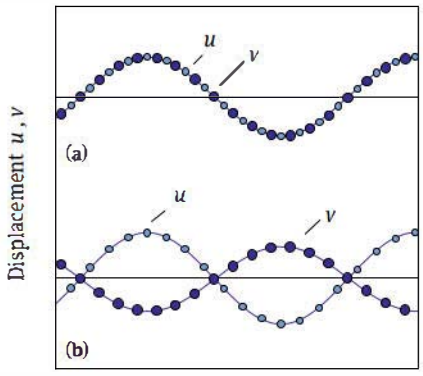


Fig. 6.12: Atomic displacement of long wavelength lattice vibrations for a diatomic basis. Light blue and dark blue dots mark the two different kinds of atom. The amplitudes are shown greatly exaggerated in comparison to the wavelength. **a)** Acoustic lattice vibrations: the neighboring atoms have almost the same phase. **b)** Optical lattice vibrations: the atoms of the two sublattices oscillate in anti-phase.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

oscillating dipole moment disappears due to the antiphase oscillation of the ions in the neighboring unit cells.

As mentioned above, if $M_1 \neq M_2$ or if the force constants differ between the individual lattice planes, a **frequency gap** opens at the boundary of the Brillouin zone. Within this **forbidden zone** there are no eigen-vibrations or modes of the lattice. If elastic waves with such a frequency are excited, they decay *exponentially* within a few wavelengths. The width of the forbidden zone is determined by the ratio of the atomic masses and the coupling constants.

The results we have obtained using linear chains can be largely transferred to real crystals if we take lattice waves propagating in directions of high symmetry. Of course, the agreement with the measured form of the dispersion curve is not satisfactory in all cases. This is especially true for optical branches, where several force constants have to be considered due to the opposite phases of the displacements. Nevertheless, in all cases, three acoustic branches are found. There is **one longitudinal** branch, and **two transverse**, the latter with perpendicular polarization directions. In directions of high lattice symmetry, branches may coincide. This is especially true for isotropic media, whose transverse branches are always degenerate.

If a primitive unit cell contains more than just one atom, further optical branches appear in addition to the acoustic ones. In general, a basis with n atoms will have $3n$ branches, 3 acoustic and $(3n - 3)$ optical. As with the acoustic branches, there are twice as many transverse as longitudinal branches in the optical case. With increasing size of the unit cell, the vibration spectrum becomes more complex and confusing as the number of optical branches increases.

Figure 6.13 shows schematically the dispersion curves for a crystal with a diatomic basis. It should be emphasized again here that in general, unless we are considering a propagation direction with high crystal symmetry, the branches are not degenerate

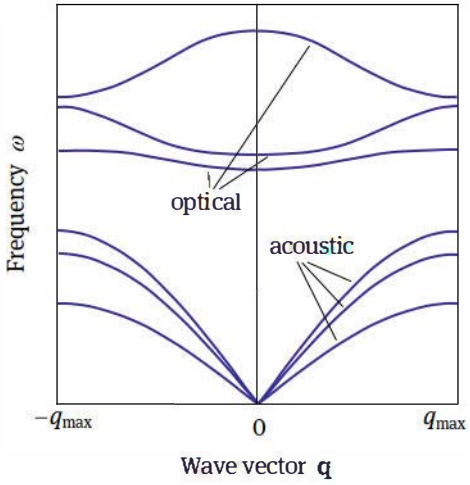


Fig. 6.13: Schematic representation of the six phonon dispersion curves for crystals with a diatomic basis. Except for directions of high lattice symmetry, the different lattice-vibration branches are not degenerate.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

and have no clearly defined polarization, so that we can then only speak of quasi-longitudinal or quasi-transverse vibrations.

6.2.3 The Equation of Motion

Now that we have familiarized ourselves with the dynamics of the crystals on the basis of the simple high-symmetry special cases, we will now briefly consider the general description of the lattice vibrations. This subsection is primarily addressed to readers interested in the formal theoretical treatment.

As a first step we set up the equation of motion of the atoms paying special attention to interatomic coupling. Unfortunately, for an unambiguous description of the atomic positions and the atomic displacements, many indices are required, which can be confusing. In the second step, making use of the translation symmetry of the lattice we arrive at a much simpler set of equations. Finally, as in the previous section, we make use of the fact that plane waves provide particularly simple solutions to the equation of motion.

Since all the atoms in a solid interact with each other, the total atomic potential energy is a function of the instantaneous position of all the atoms. We therefore introduce the **potential function** V, which represents the potential energy of all the atoms in the entire solid as a function of their momentary displacement from the equilibrium position. We expand this function into a Taylor series and truncate it after the second-order term. As we dealing with the region around equilibrium, the linear term must vanish, therefore we can calculate the potential function in the following form

$$\widetilde{V} = \widetilde{V}_0 + \frac{1}{2} \sum_{\mathbf{R}\alpha i} \sum_{\mathbf{R}'\alpha' i'} \frac{\partial^2 \widetilde{V}}{\partial u_{\mathbf{R}\alpha i} \partial u_{\mathbf{R}'\alpha' i'}} \bigg|_0 u_{\mathbf{R}\alpha i} u_{\mathbf{R}'\alpha' i'} + \dots \,. \tag{6.42}$$

where the quantity $u_{\mathbf{R}\alpha i}$ represents the component i of the displacement of atom α in the unit cell, defined by the lattice vector $\mathbf{R}$. The constant $\widetilde{V}_0$ simply defines the zero of the potential energy, and being of no importance for further discussion can therefore be omitted. If we neglect the higher order terms, then we are limited to the *harmonic approximation*.

To clarify the ensuing treatment, we are making use of the formal similarity with the potential of a harmonic oscillator. If a particle moves in a harmonic potential $\mathcal{V} = \tilde{C}x^2/2$, then the restoring force $\tilde{F} = -\tilde{C}x$ acts on this particle, where the force constant is given by $\tilde{C} = \mathrm{d}^2\mathcal{V}/\mathrm{d}x^2$. This is the basis for (6.42) and for the further procedure in dealing with the equation of motion of the atoms.

First we define the generalized force constant $C_{\mathbf{R}\alpha i, \mathbf{R}'\alpha' i'}$ with the help of the equation

$$C_{\mathbf{R}\alpha i, \mathbf{R}'\alpha' i'} \equiv \frac{\partial^2 \widetilde{V}}{\partial u_{\mathbf{R}\alpha i} \partial u_{\mathbf{R}'\alpha' i'}} \,. \tag{6.43}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In order to be able to use vector notation in the following, we denote the three spatial directions by the indices i or i' and use the shortened notation $\mathbf{C}_{\mathbf{R}\alpha,\mathbf{R}'\alpha'}$ for the force constants. The force constants can be combined into a Cartesian tensor, the **force constant tensor** [**C**].

The force acting on atom α in the unit cell with lattice vector $\mathbf{R}$ when another atom α' with lattice vector $\mathbf{R}'$ undergoes a displacement $\mathbf{u}_{\mathbf{R}'\alpha'}$ can be written:

$$\mathbf{F}_{\mathbf{R}\alpha} = -\mathbf{C}_{\mathbf{R}\alpha,\mathbf{R}'\alpha'} \cdot \mathbf{u}_{\mathbf{R}'\alpha'} \,. \tag{6.44}$$

Therefore, in the formalism used here, in principle the coupling between all the atoms are considered. If the reference atom itself is displaced the equation still applies, but $\mathbf{R}' = \mathbf{R}$ and $\alpha' = \alpha$ must be set.

Taking into account the translational symmetry of the crystal, we can make a considerable simplification of the force constant tensor. Since the unit cells are not distinguishable, the components of [**C**] do not depend on the absolute position of the cell under consideration, but are only a function of $\tilde{\mathbf{R}} = (\mathbf{R}' - \mathbf{R})$, the relative position of the interacting atoms. We can therefore write: $\mathbf{C}_{\mathbf{R}\alpha,\mathbf{R}'\alpha'} = \mathbf{C}_{\alpha\alpha'}(\tilde{\mathbf{R}})$.

Thus a consideration of the crystal symmetries reduces the number of independent constants, thereby simplifying the equations and their treatment. Since the coupling forces decrease rapidly with distance, it is often sufficient to include only the nearest neighbors in the calculation as an approximation. In other words, each atom is then simply connected to its neighbors by a spring.

With the help of (6.44) we can calculate the equation of motion of atom α with mass M_α in the unit cell with lattice vector $\mathbf{R}$ by

$$M_\alpha \ddot{\mathbf{u}}_{\mathbf{R}\alpha} = -\sum_{\tilde{\mathbf{R}}\alpha'} \mathbf{C}_{\alpha,\alpha'}(\tilde{\mathbf{R}}) \cdot \mathbf{u}_{\mathbf{R}+\tilde{\mathbf{R}},\alpha'} \,, \tag{6.45}$$

where the sum is over all atoms. If there are $\mathcal{N}$ unit cells with n atoms each, we have $3n\,\mathcal{N} = 3\,N$ coupled differential equations, where N is the number of atoms in the sample.

Decoupling into independent equations can be easily achieved by using plane waves. Allowing different amplitudes for the various atoms in a unit cell, we can write:

$$\mathbf{u}_{\mathbf{R}\alpha}(t) = \mathbf{U}_\alpha \, \mathrm{e}^{-\mathrm{i}[\omega t - \mathbf{q}\cdot\mathbf{R}]} \,, \tag{6.46}$$

where $\mathbf{U}_\alpha$ is the amplitude associated with atom α. In this approach we have not explicitly emphasized the phase factor, which depends on the position of the atom α in the unit cell. Inserting (6.46) into (6.45), we obtain:

$$\omega^2 M_\alpha \mathbf{U}_\alpha = \sum_{\alpha'} \left[\sum_{\tilde{\mathbf{R}}} \mathbf{C}_{\alpha,\alpha'}(\tilde{\mathbf{R}}) \, \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\tilde{\mathbf{R}}} \right] \mathbf{U}_{\alpha'} = \sum_{\alpha'} \mathbf{D}_{\alpha,\alpha'}(\mathbf{q}) \mathbf{U}_{\alpha'} \,. \tag{6.47}$$

The expression in brackets is the Fourier transform of the force constant tensor [**C**], for which we use the abbreviation:

$$\mathbf{D}_{\alpha,\alpha'}(\mathbf{q}) \equiv \left[\sum_{\tilde{\mathbf{R}}} \mathbf{C}_{\alpha,\alpha'}(\tilde{\mathbf{R}}) \, \mathrm{e}^{\mathrm{i}\mathbf{q}\cdot\tilde{\mathbf{R}}} \right] \,. \tag{6.48}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The matrix [**D**] with components $\mathbf{D}_{\alpha,\alpha'}(\mathbf{q})$ is often referred to as **dynamical matrix**. It contains all the information needed to describe the elastic properties of an ideal crystal. The linear system of equations (6.47) is homogeneous and only of order $3n$. For a crystal with a monatomic basis, $n = 1$, so that for each wave vector $\mathbf{q}$ we need only solve three equations for each of the three spatial directions. The resulting three solutions describe waves with different polarizations.

The eigenvalues can be obtained as in usual vibration theory: The system of equations (6.47) has non-trivial solutions only if the coefficient determinant disappears. That is, if:

$$\det\left\{\mathbf{D}_{\alpha,\alpha'}(\mathbf{q}) - \omega^2 M_\alpha \mathbf{1}_{\alpha,\alpha}\right\} = 0\,, \tag{6.49}$$

where $\mathbf{1}_{\alpha,\alpha} = \delta_{\alpha,\alpha}$ are the components of the unit tensor [**1**]. For each given wave vector $\mathbf{q}$ there are $3n$ eigenfrequencies $\omega_{\mathbf{q},j}$. The index j serves to distinguish between the different branches. As already mentioned in the discussion of sound waves, the relationship between the eigenfrequencies $\omega_{\mathbf{q},j}$ and the wave vectors $\mathbf{q}$ is known as the **dispersion relation**. If the solutions for $\omega_{\mathbf{q},j}$ are known, the amplitudes $\mathbf{U}_\alpha(\mathbf{q}, j)$ can be calculated.

The two crucial steps in deriving this important result were the exploitation of the translation symmetry of the lattice, and the subsequent decoupling of the system of equations (6.45). A look at (6.49) shows that these equations are identical to those of harmonic oscillators. Each *normal mode*, identified by $\omega_{\mathbf{q},j}$, can therefore be formally assigned to a harmonic oscillator. As we will see in the course of this chapter, these oscillators have quantum mechanical properties, i.e. their energy is quantized and thus also their oscillation amplitude.

6.3 Experimental Determination of Dispersion Curves

From the coherent elastic scattering from the rigid lattice, the *geometrical* arrangement of the atoms can be derived via the *static structure factor*. However, for information on the *dynamics* of the lattice, we have to consider the temporal changes that lead to *inelastic scattering processes*.

6.3.1 Dynamic Scattering, Phonons

Going back to equations (4.3) and (4.4), which we derived for the field strength $A_s(t)$ of the scattered radiation, we can write:

$$A_s(t) = \frac{\widetilde{A}}{R_0} e^{-i(\omega_0 t - kR_0)} \mathcal{A}(\mathbf{K}) \propto e^{-i\omega_0 t} \mathcal{A}(\mathbf{K})\,. \tag{6.50}$$

In the following discussion we omit the unimportant amplitude $\widetilde{A}$ and detector distance R_0 prefactors. For the sake of simplicity, we consider a crystal with a monatomic

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

basis and assume that the scattering centers are point-like. We then obtain the scattering amplitude (4.4) by summing up the contributions of all atoms, for whose position vectors we write $\mathbf{r}_m$, i.e. $\mathcal{A}(\mathbf{K}) \propto \sum \exp(-i\mathbf{K} \cdot \mathbf{r}_m)$. For the case of neutron scattering, which is what we are mainly concerned with here, the assumption of point-like scattering centers is well justified.

Since, owing to lattice vibrations, the positions $\mathbf{r}_m(t)$ of the atoms are time-dependent, we split the position vectors into two parts $\mathbf{r}_m(t) = [\mathbf{R}_m + \mathbf{u}_m(t)]$, where the lattice vector $\mathbf{R}_m$ marks the mean position of the atom m and $\mathbf{u}_m(t)$ is the instantaneous displacement, and obtain for the scattered signal the proportionality:

$$A_S(t) \propto e^{-i\omega_0 t} \sum_m e^{-i\mathbf{K}\cdot\mathbf{r}_m(t)} \propto e^{-i\omega_0 t} \sum_m e^{-i\mathbf{K}\cdot\mathbf{R}_m} e^{-i\mathbf{K}\cdot\mathbf{u}_m(t)} . \tag{6.51}$$

Since the atomic vibration amplitudes are small compared with the atomic spacing, then in general $(\mathbf{K} \cdot \mathbf{u}_m) \ll 1$ applies, allowing us to truncate the expansion of exponential function $e^{-i\mathbf{K}\cdot\mathbf{u}_m(t)}$ after a few terms:

$$e^{-i\mathbf{K}\cdot\mathbf{u}_m(t)} \approx 1 - i\mathbf{K} \cdot \mathbf{u}_m(t) - \frac{1}{2}[\mathbf{K} \cdot \mathbf{u}_m(t)]^2 - \ldots . \tag{6.52}$$

We first consider only the constant and the linear term of the expansion. The quadratic term will be discussed later. As a solution for the displacement we choose a superposition of plane waves of the form:

$$\mathbf{u}_m(t) = \sum_{\mathbf{q}} \mathbf{U}_{\mathbf{q}} e^{\pm i\left[\mathbf{q}\cdot\mathbf{R}_m - \omega_{\mathbf{q}} t\right]} , \tag{6.53}$$

where the amplitude $\mathbf{U}_{\mathbf{q}}$ depends on the wave vector, the lattice vibration branch concerned and the temperature. Since we have allowed both signs in the exponent, only wave vectors $\mathbf{q} > 0$ will appear.[15] The summation, which we will not carry out here, extends over all allowed wave vectors. The question of which $\mathbf{q}$-values are allowed and how they are affected by the temperature will be discussed in the following section. We insert equations (6.52) and (6.53) into equation (6.51) and obtain:

$$A_S(t) \propto \underbrace{\sum_m e^{-i\mathbf{K}\cdot\mathbf{R}_m} e^{-i\omega_0 t}}_{\text{elastic}} - \underbrace{\sum_m \sum_{\mathbf{q}} i\mathbf{K} \cdot \mathbf{U}_{\mathbf{q}} e^{-i(\mathbf{K}\mp\mathbf{q})\cdot\mathbf{R}_m} e^{-i(\omega_0 \pm \omega_{\mathbf{q}})t}}_{\text{inelastic}} . \tag{6.54}$$

The first term was described in Section 4.4 in the context of elastic scattering. There we found that the summation over the lattice vectors $\mathbf{R}_m$ only led to non-zero contributions if the diffraction condition $\mathbf{K} = \mathbf{G}$ was fulfilled. The same arguments are also valid for the inelastic scattering as described by the second term. The modified scattering condition is now $(\mathbf{K} \mp \mathbf{q}) = \mathbf{G}$. If we take the temporal average of the second term, it is

15 This simple approach for the displacement is only allowed for a monatomic basis, because there are no phase differences between the different atoms of the basis.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

only non-zero if the scattered wave additionally fulfills the condition $\omega = (\omega_0 \pm \omega_q)$. In other words, the scattered radiation is modulated by the oscillations of the atoms, and thus the emitted wave will have sidebands.

If we multiply the two diffraction conditions by Planck's quantum of action we obtain the expressions for the **conservation of energy and momentum** of interacting particles:

$$\hbar\omega = \hbar\omega_0 \pm \hbar\omega_{\mathbf{q}} , \tag{6.55}$$

$$\hbar\mathbf{k} = \hbar\mathbf{k}_0 \pm \hbar\mathbf{q} + \hbar\mathbf{G} . \tag{6.56}$$

This result is of particular importance. We can interpret the energy conservation theorem in a similar way as for atomic scattering processes, e.g. Raman scattering from molecules: An incident X-ray quantum, neutron or electron interacts with the lattice and creates or destroys a vibrational quantum of the solid, a "sound particle". In analogy with the vibrational quanta of the electromagnetic field, the photons, these vibrational quanta of the elastic field are called **phonons**.[16] Correspondingly, a phonon has energy $\hbar\omega_{\mathbf{q}}$ and momentum $\hbar\mathbf{q}$. We will discuss the properties of these particles in the following sections. The conservation laws (6.55) and (6.56) are an expression of the fact that during the scattering process either a phonon is generated (– sign) or a phonon is annihilated (+ sign). However, the law of conservation of momentum has an unusual form: as we can always add to the momentum of the particles involved a momentum $\hbar\mathbf{G}$, where $\mathbf{G}$ stands for any reciprocal lattice vector. This can be imagined as follows: a phonon is generated or absorbed during the scattering process and a Bragg reflection occurs at the same time. The momentum $\hbar\mathbf{G}$ is transferred to the crystal as a whole. Restrictively, the momentum-like quantity $\hbar\mathbf{q}$ is therefore called the **quasi momentum** or **crystal momentum**.

Figure 6.14 shows schematically the scattering process just discussed for the case of phonon absorption. A particle, e.g. a neutron, with momentum $\hbar\mathbf{k}_0$ scatters from a

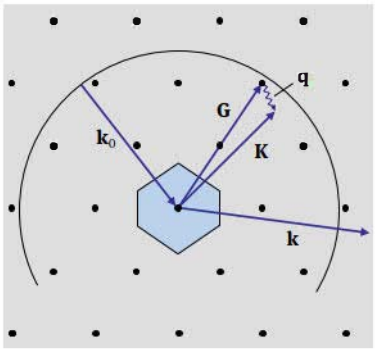


Fig. 6.14: Inelastic scattering. The incident particle with the momentum $\hbar\mathbf{k}_0$ absorbs a phonon with momentum $\hbar\mathbf{q}$. The reciprocal lattice vector involved is denoted by $\mathbf{G}$, the scattering vector by $\mathbf{K}$ and the wave vector of the scattered particle by $\mathbf{k}$. To clarify the additions involved in the diagram, $\mathbf{k}_0 + \mathbf{q} + \mathbf{G} = \mathbf{k}$. The Ewald sphere and the first Brillouin zone are also drawn in.

16 The term *phonon* was first used by *J.I. Frenkel* in his book *Wave Mechanics, Elementary Theory*, Clarendon Press, Oxford, 1932.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

phonon and absorbs it. This increases the energy of the scattered particle, the value of the wave vector increases and the wave vector ends up outside the Ewald sphere. The momentum transfer $\hbar\mathbf{K}$ is composed of $\hbar\mathbf{G}$, which is absorbed by the whole lattice, and the quasi momentum $\hbar\mathbf{q}$ of the destroyed phonon. If a phonon had been generated, rather than absorbed, the wave vector of the scattered neutron would have been inside the Ewald sphere.

An interesting result of this discussion of the scattering process is that lattice waves also have a particle character and that their energies and thus also their amplitudes are quantized. However, this is not unexpected. As mentioned above, the motion of the atoms can be broken down into normal vibrations whose equations of motion corresponds to those of harmonic oscillators. Of course, the energy of these oscillators is quantized, with eigenvalues $E_{\mathbf{q}} = (n_{\mathbf{q}} + \frac{1}{2})\,\hbar\omega_{\mathbf{q}}$. Each phonon represents a quantum of energy $\hbar\omega_{\mathbf{q}}$. The quantum number $n_{\mathbf{q}}$ indicates how many phonons with wave vector $\mathbf{q}$ are present in the solid. To understand the properties of phonons, it is important to understand that the harmonic oscillators mentioned here *cannot* be attributed to the localized vibrations of individual atoms. *Every* atom of the solid contributes to every vibrational state $\omega_{\mathbf{q}}$, i.e. to every phonon. A phonon is thus an excitation of the *whole* lattice.

In contrast to photons, phonons do not carry a real momentum. It can be shown that the true momentum of the phonons is exactly zero and therefore no mass transport is associated with lattice vibrations. Phonons behave *as if* they were carrying the momentum $\mathbf{p} = \hbar\mathbf{q}$. In contrast to photons, phonons are not fundamental particles, because they are only a consequence of the introduction of normal coordinates. For this reason, they are sometimes called *quasi particles*. In fact, the momentum of the scattered particles changes when a phonon is created or absorbed. As with elastic scattering, the corresponding momentum transfer involves the solid as a whole.

Now let us look at some numerical values. First, we estimate the mean square of the amplitude of phonons. For this purpose we consider a standing wave with an amplitude $u = U\cos qx\,\cos\omega t$. Taking a particular atom, the kinetic energy is given by $\frac{1}{2}M(\partial u/\partial t)^2$. Averaging over all N atoms of the crystal, using $\langle\cos^2 qx\rangle = \frac{1}{2}$, for the contribution of all atoms we find $\frac{1}{4}NM\omega^2U^2\sin^2\omega t$. If we now take the temporal average, we find for the kinetic energy $\frac{1}{8}NM\omega^2U^2$, since $\langle\sin^2\omega t\rangle = \frac{1}{2}$. Adding a similar quantity for the potential energy, results in the vibrational energy

$$\frac{1}{4}NM\omega^2U^2 = \left(n_{\mathbf{q}} + \frac{1}{2}\right)\hbar\omega_{\mathbf{q}}\,. \tag{6.57}$$

Taking $NM = \varrho V$, for a sample with volume V, the square of the amplitude results in the expression

$$U^2 = \left(n_{\mathbf{q}} + \frac{1}{2}\right)\frac{4\hbar}{\omega\varrho V}\,. \tag{6.58}$$

As an example we consider *one* phonon at the edge of the Brillouin zone (see Figure 6.18) with a wavelength of $\lambda = 3.6$ Å and frequency $\nu = 5$ THz in a copper cube with an edge

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

length of 1 cm and density $\varrho = 8.9\,\mathrm{g/cm^3}$. Using these values, we find a numerical value for the mean amplitude of approximately 5×10^{-21} cm. This is an extremely small value! However, the *number* of excited phonons is also of interest. At room temperature, the copper cube in question stores about 500 J (see Figure 6.32). With an energy of $\hbar\omega = 3.3 \times 10^{-21}$ J for phonons at the boundary of the Brillouin zone, we find that at room temperature the number of excited phonons is 1.5×10^{23} per $\mathrm{cm^3}$, in other words, comparable to the number of atoms. If the displacements of the atoms were all in phase, this would repesent a total amplitude of $NU \approx 7.5$ m! Since the thermal phonons are independent of one another, i.e. the displacements caused by the individual phonons have no fixed phase relationship with each other, the result is a much more reasonable mean displacement amplitude of $\sqrt{N\overline{U^2}} \approx 0.2\,\text{Å}$.

6.3.2 Coherent Inelastic Neutron Scattering

Phonon dispersion curves can be determined most easily by means of coherent inelastic scattering experiments. In Chapter 4 we have already got to know various probes for elastic scattering which can be used for structure determination. Are all probes discussed there also suitable for inelastic scattering experiments? In order to obtain a significantly measurable effect, the relative energy change of the scattered particle should be as large as possible. Since phonons have an energy of about 10^{-2} eV, this requirement is well met by thermal neutrons giving $\delta E/E \approx 10^{-1}$. Phonon dispersion curves are therefore in most cases determined with neutron scattering experiments, since the relative energy change resulting from X-ray scattering, on the other hand, is only of the order of magnitude $\delta E/E = \omega_q/\omega_0 \approx 10^{-6}$. Measurements of phonon dispersion using inelastic X-ray scattering therefore require extremely high energy resolution. That said, enormous progress has indeed been made in this area in recent years. Not only have new synchrotron sources been built that are optimized for X-ray scattering experiments, but the development of spectrometers in the X-ray range has also advanced. Both experimental techniques are now available and complement each other. Nevertheless, the detection of acoustic phonons near the Γ-point of the Brillouin zone with a very small wave number is still not possible because the energy transfer is very small. Further techniques suitable for this purpose are both ultrasonic measurements as discussed in Section 6.1 and Brillouin scattering experiments, which we will discuss in this section.

A typical setup for a scattering experiment has already been shown schematically in Figure 4.22. Figure 6.15 shows the experimental setup of a so-called *three-axis spectrometer* for neutron scattering experiments. After the neutrons from a nuclear reactor have been decelerated to thermal energies by the moderator, they impinge on a single crystal ("monochromator"), which produces a "monoenergetic" neutron beam by Bragg reflection. By rotating this crystal, the appropriate Bragg reflection, and thus

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

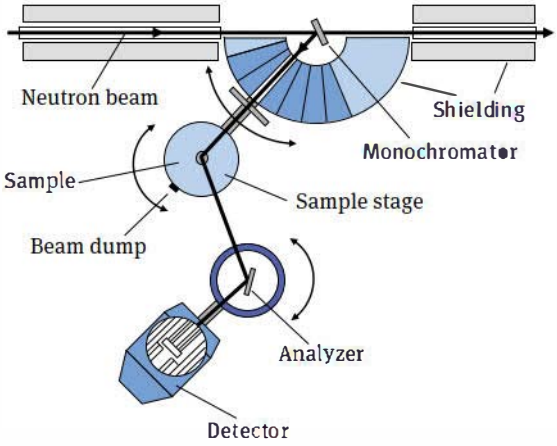


Fig. 6.15: Schematic representation of a three-axis neutron spectrometer. (The drawing is based on information from the Institute Laue-Langevin in Grenoble for the IN3 instrument.)

the desired energy, can be adjusted. After scattering in the sample, the neutrons are incident on a single crystal ("analyzer") and undergo a further Bragg reflection. As a result of the change in energy during the inelastic scattering process, a small change in the Bragg angle occurs at the analyzer crystal, which is measured. To vary the scattering angle itself, the analyzer crystal and the ^{3}He detector can be rotated around the sample position.

Of course, there are other measuring techniques where the experimental arrangements differ from that outlined here. For example, there are *time-of-flight spectrometers*, where the neutron beam is split by means of *choppers* into neutron bunches with a defined velocity. After the neutrons have gained or lost energy in the sample, they propagate with different velocities and arrive at the detectors at different times. The change in energy can then be determined directly from the differences in times of flight.

Along a particular direction of observation, neutrons scattered with the participation of phonons will be detected from different reciprocal lattice vectors. The observed neutrons therefore have discrete energy values which can be resolved with the help of the analyzer crystal or by their time-of-flight and assigned to the corresponding phonons. In Figure 6.16 this aspect of neutron scattering is shown again in momentum space, illustrating an inelastic scattering process in which the reciprocal lattice vector (120) is involved. In the process shown, a phonon is generated with wave vector $\mathbf{q}$. This reduces the energy of the scattered neutron and its wave vector $\mathbf{k}$ ends inside the Ewald sphere. Clearly the process fulfils the conservation of momentum (6.56).

The discussion so far might give the impression that the scattered neutrons have a broad energy spectrum, since interactions with phonons of different energies takes place. In fact, however, this is not the case: With a fixed reciprocal lattice vector according to (6.56), the direction of observation serves to fix the wave vector $\mathbf{q}$ of the phonons involved, which still have to be assigned to the different branches. Since energy must also be conserved as in (6.55), the phonons of each branch give rise to a

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

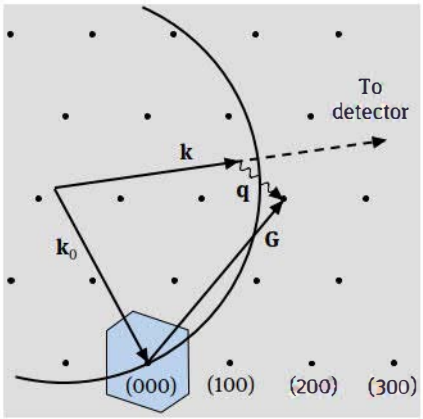


Fig. 6.16: Inelastic neutron scattering. Incident neutrons with wave vector $\mathbf{k}_0$ are scattered inelastically, generating a phonon. The quasi-momentum conservation $\mathbf{k} = \mathbf{k}_0 - \mathbf{q} + \mathbf{G}$ applies. The Brillouin zone is depicted in blue.

well-defined energy change of the scattered neutrons, which can be positive or negative depending on the scattering process. Furthermore, different reciprocal lattice vectors may be involved in the scattering.

6.3.3 Debye-Waller Factor

A detailed investigation of the scattering properties of crystals shows surprisingly that a very weak scattering intensity is observed even when the diffraction condition (4.21) is not obeyed. This "background", i.e. the scattering intensity between the defraction peaks, increases with increasing temperature.

When deriving the scattering amplitude for inelastic scattering, we truncated the expansion of the exponential factor in (6.52) after the linear term thereby restricting ourselves to *one-phonon scattering*, i.e. scattering processes in which only *one* phonon is involved. As can be shown, the number of phonons involved in the scattering process is directly related to the power of each term in the series expansion. Since the series converges rapidly, only the first terms are important, so that the probability for the occurrence of *multi-phonon scattering processes* decreases very rapidly with the number of phonons involved. While direct observation of such processes is hardly possible, they are responsible for the intensity decrease of the Bragg reflections with increasing temperature.

This is because during an inelastic scattering process, quasi-momentum and energy are transferred from the scattered particles, the incident neutrons, photons or electrons, to phonons. While in single-phonon processes the transfer has a well-defined value, in multi-phonon processes the energy and quasi-momentum can be distributed arbitrarily between the participating phonons. Thus the unique assignment of the scattering vector and the momentum of a phonon, as in the case of one-phonon scattering, no

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

longer holds. As a result multi-phonon scattering processes reduce the intensity of the observed maxima and contribute to the *background*.

We now take a closer look at the temperature dependence of the scattering. Taking $\mathbf{u}(t)$ as the instantaneous displacement of a lattice atom from its equilibrium position arising from its thermal motion, we obtain for the time average of the structure factor:

$$\langle \mathcal{S}_{hkl} \rangle = \sum_i f_i \left\langle e^{-i\mathbf{G}\cdot(\mathbf{r}_i+\mathbf{u}(t))} \right\rangle = \left(\sum_i f_i e^{-i\mathbf{G}\cdot\mathbf{r}_i} \right) \left\langle e^{-i\mathbf{G}\cdot\mathbf{u}(t)} \right\rangle . \tag{6.59}$$

Since $\mathbf{G}\cdot\mathbf{u} \ll 1$, we can cut short the expansion of $\langle \exp(-i\mathbf{G}\cdot\mathbf{u}) \rangle$ after the third term to obtain:

$$\left\langle e^{-i\mathbf{G}\cdot\mathbf{u}(t)} \right\rangle = 1 - i\langle \mathbf{G}\cdot\mathbf{u} \rangle - \frac{1}{2}\langle (\mathbf{G}\cdot\mathbf{u})^2 \rangle - \cdots \approx 1 - \frac{1}{2}G^2\langle u^2 \rangle \langle \cos^2\theta \rangle , \tag{6.60}$$

where θ is the angle between $\mathbf{G}$ and the displacement $\mathbf{u}$. Since for independently vibrating atoms $\langle \mathbf{G}\cdot\mathbf{u} \rangle = 0$ this term can also be omitted. Thus, with the spatial averaging of $\cos^2\theta$ yielding the factor $1/3$, we obtain:

$$\left\langle e^{-i\mathbf{G}\cdot\mathbf{u}(t)} \right\rangle \approx 1 - \frac{1}{6}G^2\langle u^2(t) \rangle . \tag{6.61}$$

Since these are the first two terms of the series expansion of the function $\exp(-G^2\langle u^2 \rangle/6)$, we can write to good approximation:

$$\langle \mathcal{S}_{hkl} \rangle = \left(\sum_i f_i e^{-i\mathbf{G}\cdot\mathbf{r}_i} \right) e^{-\frac{1}{6}G^2\langle u^2(t) \rangle} . \tag{6.62}$$

The temperature dependence of the scattering intensity can thus be described by:

$$I_{hkl}(T) = I_0 D_{hkl} = I_0\, e^{-\frac{1}{3}G^2\langle u^2(t) \rangle} , \tag{6.63}$$

where I_0 is the scattering intensity we would obtain for the rigid lattice. The factor D_{hkl} is the so-called **Deybe-Waller factor**[17].

We will now consider more in detail the mean square displacement $\langle u^2(t) \rangle$ which appears in equation (6.63). In the following we will assume that the atoms move in a harmonic potential. This means that we are temporarily disregarding the coupling between the atoms. It follows from the equipartition theorem that the mean value $\langle U \rangle$ of the potential energy of the atom vibrating about its equilibrium position is given by:

$$\langle U \rangle = \frac{1}{2}M\omega^2\langle u^2 \rangle = \frac{3}{2}k_B T , \tag{6.64}$$

where ω is the angular frequency of vibration and M is the mass of the oscillating atom.

Inserting the value of $\langle u^2 \rangle$ taken from (6.64) it into (6.63), we obtain an expression for the temperature dependence of the scattering intensity:

$$I_{hkl} = I_0 \exp\left[\left(-k_B T/M\omega^2 \right) G^2 \right] . \tag{6.65}$$

17 Ivar Waller, *1898 Flen, †1991 Uppsala

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6.17 shows the result of a measurement on aluminium. Since this metal has a face-centered cubic structure, there will be no (h00)-reflections with odd h (cf. exercise in Chapter 4). As just discussed above, the scattering intensity decreases with increasing indexing of the planes and with increasing temperature and these differences between the defraction peaks become more pronounced with increasing temperature. As discussed above, the loss of intensity arises from the inelastic scattering, which manifests itself as a diffuse background increasing with increasing temperature.

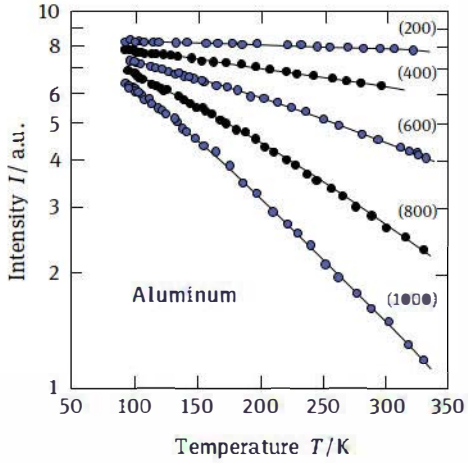


Fig. 6.17: The temperature dependence of the scattering intensity of the (h00)-reflections of aluminium. As expected reflexes for odd values of h are missing. (After R.M. Nicklow, R.A. Young, Phys. Rev. **152**, 591 (1966).)

Equation (6.65) is only valid for high temperatures. At low temperatures the zero-point motion of the atoms must be taken into account, because the mean square displacement $\langle u^2 \rangle$ does not disappear even at absolute zero. Since the atoms oscillate in a three-dimensional harmonic potential with zero-point energy $3\hbar\omega/2$, we find:

$$I_{hkl} = I_0 \exp\left[(-\hbar/2M\omega)\, G^2\right] . \tag{6.66}$$

Putting in typical numerical values ($G = 5 \times 10^{10}$ m^{-1}, $\omega = 5 \times 10^{13}$ s^{-1}, $M = 10^{-25}$ kg), we find that at $T = 0$ about 2.5 % of the radiation is scattered inelastically, with a value of about 4 % at room temperature.

6.3.4 Experimentally Determined Dispersion Curves

To check the previous theoretical considerations we will make a comparison using the experimentally-determined dispersion curves for copper, lithium fluoride and silicon. As the dispersion curves show, crystals with a monatomic basis have only acoustic branches. As an example, Figure 6.18 shows the dispersion curves for copper measured by neutron scattering. The phonon dispersion curves are plotted in the [$\bar{1}$00], [110]

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

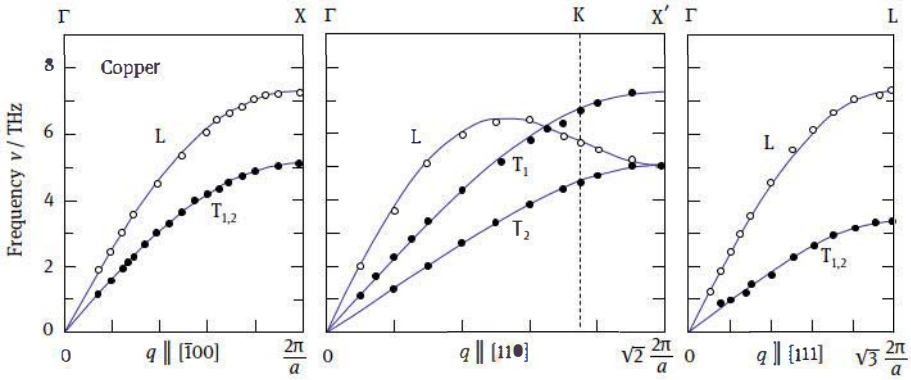


Fig. 6.18: Phonon dispersion curves for copper. The solid lines show the calculated dispersion curves, with the measured values represented by circles. The longitudinal and the two transverse branches are labelled L, T_1 and T_2 respectively. The dotted line indicates the end of the Brillouin zone in the [110] direction. (After E.C. Svensson et al., Phys. Rev. **155**, 619 (1967).)

and [111]-directions of the face-centered cubic lattice. It is clear that for these selected directions the approximation made assuming a linear chain fits well with the measured results. In the [$\bar{1}$00] and [111]-direction the two perpendicularly-polarized transverse branches are degenerate. However, if the wave propagation is not in a direction with high symmetry, three distinct branches appear as in the [110]-direction. The solid lines in the figure are theoretical dispersion curves, whose agreement with the experimental data is impressive. In the calculation, interactions up to the fourth nearest neighbors was taken into account.

In the figures we notice that the dispersion curves shown are ending at different values of the wave vector depending on the direction of propagation of the lattice waves. We can see the reason for this kind of representation by referring to Figure 6.19 which shows two Brillouin zones of a face-centered cubic lattice, shifted relative to each other.

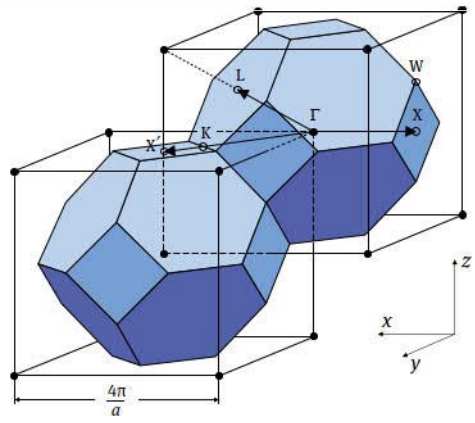


Fig. 6.19: The Brillouin zones of a face-centered cubic crystal. Two adjacent Brillouin zones and two non-primitive unit cells are shown. The arrows, Γ-X, Γ-X′ and Γ-L, indicate the directions along which the dispersion curves shown in Figure 6.18 were measured.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Also shown are two non-primitive unit cells of the reciprocal lattice of the face-centered cubic lattice. The center of the first unit cell coincides with the Γ-point, the second is shifted by the vector $\frac{4\pi}{a}(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2})$. The edge length of the cube of the reciprocal unit cell is $4\pi/a$. If we now move, as shown in the figure, in the [100]-direction starting from the Γ-point, then after a distance of $2\pi/a$ we arrive at the X-point reaching the boundary of the Brillouin zone. In [111]-direction we reach the L-point and thus the edge of the Brillouin zone at $\sqrt{3}\pi/a$. In [110]-direction one leaves the Brillouin zone at K-Point. However, the dispersion curves in Figure 6.18 for this direction are plotted up to the X-point of the Brillouin zone, where the wave vector has the value $2\sqrt{2}\,\pi/a$.

Now let us take a closer look at the propagation of lattice waves along the ΓX-direction. Since the X-point is located midway between two reciprocal lattice points, the shape of the dispersion curves when crossing the boundary of the Brillouin zones is symmetric as we know from the one-dimensional case. However, if we take the path from Γ to the X′-point, we see that the K-point is not midway between two reciprocal lattice points. The shape of the dispersion curve is therefore not symmetrical when crossing the boundary, even though the K-point is located at the boundary of the Brillouin zone. Also since this direction has a two-fold axis of rotation, the dispersion curves for the two transverse waves, T_1 and T_2, are not the same.

Figure 6.20 shows the theoretical and experimental dispersion curves for face-centered cubic lithium fluoride, with a basis consisting of the two ions Li^+ and F^-. The good agreement between theory and experiment is again remarkable, especially since only interactions up to the next-nearest neighbors was considered. Besides the acoustic branches, there are also three optical phonon branches owing to the diatomic basis. We should note the strong splitting of the optical branches at $\mathbf{q} \approx 0$, which is typical for ionic crystals. The longitudinal branch lies far above the transverse branches.

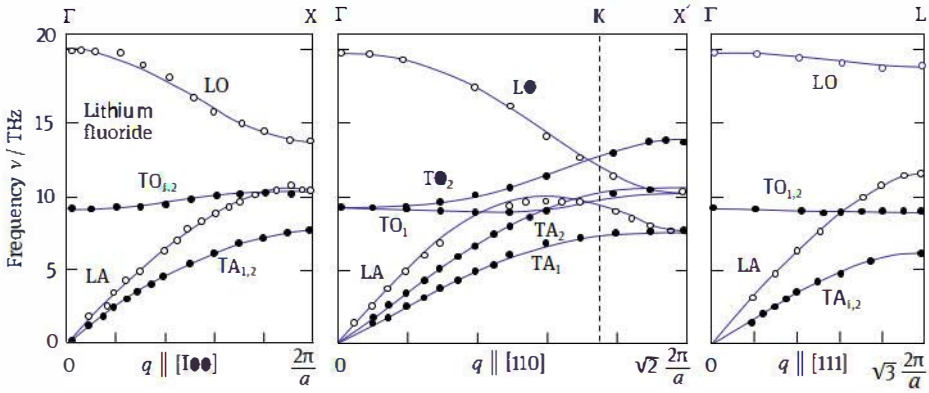


Fig. 6.20: Phonon dispersion curves for lithium fluoride. Both the theoretical curves and experimental points are shown. The branches are labeled LA, TA, LO and TO, the first letter indicating the polarization and the second representing either *acoustic* or *optical*. (After G. Dolling et al., Phys. Rev. **168**, 970 (1968).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

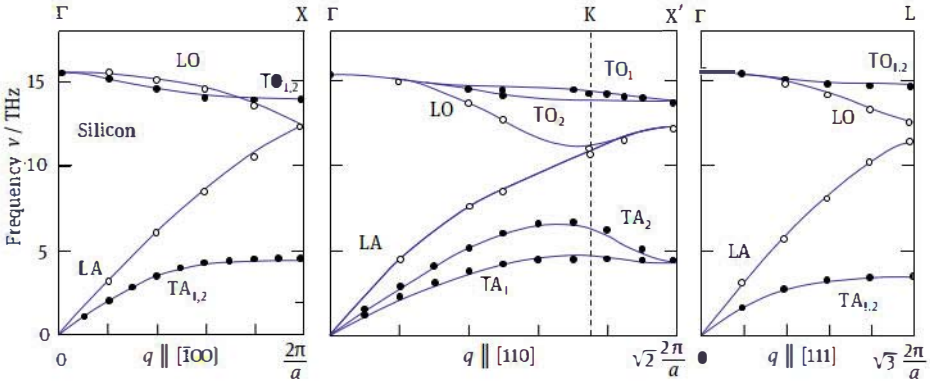


Fig. 6.21: Phonon dispersion curves for silicon. Again both the theoretical curves and experimental points are shown. The phonon branches are labelled as in Figure 6.20. (After P. Giannozzi et al., Phys. Rev. B 43, 7231 (1991).)

This arises from the Coulomb interaction between the oppositely-charged sublattices, which leads to a macroscopic polarization during longitudinal vibrations, a fact that we will discuss in more detail in Section 13.3. However, if no dipole moment is generated during lattice vibration, then in cubic crystals all optical vibrations are degenerate at the Γ-point. The high symmetry ensures that the force constants are independent of the direction of displacement. An example of this is the vibrational spectrum of silicon, shown in Figure 6.21.

6.3.5 Light Scattering

Hitherto, in discussing the scattering experiments we have always implicitly assumed that the wavelength of the incident radiation is smaller than the lattice spacing. Would we expect scattering in the converse case? As we will see, this is quite possible. For *visible light*, which we are dealing with here, the scattering condition (6.56) can be simplified. Because the energy of the photons is much greater than that of the phonons, the energy, and thus also the wave number of the photons, is only slightly changed when phonons are generated or absorbed. Since the wave number k_0 of light corresponds to approximately one thousandth of the extension of the Brillouin zone, light scattering takes place in the vicinity of the Γ-point. So we can simplify and write equation (6.56) for the light scattering

$$\mathbf{k} = \mathbf{k}_0 \pm \mathbf{q} \,. \tag{6.67}$$

In the classical description of the scattering process, the incident wave causes an oscillating dipole moment in each atom, which then itself emits a wave whose intensity increases proportionally to ω^4.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Rayleigh Scattering. If the interaction of light with the solid takes place with no phonon participation, we may additionally set $\hbar\mathbf{q} = 0$ in (6.67). This *elastic* scattering process, in which no frequency shift occurs, is called **Rayleigh scattering**. In this case, following directly from (6.67) we find the simple but astonishing relationship $\mathbf{k} = \mathbf{k}_0$, i.e., scattering occurs only in forward direction, in the other directions no scattered light should be observed. Or expressed differently: in ideal crystals, the scattering contributions of the individual atoms are exactly canceled out by the interference except in the forward direction. This is an unusual way to describe the propagation of light in crystals. The quasi-momentum conservation condition above is based on the translation symmetry of the lattice. If this is disturbed, e.g. by defects, then these can act as scattering centers, since the irregular arrangement of the scattered waves no longer averages out. Therefore, in real crystals some level of Rayleigh scattering always occurs. The same applies to amorphous solids, as the conservation of the wave vectors is not preserved.

Raman and Brillouin Scattering. If $\mathbf{q} \neq 0$ in equation (6.67), phonons are either generated or destroyed during the scattering process, as shown in Figure 6.22. If the wavelength of the scattered light shifts to longer wavelengths, we speak of a **Stokes process**[18], if it shifts to a shorter wavelength, we speak of an **anti-Stokes process**. Light scattering involving acoustic phonons is known as Brillouin scattering, while scattering involving optical phonons is known as Raman scattering.

Fig. 6.22: Wave vector diagrams for light scattering processes. **a)** Phonon generation, **b)** phonon absorption.

As mentioned above, the reciprocal lattice vector does not play a role in light scattering. We use equation (6.67) and write: $q^2 = k_0^2 + k^2 - 2kk_0 \cos\vartheta$. Because the energy of the phonons can be neglected in comparison to the photon energy, the scattering process hardly changes the energy and thus the value of the photon momentum, so that we can write $k_0^2 \approx k^2$ and $q^2 \approx 2k_0^2(1 - \cos\vartheta) = 4k_0^2 \sin^2(\vartheta/2)$. For the wave number q of the participating phonons we thus have:

$$q = 2k_0 \sin\frac{\vartheta}{2} . \tag{6.68}$$

18 George Gabriel Stokes, *1819 Skreen, County Sligo, †1903 Cambridge

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As already emphasized, owing to the small wave number of the photons involved, only phonons in the immediate vicinity of the Γ-point participate in light scattering, whereas in inelastic scattering of neutrons all wave vectors are accessible. The small frequency changes $\Delta\omega = (\omega - \omega_0) = \pm\omega_q$ are detected with high-resolution spectrometers.

Let us first take a closer look at light scattering with optical phonons, i.e. **Raman scattering**[19]. In such measurements the sample is irradiated with a laser and the scattered light is usually analyzed with a double monochromator. Since the dispersion curve for the optical phonons at the Γ-point shows an extremum, the frequency of the interacting phonons depends only weakly on the wave vector and the observed line shift is therefore practically independent of the direction of observation. Typical frequency changes $\Delta\nu$ are in the range of a few hundred to about a thousand cm^{-1}, i.e. of the order of 10 THz.[20]

The following two figures show measurements of Raman scattering in germanium and silicon. Both crystals have a diatomic basis and thus three optical phonon branches. As can be seen in Figure 6.21 using silicon as an example, the dispersion curves of the longitudinal and transverse optical phonons of non-polar cubic crystals coincide at the Γ-point, so that only one Stokes or anti-Stokes line is observed in Raman scattering. In Figure 6.23 the Stokes lines of four different germanium crystals are depicted. One of the crystals consisted of natural germanium, while the other three were isotopically pure. As equation (6.39) suggests, the frequency of optical phonons at the Γ-point and thus also the Raman shift is proportional to $M^{-1/2}$, where M is the isotopic mass.

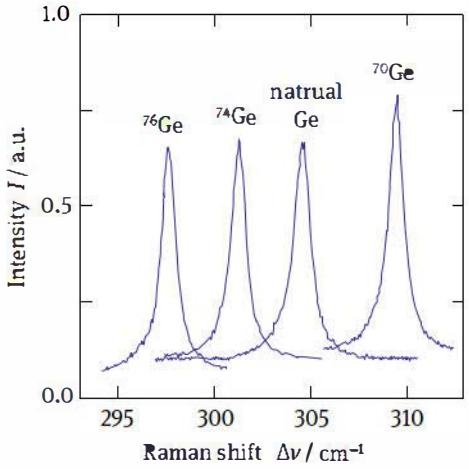


Fig. 6.23: Raman spectra of isotopically pure and natural germanium at 80 K. The shift of the Raman line is to a good approximation proportional to $M^{-1/2}$ where M is the isotopic mass. The unshifted Rayleigh line is not visible, since the zero-point of frequency shift is not shown. (After T. Ruf et al., Phys. Bl. **52**, 1115 (1996).)

19 Chandrasekhara Venkata Raman, *1888 Tiruchirappalli, †1970 Bangalore, Nobel Prize 1930

20 In this field it is common practice to indicate the frequency shift in reciprocal wavelengths, i.e. in cm^{-1}. The conversion is $\Delta\nu/c = \lambda^{-1}$, where c is the velocity of light.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6.24 shows Raman lines for silicon recorded at different temperatures. Due to the slight anharmonicity of the lattice (see Section 7.1), both the position and the width of these lines depend on the temperature. We can see the clear asymmetry in the scattering intensity between the Stokes and anti-Stokes lines. Since 15 THz corresponds to a thermal energy of about 700 K even at the highest temperature measurements at 770 K, few phonons would be excited at the high frequency required for the anti-Stokes process. At 20 K this line is no longer observable at all, because at such low temperatures phonons of this energy are completely absent.

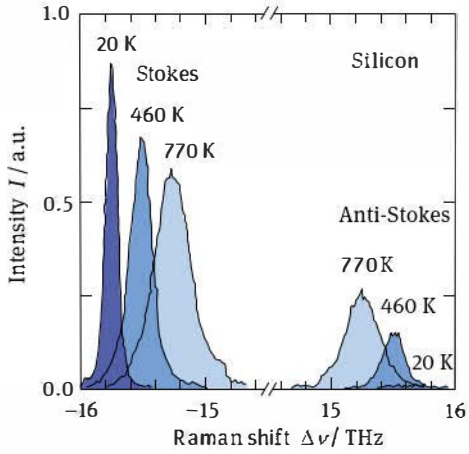


Fig. 6.24: Raman spectra of silicon at different temperatures. The zero-point of the frequency shift is suppressed in this plot. (After T.R. Hart et al., Phys. Rev. B **1**, 638 (1970).)

In Raman spectra, however, not all those lines that would be permitted by energy and momentum conservation are observed. A further prerequisite for their occurrence is a non-vanishing coupling between phonons and photons. As already mentioned, the incident light causes a polarization oscillating with the light frequency via the electrical susceptibility (see also Section 13.3) of the atoms. *Raman scattering* occurs when lattice vibrations cause a modulation of this polarization. Classically, this means that the radiation of an electromagnetic wave is associated with the sum or difference frequency of the two vibrations, whereby the scattered wave has a characteristic directional and polarization dependence. In the case of *infrared absorption*, coupling takes place via the oscillating dipole moment produced by the phonons. With the help of symmetry considerations, the selection rules for both effects can be derived. For crystals with inversion symmetry it is strictly valid that infrared absorption and Raman scattering are mutually exclusive. For example, ionic crystals with NaCl structure show strong infrared absorption but no Raman scattering, whereas the opposite is true for silicon.

If two phonons are involved in the scattering process, we need to modify the **k**-conservation condition of (6.67). The new condition for the wave vectors is then $(\mathbf{k} - \mathbf{k}_0) = (\mathbf{q}_1 + \mathbf{q}_2)$. If the wave vectors of the two phonons involved are almost oppositely directed, i.e. $\mathbf{q}_1 \approx -\mathbf{q}_2$, then the quasi-momentum conservation in

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

light scattering can be fulfilled, even if phonons with large wave numbers are involved in the scattering process. In this case, the restriction of the scattering process to the close vicinity of the Γ-point no longer holds. However, if two phonons are involved, the ensuing scattering intensities are comparatively low for reasons we mentioned in the discussion of the Debye-Waller factor. Nevertheless, *second order Raman scattering* is observable and allows, for example, the number of thermally excited phonons to be determined (cf. Section 6.4).

Since the quasi-momentum conservation argument is based on the translational invariance of the crystal lattice, it does not apply to amorphous solids. In these materials, therefore, all phonons can participate in the first order Raman process. As a result, there are no well-defined, narrow lines as in crystals, but strongly broadened maxima. In addition, there are no symmetry-based arguments to exclude the simultaneous infrared absorption and Raman scattering as in many crystals. Therefore, in amorphous solids both effects can be observed simultaneously.

We now take a closer look at **Brillouin scattering**. For reasons of quasi-momentum conservation, only long-wavelength acoustic phonons are involved. Since the dispersion relation $\omega = vq$ applies to these phonons, the frequency of the scattered phonons depends on the scattering angle ϑ. As follows from (6.68), the typical frequency changes $\Delta\nu$ are in the range of 20 GHz. To measure the relatively small frequency changes, the sample is generally irradiated with a frequency-stabilized laser and the scattered light is analyzed with a high-resolution *Fabry-Pérot interferometer*[21,22] which is operated in high order. The spectra therefore repeat periodically as a function of the plate spacing and partially overlap in most experiments. However, there is always only *one* set of lines belonging to *one* order. Figure 6.25 shows a measurement on indium antimonide in which, in addition to the Rayleigh line (R), which was suppressed by many orders of magnitude by experimental precautions, all three acoustic phonon branches LA, TA_1 and TA_2 could be observed. The lines associated with a common order are highlighted by light blue.

As brief consideration indicates, there is a close formal relationship between Brillouin scattering and Bragg reflection. According to the definition of the scattering angle θ in X-ray diffraction and ϑ in light scattering, $\vartheta/2 = \theta$. Thus equation (6.68), which is valid for light scattering, can be rewritten as follows:

$$q = 2k_0 \sin\theta \qquad \text{and} \qquad \lambda_0 = 2\lambda_q \sin\theta \,. \tag{6.69}$$

Here λ_0 is the light wavelength and λ_q the sound wavelength.[23] Comparison with the Bragg condition (4.24) shows that in the case of light scattering, the density variation

21 Maurice Paul Auguste Charles Fabry, *1867 Marseille, †1945 Paris

22 Jean-Baptiste Alfred Pérot, *1863 Metz, †1925 Paris

23 It should be noted that λ_0 refers to the wavelength of light *in the sample*. Therefore, the refractive index of the solid must also be taken into account.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

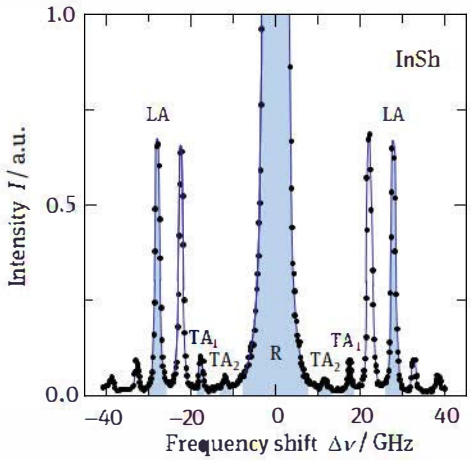


Fig. 6.25: Brillouin spectrum of indium antimonide. The measured values associated with one order are marked in light blue and labeled LA, TA_1 and TA_2. The dominant central line R is due to Rayleigh scattering. (After J.R. Sandercock, *Proc. 2nd Int. Conf. Light Scattering in Solids*, M. Balkanski, ed., Flammarion Paris, 1971.)

caused by the sound waves takes the role of the scattered density distribution. The wavefronts therefore correspond to the lattice planes in Bragg scattering. The phonons periodically modulate the refractive index, resulting in a diffraction of the incident light. In this picture, the frequency shift of the light arises from the Doppler effect, which the running sound waves produce. However, since the phonons give rise to only small variations in the refractive index, the intensity of the scattered light is very low.

6.4 The Specific Heat Capacity

The specific heat is an important thermodynamic quantity, which is largely determined by the thermally excited phonons in dielectric solids. Its correct description is a benchmark test for the theory of lattice vibrations. For experimental reasons, the specific heat of solids is measured at constant pressure, i.e. C_p is determined. However, the specific heat at constant volume, which is directly linked to the internal energy U via the equation $C_V = (\partial U/\partial T)_V$, is more suitable for a theoretical description. These two important quantities are linked by the thermodynamic relationship $(C_p - C_V) = \alpha_V^2 VTB$, where α_V is the thermal volume expansion coefficient and B the compression modulus. While the difference $(C_p - C_V)$ in gases is relatively large, C_V and C_p in solids differ only slightly due to the relatively low thermal expansion.

Both the temperature dependence as well as the absolute value of the specific heat have a number of characteristic features. The typical temperature dependence is shown in Figure 6.26, displaying measurements on diamond conducted in the 19th century. Starting at low temperatures, the specific heat of all crystals first increases

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

rapidly and then approaches a constant value, determined by the **Dulong-Petit law**[24,25] $C_V = 3N_A k_B = 3R_m \approx 24.94\,\mathrm{J\,mol^{-1}\,K^{-1}}$, where N_A is the Avogadro constant and R_m the universal gas constant.

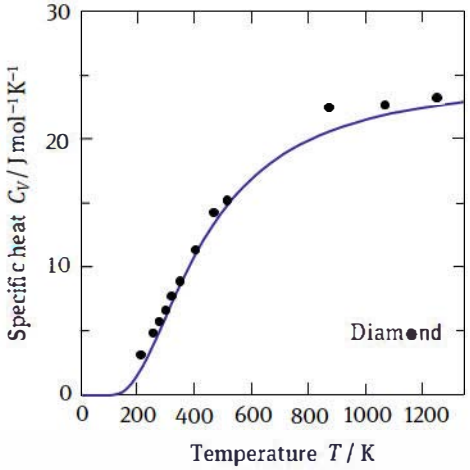


Fig. 6.26: The molar heat capacity of diamond as a function of temperature. The solid curve represents the temperature dependence as calculated by Einstein. (After A. Einstein, Ann. Phys. **327**, 180 (1907). The measured data are from F.H. Weber, Ann. Phys. **230**, 367 (1875).)

The first attempt at an explanation of the general dependence, especially the rapid fall in the specific heat with decreasing temperature, was made by *A. Einstein* in 1907. He assumed that the atoms in a solid could be approximated by uncoupled harmonic oscillators with identical eigenfrequencies. Here it is crucial that the energy of the oscillators cannot change continuously following the principle of equal distribution as in classical physics, but is quantized. In this picture, the specific heat must fall to zero at low temperatures, since the oscillators can no longer be excited when the thermal energy $k_B T$ falls below their oscillation energy. We will not discuss the *Einstein model* here in further detail, since it predicts a too small value at low temperatures. This model neglects to take into account that the atoms of the solid are coupled to one another. As we have seen in Section 6.2, the coupling causes the spectrum of lattice vibrations to extend to low frequencies, leading to a significantly improved description of the specific heat.

6.4.1 Phonon Density of States

The discussion so far may have given the impression that although the relationship between frequency and wave vector in the case of the lattice vibrations is predetermined

24 Pierre Louis Dulong, *1785 Rouen, †1838 Paris
25 Alexis Thérèse Petit, *1791 Vesoul, †1820 Paris

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

by the dispersion relations, the wave vectors themselves are continuously variable quantities. However, the finite size of a sample and the associated finite number of atoms limits the number of possible eigenmodes, the calculation of which requires knowledge of the boundary conditions. Here we first consider the case of **periodic boundary conditions** (or Born-von-Kármán[26] boundary conditions). The starting point in this case is a macroscopic crystal of finite size in the form of a parallelepiped. Now we periodically repeat this crystal in all directions. In each "copy" the same instantaneous atomic arrangement and thus the same vibrational state exists at any given time. This means that the properties of this solid repeat periodically in space. This results in an infinitely extended crystal with the full translational symmetry of an ideal crystal. This procedure means that there are no external boundary surfaces to influence the behavior.

We first consider a cubic crystal, taking a cube with edge length L. The atomic displacements $\mathbf{u}(\mathbf{r},t) = \mathbf{U}_{\mathbf{q}} e^{-i(\omega_q t - \mathbf{q}\cdot\mathbf{r})}$, generated by the phonons, must satisfy the periodic boundary condition

$$u(x,y,z,t) = u(x+L,y,z,t) = u(x,y+L,z,t) = u(x,y,z+L,t)\,. \tag{6.70}$$

If we insert the expression for the displacement of the atoms into these equations, we obtain, for example, for the x-direction:

$$e^{i\mathbf{q}\cdot\mathbf{r}} = e^{i\mathbf{q}\cdot(x+L,y,z)} = e^{i\mathbf{q}\cdot\mathbf{r}} \cdot e^{iq_x L}\,. \tag{6.71}$$

The periodicity of the atomic displacements is guaranteed if the x-component of the wave vectors satisfies the condition

$$q_x = \frac{2\pi}{L} m_x\,, \tag{6.72}$$

because then the relation $\exp(iq_x L) = 1$ applies to every integer m_x. (This is equivalent to saying that m_x wavelengths $2\pi/q_x$ fit into the cube side L_x, thus repeating the wave pattern in each length L_x.) We obtain the corresponding expressions for the other two directions in space.

As shown in the previous section, a crystal with N atoms has $3N$ solutions to the equation of motion, i.e. $3N$ normal or eigenmodes, which are distributed among the individual branches. If our cube consists of $\mathcal{M}$ unit cells per edge length, with each unit cell containing p atoms, and we have $\mathcal{N}$ unit cells in total, then $3N = 3p\,\mathcal{N} = 3p\mathcal{M}^3$. Each phonon branch thus contains $\mathcal{N}$ vibrational states. This leads to the following constraint for the quantum numbers m_i

$$-\frac{\mathcal{M}}{2} < m_i \leq \frac{\mathcal{M}}{2}\,. \tag{6.73}$$

26 Theodore von Kármán, *1881 Budapest, †1963 Aachen

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The exact number of vibrational states per branch $\mathcal{N}$ is only obtained if we include only one of the two wave vectors with the quantum numbers $\pm\mathcal{M}/2$. The reason for this is that, as is easily shown by insertion, both signs lead to identical atomic displacements.

An obvious, but perhaps not a very realistic procedure is to define the state of motion of the boundary atoms, e.g. to regard them as immobile or freely movable, and take into account *only* our original cube of edge length L. This leads to the so-called **fixed boundary conditions**. The most important difference between periodic and fixed boundary conditions is that in the first case propagating waves and in the second standing waves are allowed. Since for standing waves the periodicity is given by half the wavelength, the relationship $q_i = \pi m_i/L$ applies instead of (6.72). Furthermore, only positive signs now occur, so that instead of (6.73) the condition $0 < m_i \leq \mathcal{M}$ results. But since, as we will see in a moment, it is primarily the number of states within a frequency interval that matters, this distinction between the two types of boundary conditions is irrelevant for macroscopic samples and their properties. However, if the number of atoms in a sample is small, the exact boundary conditions for the lattice vibrations play an important role. In the case of nanostructures or thin films, for example, the boundary conditions have a large influence on the properties but their exact specification usually proves to be problematic.

Now we go back to the periodic boundary condition case and consider non-cubic lattices. Taking as the starting point a finite crystal with $\mathcal{M}$ unit cells along each of the directions of the three basis vectors. To satisfy the periodic boundary conditions, the wave vectors must satisfy the condition (analogous to (6.71)):

$$e^{i\mathcal{M}\mathbf{q}\cdot\mathbf{a}_1} = e^{i\mathcal{M}\mathbf{q}\cdot\mathbf{a}_2} = e^{i\mathcal{M}\mathbf{q}\cdot\mathbf{a}_3} = 1\,. \tag{6.74}$$

Making a comparison with the definition of the reciprocal lattice we find that for the allowed wave vectors the relation

$$\mathbf{q} = \sum_{i=1}^{3} \frac{m_i\,\mathbf{b}_i}{\mathcal{M}} \tag{6.75}$$

holds, where $\mathbf{b}_i$ are the basis vectors of the reciprocal lattice. The maximum value of m_i is again given by $\mathcal{M}/2$. Linked to this are the maximum wave vectors $\mathbf{q}_i^{\max} = \mathbf{b}_i/2$ in the direction of the coordinate axes of the reciprocal lattices, which also applies to wave vectors with negative signs. The basis vectors $\mathbf{b}_i$ define a unit cell of the reciprocal lattice, which we place so that its center coincides with the lattice vector (000). In this parallelepiped, the allowed states are *evenly distributed*, as shown in Figure 6.27.

Since all allowed wave vectors are located within *one* unit cell of the reciprocal lattice, their density $\varrho_{\mathbf{q}}$ in reciprocal space is given by the ratio of the number of allowed states $\mathcal{N}$ to the volume $(2\pi)^3/V_C$ of the unit cell of the reciprocal lattice (cf. equation (4.14)). We therefore obtain the important relationship for the density of allowed states:

$$\varrho_{\mathbf{q}} = \frac{\mathcal{N}}{(2\pi)^3/V_C} = \frac{V}{(2\pi)^3}\,. \tag{6.76}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

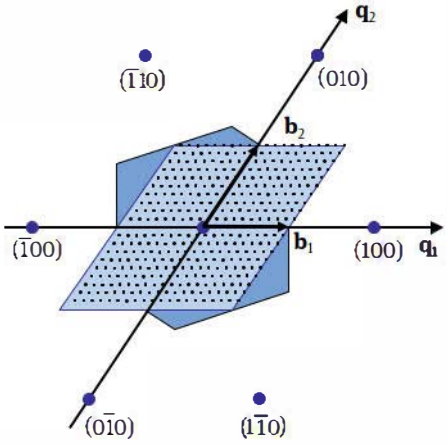


Fig. 6.27: A section, in the b_1-b_2 plane, through the reciprocal lattice of a "monoclinic crystal" with 20 unit cells per edge length. The large points reflect the reciprocal lattice, the smaller ones the allowed $\mathbf{q}$-vectors. The unit cell of the reciprocal lattice with the basis vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ is shown in light blue, the Brillouin zone in the background in dark blue.

It is remarkable that the **density of states in the reciprocal space** depends only on the sample volume and not on the crystal structure. Thus nothing changes regarding the physics if we choose another cell of the same size and place the permitted states there. Usually taking the first Brillouin zone, also shown in Figure 6.27, proves to be the most advantageous choice.

We start by giving the density of states in reciprocal space for *low-dimensional systems*. The derivation of ϱ_q shows directly how the dimensionality of the system enters, so that we only present the result here. For a linear chain of the length L or for a two-dimensional systems with area A we find:

$$\varrho_q^{(1)} = \frac{L}{2\pi} \qquad \text{and} \qquad \varrho_q^{(2)} = \frac{A}{(2\pi)^2} \,. \tag{6.77}$$

In the next step we determine the density of states of three-dimensional systems in frequency space. This quantity, the **density of states** $\mathcal{D}(\omega)$, is the important quantity in the calculation of the specific heat, which is determined by the number of excitations with energies comparable to the thermal energy. Thus, we need to determine the number of permitted states in the frequency interval $\mathrm{d}\omega$ as shown in Figure 6.28 for the two-dimensional case. These are the states between the two surfaces $S(\omega)$ and $S(\omega + \mathrm{d}\omega)$. In principle, we have to sum up all states in this frequency interval. However, because the states in the reciprocal space are very closely packed, we can solve this problem by simply replacing the summation by an integration. The number of permitted $\mathbf{q}$-values in the frequency interval under consideration is therefore:

$$\mathcal{D}(\omega)\,\mathrm{d}\omega = \varrho_q \int_{\omega = \text{const.}}^{\omega + \mathrm{d}\omega = \text{const.}} \mathrm{d}^3 q \,. \tag{6.78}$$

To use this general formula in practice, we can express the volume element by $\mathrm{d}^3 q = \mathrm{d}q_\perp \mathrm{d}S_\omega$. Here $\mathrm{d}S_\omega$ represents an element of the surface $S(\omega)$ and $\mathrm{d}q_\perp$ the distance

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

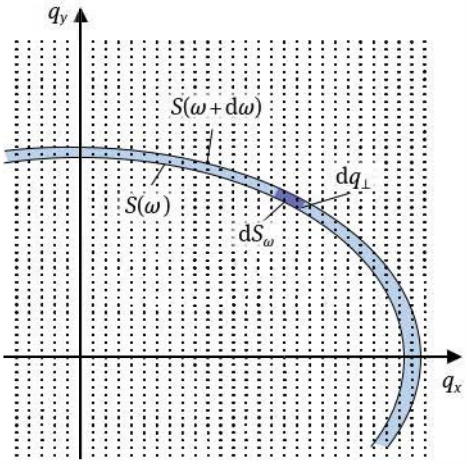


Fig. 6.28: To calculate the density of states. The density of states is obtained by summing the states in the light blue area between $S(\omega)$ and $S(\omega + d\omega)$. The area element $dq_\perp\, dS_\omega$ is marked in dark blue.

between the two surfaces $S(\omega)$ and $S(\omega + d\omega)$. To transform $dq_\perp$ we use the definition of group velocity. Since $v_g = |d\omega/d\mathbf{q}| = |\text{grad}_\mathbf{q}\,\omega| = |d\omega/dq_\perp|$, we find $|dq_\perp| = d\omega/v_g$. Thus following from (6.78) we find for the density of states:

$$\mathcal{D}(\omega)\,d\omega = \varrho_\mathbf{q}\,d\omega \int\limits_{\omega=\text{const.}} \frac{dS_\omega}{|\text{grad}_\mathbf{q}\,\omega|} = \frac{V}{(2\pi)^3}\,d\omega \int\limits_{\omega=\text{const.}} \frac{dS_\omega}{v_g}\,. \tag{6.79}$$

In the simplest case, that is in isotropic solids, the surface of constant frequency in reciprocal space is the surface of a sphere on which the group velocity is constant. If q is the radius of this sphere, then the density of states can be written:

$$\mathcal{D}(\omega)d\omega = \varrho_q\,d\omega\,\frac{4\pi q^2}{v_g} = \frac{V}{2\pi^2}\,\frac{q^2}{v_g}\,d\omega\,. \tag{6.80}$$

The distribution of the states in frequency space is thus clearly determined by the group velocity. The less steep the dispersion curve, the denser are the allowed states in frequency space. This situation is illustrated in Figure 6.29. Since the two selected ranges dq have the same size, they contain the same number of allowed states. However, it is obvious that these states are located in frequency ranges of different sizes.

In general, however, the directional dependency must also be taken into account. The flatter the dispersion curve $\omega(\mathbf{q})$, the greater the density of states, since this is accompanied by a reduction in the group velocity. In phonon spectra there are always frequency ranges in which the dispersion curve is horizontal (cf. Section 6.3) and thus the group velocity vanishes. These critical points are known as **Van-Hove**[27] **singularities**. In a linear chain the density of states diverges at these points. In three-dimensional solids, however, this is not the case. The integrand diverges in (6.79),

27 Léon Van Hove, *1924 Brussels, †1990 Geneva

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

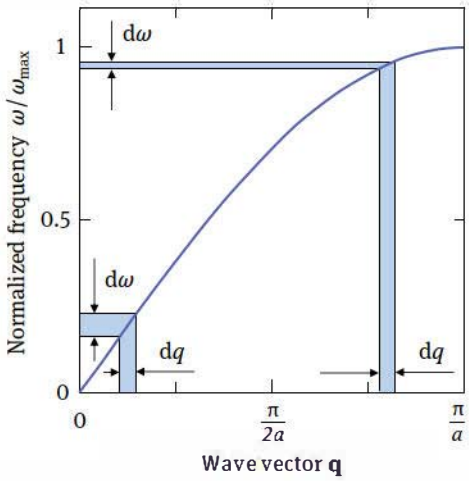


Fig. 6.29: Illustration of the increase in the density of states with decreasing slope of the dispersion curve. The two blue areas with identical interval dq contain the same number of allowed states, which are located in differently spaced frequency intervals.

but not the integral. This can be shown by taking the dispersion curve $\omega(\mathbf{q})$ in the vicinity of a critical point at frequency ω_c, expanding it in a Taylor series in ω and then performing the integration. In the case of a local maximum in the dispersion curve, the calculation shows that the density in the vicinity of the maximum has a parabolic dependence $\mathcal{D}(\omega) \propto \sqrt{\omega_c - \omega}$. As expected, it is not the density of states that diverges, but its derivative $d\mathcal{D}(\omega)/d\omega \propto (\omega_c - \omega)^{-1/2}$. There are four different types of critical points, local maxima, local minima and two different types of saddle points. These are described by the following equations,

$$\mathcal{D}(\omega) = \underbrace{C\sqrt{(\omega - \omega_c)}}_{\text{minimum}}, \qquad \mathcal{D}(\omega) = \underbrace{C\sqrt{(\omega_c - \omega)}}_{\text{maximum}}, \tag{6.81}$$

and

$$\mathcal{D}(\omega) = \underbrace{D_0 - C\sqrt{(\omega - \omega_c)}}_{\text{saddle point 1}}, \qquad \mathcal{D}(\omega) = \underbrace{D_0 - C\sqrt{(\omega_c - \omega)}}_{\text{saddle point 2}}. \tag{6.82}$$

The above four types of Van Hove singularities are shown schematically in Figure 6.30a with the special points labeled by arrows.

It should be pointed out that we have so far only considered the density of states of one branch. When discussing the density of states of three-dimensional samples, we must of course sum up the contributions of the individual phonon branches. This means that the actual density of states $D(E)$ is given by $D(E) = \sum_i \mathcal{D}_i(E)$, where $\mathcal{D}_i(E)$ is the contribution of each of the individual branches.

Figure 6.30b shows the density of states as calculated from the phonon dispersion curves of silicon shown in Figure 6.21. Since the contributions of six phonon branches overlap, the mapping of the particular features in the density of states and the dispersion curve is not always easy. At low frequencies, the density of states for silicon is

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

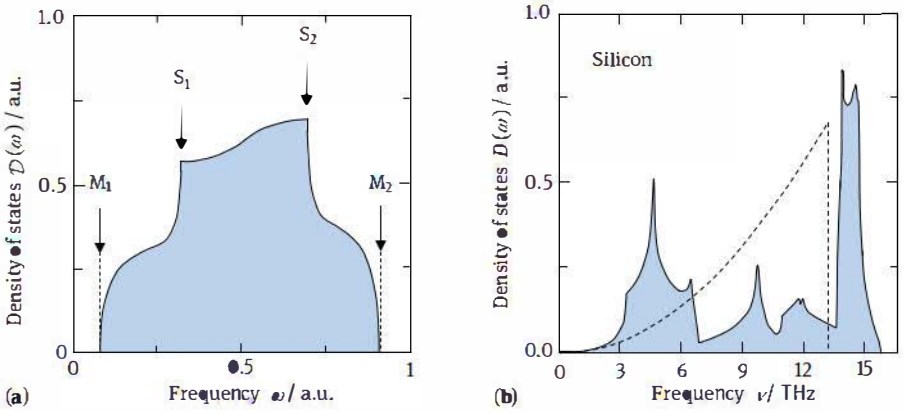


Fig. 6.30: **a)** Schematic representation of the various Van Hove singularities. The positions of a minimum M_1, a maximum M_2, a saddle point S_1 and a saddle point S_2 are indicated. **b)** The actual phonon density of states $D(\omega)$ for silicon. The density of states as calculated for the Debye approximation is shown for comparison by the dashed curve. (After W. Weber, Phys. Rev. B 15, 4789 (1977).)

determined by acoustic phonons and starts with a quadratic relation. We will discuss this typical dependence in more detail in the following section. The high density of states above 13 THz is due to the optical phonons, which only extend over a relatively narrow frequency band. The Van-Hove singularity is clearly visible when the density of states decreases at the highest frequencies. In addition, also shown in the figure is the density of states based on the so-called *Debye approximation*, which we will now discuss.

The density of states can be determined directly by various experimental methods, e.g. by second-order Raman scattering or by incoherent inelastic neutron scattering, or, if the dispersion curves are fully known, they can be calculated by averaging over all branches and spatial directions.

6.4.2 The Specific Heat in the Debye Approximation

We now use the concept of the density of states to calculate the specific heat of solids in the **Debye approximation**. This approximation, which was proposed by *P. Debye*, referred to *isotropic* solids with a monatomic basis, i.e. solids that have only acoustic phonons. Furthermore, the simple form of the dispersion relation $\omega = vq$ is assumed for all phonons, although it actually is valid only for long-wavelengths phonons. In other words: we use the approximation of the elastic continuum up to the highest frequencies. Under these assumptions, the equation for the density of states (6.80) simplifies to:

$$\mathcal{D}(\omega)\,\mathrm{d}\omega = \frac{V}{2\pi^2}\,\frac{\omega^2}{v^3}\,\mathrm{d}\omega\,, \tag{6.83}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

where here $\mathcal{D}(\omega)$ is the density of states *per phonon branch*. It increases quadratically with frequency and therefore must have an upper *cut-off frequency* $\omega_{\max}$ since the number of vibrational states is finite. In the Debye approximation, this cut-off frequency is defined by the number of allowed states per branch. Therefore the number of atoms N and the cut-off frequency $\omega_{\max}$ are related by

$$N = \int_0^{\omega_{\max}} \frac{V}{2\pi^2} \frac{\omega^2}{v^3} \, \mathrm{d}\omega. \tag{6.84}$$

From this, we find for the cut-off frequency

$$\omega_{\max} = v \sqrt[3]{\frac{6\pi^2 N}{V}} = \frac{v}{a} \sqrt[3]{6\pi^2}\,. \tag{6.85}$$

The second expression is obtained by using the relation $V/N = a^3$, where a is the lattice constant.[28] The cut-off frequency deviates from the value that one would derive from equation (6.73). This is understandable, since in the Debye approximation the allowed states fill a sphere in $\mathbf{q}$ space, whereas under the originally chosen boundary conditions (6.73) they fill a cube in reciprocal space. Since the density of the states in $\mathbf{q}$ space is the same in both cases, sphere and cube have the same volume.

Since there is one longitudinal and two (degenerate) transverse phonon branches in isotropic solids, we can calculate the **Debye density of states**, by introducing an "average" Debye velocity v_{D}:

$$\frac{3}{v_{\mathrm{D}}^3} = \frac{1}{v_{\ell}^3} + \frac{2}{v_{\mathrm{t}}^3} \tag{6.86}$$

The factor 3 in the expression takes into account the fact that $3N$ vibrational states are present. For many solids we can approximately set $v_{\ell} \approx 3v_{\mathrm{t}}/2$, whereby v_{D} takes the value $v_{\mathrm{D}} \approx 1.2\, v_{\mathrm{t}}$.

Using this form of the Debye velocity we find for the density of states:

$$D(\omega) = \frac{V\omega^2}{2\pi^2} \left(\frac{1}{v_{\ell}^3} + \frac{2}{v_{\mathrm{t}}^3} \right) = \frac{3V}{2\pi^2} \frac{\omega^2}{v_{\mathrm{D}}^3}\,. \tag{6.87}$$

The density of states at low frequencies is thus primarily determined by the transverse lattice waves.

Although there must be a different upper cut-off frequency for longitudinal and transverse phonons due to their different sound velocity, it is common practice to introduce only one cut-off frequency, the **Debye frequency** ω_{D}, in the approximation discussed here. The value of the Debye frequency is calculated from equation (6.85), if

28 As we have seen, the number of allowed vibrational states per branch is determined by the number of unit cells $\mathcal{N}$ and not by the number of atoms N. In the present case, however, we are looking at crystals with a monatomic basis, for which $\mathcal{N} = N$ holds.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the velocity of sound v is replaced by the Debye velocity v_D. This simple theory is also often applied to crystals with polyatomic bases with no distinction between acoustic and optical branches. In these cases the Debye frequency is also determined simply by the particle number density.

If the phonon density of states is known, the temperature-dependent contribution of the phonons to the internal energy U and thus to the specific heat of the solid can be calculated. For the internal energy we find the simple expression:

$$U = \int_0^{\omega_D} \hbar\omega\, D(\omega)\langle n(\omega, T)\rangle \mathrm{d}\omega\,, \tag{6.88}$$

where

$$\langle n(\omega, T)\rangle = \frac{1}{\mathrm{e}^{\hbar\omega/k_B T} - 1} \tag{6.89}$$

is the *Bose-Einstein factor*[29], which reflects the mean thermal occupation of the eigenstates. For harmonic oscillators this factor indicates how many energy levels are occupied on average. In our case, $\langle n(\omega, T)\rangle$ tells us how many phonons with frequency ω are excited in the thermal average per allowed wave vector.

When integrating (6.88), we do not need to take into account the zero-point energy, since it does not contribute to the specific heat. We will discuss its magnitude separately in the next subsection. For the internal energy we find[30]

$$U(T) = \frac{9N}{\omega_D^3} \int_0^{\omega_D} \frac{\hbar\omega^3}{\mathrm{e}^{\hbar\omega/k_B T} - 1} \mathrm{d}\omega\,. \tag{6.90}$$

The only material-dependent parameter is the Debye frequency ω_D. Now we introduce the **Debye temperature** Θ via the relationship $k_B\Theta = \hbar\omega_D$ and use the abbreviations $x = \hbar\omega/k_B T$ and $x_D = \hbar\omega_D/k_B T = \Theta/T$. This leads to the final result for the specific heat

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V = 9Nk_B \left(\frac{T}{\Theta}\right)^3 \int_0^{x_D} \frac{x^4 \mathrm{e}^x}{(\mathrm{e}^x - 1)^2} \mathrm{d}x\,. \tag{6.91}$$

This is the famous formula of *P. Debye*, which provides a universal representation of the temperature dependence of specific heat with only *one* material-specific parameter. The following Table 6.3 contains the Debye temperatures Θ of some crystalline materials.

Let us now take a look at the limiting cases of high and low temperatures in the Debye approximation. At low temperatures, we expect a rapid increase in specific heat, but this increase should be less steep than in Einstein's theory. At high temperatures, the Debye theory should agree with the results of classical thermodynamics, in other words it should reproduce the Dulong-Petit law.

---

29 Jagadish Chandra Bose, *1858 Bikrampur, †1937 Giridih

30 We note here and below that the number of degrees of freedom is given by $3\rho\mathcal{N} = 3N$.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Tab. 6.3: Debye temperatures of some selected materials. (Various sources.)

Element	Θ(K)	Element	Θ (K)	Compound	Θ(K)	Compound	Θ(K)
Ag	227	K	91	LiF	670	RbF	267
Al	433	Li	344	LiCl	420	RbCl	194
Au	162	Na	156	LiBr	340	CsF	245
C_{dia}	2 250	Pb	105	NaF	445	CsCl	175
C_{gra}	431	Pt	237	NaCl	297	CsBr	125
Cs	41	Se	152	NaBr	238	CsI	102
Cu	347	Si	645	NaI	197	AgCl	180
Fe	477	Sn	200	KF	335	AgBr	140
Ge	373	W	383	KCl	240	BN	600
He_{solid}	25	Zn	329	KJ	173	SiO_2	255

For *high temperatures* the value of x in (6.91) goes to zero. In this case we can simplify and evaluate the integral as follows:

$$\int_0^{x_D} \frac{x^4 e^x}{(e^x-1)^2}\,dx \approx \int_0^{x_D} \frac{x^4 \cdot 1}{(1+x-1)^2}\,dx = \int_0^{x_D} x^2\,dx = \frac{1}{3}\left(\frac{\Theta}{T}\right)^3 . \tag{6.92}$$

If we insert this result in (6.91), we find $C_V = 3N_A k_B = 3R_m$ for the molar heat capacity in agreement with the Dulong-Petit law, where R_m is again the universal gas constant.

For *low temperatures* the upper integration limit $x_D \to \infty$. In this case the integral can be solved analytically to give:

$$\int_0^{\infty} \frac{x^4 e^x}{(e^x-1)^2}\,dx = \frac{4\pi^4}{15}, \tag{6.93}$$

yielding for the specific heat:

$$C_V = \frac{12\pi^4}{5} N k_B \left(\frac{T}{\Theta}\right)^3 . \tag{6.94}$$

This is the famous **T^3-law** for the specific heat of non-metallic solids at low temperatures. Good agreement between the theoretical description (6.94) and experiment is shown in Figure 6.31 for the case of argon. Similar good agreement is observed for other pure insulating crystals.

Since the sound velocities of solids vary over a wide range and the Debye temperature is proportional to the sound velocity, the specific heat also varies considerably at low temperatures for different materials. The extreme cases are diamond and solid helium, whose Debye temperatures, according to Table 6.3, differ by almost two orders of magnitude and whose specific heats therefore differ by about six orders of magnitude!

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

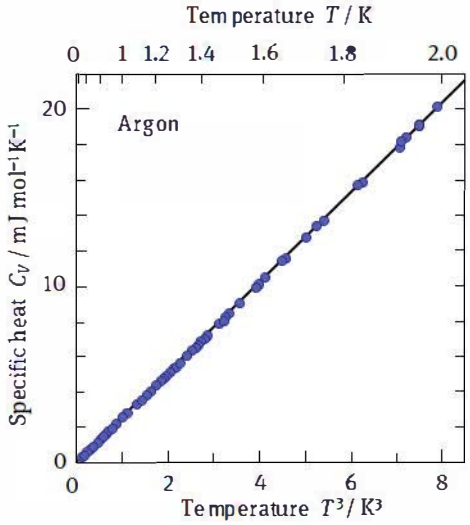


Fig. 6.31: The specific heat of crystalline argon at low temperatures, plotted versus T^3. (After L. Finegold, N. E. Phillips, Phys. Rev. 177, 383 (1964).)

How good is the Debye approach? Figure 6.31 demonstrates an impressive agreement at low temperatures. A comparison between theory and experiment at higher temperatures for many different materials is shown in Figure 6.32 and is also convincing. However, if we carry out a more precise analysis of the data for individual substance, it becomes apparent that discrepancies do occur in the transition region between low and high temperatures. In general, the Debye approximation is good for $T < \Theta/10$ at low temperatures and for $T > \Theta/5$ at high temperatures. At low frequencies, which are important at low temperatures, the density of phonon states is very well described by the quadratic frequency dependence of (6.87). At high temperatures, the details of the spectrum become unimportant because all vibrations are excited. Comparing

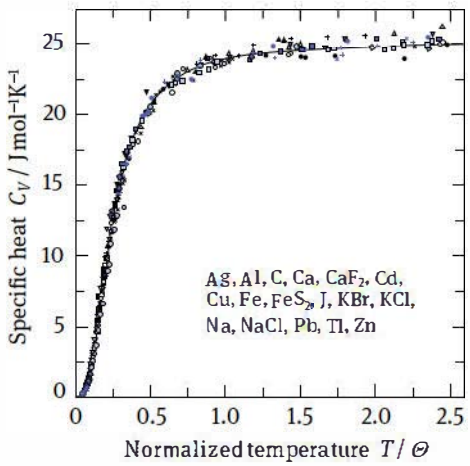


Fig. 6.32: Molar heat capacity of different substances as a function of the reduced temperature T/Θ. Due to the normalization to the Debye temperature all measured data fall on one curve. (After E. Schrödinger, *Handbuch der Physik*, volume X, H. Geiger, K. Scheel, eds., Springer, 1926)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the actual density of state of silicon in 6.30b with the density of state in the Debye approximation, the agreement between theory and experiment is nevertheless surprising. In principle, an improvement of the theory at higher temperatures could be achieved by the separate treatment of acoustic and optical phonons. Given that the density of states does not rise steadily, as assumed in the Debye approximation, hut is relatively low in the middle frequency interval, the Debye approximation can be used for the acoustic phonon. However, for the description of the contribution of optical phonons, the Einstein model is better suited, since the assumption of fixed oscillator frequencies comes closer to the real situation. Since we have not discussed the Einstein model we will not pursue these "subtleties" any further.

The discrepancy between the Debye theory and the experimental results in the intermediate temperature range is clearly visible in Figure 6.33, which shows measurements on diamond but with a higher resolution and a wider temperature range than those of Figure 6.26. In addition to newer data, the data points which were available to *A. Einstein* are also shown, which were measured by *F.H. Weber* in 1872. It is clear that the Einstein approximation only describes the data well at higher temperatures, whereas the Debye approximation also leads to a very good agreement at low temperatures. The discrepancy in the middle temperature region results from the difference between the real phonon density of states and the simple approximation assumed by Debye, as shown in Figure 6.30.

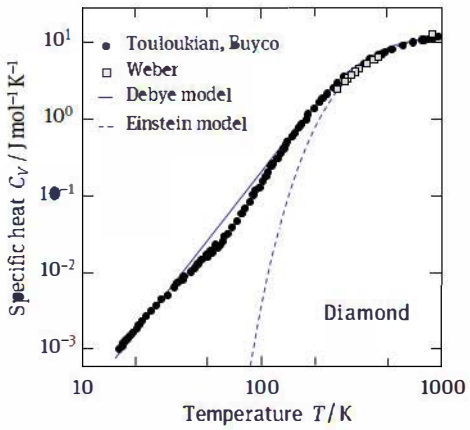


Fig. 6.33: Specific heat of diamond. (After Y.S. Touloukian, E.H. Buyco, Thermophysical Properties of Matter, Volume V, IFI/ Plenum, 1970.) The data represented by squares are from F.H. Weber, Ann. Phys. **147**, 311 (1872). Also drawn are the theoretical curves according to the models of *A. Einstein* and *P. Debye*.

6.4.3 The Specific Heat in Low-dimensional Systems

Having discussed the specific heat of crystals, we now briefly consider the density of states of one-dimensional or isotropic two-dimensional systems (cf. equation (6.77)).

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The group velocity v_g in these simple cases is independent of direction and can be pulled out of the integral. Thus, we have $\int dS_\omega$, which is reduced to a line integral in two-dimensional systems, for which we find the circumference $2\pi q$ in isotropic q-space. For the density of states $\mathcal{D}_2(\omega)$ per branch we obtain the expression

$$\mathcal{D}^{(2)}(\omega) = \frac{A}{4\pi^2}\frac{2\pi q}{v_g} = \frac{A}{2\pi}\frac{q}{v_g}\,, \tag{6.95}$$

and thus a linear increase with the wave number, if v_g is independent of q.

When deriving the density of states of a linear chain, we must take into account that due to the periodic boundary conditions, the "surface integral" results in a factor of two:

$$\mathcal{D}^{(1)}(\omega) = \frac{L}{2\pi}\frac{2}{v_g} = \frac{L}{\pi v_g}\,. \tag{6.96}$$

The density of states is independent of the wave number q, so it is constant for linear dispersion.

If we do the calculation of specific heat for a two-dimensional system, we find a quadratic dependence. This is due to the linear increase of the density of states with the wave number in the limiting case of low temperatures:

$$C_V \propto T^2\,. \tag{6.97}$$

A good example of such a system is provided by gas atoms condensed on graphite. The Van der Waals forces bind the atoms to the substrate, while allowing them to move laterally across the substrate surface. As a result, two-dimensional crystals can be formed which can be investigated by scattering methods as well as via the specific heat. The result of such a series of measurements on ^{3}He layers with thicknesses of less than one monolayer is shown in Figure 6.34. That the specific heat is indeed proportional to T^2 in all cases is clearly seen from the data. As the occupancy increases, the distance

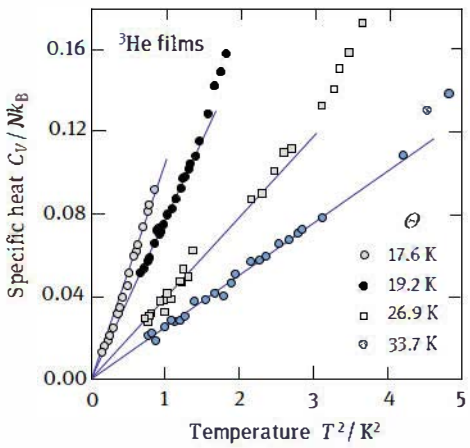


Fig. 6.34: The specific heat of several monolayer ^{3}He films on graphite plotted as a function of T^2, with the density of helium atoms increasing from 0.078 to 0.092 per Å^2 from left to right. The density increase also increases the Debye temperature Θ, which is used to label the curves. (After S.V. Hering et al., J. Low Temp. Phys. **25**, 793 (1976).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

between the atoms decreases and the interaction between the atoms becomes stronger. Therefore, with increasing density, the two-dimensional crystal on the substrate surface becomes stiffer. As a result the Debye temperature increases leading to a corresponding decrease in the specific heat.

6.4.4 The Zero-point Energy and Number of Excited Phonons

We have already faced the question of the size of the zero-point energy U_0 in connection with the binding energy of noble gas crystals. There we only dealt with the zero-point energy qualitatively, now we wish to actually calculate it in the context of the Debye approximation. The zero-point vibrations have half the energy of the thermally excited phonons at the same frequency. Since they are always present, independent of the temperature, the average occupation number of eigenmodes has the value of 1. For the internal energy of the zero-point vibrations, instead of equation (6.88) we have the following expression:

$$U_0 = \int_0^{\omega_D} \frac{\hbar\omega}{2} D(\omega)\, d\omega = \frac{9}{8} N k_B \Theta \,. \tag{6.98}$$

In other words, the zero-point energy increases linearly with the Debye temperature. The numerical values for noble gas crystals are given in Table 2.2. What was initially surprising about these values was that the zero-point energy U_0 does not decrease monotonically with the mass of the noble gas atoms. From our point of view here, however, this is not surprising, since it is not the mass of the individual atoms that is important, but the Debye temperature, into which the density and strength of the coupling between the atoms is build in. With $\Theta = 92\,\mathrm{K}$, solid argon has the highest Debye temperature of all noble gas crystals and thus also the highest zero-point energy.

A comparison with thermal energy is interesting. If we compare the zero-point energy and the thermal energy at the Debye temperature, we find roughly $U(T = \Theta) \approx 2U_0$. The zero-point energies and thermal energies of most substances are comparable at room temperature!

At the end of this section we will consider how the number N_{ph} of thermally excited phonons increases with temperature, since in the following chapters we will repeatedly need the relationship between the number of phonons and temperature. In the context of the Debye approximation, this quantity is defined by:

$$N_{ph} = \int_0^{\omega_D} D(\omega)\langle n(\omega, T)\rangle\, d\omega = \frac{3V}{2\pi^2 v_D^3} \left(\frac{k_B T}{\hbar}\right)^3 \int_0^{x_D} \frac{x^2}{e^x - 1}\, dx \,. \tag{6.99}$$

For the limiting case when $T \ll \Theta$, $x_D \to \infty$. Therefore the integral becomes constant and $N_{ph} \propto T^3$. For the other limiting case $T \gg \Theta$, the evaluation is the same as for equation (6.92). A dependence proportional to $(\Theta/T)^2$ is found for the integral and

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

thus $N_{ph} \propto T$. We summarise both results:

$$N_{ph} \propto \begin{cases} T^3 & \text{for } T \ll \Theta, \\ T & \text{for } T \gg \Theta. \end{cases} \tag{6.100}$$

The number of phonons increases rapidly at low temperatures. Around $T \approx \Theta$ the increase flattens out and becomes linear. For many phenomena that we will discuss later, the number of thermally-excited phonons plays an important role.

6.5 Vibrations in Amorphous Solids

We now turn to the vibration spectrum of amorphous solids, which is far less well known than that of crystals. First of all, we like to discuss how the dispersion curves for vibrational excitations look in this case. Figure 6.35 shows an attempt to illustrate the vibration spectrum of amorphous solids.

Of course, sound waves can be excited in glasses. For example, the ultrasonic experiment shown in Figure 6.5 was performed on a glass sample. However, the original definition of a phonon is based on the existence of a periodic lattice. For amorphous solids there is no periodic lattice so the definition of a phonon can only be strictly valid for long wavelengths where the details of the lattice periodicity become unimportant. In measurements on thin amorphous films, in which superconducting junctions were used to generate and detect phonons (see Section 11.2), it could be shown that well-defined phonons with frequencies up to about 500 GHz and wavelengths down to about 100 Å do exist. At least up to these frequencies, the observed dispersion curves are linear and the sound velocity is therefore frequency independent. With $q < \pi/50a$ the corresponding wave vectors are relatively small. To what extent phonon-like vibrations at higher

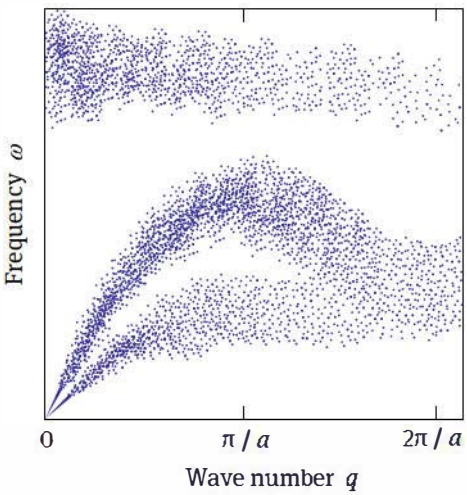


Fig. 6.35: Schematic illustration of the dispersion curves of amorphous materials. With increasing frequency, the phonon branches become less well-defined.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

frequencies are spatially localized or can move like lattice waves in a crystal is still largely unknown and the subject of both experimental and theoretical investigations.

Since amorphous materials lack periodicity in the atomic arrangement, neither a reciprocal lattice vector nor a Brillouin zone can be defined. A reduction of the dispersion curves to the first Brillouin zone is not possible, so that excitations with wave numbers $q > \pi/a$ may also exist. Furthermore, the disordered structure gives rise to spatial fluctuations of density and force constants and thus also a variation in vibrational frequencies. In Figure 6.35 the associated blurring of the dispersion curves is suggested by "point clouds".

Based on the experimental results, theoretical considerations and numerical simulations have shown that there are small areas in amorphous materials where the reactive forces are much smaller than in the rest of the sample or in crystals of the same composition. As a result, "soft", spatially localized vibrational states can occur. Despite their very small vibrational energies these states are not able to propagate as long-wave phonons.

Already during the discussion of the structure of amorphous solids, we have seen that the conventional methods of structure determination cannot simply be transferred from crystals to amorphous solids. This also applies to the interpretation of inelastic neutron scattering, which can be used to determine the vibrational spectrum of crystals. Of course, the expression for the conservation of energy for the scattering of neutrons on vibrations of amorphous solids can still be applied. However, because the lattice periodicity is missing and thus a reciprocal lattice vector cannot be defined, the conservation law (6.56) for the quasi-momentum does not hold.

Great efforts have been made to determine the spectrum of elastic excitations, i.e. the vibration spectrum of amorphous solids. Besides measurements of coherent inelastic neutron and Raman scattering, investigations of incoherent neutron scattering in particular have contributed to this. Figure 6.36 shows the density of states of vitreous silica at low vibrational energies as derived from various neutron scattering measurements at low and very high temperatures. To allow a comparison with the Debye model, the normalized density of states $D(\nu)/\nu^2$ is plotted as a function of the frequency ν. In this type of representation, the Debye density of state is represented by a horizontal straight line. The straight lines drawn have been calculated from elastic data using equation (6.87). Obviously at low frequencies the measured density of states considerably exceeds the Debye density of states, it is also much higher than that of crystalline quartz. In this type of plot we also see at about 1 THz a distinct maximum, the so-called **boson peak**. This peak can be found in a similar form for all amorphous solids. The cause of these low-energy vibrational states is not yet clear and is the subject of much speculation. At very low energies the reduced density of states $D(\nu)/\nu^2$ increases steeply with decreasing energy, but this energy range is not accessible for neutron scattering and therefore not shown in the figure. This part of the vibrational spectrum will be discussed in more detail in the next section.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

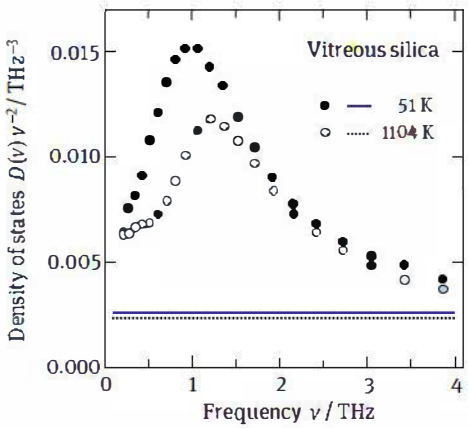


Fig. 6.36: Normalized density of state of the vibrational excitations of vitreous silica at relatively low frequencies. The neutron scattering measurements were performed at 51 K and 1104 K. The straight lines show the reduced Debye density of states for the specified temperatures. (After A. Wischnewski et al., Phys. Rev. B, 57, 2663 (1998).)

The surprising thing here is the strong temperature dependence of the measured density of states. It is known from crystals that the temperature dependence of the vibrational spectrum is caused by the anharmonicity of the lattice potential, which we will discuss in the next chapter. Obviously, anharmonic effects are particularly pronounced in amorphous solids.

6.5.1 Heat Capacity of Glasses

For crystals, elastic waves with long wavelengths can be described very well in the framework of an elastic continuum. The concept of phonons should be easily transferable to amorphous solids as long as the wavelengths are long enough for the medium to be considered homogeneous. As mentioned above, the existence of well-defined phonons has been experimentally proven in the long-wavelength limit up to about 500 GHz. Consequently, it can be assumed, although it has not yet been proven, that even for amorphous solids at low temperatures, the Debye theory provides a good description of the phonon contribution to the specific heat.

As we saw when discussing the specific heat of crystals, conclusions can be drawn on the vibration spectrum of the material from the temperature dependence. At high temperatures, the specific heat of glasses is described by the Dulong-Petit law, which expresses the fact that all vibrational degrees of freedom are excited. However, with decreasing temperature the specific heat in glasses decreases less steeply. This means that in amorphous solids, in addition to phonons, there must be other excitations. This observation is consistent with the measured density of states of lattice vibrations of vitreous silica at relatively low frequencies as shown in Figure 6.36.

The contribution of the additional vibrational states becomes particularly obvious at very low temperatures. To show this, Figure 6.37 compares the heat capacities of the two modifications of SiO_2, vitreous silica and crystalline quartz. It can be seen that

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

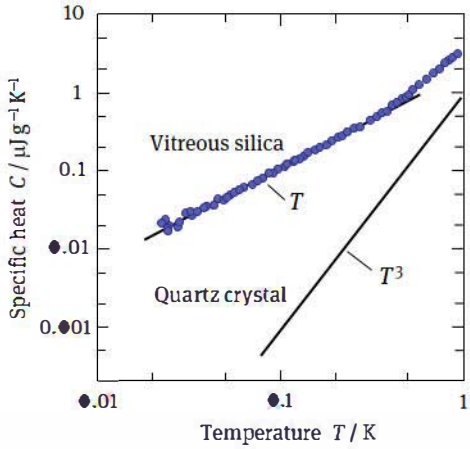


Fig. 6.37: The low-temperature dependence of the specific heats of vitreous silica and crystalline quartz. (After S. Hunklinger, Festkörperprobleme **17**, 1 (1977). The data are taken from the work of R.C. Zeller, R.O. Pohl, Phys. Rev. B 4, 2029 (1971) and J.C. Lasjaunias et al., Solid State Commun. **17**, 1045 (1975).)

below 1 K the two specific heats have completely different temperature dependences. While the crystalline material shows the classic T^3-temperature dependence, the glass exhibits an approximately linear dependence. At the lowest measuring temperature of about 25 mK the specific heat of the glass is more than 1000 times higher than that of the quartz crystal!

The linear term of the specific heat of glasses is described by the phenomenological **tunneling model**, which we discuss briefly here. In this model, it is assumed that even at very low temperatures (down to below 1 mK) structural rearrangements, based on quantum-mechanical tunneling processes may still occur in amorphous structures. This is possible if single atoms or small groups of atoms lack a unique equilibrium

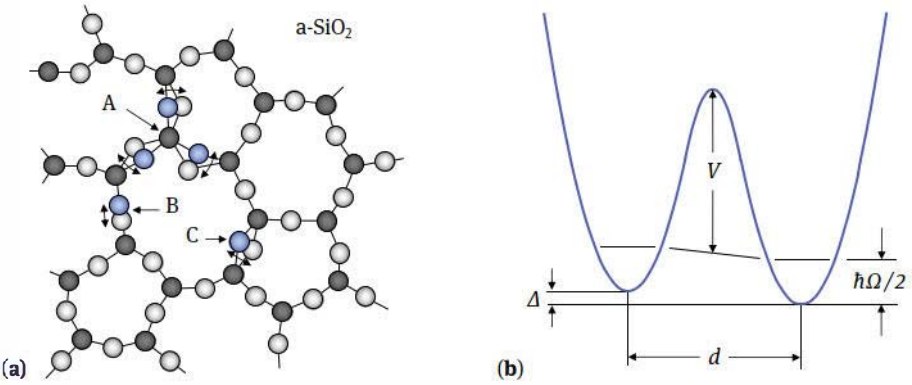


Fig. 6.38: Tunneling systems in vitreous silica. **a)** Two-dimensional illustration of the structure of vitreous silica. Individual atoms or small groups of atoms (denoted by A, B and C) do not occupy a clear equilibrium position. **b)** Double-well potential. The potential of the tunneling *particles* is characterized by the asymmetry energy Δ, the distance d and the potential barrier V. The ground state energy $\hbar\Omega/2$ of the isolated wells is also shown.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

position, with possible alternative positions being indicated in Figure 6.38a for vitreous silica. These groups of atoms, which we will simply call particles in the following, move in a double-well potential as shown in Figure 6.38b, in a simple approximation.

At high temperatures, the particles can simply jump over the barrier (cf. Section 5.1) by thermal activation. As the temperature decreases, the thermal process becomes less probable, but even at the lowest temperatures the particles can still tunnel between the two wells. They are therefore not localized in one well, as would be expected classically. From the ground states of the two individual wells, a common doublet, a **two-level system**, emerges. This splitting which is crucial to explaining the specific heat at low temperatures, can be calculated to good approximation by a linear combination of the wave functions ψ_a and ψ_b of the uncoupled states in well a or well b, as discussed in Section 2.4 when we treated the covalent bond. We therefore use equation (2.22), which we rewrite here:

$$(H_{aa} - E)(H_{bb} - E) - (H_{ab} - ES)^2 = 0\,. \tag{6.101}$$

H_{aa} and H_{bb} are the ground state energies which the particle would take up if there were no second well. We place the zero energy point at the midpoint between the two potential minima. Then $H_{aa} = (\hbar\Omega + \Delta)/2$ and $H_{bb} = (\hbar\Omega - \Delta)/2$. Furthermore we assume that the overlap of the two wave functions is only small, so that we may set the overlap integral $S \approx 0$. We then obtain the following:

$$E_{\pm} = \frac{1}{2}\left(\hbar\Omega \pm \sqrt{\Delta^2 + 4H_{ab}^2}\right)\,. \tag{6.102}$$

The ground state thus splits into a two-level system, with the energy difference between the two states determined by

$$E = E_+ - E_- = \sqrt{\Delta^2 + 4H_{ab}^2} = \sqrt{\Delta^2 + \Delta_0^2}\,. \tag{6.103}$$

With the help of perturbation theory H_{ab} can be calculated and we obtain approximately $-2H_{ab} = \Delta_0 = \hbar\Omega\, e^{-\lambda}$ where Δ_0 is referred to as the **tunneling splitting**. For the tunneling parameter λ, which is given by the potential shape and the mass m of the tunneling particles, the expression $\lambda^2 = mVd^2/2\hbar^2$ can be found as an approximation using the WKB method. The precise microscopic nature of the tunneling "particle" is still open. Certainly, Figure 6.38a in which the tunneling of individual atoms is suggested, is too simple, because it leads to ground state splittings E that are too large. Furthermore, individual atoms or groups of atoms cannot move completely independently of their surroundings because of their coupling to neighboring atoms. The tunneling motion of the particle probably consists of the joint motion of several atoms, whereby the amplitude of the displacement from the center of the tunneling system decreases rapidly with distance. The motion itself will not be a pure translation or rotation motion. The distance d between the wells is therefore understood as a so-called *configuration coordinate* and m as an *effective mass* which depends on the detailed nature of the particle motion. The concrete solution of the problem for real systems is difficult. The

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

reason for this is not only the lack of knowledge of the amorphous structure, but also the fact that the lack of symmetry makes it extremely difficult to treat the local motion of the atoms.

Owing to the irregularities of the amorphous structure, the double-well potentials distributed around the material are clearly not identical and we need to introduce a distribution which takes into account the statistical fluctuations of the relevant parameters.

In the tunneling model it is assumed that the parameters λ and Δ are independently and uniformly distributed. This can be described by the simple distribution function:

$$P(\Delta,\lambda)\mathrm{d}\Delta\mathrm{d}\lambda = \overline{P}\,\mathrm{d}\Delta\,\mathrm{d}\lambda \tag{6.104}$$

where $\overline{P}$ is a constant. Taking (6.103) into account, the distribution function $P(\Delta, \lambda)$ can be transformed into the distribution function:

$$P(E,\lambda)\mathrm{d}E\mathrm{d}\lambda = P(\Delta,\lambda)\,\frac{\mathrm{d}\Delta}{\mathrm{d}E}\,\mathrm{d}E\mathrm{d}\lambda = \overline{P}\,\frac{E}{\sqrt{E^2 - (\hbar\Omega\,\mathrm{e}^{-\lambda})^2}}\,\mathrm{d}E\mathrm{d}\lambda\,. \tag{6.105}$$

As in the case of phonons, the specific heat of the tunneling systems is also determined by the density of states $D(E)$ of the relevant excitations. To calculate the density of states of the two-level systems, we integrate over λ to obtain:

$$D(E) = \int_0^{\lambda_{\max}} P(E,\lambda)\mathrm{d}\lambda = \overline{P}\,\lambda_{\max}\,\mathrm{In}\,\frac{\hbar\Omega}{2E}\,. \tag{6.106}$$

The occurrence of an unphysical divergence at the upper limit of the integral is avoided by introducing an upper bound $\lambda_{\max}$ for the tunneling parameter λ. In real systems, of course, the distribution function $P(\lambda)$ does not stop abruptly, but gradually approaches zero. However, very large λ-values are of no physical relevance since the tunneling probability in tunneling systems with a large λ is so small that the tunneling particles are effectively confined in one well and no longer influence the dynamics of the lattice.

Although the numerical value of the ground state energy $\hbar\Omega$ is not known, we can assume that $\hbar\Omega \gg 2E$. Since the energy E varies only within relatively narrow limits, and the logarithmic dependence thus causes only a slight variation of the density of states with energy, we can take $D(E)$ to be constant to good approximation. In amorphous solids, we have found that at low temperatures the density of states D_0 is approximately constant. Since the mean thermal population of two-level systems is given by $f(E,T) = [\exp(E/k_\mathrm{B}T) + 1]^{-1}$,[31] with $x = E/k_\mathrm{B}T$, we find for the internal energy:

$$U = \int_0^\infty \frac{D_0 E}{\exp(E/k_\mathrm{B}T) + 1}\,\mathrm{d}E = D_0 k_\mathrm{B}^2 T^2 \int_0^\infty \frac{x}{\mathrm{e}^x + 1}\,\mathrm{d}x = \frac{\pi^2 D_0 k_\mathrm{B}^2 T^2}{12}\,. \tag{6.107}$$

31 When calculating the specific heat of two-level systems, it must be taken into account that the thermal occupation of the two levels is *not* represented by the Bose-Einstein factor which is derived on the basis of the level scheme of a harmonic oscillator. For two-level systems the mean thermal

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

From this the specific heat can be calculated to be

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V = \frac{1}{6}\pi^2 D_0 k_B^2 T \,, \tag{6.108}$$

which is in relatively good agreement with the temperature dependences as observed for vitreous silica and other amorphous solids.

It is an interesting question how many tunnelling systems there actually are. Definite quantitative statements are not possible, because above a few Kelvin, the contribution of tunneling systems to the specific heat is increasingly influenced by other contributions that cannot be separated out. However, it can be roughly deduced from experimental data that in the range $0\,\mathrm{K} < E/k_B \leq 1\,\mathrm{K}$ about $10^{17} - 10^{18}$ tunneling systems/cm^3 are present in amorphous solids.

Although the number of tunneling systems is small compared to the total number of atoms, at temperatures below 1 K they determine most of the properties of amorphous solids. At higher temperatures, localized excitations, which are closely related to the tunneling systems probably also play a role. In principle, these could be similar atomic configurations, but lacking the barrier between the two "equilibrium positions". The particles then move in a slightly curved potential and are only exposed to weak restoring forces. This means that the vibrational frequencies of these groups of atoms are small compared to the Debye frequency typical for crystalline solids.

From the multitude of amorphous solids, we have chosen vitreous silica as an example, since this material has been extensively studied. However, the explanations given here also apply, with slight modifications, to most other amorphous solids.

6.6 Exercises and Problems

1. Vibrating Aluminum Cylinder. A 20cm long polycrystalline aluminum rod is excited to longitudinal vibrations. At $\nu_0 = 12.9$ kHz resonant behavior is observed. Calculate the velocity of sound.

The sound velocity is measured on a 1 cm long sample with the help of longitudinal ultrasound. The echoes observed have a time interval of 3.22 µs. Compare the measured sound velocities. The following parameters are given: Density $\varrho = 2700\,\mathrm{kg/m^3}$, modulus of elasticity $E = 70.2\,\mathrm{GPa}$ and Poisson's ratio $\nu = 0.33$. Why do the results differ?

2. Inelastic Neutron Scattering. To study the dispersion relation of the phonons of copper, neutrons of the wavelength $\lambda_0 = 2.178$ Å are scattered on a copper single crystal

occupation is given by $f(E,T) = [1 + \exp(E/k_B T)]^{-1}$. This factor formally corresponds to the Fermi-Dirac distribution, which we discuss in Section 8.1. The different distributions have no influence on the temperature dependence of the specific heat, since both functions each contain the quantity $E/k_B T$ as an argument.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

(face-centered cubic, lattice constant $a = 3.615$ Å). The sample, the detector and the neutron source span a plane parallel to the (001) lattice planes of the crystal. If the neutron beam is incident parallel to the [100]-direction, one observes under an angle of $2\theta = 34.78°$ scattered neutrons of the wavelength $\lambda = 1.375$ Å.

(a) Are phonons created or annihilated?
(b) What is the frequency ν of the phonons involved in the scattering process?
(c) Calculate the scattering vector and sketch the scatter process in reciprocal space.
(d) What is the value and direction of the wave vector of the phonons involved?
(e) Does the result of the measurement agree with the published values of the dispersion relation of copper in Figure 6.18?

3. Brillouin Scattering. Light from an argon ion laser ($\lambda_L = 514.5$ nm) is scattered by a NaCl crystal (refractive index $n' = 1.54$). The scattered light is analyzed perpendicular to the beam direction. Two pairs of lines are observed. The frequencies of these lines are (with respect to the laser) shifted by $\Delta\nu = \pm 19.26$ GHz and $\Delta\nu = \pm 10.25$ GHz, respectively.

(a) Calculate the sound velocity of the phonons observed in this experiment.
(b) Which phonons (frequency and wave vector) can be observed for scattering angles between backward scattering ($\vartheta = 180°$) and forward scattering ($\vartheta = 0°$)? Which part of the Brillouin zone is covered by this?

4. Acoustic and Thermal Phonons. In an ultrasonic experiment (Figure 6.5), sound pulses at the frequency $\nu = 100$ MHz are coupled into a cylindrical silicon sample (diameter 1 cm, length 4 cm) at 4.2 K. The pulse duration is 1 µs, the sound intensity 5 mW/cm^2. The lattice constant and the Debye temperature are given by $a = 5.43$ Å and $\theta = 645$ K, respectively.

(a) How many phonons does a single pulse produce?
(b) After the excited phonons are thermalized by inelastic scattering what is the final temperature?
(c) How many phonons were generated as a result of thermalization in the frequency interval $\Delta\nu$ between 100 and 101 MHz?

5. One-dimensional Systems with a Diatomic Basis. Consider a linear chain of diatomic molecules similar to the one studied in Section 6.2. To simplify, we assume that the masses of the atoms involved do not differ, but the force constants do. Derive the dispersion relation and determine the frequency of the phonons at the wave vectors $q = 0$ and $q = \pi/a$.

6. Two-Dimensional Lattice. We consider a monatomic layer whose atoms with mass m are so weakly coupled to the substrate that they can be considered isolated. Furthermore, we assume that a square lattice with the lattice constant a has formed.

(a) Which expression describes the density of states of this layer in the Debye approximation?

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

(b) What is the relationship between the lattice constant and the Debye temperature?
(c) Show that at low temperatures the specific heat increases quadratically with temperature.

7. Excitations in Glasses. At low temperatures, both phonons and two-level systems are thermally excited in glasses. Using Figure 6.37, estimate the number of thermally excited two-level systems for vitreous silica at the temperatures 10 mK and 1 K and compare these values with the number of phonons present in each case. Note: The sound velocities of vitreous silica can be found in Section 6.1. Furthermore, the numerical values $\int_0^\infty x^2(e^x - 1)^{-1}\,dx = \Gamma(3)\zeta(3) \approx 2.404$ and $\varrho = 2201\ \mathrm{kg/m^3}$ are given.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

7 Lattice Anharmonicity

When discussing the lattice dynamics of crystals above, we have implicitly assumed that atoms move in a strictly harmonic potential and therefore perform harmonic oscillations. In this idealized concept, the coupled motion of the atoms can be broken down into normal modes that do not interact with each other. If this were indeed the case, it would not be possible to bring a non-equilibrium distribution of the phonons into equilibrium. It is only the anharmonicity of the crystal lattice that enables thermal equilibrium to be reached. Furthermore, the anharmonicity also manifests itself in many other solid state properties, for example in thermal expansion, finite thermal resistance, ultrasonic absorption or the various values of adiabatic and isothermal constants.

In principle, the non-linear properties can be described by taking into account higher-order terms in the expansion of the potential function of the lattice. This typically results in expressions that are so complicated and confusing that it is difficult to calculate concrete properties. Furthermore, the large number of material constants that follow would also be difficult to determine experimentally. Nevertheless, in order to develop a certain "feeling" for the important property of *anharmonicity*, we will go ahead and describe the thermal expansion and the interaction between the phonons phenomenologically, avoiding the analytical path via the development of the potential function to higher orders.

7.1 Thermal Expansion and the Equation of State

Thermal expansion implies the deviation of the lattice potential from the purely harmonic form because, as we shall see, the volume of a solid would be independent of temperature without the anharmonicity. The link between the sample volume and temperature means that important thermodynamic quantities such as internal or free energy also depend on sample volume.

Here we take a closer look at thermal expansion. First, we derive the *equation of state* for solids. In the absence of adequate microscopic models, we use a phenomenological approach proposed by *E. Grüneisen*[1]. By differentiating the equation of state with respect to volume, we will establish a link between the thermal expansion, the compressibility, and the specific heat.

We start with the free energy, which in the absence of defects and electronic or magnetic excitations is determined exclusively by the lattice vibrations. We have already made repeated use of the fact that each lattice vibration with wave vector $\mathbf{q}$ can formally be assigned a harmonic oscillator with frequency $\omega_{\mathbf{q}}$. As can be seen from *statistical*

1 Eduard Grüneisen, *1877 Giebichenstein, †1949 Marburg

https://doi.org/10.1515/9783110666502-007

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

mechanics, the expectation value of the free energy F of a harmonic oscillator can be calculated from the partition function

$$Z_{\text{osc}} = \sum_n e^{-E_n/k_B T} = \sum_n e^{-\hbar\omega(n+1/2)/k_B T} = \frac{e^{-\hbar\omega/2k_B T}}{1 - e^{-\hbar\omega/k_B T}} . \tag{7.1}$$

The summation is carried out over all eigenstates E_n. This results in the free energy given by:

$$F_{\text{osc}} = -k_B T \ln Z = k_B T \ln\left(1 - e^{-\hbar\omega/k_B T}\right) + \frac{1}{2}\hbar\omega . \tag{7.2}$$

The second, temperature-independent term on the right reflects the zero-point energy. The free energy of all the phonons is obtained by adding the contributions of all oscillators, i.e. by summing over all wave vectors $\mathbf{q}$ and all phonon branches j. To this, we add the elastic energy associated with the volume change of the sample $\delta V = (V - V_0)$, where V_0 denotes the initial volume. Since the compression modulus B depends only weakly on temperature, this contribution is, to a good approximation, independent of temperature. Thus the free energy F of a solid is given by:

$$F = \frac{BV_0}{2}\left(\frac{\delta V}{V_0}\right)^2 + \sum_{\mathbf{q},j}\left[k_B T \ln\left(1 - e^{-\hbar\omega_{\mathbf{q},j}/k_B T}\right) + \frac{1}{2}\hbar\omega_{\mathbf{q},j}\right] . \tag{7.3}$$

By differentiating with respect to the volume, we obtain the relationship between pressure and volume:

$$p = -\left(\frac{\partial F}{\partial V}\right)_T = -B\left(\frac{\delta V}{V}\right) - \hbar\sum_{\mathbf{q},j}\frac{\partial\omega_{\mathbf{q},j}}{\partial V}\left[\frac{1}{e^{\hbar\omega_{\mathbf{q},j}/k_B T} - 1} + \frac{1}{2}\right] . \tag{7.4}$$

Since only small volume changes occur, V_0 can be replaced by V in the last equation.

The anharmonicity is noticeable in that the change in volume leads to a change in the oscillation frequencies. If the solid were perfectly harmonic, then $(\partial\omega_{\mathbf{q},j}/\partial V) = 0$. As a result, the summation can only be carried out after some significant simplification that the relative frequency change $\delta\omega_{\mathbf{q},j}/\omega_{\mathbf{q},j}$ of the lattice vibrations is independent of frequency and proportional to the relative volume change $\delta V/V$. It is also usually assumed that all phonon branches behave in the same way. These assumptions can be expressed by setting:

$$\frac{\delta\omega_{\mathbf{q},j}}{\omega_{\mathbf{q},j}} = -\gamma\frac{\delta V}{V} , \tag{7.5}$$

where γ is the dimensionless **Grüneisen parameter**, a constant characteristic of the substance under consideration. We can also write In differential form:

$$\gamma = -\frac{\partial\left(\ln\omega_{\mathbf{q},j}\right)}{\partial\left(\ln V\right)} . \tag{7.6}$$

Typical values of the Grüneisen parameter fall in the range $\gamma \approx 1 - 3$. This means that phonon frequencies and volumes change approximately similarly under pressure. The

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

assumption that $\delta\omega_{\mathbf{q},j}/\omega_{\mathbf{q},j}$ does not depend on the frequency can only be a rough approximation of the actual situation. Even more problematic is the assumption that all phonon branches behave in the same way. Since it is known that longitudinal and transverse vibrations react differently to volume changes, improved agreement between theory and experiment can be achieved by using specific Grüneisen parameters γ_j for each phonon branch.

With the introduction of the Grüneisen parameter, equation (7.4) simplifies to:

$$p = -B\left(\frac{\delta V}{V}\right) + \frac{\gamma}{V}\sum_{\mathbf{q},j}\hbar\omega_{\mathbf{q},j}\left[\frac{1}{e^{(\hbar\omega_{\mathbf{q},j}/k_B T)}-1} + \frac{1}{2}\right] . \tag{7.7}$$

The summation is made over the product of all phonon energies and their occupation probability. This is just the energy of all the excited phonons, i.e. their internal energy U including the zero-point energy. If we multiply this equation by the volume, we obtain the **equation of state**:

$$pV = -B\delta V + \gamma U(T) . \tag{7.8}$$

We derive the equation of state at constant volume as a function of temperature, taking into account that the specific heat is given by $C_V = (\partial U/\partial T)_V$ and obtain $(\partial p/\partial T)_V = \gamma C_V/V$. If we use the chain rule $(\partial p/\partial T)_V \cdot (\partial T/\partial V)_p \cdot (\partial V/\partial p)_T = -1$, together with the definition of the thermal volume expansion coefficient[2] $\alpha_V = V^{-1}(\partial V/\partial T)_p$ and the definition of the bulk modules $B = -V(\partial V/\partial p)_T^{-1}$, we obtain the relationship $(\partial p/\partial T)_V = \alpha_V B$.

With this we find the **Grüneisen relation** from (7.8)

$$\alpha_V = \frac{\gamma C_V}{BV} . \tag{7.9}$$

In this approximation, the thermal expansion coefficient α_V is given by the Grüneisen parameter γ, the specific heat C_V and the bulk modules B. Since the bulk modules is only weakly dependent on temperature, the temperature dependence of the expansion coefficient is almost completely determined by the specific heat. At high temperatures, i.e. at room temperature or above, the specific heat C_V and thus also the expansion coefficient α_V are practically constant. At low temperatures, on the other hand, both quantities are strongly dependent on the temperature. Figure 7.1 shows the temperature dependence of the thermal expansion coefficient and the suitably normalized specific heat of aluminum.[3] The agreement between the two quantities is convincing and confirms the validity of the Grüneisen relation.

2 The thermal volume expansion coefficient α_V is related to the thermal linear expansion coefficient α_L via the relationship $\alpha_V = 3\alpha_L$.

3 The free electrons of aluminum also contribute to specific heat and thus to the thermal expansion. As we will see in the next chapter, since at higher temperatures their contribution is negligible compared to that of the lattice, we will therefore neglect it here.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

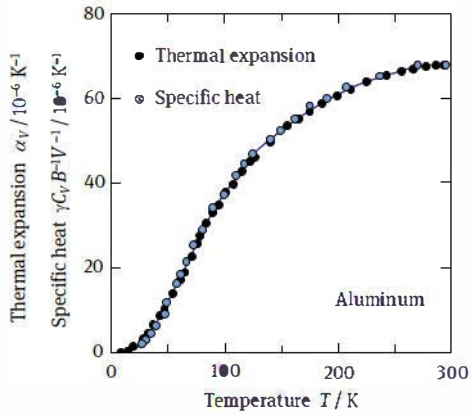


Fig. 7.1: Thermal expansion coefficient and reduced specific heat of aluminum. Here the value of the Grüneisen parameter γ is taken as 2.2. (After Y.S. Touloukian, E.H. Buyco, *Thermophysical Properties of Matter*, Volume IV, IFI/ Plenum, 1970.)

Since, as already mentioned above, the derivation of the Grüneisen relation is very simplified, some additional comments are appropriate here. Volume changes generally affect the various phonon branches differently because the atoms in crystals usually vibrate in an anisotropic environment. For example, we usually find a much higher value for the Grüneisen parameter for longitudinal waves than for transverse waves. The Grüneisen parameter can also have negative values and different signs for different branches and crystal directions. This of course also has an effect on the thermal expansion. An example of this unusual behavior is that of hexagonal tellurium, where the expansion coefficient is negative in the c axis direction, but positive along the other two axes.

The sign of the Grüneisen parameter can also depend on the temperature. For example, for many crystals with a zinc-blende or diamond structure, such as silicon, the Grüneisen parameter is positive at room temperature, but negative at low temperatures. Figure 7.2 shows a measurement of the *linear* thermal expansion coefficient α_L of silicon

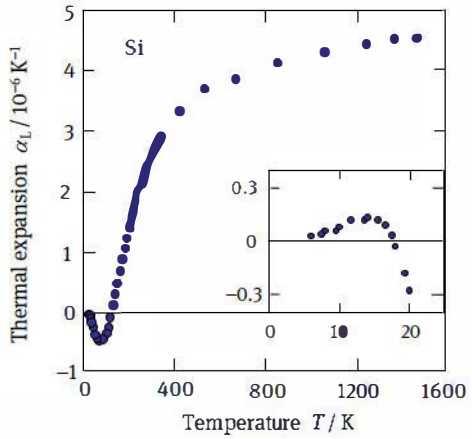


Fig. 7.2: The linear coefficient of expansion α_L of silicon. The inset shows an enlargement of the behavior at low temperatures. (After K.G. Lyon et al., J. Appl. Phys. **48**, 865 (1977) and Y. Okada, Y. Tokumaru, J. Appl. Phys. **56**, 314 (1984).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

over a wide temperature range. At the higher temperatures, α_L decreases only slightly with decreasing temperature as expected. However, at around room temperature α_L starts to show a steeper drop which is caused by the decrease in specific heat. Surprisingly, the thermal expansion coefficient changes its sign at 125 K, passes through a minimum at 75 K and becomes positive again below 20 K. Finally, α_L decreases again, with the variation at low temperatures being consistent with the T^3 variation of the specific heat. The negative value of the Grüneisen parameter is caused by transverse acoustic phonons.

Although equation (7.9) was derived for phonons only, the applicability does not depend on whether the free energy is determined by phonons or by other excitations. Lattice defects usually react very sensitively to external influences, thus at low temperatures they can often play a considerable role in the anharmonic properties. As an example, we consider here the thermal expansion of *amorphous solids* at low temperatures. In Section 6.5 we have seen that in these materials at temperatures below 1 K the specific heat and thus the internal energy is primarily attributable to tunneling systems and not to phonons. Therefore, if an external force acts on an amorphous sample, its volume will change as will the energy splitting E of the tunneling systems. According to equation (7.5) this change should be given by $\delta E/E = -\gamma\,(\delta V/V)$. We therefore expect phonons *and* tunneling systems to contribute to thermal expansion at low temperatures. Equation (7.9) states that there is proportionality between the expansion coefficient and the specific heat $\alpha_L \propto C_V$. Since at low temperatures the specific heat of amorphous solids can be expressed as the sum of a linear and a cubic term, we would expect an expression $\alpha_L = aT + bT^3$ for the thermal expansion, where the parameters a and b are material-specific constants determined by the respective values of γ and C_V. As Figure 7.3 shows, this temperature dependence has indeed been found in amorphous solids. In order to separate the linear and the cubic term graphically, the quantity α_L/T is plotted in this figure as a function of T^2. In all cases, a straight line is found within

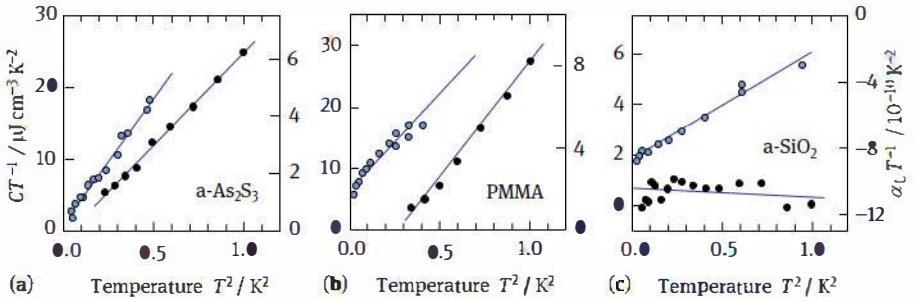


Fig. 7.3: The specific heat (blue points) and linear thermal expansion coefficient (black points) of amorphous materials. The quantities C/T or α_L/T are plotted as a function of T^2. **a)** Semiconducting glass a-As_2S_3, **b)** Polymer PMMA, **c)** Vitreous silica a-SiO_2. (After D.A. Ackerman et al., Phys. Rev. B **29**, 966 (1984).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the accuracy of the measurement in this type of plot: the axis intercept provide the constant a and the slope determines the constant b.

The temperature dependence of the thermal expansion of vitreous silica, where the value $a = -1 \times 10^{-9}\,\mathrm{K}^{-2}$ is found, is remarkable. This results in a very large negative value $\gamma = -65$ for the Grüneisen parameter of tunneling systems which are responsible for the linear term in the specific heat. The energy splitting of the tunneling systems, and thus their contribution to the internal energy, is obviously extremely sensitive to volume or pressure changes. Additionally, we should note that the thermal expansion of amorphous materials is extremely small at temperatures below 1 K despite the large absolute value of their Grüneisen parameter, because the specific heat is negligible compared to its value at room temperature. This also explains the large scattering of the data points in Figure 7.3 despite the fact that the experimental accuracy of the measuring device was 2×10^{-4} Å.

From the discussion here, we would expect that Grüneisen parameters would normally be determined by measuring the thermal expansion and the specific heat. However, this is not the case, since the Grüneisen parameter of individual phonon branches can be measured "directly" by changing the volume which thus, according to (7.5), also changes the phonon frequencies of the sample. For example, the Grüneisen parameter of optical phonons can be determined relatively easy by the pressure shift of Raman or infrared spectra. Ultrasonic experiments allow for the study of the anharmonicity of acoustic branches. Here, the pressure dependence of the sound velocity is measured and the corresponding Grüneisen parameter is calculated. However, both methods only allow the investigation of long-wavelength phonons. If we are interested in the behavior of short-wavelength phonons, the pressure dependence of inelastic neutron scattering must be used.

7.2 Phonon-Phonon Scattering

As mentioned at the beginning of the chapter, there is no interaction between the normal modes in the harmonic approximation. However, the existence of non-quadratic terms in the lattice potential eliminates the independence of the lattice vibrations. The interaction of the phonons with one another is therefore a particularly important consequence of the anharmonicity of the lattice potential. Due to the complexity involved, however, there is no fully developed theory. Therefore, we will just pick out some important processes and discuss them qualitatively.

7.2.1 Three Phonon Processes

The simplest and most important interaction process is the **three-phonon process** which can be impressively demonstrated in an ultrasound experiment, the principle of

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

which is shown in Figure 7.4. The sound waves produced by transducers (1) and (2) meet in the superposition zone and, owing to the lattice anharmonicity, produce a new wave which can be detected by transducer (3). In the experiment illustrated the interacting sound waves had different frequency and polarization, the reasons for which are briefly discussed in the following section.

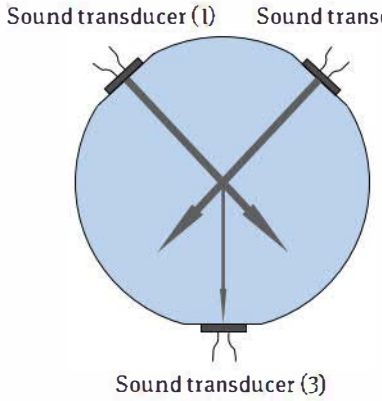


Fig. 7.4: Measuring setup for the investigation of the phonon-phonon interaction. Sound transducer (1) generates transverse sound waves at 10 MHz, and transducer (2) longitudinal sound waves at 15 MHz. The transducer (3) detects longitudinal sound waves at 25 MHz. (After F.R. Rollins, Jr., L.H. Taylor, P.H. Todd, Jr., Phys. Rev. **136**, A597 (1964).)

From the selected frequencies and from the geometry of the experimental setup, it can be seen in this experiment that energy and quasi-momentum conservation are fulfilled for the interacting phonons and the following equations apply:

$$\hbar\omega_1 + \hbar\omega_2 = \hbar\omega_3 \qquad \text{and} \qquad \hbar\mathbf{q}_1 + \hbar\mathbf{q}_2 = \hbar\mathbf{q}_3 \,. \tag{7.10}$$

These relations are known as the **selection rules** for the scattering processes.

7.2.2 Ultrasonic Absorption in Crystals

The propagation of phonons is hindered by scattering processes. In dielectric crystals two main mechanisms contribute: scattering by defects, and the interaction of the phonons with each other. In ultrasonic experiments, the scattering processes manifest themselves via their effect on the attenuation. In Figure 6.5 we have already sketched the basic measurement setup for such experiments. The attenuation coefficient or the mean free path l of the ultrasonic waves can be determined directly from the decrease in signal amplitude with increasing propagation distance or with the propagation time.

In the treatment of phonon scattering, we use the particle picture, which makes the processes occurring very clear.[4] We start from the relationship generally valid for

4 Ultrasonic propagation is a transport phenomenon that is best described by *Boltzmann's transport equation*. As in the discussion of heat conduction, we will take a simplified approach here. The transport equation is discussed in Section 9.2 in connection to electrical conductivity.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

scattering processes:

$$l = \frac{1}{n\sigma} . \tag{7.11}$$

The mean free path of the scattered particles is inversely proportional to the **scattering cross section** σ of the individual scattering centers and to their density n. The scattering cross section in turn is proportional to the square $|\mathcal{A}|^2$ of the scattering amplitude, which we introduced in Section 4.2.

First, we will look at the elastic scattering of phonons by point defects, where the basic concept can be clearly illustrated. Then we will consider *phonon-phonon scattering*, which is inelastic and caused by the anharmonicity of the lattice. Part of the results obtained here will also be used later in Section 7.3 when we discuss thermal conductivity.

Scattering at Point Defects. Let us assume that we have a crystal containing a significant number of point defects, for example vacancies or interstitial atoms. Since these violate the translational symmetry of the lattice, ultrasonic waves are *elastically scattered* by them. Labelling the phonons before and after scattering with the indices 1 and 2, the energy and quasimomentum conservation can be written: $\omega_1 = \omega_2$ and $\mathbf{q}_1 = \mathbf{q}_2 + \mathbf{K}$. As a result of the scattering process, the quasimomentum $\hbar\mathbf{K}$ is transferred to the defect. The value of the transferred quasimomentum lies in the range $0 < K \leq 2q$ where the equal sign implies backscattering. As with the scattering of light on small particles, the scattering amplitude or the scattering cross section can be calculated with the help of *Rayleigh scattering*. Assuming that the wavelength is large compared to the size of the defect, i.e. $\lambda \gg a$, one finds the relation $|\mathcal{A}|^2 \propto \sigma \approx \pi a^2 (aq)^4$, where a is the size of the defect, which for atomic point defects is of the order of an atomic spacing. The exact value of the scattering cross section depends both on the elastic properties of the crystal and the specific properties of the defect.

We are not interested here in the absolute value of the effect, but rather in its frequency and temperature dependence. Since $\omega \propto q$ applies to long-wavelength phonons and the scattering center density n is determined by the point defect density n_D, equation (7.11) leads to the following expression for the inverse mean free path for ultrasonic phonons

$$l^{-1} \propto n_D \omega^4 . \tag{7.12}$$

Obviously, phonon scattering at point defects are especially important at high frequencies. Since non-resonant scattering provides a temperature-independent contribution, it only causes a constant background when measuring the temperature response of ultrasound attenuation. However, in crystals with large impurity levels, scattering by point defects can be the largest contribution to the attenuation.

Phonon-Phonon Scattering. Figure 7.5 shows a measurement of the attenuation of transversal sound waves in a pure quartz crystal. The temperature dependence is

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

typical of that for pure dielectric crystals:[5] at low temperatures the attenuation in good crystals is negligible, rising rapidly with increasing temperature and finally saturating. This arises from the interaction of ultrasonic phonons with thermal phonons. Put simply, an ultrasonic phonon is annihilated by scattering with a thermal phonon.

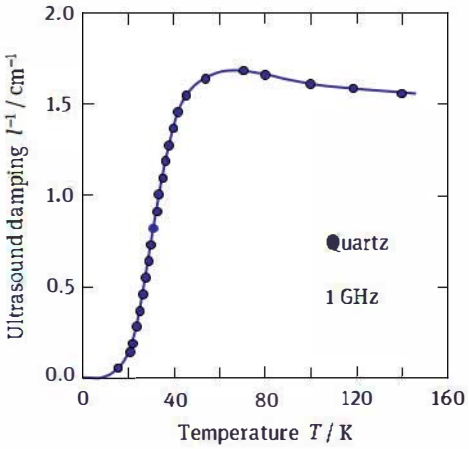


Fig. 7.5: The temperature dependence of the attenuation of transverse sound waves in a quartz crystal at 1 GHz. (After H.E. Bömmel, K. Dransfeld, Phys. Rev. 117, 1245 (1960).)

The full treatment of phonon interactions is a challenging theoretical problem requiring extensive calculation. We limit ourselves here to the main features of the theory and assume the validity of the Debye approximation. It is important for further discussion that the strength of the phonon-phonon interaction, and thus the scattering cross section, is not determined by the displacement of the atoms but by the change in the interatomic distance, as described by the *strain tensor* [**e**] defined in (6.4), since for common displacements of the atoms, the interatomic distances remain constant and thus anharmonicities do not come into play.

We first consider the scattering cross section σ. Without going through the necessary steps in detail, we will simply state that the scattering amplitude $\mathcal{A}$ is proportional to $\prod_i e_{\mathbf{0},i}$ where $e_{0,i}$ represents the amplitude of the strain caused by the phonons. Therefore, for the scattering cross section, the following proportionality applies to the three-phonon process:

$$\sigma \propto |\mathcal{A}|^2 \propto (e_{0,1}\, e_{0,2}\, e_{0,3})^2 \propto \omega_1\, \omega_2\, \omega_3 \,. \tag{7.13}$$

The relation $e_{\mathbf{0}}^2 \propto \omega$ follows directly from the fact that the energy of elastic waves $\hbar\omega$ increases quadratically with the distortion. The absolute value of the scattering cross section is determined by terms which include the second and third order derivatives of the potential function $\tilde{V}$, which was introduced in Section 6.2 and is given by the

5 In metals, the damping is usually dominated by the interaction between phonons and free electrons.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Taylor expansion of (6.42). In Section 6.2 we truncated the expansion after the second term since we were only interested in the harmonic approximation. The constants in this expression depend on the crystal symmetry and differ for the different phonon branches.

In the following discussion we will use the index "us" instead of "1", because the incident quasiparticles are ultrasonic phonons. The expression (7.11) for the mean free path includes the scattering cross section and the density of the scattering centers, which we will now look at more closely. It is obvious that we cannot simply calculate the total number of thermally excited phonons since the scattering cross section depends on the phonon frequency. The relevant quantity is therefore the number of phonons excited at a certain frequency, i.e. the quantity $D(\omega)\langle n(\omega, T)\rangle$, the product of the density of states and the average thermal population. What is also less obvious is that *six* different processes are hidden behind the conservation laws (7.10). We will pick just two of them. The remaining four can be easily taken into account when the order of phonons is cyclically interchanged for the interaction process. The scattering process that has been discussed so far is the scattering of an ultrasonic phonon $\mathbf{q}_{us}$ with a thermal phonon $\mathbf{q}_2$, whereby a further phonon $\mathbf{q}_3$ is generated. This process can also take place in the reverse direction: a thermal phonon $\mathbf{q}_3$ can decay under the influence of the ultrasonic wave into an ultrasonic phonon $\mathbf{q}_{us}$ and a thermal phonon $\mathbf{q}_2$. If there were the same number of phonons with wave vectors $\mathbf{q}_2$ and $\mathbf{q}_3$, both processes would take place with the same probability. The crucial variable is therefore the difference between the number of phonons with wave vectors $\mathbf{q}_2$ and $\mathbf{q}_3$, $\delta n = D(\omega_2)\langle n(\omega_2)\rangle - D(\omega_3)\langle n(\omega_3)\rangle$. The two frequencies ω_2 and ω_3 are only slightly different since $\omega_{us} = (\omega_3 - \omega_2)$ holds. Thus we can approximate by setting $D(\omega_3) \approx D(\omega_2)$ and expand the mean occupation number difference in a Taylor series, which we truncate after the first term, i.e.:

$$\delta n \approx D(\omega_2)\left\{\langle n(\omega_2)\rangle - \left[\langle n(\omega_2)\rangle + \frac{\partial\langle n(\omega)\rangle}{\partial\omega}\bigg|_{\omega_2}\omega_{us} + \ldots\right]\right\} \approx -\omega_{us}\, D(\omega_2)\frac{\partial\langle n(\omega)\rangle}{\partial\omega}\bigg|_{\omega_2} . \quad (7.14)$$

We will only briefly outline the further relatively complex derivation of the damping coefficient but rather explain the result. First, the four additional scattering channels mentioned above must be included. Since all thermally excited phonons contribute to the scattering, it is still necessary to average over all the wave vectors and spatial directions, taking into account energy conservation. This results in a factor $1/\omega_{us}$, which compensates the factor ω_{us} in (7.14). In equation (7.13) for the calculation of the scattering cross section, if we set $\omega_2 \approx \omega_3 \approx \omega$ and assume that $D(\omega) \propto \omega^2$, we obtain the following relationship for the inverse mean free path:

$$\begin{aligned} l^{-1} &\propto \int \sigma(\omega)\, D(\omega)\frac{\partial\langle n(\omega,T)\rangle}{\partial\omega}\,\mathrm{d}\omega \propto \omega_{us}\int \omega^4\frac{\partial\langle n(\omega,T)\rangle}{\partial\omega}\,\mathrm{d}\omega , \\ &\propto \omega_{us}\,T^4\int_0^{x_D} x^4\frac{\partial\langle n(x)\rangle}{\partial x}\,\mathrm{d}x \propto \omega_{us}\,T^4\int_0^{x_D}\frac{x^4\mathrm{e}^x}{(\mathrm{e}^x-1)^2}\,\mathrm{d}x . \end{aligned} \quad (7.15)$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As with the calculation of the specific heat in Section 6.4, we have used the abbreviations $x = \hbar\omega/k_B T$ and $x_D = \hbar\omega_D/k_B T$. Since the same integral appeared there, we can assume the result directly. For low temperatures, $x_D \to \infty$, the integral yields a constant value. This gives us the expression for the frequency and temperature dependence:

$$l^{-1} \propto \omega_{us} T^4 . \tag{7.16}$$

This relationship is known as **Landau-Rumer damping.**[6,7] As can be seen in Figure 7.6, the expected temperature dependence of the damping is indeed observed in the experiment.

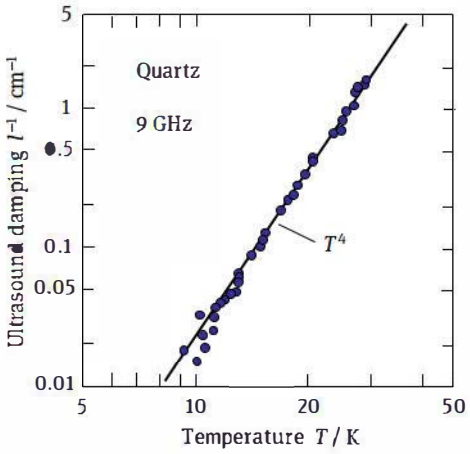


Fig. 7.6: Logarithmic plot of the damping of transverse sound waves in quartz at 9 GHz as a function of temperature. The T^4 dependence is indicated by the straight line. (After M.F. Lewis, E. Patterson, Phys. Rev. **159**, 703 (1967).)

Referring briefly back to Figure 7.5, from which we saw that at higher temperatures the damping in a quartz crystal flattens out. In general, in this temperature range the ultrasound damping depends on the sample material, the direction of sound propagation and the polarization of the sound waves. However, in the simplest approximation, the damping can be assumed to be constant at these temperatures. We should note that at such high temperatures the simple particle picture is no longer relevant in describing the interaction between ultrasonic phonons and thermal phonons. The reason being that due to the many interactions between the thermal phonons, their average lifetimes and, thus their mean free paths, decrease rapidly with increasing temperature. As soon as the latter become comparable to the wavelength of the ultrasound, this simple particle concept loses its validity. We will not pursue this topic any further here, although it is highly relevant to the technical applications of ultrasound.

Surprisingly, (7.16) does not apply to longitudinal waves, for which an even steeper rise of the attenuation with temperature is observed at low temperatures. The reason for

6 Lev Davidovich Landau, *1908 Baku, †1968 Moscow, Nobel Prize 1962
7 Yuri Borisovich Rumer, *1901 Moscow, †1985 Novosibirsk

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

this is that conservation of energy and quasimomentum forces certain *selection rules*, which we have already pointed out in the description of Fig. 7.4 without considering them in the derivation of (7.16). Even a simple thought experiment shows that scattering processes involving phonons of only *one* branch must be largely suppressed. The two conservation laws (7.10) can only be fulfilled simultaneously if the dispersion curve is strictly linear and the scattering is such that the wave vectors of all phonons involved lie on a straight line. Since the dispersion curves are always nonlinear, the two conservation laws cannot be fulfilled simultaneously collisions between phonons of the same branch. Therefore, phonon-phonon interactions must take place between phonons of different branches. The selection rules must be considered when deriving the ultrasonic attenuation and lead to modifications of the simple considerations given above. A weak softening of the selection rules is caused by the fact that the dispersion curves have a finite width due to the finite phonon lifetime. This effect allows for a limited interaction between the phonons of the same branch. This effect plays a crucial role in the attenuation of longitudinal sound waves at low temperatures.

7.2.3 Spontaneous Phonon Decay

Now we turn to the interesting question of whether phonons have indefinitely long lifetimes in a defect-free crystal at absolute zero. At first sight, this might be expected, since there are no thermally excited phonons with which the generated phonons could interact. The answer is somewhat surprising: phonons experience spontaneous decay through their interactions with zero-point vibrations of the lattice.

Figure 7.7 shows that the high-frequency phonons in CaF_2 decay relatively rapidly at 2 K and that their average lifetime is strongly dependent on the frequency. Roughly speaking, a phonon with frequency ω decays into two phonons with about half that

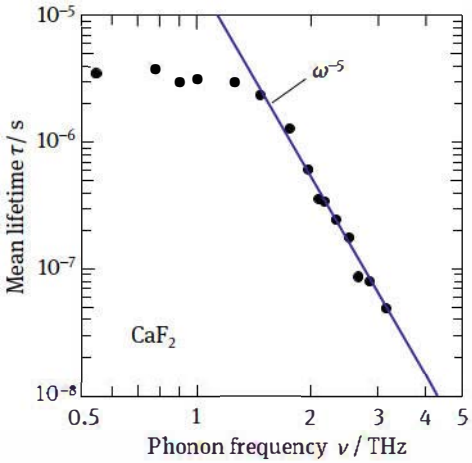


Fig. 7.7: Average lifetime of terahertz phonons in CaF_2, measured at 2 K. (After R. Baumgartner et al., Phys. Rev. Lett., **47**, 1403 (1981).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

frequency, since the phase space for this final state is largest. Since this is also a three-phonon process, the frequency dependence of the scattering cross section can be calculated in the same way as for normal ultrasonic attenuation. For this purpose we use equation (7.13) and set $\omega_1 = \omega$ and $\omega_2 = \omega_3 = \omega/2$. This gives us the proportionality $\sigma \propto \omega^3$ for the scattering cross section. Since the phonon density of states increases quadratically, the number of scattering partners also rises quadratically with frequency. Thus it follows using (7.11) that $l \propto \omega^{-5}$. Taking the mean life time $\tau = l/v$ the following relationship is then obtained

$$\tau \propto \frac{1}{\omega^5} \,. \tag{7.17}$$

The strong frequency dependence of the process means that spontaneous decay only makes a noticeable contribution to the damping coefficient for very high-frequency phonons and not for those with the usual ultrasonic frequencies. In the measurement shown in Figure 7.7, the frequency range was chosen so that $\hbar\omega \gg k_B T$ and thus even at 2 K the decay and not the scattering of high-frequency phonons dominates. The levelling off of the curve below 1.5 THz is due to the measurement technique, in which Eu^{2+} ions were used to generate phonons, which in turn shortens the lifetime of the phonons through resonant absorption.

7.2.4 Ultrasound Absorption in Amorphous Solids

In crystals and amorphous solids, the mechanisms responsible for ultrasonic attenuation, are very different. Although attenuation in amorphous materials at higher temperatures is not yet fully understood, it is certain that phonon scattering processes, which determine attenuation in pure dielectric crystals, play only a minor role. The difference in behavior becomes particularly clear at low temperatures: While in crystals the attenuation disappears or becomes very small, in amorphous materials at low temperatures it increases strongly.

We now take a closer look at the ultrasonic absorption of glasses at low temperatures. The reason for the strong attenuation or the short mean free path of the phonons is the **resonant interaction** between the sound waves and the atomic tunneling systems, which we learned about in Section 6.5 when discussing the specific heat of amorphous solids. In principle, this process proceeds as follows: if the energy $\hbar\omega$ of the ultrasonic phonons and the energy splitting E of tunneling centers coincide, transitions between the two levels can be induced, as shown in Figure 7.8.

Here too, the scattering process can run in both directions: either an ultrasonic phonon is absorbed or a phonon is emitted. How effective the respective process is depends on the relative occupation of the two levels. At $T \to 0$, the mean free path becomes minimal, since all tunneling systems are in their ground state, only the process in which phonons are absorbed remains. If $k_B T \gg E$, the two levels are occupied approximately equally, such that transitions are induced with equal rate in both di-

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Fig. 7.8: Schematic representation of the transitions for resonant interaction between phonons and two-level systems (tunneling systems).

rections. On average, the sound wave experiences no resonant damping due to the presence of the tunneling systems. Analogous to the three-phonon process, the occupation number difference of the two levels is important for the calculation of the mean free path. If n is the number density of the tunneling systems with the energy splitting E; n_1 and n_2 the number densities of the systems in the upper and lower levels; $n = (n_1 + n_2)$; and $n_1/n_2 = \exp(-E/k_B T)$ holds, then for an occupation number difference of $\delta n = (n_2 - n_1)$ for the two-level systems, we find:

$$\delta n = n \tanh \frac{E}{2k_B T} . \tag{7.18}$$

Since in the resonant process $E = \hbar\omega_{us}$, we get with (7.11)

$$l^{-1} = \sigma \delta n = \tilde{C}\omega_{us} \tanh \frac{\hbar\omega_{us}}{2k_B T} . \tag{7.19}$$

According to (7.13) the absorption is proportional to the ultrasonic frequency, since only one ultrasonic phonon is involved in the absorption process, thus $\sigma \propto e_0^2 \propto \omega_{us}$. The prefactor $\tilde{C} = \bar{P}\,\bar{\gamma}^2/\varrho v^3$ in addition, to the mass density ϱ and the sound velocity v, also contains the density of states $\bar{P}$ of the tunneling systems and a parameter $\bar{\gamma}$, reflecting the strength of the coupling between phonons and tunneling systems. From the specific heat of the amorphous solids (cf. Section 6.5) we know that there is a broad distribution of energy splittings of tunneling systems. We therefore expect attenuation at *all* measuring frequencies, with a temperature dependence being given by the thermal population of the two levels. Such a curve is shown in Figure 7.9.

The limiting cases for high and low temperatures can be easily specified. At very low temperatures $k_B T \ll \hbar\omega_{us}$ and thus $\tanh(\hbar\omega_{us}/2k_B T) \approx 1$. The damping therefore has its largest, almost temperature-independent value at very low temperatures, where $l^{-1} \propto \omega_{us} T^0$. At high temperatures $\hbar\omega_{us} \ll k_B T$, the hyperbolic tangent can be approximated by its argument $(\hbar\omega_{us}/2k_B T)$. The damping then shows a pronounced frequency and temperature dependence, which is given by $l^{-1} \propto \omega_{us}^2/T$.

The ultrasonic damping by tunneling systems has another very characteristic feature. According to (7.19), the attenuation is proportional to the difference in occupation number δn, but the relationship (7.18) is only valid in thermal equilibrium. However, the absorption process itself causes a change in the occupation: with increasing sound

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

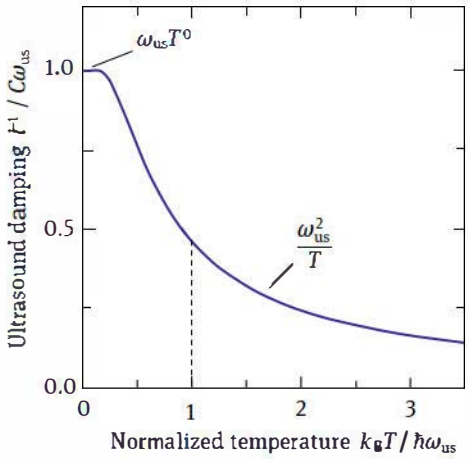


Fig. 7.9: Temperature dependence of the resonant ultrasonic damping normalized to the damping at absolute zero as a function of the normalized temperature $k_B T/\hbar\omega_{us}$.

intensity J, the occupation of the upper levels increases, the occupation number difference δn decreases and the damping due to resonant processes also decreases. An experimental confirmation of this **saturation effect** is shown in Figure 7.10 where a measurement on a borosilicate glass at about 0.5 K and 940 MHz is depicted. The curve of the inverse mean free path can be described very well with the theory discussed here only in its basic aspects.

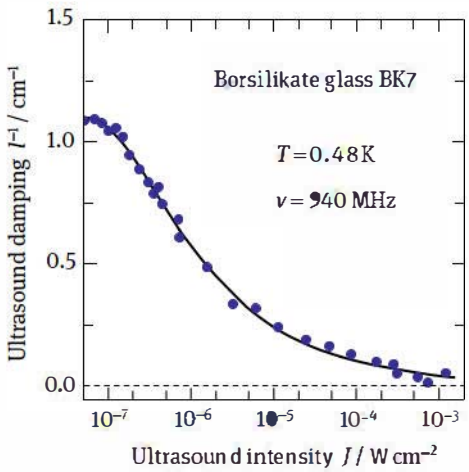


Fig. 7.10: Ultrasonic damping in borosilicate glass as a function of sound intensity. The solid line shows the theoretically-expected dependence. (After S. Hunklinger, Cryogenics **28**, 224 (1988).)

Measurements of the temperature dependence of ultrasonic attenuation in vitreous silica at relatively high and very low sound intensities are shown in Figure 7.11. In addition to the resonant damping at temperatures below 1 K just discussed, there is also a sharp increase at higher temperatures. This arises from another damping mechanism, namely **relaxation absorption**. This damping is based on the fact that sound waves

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

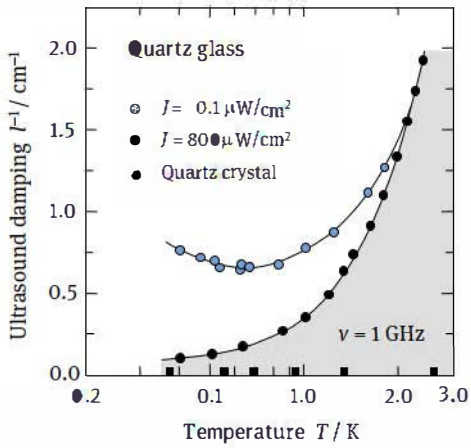


Fig. 7.11: The temperature dependence of the ultrasound attenuation in quartz glass at 1 GHz and two different sound intensities J. In crystalline quartz the attenuation over this temperature range is negligible. (After S. Hunklinger, Festkörperprobleme **17**, 1 (1977).)

not only interact resonantly with tunneling centers of suitable energy splitting, but also modulate the energy splitting of all systems and thus affect their thermal equilibrium occupations. This mechanism causes a strong increase in absorption above 1 K. Since at higher sound intensities the contribution of the resonant interaction falls owing to the saturation effect, only the relaxation contribution remains. This is indicated in grey in Figure 7.11. Therefore, by measuring at very low and at relatively high sound amplitudes, the contribution of the resonant damping can be separated from that of the relaxation effect. However, we will not go into the further details of relaxation damping here, since we will consider it in more detail in Section 13.3 in connection with the dielectric properties of solids. In addition, we should note that at low temperatures the resonant interaction is not only of great importance in glasses, but also in crystals, where point defects often give rise to the formation of tunneling centers (see Section 13.3).

7.3 Heat Transport in Dielectric Crystals

In dielectric crystals, heat is transported by lattice vibrations, i.e. by phonons. Here we can distinguish between classical *heat conduction* and the *ballistic propagation* of the phonons. A characteristic of classical heat conduction is that the phonons diffuse through the sample, whereas in ballistic propagation they pass through the sample without major interactions. We first discuss ballistic phonons, which can be observed at low temperatures. We will then deal with classical thermal conduction and finally with heat transport in one-dimensional systems.

In Figure 7.12, two typical sample geometries are illustrated. Figure 7.12a shows a typical setup for measuring ballistic phonon transport. Here, the heater is switched on for a short time and the time variation of the temperature on the other side of the thin specimen plate is recorded by means of a bolometer, i.e. a temperature-sensitive

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

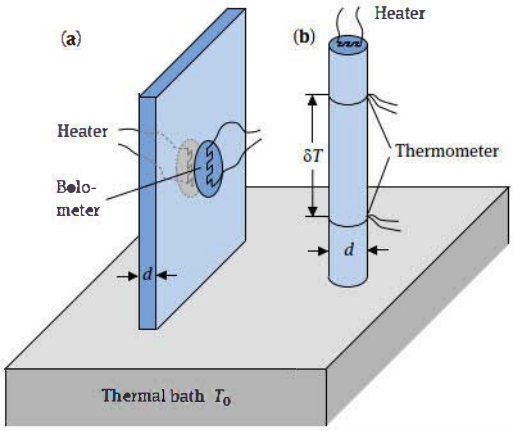


Fig. 7.12: Schematic of the measurement arrangements **a)** for investigating the ballistic propagation of phonons where the heat transport through a sample in the form of a thin plate is investigated and **b)** for measuring the thermal conductivity via the thermal transport along a long thin rod.

resistor with a small heat capacity. Figure 7.12b shows a classical heat conduction experiment, a long, thin rod is fixed on one end to a heat bath, and is heated at the other free end. The heat bath at the fixed end absorbs the heat and simultaneously provides the measurement temperature T_0. Two thermometers, placed a distance apart along the rod define the temperature difference δT, which should be much smaller than the bath temperature.

7.3.1 Ballistic Propagation of Phonons

Since at low temperatures, phonons pass through the sample without interacting with each other, **heat pulse experiments** can be performed. A corresponding experimental setup is shown in Figure 7.12a. As discussed above, a short electrical pulse to the heater genersates a heat pulse which passes through the crystal and is detected by a bolometer on the other side. The phonon transit time $t = d/v$ from heater to detector depends on the phonon polarization. The result of such an experiment on InSb samples is shown in Figure 7.13. In the pure sample the measurement signal has two distinct maxima which can be identified with the arrival of the longitudinal and transverse phonons. Since the heat pulse propagated in the [111] direction in this experiment, the two transverse branches were degenerate and thus cannot be distinguished.

We should note here that the main purpose of this experiment was to investigate the coupling between phonons and electrons. Therefore, in a second experiment a doped and therefore electrically conductive sample was used. Clearly the conduction electrons mainly couple to the longitudinal phonons, thereby strongly attenuating the heat pulse.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

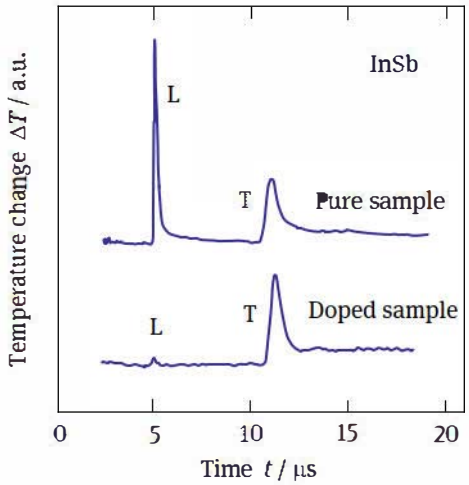


Fig. 7.13: Propagation of heat pulses in pure and doped InSb at 1.6K. The heat pulses were generated by a thin gold-film resistor and detected by a superconducting bolometer. (After J.P. Maneval et al., Phys. Rev. Lett. **27**, 1375 (1971).)

7.3.2 Thermal Conductivity

If the phonons do not pass through the sample without scattering as in ultrasonic or heat pulse experiments, but propagate by diffusion, the energy transport is determined by the temperature gradient. Under stationary conditions the **Fourier law**[8] applies

$$\mathbf{j} = -\Lambda \operatorname{grad} T\,, \tag{7.20}$$

where $\mathbf{j}$ is the heat flux density and Λ is the coefficient of thermal conductivity. A simple setup for the determination of classical heat conduction is shown schematically in Figure 7.12b.

Figure 7.14 shows the temperature dependence of the thermal conductivity coefficient[9] of a very pure sodium fluoride crystal, which is typical of crystalline dielectrics. The maximum thermal conductivity of 2.4×10^4 W/m K at 16.5 K was the highest value measured in a solid for many years. In less pure crystals, the conductivity maximum is not as pronounced. It is surprising that metals do not reach such a high level of thermal conductivity. For comparison: the thermal conductivity of pure copper at room temperature is about 400 W/m K, much higher than that of dielectrics, but it reaches its maximum at low temperatures at only about 1.8×10^4 W/m K.

For the sake of simplicity, we base our description of heat conduction on the analogy with kinetic gas theory first used by *P. Debye* and consider the phonons to approximate an ideal gas. According to the kinetic gas theory, the coefficient of thermal

8 Jean-Baptiste Joseph Fourier, *1768 Auxerre, †1830 Paris

9 In the following, the coefficient of thermal conductivity is often referred to only as *thermal conductivity*.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

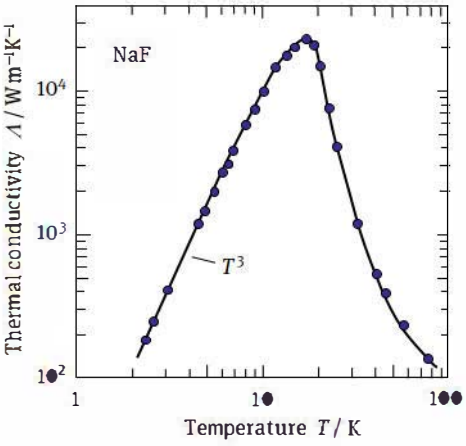


Fig. 7.14: The thermal conductivity of a high purity sodium fluoride single crystal as a function of temperature. (After H.E. Jackson et al., Phys. Rev. Lett. **25**, 26 (1970).)

conductivity Λ of gases is given by:

$$\Lambda = \frac{1}{3} C v l \,. \tag{7.21}$$

where C is the specific heat of the gas, v the average velocity of the atoms and l their mean free path. We use this equation to describe the thermal conductivity of solids, where C is the specific heat, v the velocity of sound and l the phonon mean free path.

In principle, we should take into account that all quantities in (7.21) depend on frequency and that the various phonon branches contribute differently to the heat conduction. Therefore, we would need to intergrate over all phonon frequencies and sum over all phonon branches j. Instead of the simple equation (7.21) the following would then apply

$$\Lambda = \frac{1}{3} \sum_j \int_0^{\omega_{max}} c_j(\omega)\, v_j(\omega)\, l_j(\omega)\, \mathrm{d}\omega \,, \tag{7.22}$$

where $c_j(\omega) = \mathrm{d}C_j/\mathrm{d}\omega$ represents the *spectral specific heat*. This equation is usually simplified in two ways. By introducing *one effective* phonon branch with linear dispersion $v = \omega/q = \text{const.}$ according to the Debye model, we can "omit" the summation over the phonon branches. Integration is avoided by applying the **dominant phonon approximation**. Here, only that part of the spectrum that contributes the most to the heat transport is taken into account. Phonons with energies comparable to the thermal energy $k_B T$ dominate the heat transport, and as such are used to model the response. This approximation is useful in the large number of cases where the exact frequency dependence of the quantities in (7.22) is unknown.

The dominant phonon approximation is not only used to describe heat conduction, but has many other applications. For example, specific heat data or ultrasonic attenuation is often analyzed within the framework of this approximation. A more detailed consideration shows that there is a relation $\hbar\bar{\omega} = p\, k_B T$ between the frequency $\bar{\omega}$ of the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

dominant phonons and the temperature. The proportionality factor p depends on the phenomenon under consideration and varies between 1 and 3. The value of p can be calculated by averaging the observed quantities.

As we will see shortly, thermal phonons in dielectric crystals experience two important scattering mechanisms: either the phonons interact with one another, or they are scattered by defects. If several scattering mechanisms A, B, C, ... are effective simultaneously and independent of one another, the individual scattering rates or the inverse mean free paths can be summed. The effective inverse mean free path l^{-1} is thus obtained the following:

$$\frac{1}{l} = \frac{1}{l_A} + \frac{1}{l_B} + \frac{1}{l_C} + \dots \tag{7.23}$$

In other words, the process that yields the greatest scattering, i.e. producing the shortest mean free path, dominates the thermal resistance.

7.3.3 Phonon-Phonon Scattering

In Section 7.2 we already discussed three-phonon processes in connection with ultrasonic attenuation and it is clear that these inelastic processes must play a similar role in the thermal resistance. As with ultrasonic damping, the energy and quasi-momentum must be conserved:

$$\hbar\omega_1 \pm \hbar\omega_2 = \hbar\omega_3 \,, \tag{7.24}$$

$$\hbar\mathbf{q}_1 \pm \hbar\mathbf{q}_2 = \hbar\mathbf{q}_3 + \hbar\mathbf{G} \,. \tag{7.25}$$

Depending on the sign, a phonon is generated or annihilated by the scattering process. As we saw in Section 6.3 when discussing the inelastic scattering of waves in crystals, a reciprocal lattice vector $\mathbf{G}$ can crop up in the quasi-momentum conservation equation. In (7.10) this was neglected, since reciprocal lattice vectors are large compared to the wave vector $\mathbf{q}_{us}$ of the ultrasonic phonons and thus do not contribute to the momentum balance. However, as we will see, the reciprocal lattice vector plays a major role in heat transport. Depending on whether such a reciprocal lattice vector is involved in the scattering process or not, we speak of **umklapp** or **U-processes** and **normal** or **N-processes** respectively. Figure 7.15 illustrates these processes for the case of phonon annihilation. If the wave vectors of the involved phonons are relatively small, the scattering process takes place within the first Brillouin zone. The sum of the quasi-momenta of all phonons involved is preserved. This also applies to the reverse process, in which a thermal phonon decays into two phonons. However, if the resulting wave vector $\mathbf{q}_3$ ends outside the first Brillouin zone, then by adding a reciprocal lattice vector $\mathbf{G}$, the resulting wave vector $\mathbf{q}_3'$ can lie within the first Brillouin zone. This changes the sum of the quasimomenta of the phonons involved.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

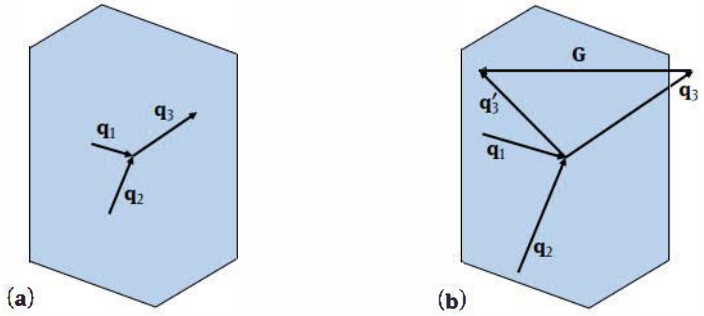


Fig. 7.15: Three-phonon process. **a)** Normal process: All wave vectors lie within the first Brillouin zone (shown in blue). **b)** Umklapp process: The values of the wave vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ have been doubled compared to those of the left-hand figure. Therefore the resulting vector $\mathbf{q}_3$ falls outside the first Brillouin zone, and thus an umklapp process occurs.

As mentioned in the previous section, there exist *selection rules* for phonon-phonon scattering. They have the effect that essentially only phonons with different polarizations and thus of different velocities are involved in this type of scattering process. However, the effects of these "subtleties" are not important for the following qualitative discussion.

Normal Process. In N-processes, the sum $\mathbf{P}$ of the quasimomenta of the phonons involved is conserved, i.e,

$$\mathbf{P} = \sum_{\mathbf{q}} N_{\mathbf{q}} \hbar \mathbf{q} = \text{const.} \tag{7.26}$$

where $N_{\mathbf{q}}$ is the number of phonons with wave vector $\mathbf{q}$. Since neither the momentum flow, nor the associated energy transport, is affected by these scattering processes, they do *not* contribute to the thermal resistance. If there were N-processes alone, a distribution of "hot" phonons with a total momentum $\mathbf{P}$, different from zero, would pass through the sample with no change in $\mathbf{P}$, i.e. the thermal conductivity would be infinitely large. For there to be a finite thermal resistance, there must be processes which oppose the conservation of the phonon momentum flow.

Umklapp Process. In the case of umklapp processes, the momentum conservation is completely different despite three phonons also being involved. A prerequisite for the umklapp process is that after scattering, the wave vector of the resulting phonon has to end outside the first Brillouin zone. Figure 7.16 illustrates the consequences of such a scattering process with an example. The two transverse phonons T_1 and T_2 with wave vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ scatter and produce a longitudinal phonon L with wave vector $\mathbf{q}_3$ in the second Brillouin zone. The crucial factor here is that the group velocities of the scattered phonons and the newly created phonon have opposite signs. The new phonon no longer transports energy in the original direction, but in the *opposite* direction.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

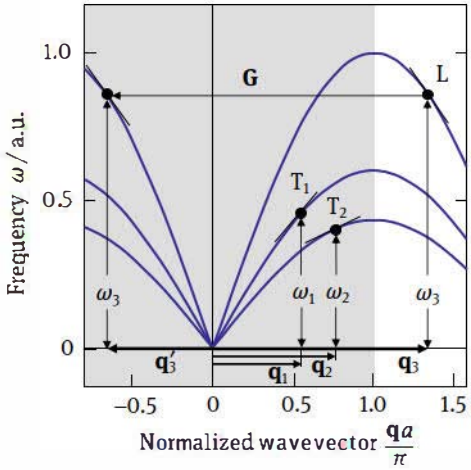


Fig. 7.16: Umklapp process. The two phonons with wave vectors q_1 und q_2 scatter and generate a phonon with wave vector q_3. By adding the reciprocal lattice vector $G = -2\pi/a$ the wave vector q_3' can be relocated in the first Brillouin zone. The sign of the group velocity, indicated by the corresponding slopes, thus reverses in this scattering process.

At *high temperatures*, umklapp processes dominate, since the majority of the excited phonons have frequencies comparable to the Debye frequency ω_D, meaning that their wave vectors are close to the boundary of the first Brillouin zone. Thus practically every scattering process leads to a momentum state outside the first Brillouin zone and is therefore an umklapp process. Here, we can directly apply the discussion on frequency and temperature dependence from Section 7.2 for the three-phonon process in ultrasonic experiments. If we evaluate the integral (7.15) for high temperatures and replace the ultrasonic frequency ω_{us} by the frequency ω_D of the Debye phonons, we find $l^{-1} \propto T$. Since at high temperatures the specific heat is approximately constant, the thermal conductivity, following equation (7.21), should decrease proportionally with the inverse temperature. In fact, for $T > \Theta$ the prediction

$$\Lambda \propto \frac{1}{T} \tag{7.27}$$

is indeed in good agreement with the observations.

At *intermediate temperatures* $T \leq \Theta$ the number of phonons with momenta sufficient to undergo an umklapp process depends strongly on the temperature. For the resulting wave vector of a scattering process to fall outside of the first Brillouin zone, as a rough approximation the energy of the involved phonons should obey $\hbar\omega > \hbar\omega_D/2$. The probability of finding phonons with this energy depends on the distribution function. For $\hbar\omega_D > k_B T$ we find the approximate relationship: $\langle n(\omega, T)\rangle = [\exp(\hbar\omega_D/2k_BT) - 1]^{-1} \approx \exp(-\hbar\omega_D/2k_BT) = \exp(-\Theta/2T)$.

On the other hand, the scattering cross section and the specific heat depend only relatively weakly on temperature. Neglecting these effects and considering only the temperature dependence of the mean free path, for $T < \Theta$ we can therefore write:

$$\Lambda \propto e^{\Theta/2T}. \tag{7.28}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As can be seen in Figure 7.14, below half the Debye temperature ($\Theta_{NaF} \simeq 445$ K) the thermal conductivity actually increases exponentially with decreasing temperature. However, we will not undertake a quantitative analysis of the temperature dependence here, but instead turn to low-temperature behavior.

7.3.4 Influence of Defects

At low temperatures, umklapp processes die out because the decreasing phonon frequencies ensure that any scattering takes place completely within the first Brillouin zone. However, since experiments show that the thermal conductivity decreases at low temperatures, there must be another scattering mechanism which limits the heat transport. The origin of this heat transport resistance is the surface where phonons are repeatedly scattered on their way from one end of the sample to the other. Here, their total momentum changes in a similar way to that in umklapp processes. In the equation for the thermal conductivity, the sample diameter d becomes the determining length scale for the mean free path. If we insert $l \approx d$ in equation (7.21), we obtain

$$\Lambda \approx C_V v d \propto T^3 d \,. \tag{7.29}$$

This temperature range, in which the thermal conductivity depends on the sample geometry, is known as the **Casimir[10] regime**. The influence of the sample dimensions is clearly shown in Figure 7.17a, which displays the thermal conductivity of various LiF crystals with various cross sections. As expected, the thermal conductivity decreases as the sample cross section is reduced.

In the Casimir regime, the effective mean free path also depends on the surface condition. With well-polished surfaces, specular reflection occurs where the phonon momentum component parallel to the surface does not change. The effective mean free path is therefore larger than the sample thickness. The crystal dimension therefore only gives the order of magnitude of the mean free path. The influence of sample preparation is shown in Figure 7.17b. For the first measurement the surface of the silicon crystal was very well polished. It was then roughened by sandblasting and chemical etching and measured again. The result for the polished crystal is significantly different from that of the roughened crystal in the 1 K – 20 K range. Below 1 K the thermal conductivities differ by a factor of about 500. In this temperature range, the effective mean free path of the polished sample proves to be extremely large, turning out to be about 7 cm and mainly determined by the *length* of the sample. This means that very well-polished surfaces hardly impede heat transport.

Which influence do the point defects or dislocations discussed in Chapter 5 have on the thermal conductivity? It turns out that both types of defects generally only contribute noticeably to the thermal resistance at very high concentrations. This does not

10 Hendrik Brugt Gerhard Casimir, *1909 The Hague, †2000 Heeze

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

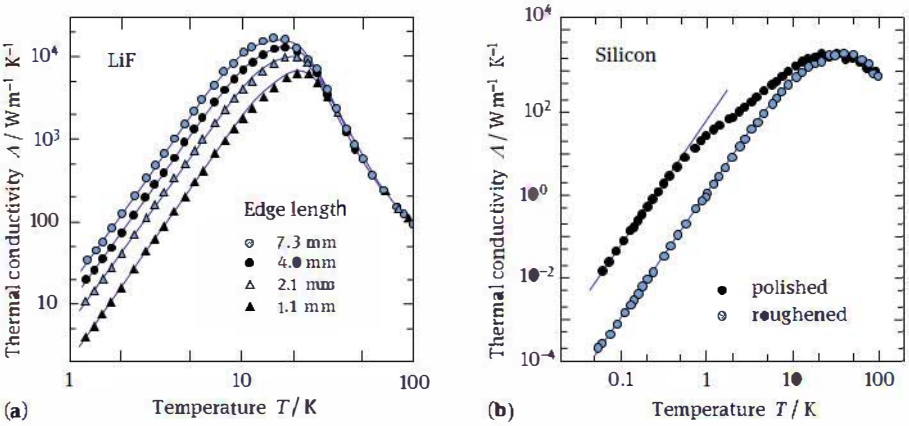


Fig. 7.17: **a)** Thermal conductivity of LiF crystals rods with almost perfectly quadratical cross sections as a function of temperature for various edge lengths. (After P.D. Thacher, Phys. Rev. 156, 975 (1967).) **b)** Thermal conductivity of a silicon crystal with very well-polished or roughened surface. (After V. Röhring, private communication.)

apply to crystals with defects that interact resonantly with phonons. When discussing the ultrasonic damping of amorphous solids, we have seen that this interaction can very effectively hinder phonon propagation. As we will see in the next section, this interaction determines the thermal resistance of these materials at low temperatures. Here we consider point defects, which act as *geometric* barriers for phonons and thus give rise to Rayleigh scattering. According to (7.12) the mean free path is then given by $l^{-1} \propto \omega^4$. Since, according to the dominant phonon approximation $\bar{\omega} \propto T$, we can write $l \propto T^{-4}$ for the mean free path. Thus point defect scattering is particularly effective at high temperatures, but in this temperature range in general the phonon-phonon interaction arising from the large number of excited phonons dominates. Nevertheless, the influence of point defect scattering on heat conduction can be observed when umklapp processes are still relatively rare, i.e. at temperatures near the maximum of the thermal conductivity.

A particularly impressive special case of point defect scattering is **isotope scattering**, where the periodicity of the crystal is degraded simply by the difference in mass between the two isotopes. The effects of this scattering process can be seen in Figure 7.18a. Starting from an almost isotopically pure ^{7}LiF crystal, by adding ^{6}Li the thermal conductivity is reduced by about a factor of ten at maximum! The small differences in the Casimir regime were caused by the slightly different sample cross sections.

Measurements of the thermal conductivity on an isotopically pure ^{28}Si crystal and on a silicon crystal with the composition of the natural isotopic abundance are shown in Figure 7.18b. With $\Lambda = 2.9 \times 10^4$ W/m K at the maximum at 26.5 K the conductivity is extremely high. The surprisingly sharp change of slope at the maximum is caused by the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

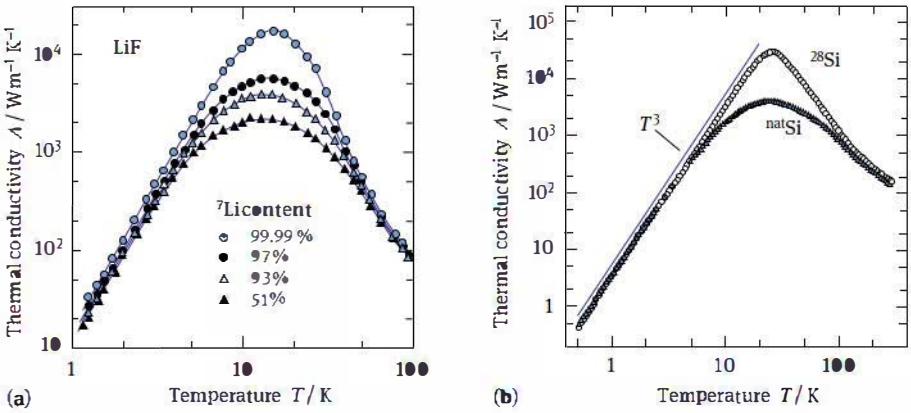


Fig. 7.18: **a)** Thermal conductivity of LiF for various $^6Li/^7Li$ ratios. (After P.D. Thacher, Phys. Rev. 156, 975 (1967).) **b)** Thermal conductivity of isotopically pure ^{28}Si and natural silicon. (After A.V. Inyushkin et al., phys. stat. sol. (c), 1, 2995 (2004).)

relatively abrupt transition from surface scattering to umklapp scattering. In contrast to this, the measured curve for crystals of natural silicon is much flatter at the maximum, where the conductivity is strongly reduced by isotope scattering. The highest value of the thermal conductivity coefficient of any solid was measured with $\Lambda = 4.1 \times 10^4$ W/m K on an isotopically pure ^{12}C diamond at 104 K.

7.3.5 Heat Transport in One-dimensional Samples

New interesting phenomena occur in heat transport when the lateral dimensions of the samples become so small that they can be regarded as one-dimensional structures. Such experiments have become possible in recent years owing to advances in the field of micro- and nanofabrication technology. In this context, we describe an experiment to investigate the heat flow from a resistive heater through four thin connecting bars. An image of the device taken with a scanning electron microscope is shown in Figure 7.19. In the center of the left picture, a 4 µm × 4 µm "island" consisting of a 60 nm thick, free-floating silicon nitride membrane can be seen, supported on four narrow strips between the dark regions indicating areas where the membrane has been completely removed. The bright areas on the central island are gold films serving as heaters and thermometers. These are connected to the contact pads (not shown in the picture) by thin niobium wires that have been vapor-deposited on the thin bars.[11] The right-hand image shows an enlargement of one of the funnel-shaped connecting bars, whose

11 As we will see in Chapter 11, niobium becomes superconducting at low temperatures and therefore has a negligible thermal conductivity, thus not affecting the heat flow via the connecting bars.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

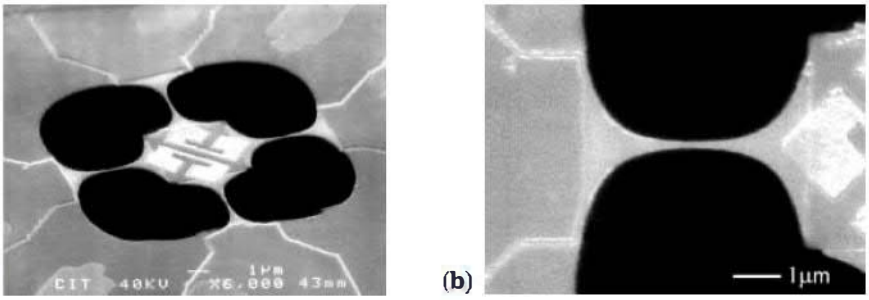


Fig. 7.19: SEM images of the experimental arrangement for investigating the heat transport in one-dimensional samples. **a)** In the center can be seen the freely floating 4 μm × 4 μm island based on a silicon nitride membrane, suspended from four thin bars. In the kidney-shaped dark areas indicate where the membrane has been completely etched away. The two gold films in the center serve as heaters and thermometers. They are connected to the contact pads by thin niobium wires. **b)** Close-up of a connecting bar with 200 nm width at the narrowest point. (After K. Schwab et al., Nature **404**, 974 (2000).)

width w is about 200 nm at the narrowest point. The special shape of the connecting bars was chosen to ensure the most effective coupling of the phonons.

Before we discuss the experimental results, let us take a look at the theoretical prediction. For this purpose, we imagine that a "warm" thermal reservoir on the left-hand side of a one-dimensional sample of length L is connected to a "cold" heat sink on the right-hand side. The energy flow J in the sample is

$$J = \frac{1}{L} \sum_{\mathbf{q}} \hbar \omega_{\mathbf{q}} \, v_{\mathbf{q}} \,, \tag{7.30}$$

where $v_{\mathbf{q}}$ is the phonon velocity. The summation is carried out over all thermally excited phonons. The energy flow consists of two components: that carried by those phonons with $q > 0$ coming from the "warm" reservoir on the left-hand side, and those phonons with $q < 0$ propagating from the "cold" sink in the opposite direction. We simplify the calculation by replacing the summation by an integration to get:

$$J = \sum_i \frac{1}{L} \int_0^\infty \varrho_i^{(1)} \hbar \omega_i \, v_{g,i} \left[\langle n_w(\omega, T) \rangle - \langle n_c(\omega, T) \rangle \right] \mathrm{d}q \,, \tag{7.31}$$

where the index i indicates the particular phonon branch, $\varrho_i^{(1)}$ the density of states (6.77) in one-dimensional $\mathbf{q}$ space and $v_{g,i}$ the group velocity of the phonons. The Bose-Einstein distribution $\langle n(\omega, T) \rangle = (\exp(\hbar\omega / k_B T - 1)^{-1}$ reflects the thermal occupation of the lattice vibrations, where the indices "w" and "c" indicate the "warm" and "cold" thermal baths, respectively. The integration is taken only over positive q values, since we have already taken into account the two directions of phonon propagation by the signs of the two terms. Furthermore, we have set to unity the transmission coefficient, which characterizes the coupling between the connecting bars and the heat reservoirs.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

We set $\varrho_i^{(1)} = L/2\pi$ and make a change in the variable from the wave number q to the frequency ω. This introduces the factor $\partial q/\partial\omega$, which, fortuitously, compensates for the group velocity $v_g = \partial\omega/\partial q$. Assuming that the temperature difference ΔT between the reservoirs is small, we may expand $[\langle n_w(\omega, T)\rangle - \langle n_k(\omega, T)\rangle]$ into a Taylor series and cut off after the linear term. With the usual abbreviation $x = \hbar\omega/k_B T$, the calculation for the **thermal conductance** G results in the expression:

$$G = \frac{J}{\Delta T} = \frac{k_B^2 T}{h} \sum_i \int_0^\infty \frac{x^2 e^x}{(e^x - 1)^2}\, dx = \sum_i \frac{\pi^2}{3} \frac{k_B^2 T}{h} = N_i G_0 \,. \tag{7.32}$$

Since the connecting bar can perform four different vibrational modes, namely one dilatational, one torsional and two bending modes, $N_i = 4$. It is remarkable that under ideal conditions, regardless of the sample dimensions, each vibrational mode makes the same contribution to the conductance, namely

$$G_0 = \frac{\pi^2 k_B^2 T}{3h} = \left[9.46 \times 10^{-13}\, \frac{W}{K^2}\right] T \,. \tag{7.33}$$

The data in Figure 7.20 confirm the theoretical prediction. In this figure, the thermal conductance of the four connecting bars is plotted as a function of temperature. The measured conductance was normalized to $16G_0$, since four vibrational modes can be excited in each of the four legs. At higher temperatures, i.e. above about 1 K, the bars behave as three-dimensional samples. As expected for the Casimir regime, the thermal conductance is then proportional to T^3. From the data in this temperature range we find an effective mean free path of $l_{eff} \approx 0.9\,\mu m$.

In considering a long sample with one-dimensional properties, what role does the transverse dimension play. We will discuss this question in detail in Section 8.1 in connection with the electrical properties of metals. Here we simply give the plausible

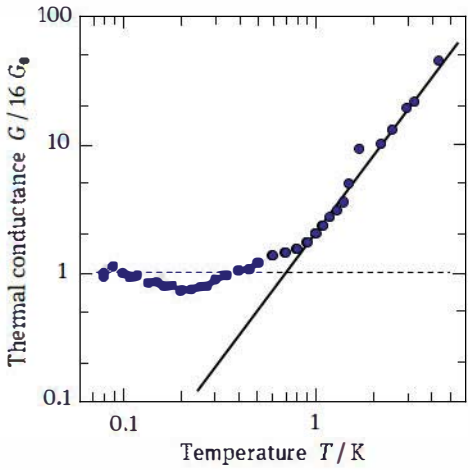


Fig. 7.20: The thermal conductance of the four connecting bars as a function of temperature. The measured conductance is normalized to $16G_0$. Below 0.8 K the connecting bars behave as one-dimensional samples. As expected, the thermal conductance increases as T^3 at higher temperatures. (After K. Schwab et al., Nature **404**, 974 (2000).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

explanation that the system behaves one-dimensionally if the wavelength λ_{th} of the thermally excited phonons is larger than the transverse dimension of the sample. In our case, this means that the width w of the connecting bars must satisfy $w < \lambda_{th}/2$. This results in a value $T_{co} \approx hv/(2wk_B)$ for the temperature of the transition to one-dimensional behavior.

If we use the values $w \approx 200$ nm and $v \approx 6000$ m/s for the width w and the mean sound velocity v, we obtain the transition temperature $T_{co} \approx 0.8$ K. In fact, the temperature dependence of the thermal conductance flattens out with decreasing temperature at 1 K and approaches the value $16\,G_0$ as expected.

7.4 The Thermal Conductivity of Amorphous Solid

The thermal conductivity of amorphous solids differs fundamentally from that of crystals. This is clearly shown by the comparison between vitreous silica (quartz glass) and crystalline quartz in Figure 7.21, where the curves are typical of these two classes of material. In pure dielectric crystals the conductivity passes through a maximum, as discussed in Section 7.3. In amorphous solids, on the other hand, the thermal conductivity decreases steadily with decreasing temperature falling significantly below that of the corresponding crystals. As indicated in Figure 7.21, three temperature ranges can be distinguished for glasses, which we will briefly discuss below using the dominant phonon approximation.

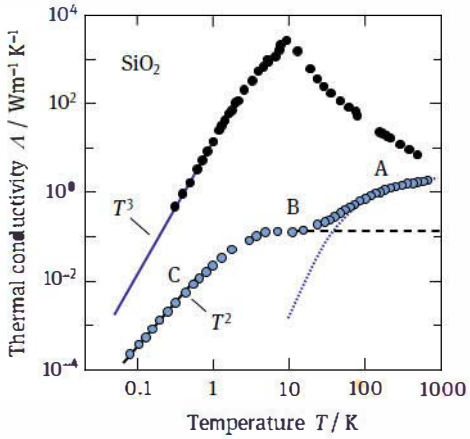


Fig. 7.21: Comparison of the thermal conductivity of quartz crystal (black points) and vitreous silica (light blue points). For vitreous silica, the subdivision into three temperature ranges A, B and C is indicated. (After R.C. Zeller, R.O. Pohl, Phys. Rev. B 4, 2029 (1971).)

High Temperatures. At high temperatures, the frequency of the dominant phonons is around 10^{13} Hz and their wavelength is 10 Å. From the thermal conductivity data, values of about 5 Å result for the mean free path meaning that the wavelength and

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the mean free path are comparable. The phonons, if they exist at all in amorphous solids at these wavelengths, are overdamped and can no longer defined as elementary excitations, and thus any description of heat transport in the phonon picture is not meaningful under these circumstances. The idea that the vibrational energy diffuses from atom to atom is more suitable, as *A. Einstein* already suggested in 1911 (but for crystals). Only recently have theories of heat transport in amorphous materials been developed for the higher temperature range and corresponding numerical simulations have been carried out. However, so far no final explanation has been arrived at.

Intermediated Temperatures (Plateau Range). In the temperature range between 1 K and 10 K for vitreous silica a *plateau* in Λ is observed. This phenomenon has nothing to do with the maximum in the thermal conductivity of the crystals, although the data in Figure 7.21 might suggest this. While the position of the maximum depends on the dimension of the sample in crystals, the shape of the vitreous sample has no influence on the position of the plateau. Since the thermal conductivity is almost independent of temperature over the extent of the plateau, the mean free path of the phonons there must be highly dependent on temperature or frequency to compensate for the rapid increase in the phonon contribution to the specific heat. Various mechanisms have been proposed to explain the strongly varying mean free path in this temperature range: e.g. scattering by point defects, spatial localization of phonons, or scattering by *soft vibrational states* (see Section 6.5). However, no explanation has yet been generally accepted.

Low Temperatures. Two properties characterize the thermal transport of glasses below 1 K: the temperature dependence of the thermal conductivity is quadratic and the conductivities of different amorphous substances are similar to one another. As we have seen in Section 6.5, long-wavelength phonons also exist in amorphous solids. At low temperatures, as in crystalline dielectrics, they are responsible for thermal transport, but with the usual sample geometries the Casimir range is not reached even at the lowest temperatures. Phonon scattering in this temperature range does not occur at the surface as in crystals, but within the sample by the resonant interaction between phonons and tunneling centers, which we discussed in Section 7.2 for the case of ultrasonic propagation. There we obtained the expression (7.19) for the inverse mean free path:

$$l^{-1} = \delta n\, \sigma = \tilde{C}\omega \tanh \frac{\hbar\omega}{2k_B T} \,. \tag{7.34}$$

As emphasized there, in amorphous solids resonant interaction can occur for all phonon frequencies because the energy splitting of the tunneling centers is uniformly distributed. To understand the temperature dependence of the thermal conductivity, we also consider the behavior of the dominant phonons with the frequency $\bar{\omega}$, to which $\hbar\bar{\omega} \propto k_B T$ applies. The occupation number difference $\tanh(\hbar\bar{\omega}/2k_B T)$ of the two-level systems, which interact with the dominant phonons, is thus constant. For the inverse mean free path, there follows the simple relationship $l^{-1} \propto \bar{\omega} \propto T$. Since the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

specific heat of the phonons according to the Debye model is proportional to T^3 at low temperatures, we find:

$$\Lambda = \frac{1}{3} Cvl \propto T^2 \tag{7.35}$$

in good agreement with the experimental results. From the known velocities of the longitudinal and transverse sound waves, the specific heat of the phonons can be calculated using the Debye modeL If these numerical values are used in equation (7.21), there is good agreement between the theoretical calculation and the experimental value.

7.5 Exercises and Problems

1. Griineisen Parameter. Potassium iodide has NaCl structure. Estimate the Grüneisen parameter of KJ by using the lattice constant $a = 7.06$ Å and the thermal expansion coefficient $\alpha_V = 1.23 \times 10^{-4}\,\mathrm{K}^{-1}$.

2. Three-phonon Processes. Consider the following three-phonon processes in an isotropic, crystalline solid with a monatomic basis: *i*) T ↔ L, L, *ii*) T ↔ T, L, *iii*) T ↔ T, T, *iv*) L ↔ L, L, *v*) L ↔ T, L *vi*) L ↔ **T, T**, where T indicates transverse branch and L longitudinal branch.

(a) Which processes are actually allowed in the framework of the Debye approximation, taking into account energy and quasimomentum conservation?

(b) Do the selection rules for non-linear dispersion curves change?

3. Ultrasound Damping by Point Defects. Estimate the density of point defects that still cause measurable attenuation of sound waves by Rayleigh scattering at frequencies at 1 MHz, 1 GHz and 1 THz. For this discussion, take the parameters for germanium, i.e. lattice constant $a = 5.66$ Å and transverse sound velocity $v = 2420$ m/s.

4. Thermal Conductivity of Germanium. What is the approximate thermal conductivity of a cylindrical germanium sample of 3 mm diameter at 1 K?

5. Thermal Conductivity in the Casimir Limit. Use Figure 7.17a to estimate the Debye temperature of LiF. LiF crystals have the same structure as sodium chloride and have a density of $\varrho = 2640\,\mathrm{kg/m^3}$.

6. Thermal Conductivity at Low Temperatures. The lower ends of three 1 cm long cylinders made of silicon, vitreous silica and copper respectively are firmly connected to a thermal bath at $T_0 = 0.5$ K. The upper ends of each are supplied with a heat input of 10 μW, resulting in the same temperature difference of 1 mK for each sample.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

At 0.5 K, the appropriate coefficients of thermal conductivity are: $\Lambda_{Si} = 1 \times 10^{-2}$ W/cm K, $\Lambda_{a\text{-}SiO_2} = 5 \times 10^{-5}$ W/cm K and $\Lambda_{Cu} = 4$ W/cm K.

(a) What is the cross-section of each sample?

(b) When the heating power is increased to 10 mW, the temperature gradients change differently. In which sample is the temperature gradient the largest, and in which the smallest?

(c) Now the experiment is repeated at 5 mK. What heating power is required to produce a temperature difference of 500 μK?

7. Mean Free Path of Thermal Phonons in Vitreous Silica. Estimate the mean free path at $T = 5$ K for thermal phonons in vitreous silica (density $\varrho = 2.20$ g/cm^3) from the thermal conductivity and sound velocity data.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

8 Electrons in Solids

In the previous two chapters, we discussed the motion of atoms about their equilibrium position. The electrons did not play a role here, because they are "instantaneously" able to follow the motions of the atomic nuclei. In this chapter, we will specifically focus on the electrons and study the electronic properties of solids. Given that the lattice moves very slowly compared to the electrons, we once again make use of the adiabatic approximation. The effects of the lattice vibrations are subsequently taken into account as *electron-lattice interactions* when discussing transport phenomena.

We assume that the electrons are located in a quasi-rigid lattice and pick out *one* electron. In crystals, the effective potential $\widetilde{V}(\mathbf{r})$ in which this electron moves is constant in time and periodic in space. This effective potential is generated by all the other electrons and the atomic cores which we assume are in their equilibrium positions. The interactions between the atomic cores as well as the interactions between the core electrons and the valence electrons is already taken into account. This approach, where only the behavior of *one* electron is investigated and make the approximation that the remaining charges only contribute to a single effective potential, is called the **one-electron approximation**. Although in this approximation the electrons do not interact directly with each other, they are subject to the Pauli principle, which states that two electrons cannot occupy the same quantum mechanical state. Correlations between the electrons, as happens in magnetism or superconductivity, are not considered here.

We discuss the fundamental questions of the electron motion in two stages. First, we deal with properties where the influence of the periodical lattice potential can be neglected. We then take a brief look at two phenomena in which *collective behavior* of electrons within the electron gas plays an important role. In the following sections we will then discuss the effects of the influence of the periodic structure of crystals on the electronic properties. In Chapter 9 we deal with the electronic transport properties of solids, e.g. the electrical and thermal conductivity of metals.

8.1 Free Electron Gas

We start with metals whose electronic properties can be attributed surprisingly well to the behavior of their *free electrons*. In this simple approximation, the electrons are assumed to move in a *constant potential*, with only at the edge of the sample a potential barrier confining the electrons to the solid. This box-potential is sketched in Figure 8.1. An electron with energy E is bound to the solid if $E < W$ where W is the depth of the potential well. At zero temperature the potential well is filled up to the "Fermi energy" E_F (see Section 8.1.2), which denotes the maximum energy of the bound electrons. Thus the minimum energy required to remove an electron from the metal is the *work function* $\Phi = (W - E_F)$. This energy plays an important role in the

https://doi.org/10.1515/9783110666502-008

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

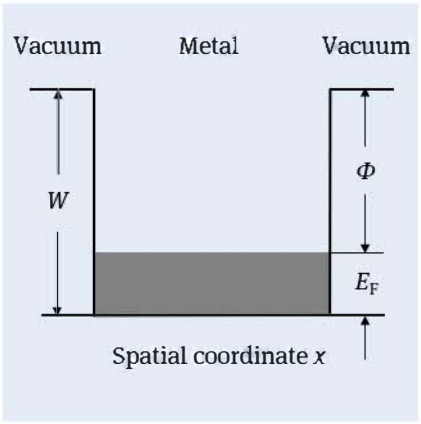


Fig. 8.1: The simple potential well assumed in the free electron model. The depth of the well is W, the work function Φ and the Fermi energy E_F.

photoelectric effect and the thermionic emission, but is of no importance here in our further discussion.

We speak of an **electron gas** or **Fermi gas**[1] if the conduction electrons behave, to a good approximation, as a classical gas. However, there is an important difference: electrons are fermions and therefore subject to the Pauli principle. This extremely simple, but successful approximation goes back to *A. Sommerfeld*[2] and is therefore called the **Sommerfeld theory**. At first sight the theory seems to be a very crude simplification of the actual situation. However, when discussing the metallic bond in Section 2.6, we saw that valence electrons avoid the atomic cores. This fact was taken into account by introducing the *pseudopotential*, which reflects the fact that the valence electrons experience only a weakly varying potential. In other words, the conduction electrons do not "see" the "bare " Coulomb potential of the atomic cores but a much weaker modulated pseudopotential. Figure 8.2 illustrates the two potential landscapes which are not based on the simple pseudopotential of Figure 2.13, but on one closer to the real situation. The right-hand edge of the image was chosen to run between the atomic cores and the left-hand edge along the sites of the atomic nuclei, where the modulation is the strongest. It is remarkable that in the case of the Coulomb potential, the potential landscape shows, as expected, a minimum at the locations of the atomic nuclei, whereas in the case of the pseudo-potential, a small hump appears at these locations.

It turns out that this considerable simplification leads to very good results with the alkaline metals such as lithium, potassium and sodium and other simple metals such as copper, silver and gold. In these materials, the only electrons present besides the delocalized *s*-shell electrons are those in closed shells. These closed-shell electrons play a minor role in determining the electronic properties of solids. On the other hand,

1 Enrico Fermi, *1901 Rome, †1954 Chicago, Nobel Prize 1938
2 Arnold Johannes Wilhelm Sommerfeld, *1868 Königsberg (Kaliningrad), †1951 Munich

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

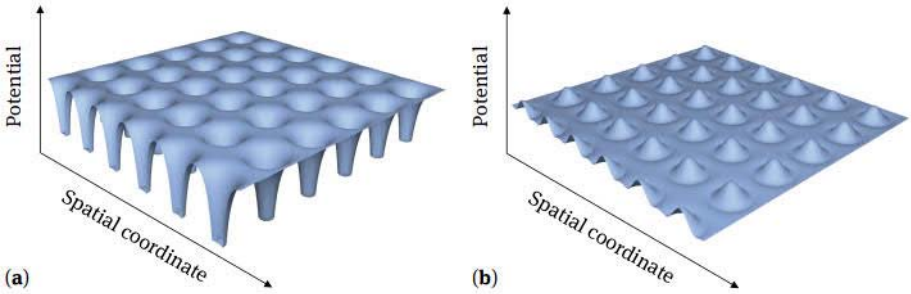


Fig. 8.2: Illustrative representation of the potential landscape. The edges of the image are chosen so that one runs through the atomic nuclei, the other between the atomic cores. **a)** Coulomb potential, **b)** pseudo potential.

the model works less well for more complex metals. For many transition metals. the assumption of quasi-free electrons is only partially fulfilled, since in addition to the *s*-electrons, there are also *d*- and/or *f*-electrons which occur in partially filled shells with partly overlapping orbitals, largely losing the characteristic properties of a free electron gas.

8.1.1 Density of States

As the first step we derive the density of states of the free electron gas. We will not only deal with three-dimensional systems, but also with two- and one-dimensional systems. We need their density of states in our treatment of the free electron gas. With the help of the density of states, we can calculate the specific heat of the free electron gas, whose small experimentally observed room-temperature value was a mystery for a long time until it was explained by *A. Sommerfeld*.

As we are assuming a constant, position-independent potential, we can ignore any anisotropy of the crystals, and we can therefore proceed in the same way as we did when considering phonons in an elastic-isotropic medium. When counting the allowed states, we assume a cube with edge length L, containing N free electrons. We set the zero point of the potential so that at the bottom of the box-potential $\bar{V} = 0$, ensuring that the electrons only have kinetic energy, and thus the stationary Schrödinger equation for an electron takes the simple form:

$$-\frac{\hbar^2}{2m}\Delta\psi(\mathbf{r}) = E\,\psi(\mathbf{r})\,. \tag{8.1}$$

As a solution, we choose a plane wave for the wave function ψ

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{V}}\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \tag{8.2}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

where the wave is characterized by its wave vector **k**, and the volume of the cube V is used for normalization. With this approach, for the energy eigenvalues E of free electrons we obtain the simple solution:

$$E = \frac{\hbar^2 k^2}{2m} . \tag{8.3}$$

The number of allowed wave vectors **k** is limited by the boundary conditions. As in the calculation of the phonon density of state in Section 5.4, we use periodic boundary conditions

$$\psi(x, y, z) = \psi(x + L, y, z) = \psi(x, y + L, z) = \psi(x, y, z + L) . \tag{8.4}$$

Thus the components of the wave vectors are:

$$k_i = \frac{2\pi}{L} m_i \tag{8.5}$$

with $i = (x, y, z)$ and the integer quantum numbers m_i. As with phonons, the allowed wave vectors are evenly distributed in momentum space with density $\varrho_k' = V/(2\pi)^3$. Following the Pauli principle, each state can be occupied by no more than two electrons (one for each up/down spin orientation). Therefore, for the electronic density of states in momentum space, we find:

$$\varrho_k = \frac{2V}{(2\pi)^3} . \tag{8.6}$$

We calculate the density of states $\mathcal{D}(E)$ in energy space with (6.78) or (6.79) and write

$$\mathcal{D}(E)\,\mathrm{d}E = \varrho_k \int_E^{E+\mathrm{d}E} \mathrm{d}^3k = \frac{\varrho_k}{\hbar}\,\mathrm{d}E \int_{E=\text{const.}} \frac{\mathrm{d}S_E}{v_g} . \tag{8.7}$$

For the free electron gas, the group velocity $v_g = \partial E/\partial(\hbar k) = \hbar k/m$ of the electrons is independent of direction. Thus the surface of constant energy (in energy space) is a sphere with surface integral $\int \mathrm{d}S_E = 4\pi k^2$. Thus it follows that the density of states can be written:

$$\mathcal{D}(E) = \frac{2V}{(2\pi)^3\hbar} 4\pi k^2 \frac{m}{\hbar k} = \frac{V}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\sqrt{E} \tag{8.8}$$

and thus for the electronic density of state per volume $D(E) = \mathcal{D}(E)/V$ the important final result is

$$D(E) = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\sqrt{E} . \tag{8.9}$$

Although the density of states of electrons and phonons is the same in reciprocal space, the densities of states in energy space differ because of their different dispersion relations. The absolute value of the density of states is determined only by the energy and the electron mass. Figures 8.3a and 8.3b show the dispersion relation (8.3) and the density of states (8.9) for three-dimensional samples. The occupied states are indicated

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

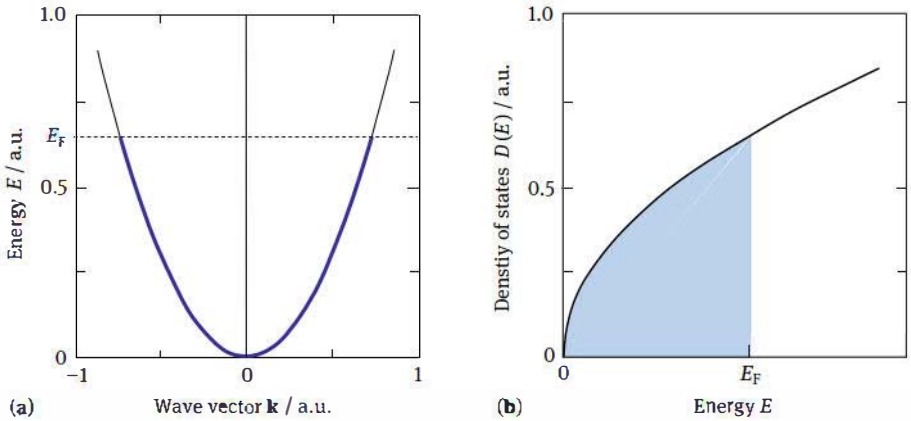


Fig. 8.3: The free electron gas in three dimensions. **a)** The energy dispersion curve $E(k)$. Occupied states are drawn in blue, unoccupied states in grey. **b)** The density of states. At $T = 0$ all states up to the Fermi energy E_F are occupied, above E_F all states are empty.

in both figures. We will return to the distinction between occupied and unoccupied states shortly.

Now we take a look at low-dimensional electron systems to see how the energy spectrum $E(k)$ and density of states $D(E)$ change with dimensionality. In Section 6.4 we already mentioned that the density of states in momentum space depends on the dimension of the system under consideration. According to (6.77), for isotropic systems we can write:

$$\varrho_k^{(\alpha)} = 2\left(\frac{L}{2\pi}\right)^{\alpha} \tag{8.10}$$

where L is the characteristic length, α is the dimensionality of the system concerned, and the factor 2 takes into account the spin states.

For two-dimensional systems, the integration over the surface in (8.7) is reduced to a line integral, which in the case of isotropic k-space yields the circumference of the circle $2\pi k$. Taking into account the two possible spin states, we find for the density of states $D^{(2)}$ per surface area A the following expression:

$$D^{(2)}(E) = \frac{\varrho_k^{(2)}}{A\hbar}\frac{2\pi k}{v_g} = \frac{m}{\pi\hbar^2}\,. \tag{8.11}$$

The density of states of a two-dimensional electron gas is thus energy independent, i.e. constant!

As an example of a *two-dimensional system*, we consider a thin *metal film* whose thickness d in the z-direction is in the nanometer range. Once again we use plane waves as solutions to the Schrödinger equation. For simplicity, we define the boundary condition in the z-direction so that only standing waves with wavelengths $\lambda_z = 2d/j$ are allowed, where j is a positive integer. Making an educated guess at the wavefunction

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

in equation (8.1):

$$\psi(x,y,z) = \frac{1}{\sqrt{V}} \sin\left(\frac{j\pi z}{d}\right) e^{ik_x x} e^{ik_y y} \tag{8.12}$$

we obtain the eigenvalues

$$E = \frac{j^2 h^2}{8md^2} + \frac{\hbar^2 k^2}{2m} = E_j + \frac{\hbar^2 k^2}{2m}\,, \tag{8.13}$$

where the wave vektor **k** lies in the xy-plane and its magnitude is given by $k^2 = (k_x^2 + k_y^2)$. For layer thicknesses in the nanometer range, the *transverse energy* E_j (i.e. the energy associated with the transverse waves) is relatively large owing to its $1/d^2$ dependence. The energy spectrum is therefore discrete with respect to the z-coordinate, but quasi-continuous in the x- and y-directions. Figure 8.4a shows the electron energy as a function of the wave vector **k**. For samples with small thicknesses in the z-direction, the parabolic dispersion curve for the free electrons splits into *sub-bands*, with different eigenvalues for the same wave vector **k** by $\Delta E_j = (E_j - E_{j-1})$.

The total electronic density of states $D(E)$ shown in Figure 8.4b, is the sum of the densities of states of the subbands, which are given by equation (8.11). We therefore have:

$$D(E) = \sum_j D_j^{(2)}(E) \tag{8.14}$$

with

$$D_j^{(2)}(E) = \begin{cases} \dfrac{m}{\pi\hbar^2} & \text{for} \quad E \geq E_j\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{8.15}$$

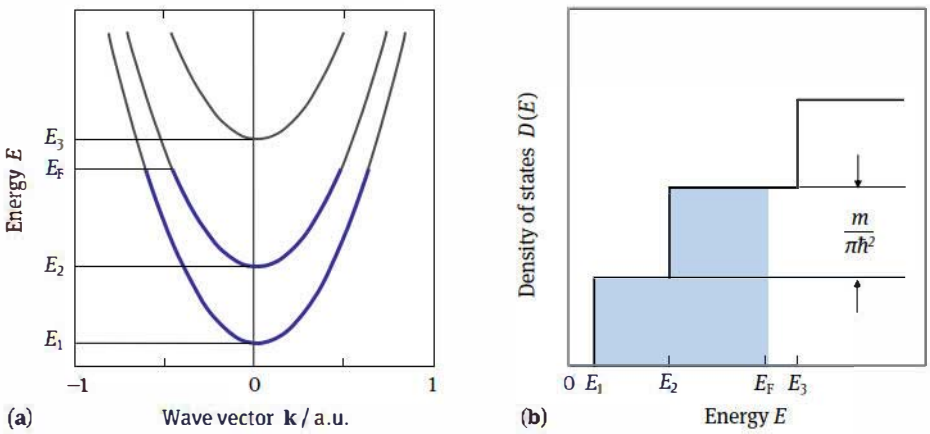


Fig. 8.4: **a)** The dispersion relation and **b)** the density of states for a two-dimensional electron gas. Occupied states are highlighted in blue.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Proceeding in the same way for one-dimensional systems, we find for the density of states $D^{(1)}(E)$ per length L:

$$D^{(1)}(E) = \frac{\varrho_k^{(1)}}{L\hbar}\frac{2}{v_g} = \frac{1}{\pi\hbar}\sqrt{\frac{2m}{E}} \,. \tag{8.16}$$

As an example, let us consider a thin wire with a rectangular cross-section. We assume that the dimensions of the cross-section fall in the nanometer range, while the wire is extended in the x-direction. For the wave function we choose an approach analogous to (8.12), taking into account the boundary conditions in the y- and z-directions:

$$\psi(x, y, z) = \psi_{i,j}(y, z)\mathrm{e}^{\mathrm{i}k_x x} \,. \tag{8.17}$$

This results in the eigenvalues:

$$E = E_{i,j} + \frac{\hbar^2 k_x^2}{2m} \,, \tag{8.18}$$

where the quantum numbers i and j denote the eigenstates in the yz-plane. Writing the transverse energies $E_{i,j}$ in this form implies very small wire cross-sections. For the density of states we find the following:

$$D(E) = \sum_{i,j} D_{i,j}^{(1)}(E) \tag{8.19}$$

with

$$D_{i,j}^{(1)}(E) = \begin{cases} \dfrac{1}{\pi\hbar}\sqrt{\dfrac{2m}{E - E_{i,j}}} & \text{for} \quad E \geq E_{i,j} \,, \\ 0 & \text{otherwise} \,. \end{cases} \tag{8.20}$$

The energy spectrum and the density of states are shown in Figures 8.5a and 8.5b. Note that the density of states diverges at the minima of each subband, i.e. there is a *Van-Hove singularity* at each of these points, which we have already discussed in Section 6.4 in connection with the density of states for lattice vibrations.
Having dealt with three-dimensional, two-dimensional and one-dimensional systems we can now reduce the third dimension. The resulting samples are called **quantum dots** or *quasi-zero-dimensional systems*. The spectrum is discrete and depends on the shape of the particular quantum dot. The density of states is also discrete being made up of a series of δ-functions.

8.1.2 The Fermi Energy

After the short excursion into low-dimensional systems, we return to three-dimensional samples. For ensembles of particles with half-integer spin, the Pauli principle has to be obeyed. This means that the occupation of the states is determined by Fermi-Dirac

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

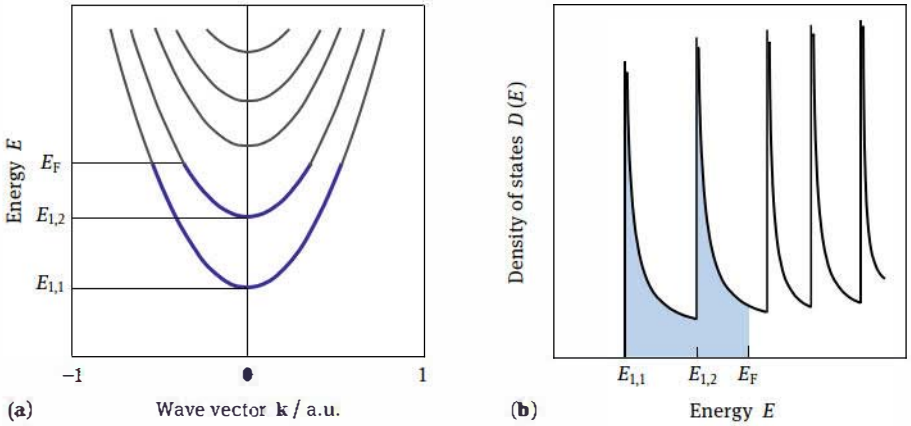


Fig. 8.5: a) Dispersion relation and **b)** density of states of an one-dimensional electron gas.

statistics.[3] The occupation probability is therefore governed by the **Fermi distribution**:

$$f(E) = \frac{1}{e^{(E-\mu)/k_B T} + 1} \,. \tag{8.21}$$

This distribution function reduces to the classical Boltzmann distribution if the probability of occupation of the states considered is very much smaller than one, i.e. when $[(E - \mu)/k_B T] \gg 1$.

The **chemical potential** μ, which appears in the distribution function expression defines the link between the free energy F and the particle number N according to:

$$\mu = \left(\frac{\partial F}{\partial N}\right)_{T,V} \,. \tag{8.22}$$

At absolute zero, the Fermi-Dirac distribution takes the values

$$f(E, T = 0) = \begin{cases} 1 & \text{for} \quad E < \mu \,, \\ \frac{1}{2} & \text{for} \quad E = \mu \,, \\ 0 & \text{for} \quad E > \mu \,. \end{cases} \tag{8.23}$$

At $T = 0$, all states are occupied for $E < \mu$, where according to the Pauli principle, only two electrons (with the two possible spin orientations) are allowed per state. The energy up to which the states are completely filled is called the **Fermi energy** E_F. Since the chemical potential indicates the smallest energy necessary for adding an electron to the Fermi gas, and since this can only happen at $T = 0$ with the Fermi energy, the chemical potential μ and the Fermi energy E_F are identical at this temperature, i.e. $E_F = \mu\,(T = 0)$.

3 Paul Adrien Maurice Dirac, *1902 Bristol, †1984 Tallahassee, Nobel Prize 1933

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The Fermi energy is defined by the electron density $n = N/V$: If we integrate over all occupied states, we simply get the particle number density, since

$$n = \frac{N}{V} = \int_0^\infty D(E) f(E, T = 0)\, dE = \int_0^{E_F} D(E)\, dE = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} \frac{2E_F^{3/2}}{3} . \tag{8.24}$$

If we solve for E_F, we see that the expression for the Fermi energy, apart from constants, only involves the mass and the number density of the electrons:

$$E_F = \frac{\hbar^2}{2m}(3\pi^2 n)^{2/3} \qquad \text{Fermi energy} . \tag{8.25}$$

Figure 8.3b already showed the parabolic curve of the density of states for three-dimensional samples. At absolute zero, the states up to the Fermi energy E_F are occupied with electrons, the states above are empty.

Since the Fermi energy is an important quantity in determining the behavior we normally define the following quantities at this energy; the **Fermi wave vector $\mathbf{k}_F$, Fermi momentum $\hbar\mathbf{k}_F$, Fermi velocity $\mathbf{v}_F$** and the **Fermi temperature** T_F of electrons can be defined as follows:

$$k_F = (3\pi^2 n)^{1/3} \qquad \text{Fermi wave vector,} \tag{8.26}$$

$$v_F = \frac{\hbar}{m}(3\pi^2 n)^{1/3} \qquad \text{Fermi velocity,} \tag{8.27}$$

$$T_F = \frac{E_F}{k_B} \qquad \text{Fermi temperature.} \tag{8.28}$$

Another frequently occurring quantity is the density of states $D(E_F)$ at the *Fermi energy*

$$D(E_F) = \frac{3}{2}\frac{n}{E_F} . \tag{8.29}$$

Table 8.1 shows values for these quantities for a number of metals, calculated using the known values for the density and the free electron mass. In all cases, the Fermi temperature T_F is much higher than the melting temperature of the metals concerned. Thus for solid metals the electrons behave as if the metal were near absolute zero at all temperatures!

At $T = 0$ the occupation probability is given by a rectangular function. For energies below E_F the states are all filled, and above E_F they are all empty with a sharp edge at the Fermi energy. At finite temperature, electrons become excited to higher energies, so that states just below the Fermi energy become emptied and states just above E_F become occupied. The states are thus populated according to equation(8.21), so that there is a "softening" of the Fermi edge with a width of about $2k_BT$. Only a small fraction of the electrons, of the order of T/T_F, are thermally excited. All the other electrons, i.e. those with energies more than k_BT from the Fermi energy, remain in the state of lowest energy and play no role in the majority of solid state phenomena. Figure 8.6 illustrates

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

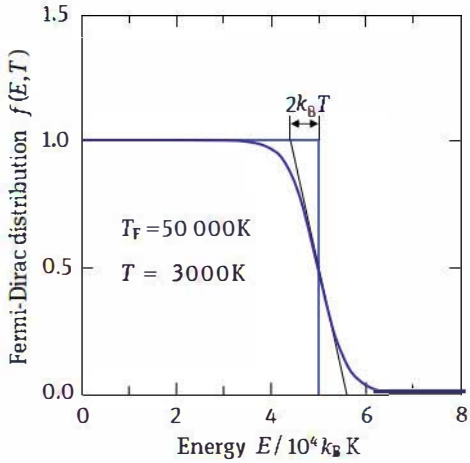


Fig. 8.6: The Fermi-Dirac distribution as a function of energy at $T = 0$ K and $T = 3000$ K. Such a high temperature was chosen because the curve for room temperature hardly differs from the curve at $T = 0$. As Fermi temperature $T_F = 50\,000$ K is assumed.

the form of the Fermi-Dirac function for temperatures $T = 0$ K and 3 000 K for a given Fermi temperature of $T_F = 50000$ K.

The value of the chemical potential is determined by the condition $f(E,T) = \frac{1}{2}$ and decreases slightly with increasing temperature. As long as $T \ll T_F$, the temperature dependence can be approximated by the *Sommerfeld expansion*

$$\mu(T) \approx E_F \left[1 - \frac{\pi^2}{12}\left(\frac{T}{T_F}\right)^2\right] . \tag{8.30}$$

As can be seen immediately from this equation and the values for T_F in Table 8.1, in metals the shift in chemical potential is irrelevant for all experimentally accessible temperatures. In Chapter 10 we will find that in semiconductors, on the other hand, the shift can substantial.

Tab. 8.1: The electron density, the magnitude of the Fermi vector, the Fermi velocity, the Fermi energy and the Fermi temperature for a number of metals. The mass of the free electrons was used to calculate the numerical values.

Element	$n/10^{28}$ m^{-3}	$k_F/$Å^{-1}	$v_F/10^6$ ms^{-1}	E_F/eV	T_F/K
Li	4.62	1.11	1.11	4.69	54400
Na	2.54	0.91	1.05	3.16	36700
Al	18.07	1.75	2.03	11.67	135400
Cu	8,49	1.36	1.57	7.04	81700
Ag	5.86	1.20	1.39	5.49	63700
Au	5.90	1.20	1.38	5.51	63900
Pb	13.20	1.57	1.81	9.37	108700

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In our simple model of free electrons, the electrons move in an isotropic environment, i.e., the Fermi wave vector $\mathbf{k}_F$ has a fixed, direction-independent magnitude. At $T = 0$ the occupied electron states in momentum space are located within the **Fermi sphere**, whose radius is determined by the density of electrons. The surface of this sphere, which plays a central role in determining the electrical properties of metals, is called the **Fermi surface**. Figure 8.7 illustrates the Fermi sphere at absolute zero. Its surface is "softened" at finite temperatures.

It should also be noted that when using fixed boundary conditions, we only consider standing waves whose wave vectors are positive. Thus, the positive octant of the Fermi sphere contains all of the physically relevant states. However, the number of vectors allowed in each of the three directions in space is twice as large as for periodic boundary conditions, so that the total number of states remains unchanged.

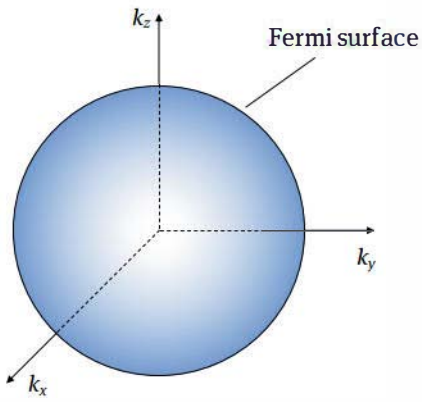


Fig. 8.7: The Fermi sphere. At absolute zero, all electrons occupy states within the Fermi sphere and the boundary of the sphere is therefore sharp.

8.2 Specific Heat

To derive the specific heat of the conduction electrons we now calculate the internal energy of the Fermi gas. At absolute zero, the internal energy per volume $u_0 = U/V$ can be written:

$$u_0 = \int_0^\infty E\,D(E)\,f(E, T = 0)\,\mathrm{d}E = \int_0^{E_F} E\,D(E)\,\mathrm{d}E = \frac{3n}{5}E_F = \frac{3n}{5}k_B T_F\,. \qquad (8.31)$$

Owing to the high Fermi temperature, even at $T =$ ● the internal energy of the electrons is much higher than that of a classical gas at room temperature!

For the specific heat, however, it is not the absolute value of the internal energy that is crucial, but the temperature-dependent component $\delta u(T) = [u(T) - u_0]$, which we briefly estimate here. As we see in Figure 8.6, the fraction of electrons that can

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

absorb thermal energy $k_B T$ is roughly given by T/T_F. Thus $\delta u \approx n k_B T^2/T_F$ and we obtain for the contribution of the electrons to the specific heat (per volume) $c_V^{el} = C_V^{el}/V$ approximately

$$c_V^{el} = \left(\frac{\partial u}{\partial T}\right)_V \approx \frac{2nk_B T}{T_F} . \tag{8.32}$$

Compared with a classical gas, here there is a drastic reduction in the specific heat by the factor T/T_F, arising from the small fraction of electrons involved. For a more precise calculation, we would have to solve the Fermi-Dirac integral:

$$u = \int_0^\infty E D(E)\, f(E,T)\, \mathrm{d}E = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} \int_0^\infty \frac{E^{3/2}}{e^{(E-\mu)/k_B T} + 1}\, \mathrm{d}E \tag{8.33}$$

which cannot be solved analytically. An approximate solution for $k_B T \ll E_F$ results in

$$u \approx u_0 + \frac{\pi^2}{6} D(E_F)(k_B T)^2 \tag{8.34}$$

and thus for the heat capacity:

$$c_V^{el} \approx \frac{\pi^2}{3} D(E_F) k_B^2 T = \frac{m k_B^2 T}{\hbar^2}\left(\frac{\pi^2 n}{9}\right)^{1/3} = \frac{\pi^2 T}{3T_F}\cdot\frac{3nk_B}{2} = \gamma T . \tag{8.35}$$

This more complete treatment yields a result which differs from our rough estimate (8.32) by only a small numerical factor. For all experimentally accessible temperatures, the specific heat increases linearly with temperature. The factor $\pi^2 T/3T_F$ indicates the reduction compared to the specific heat of a classical gas. The proportionality factor γ, often called **Sommerfeld coefficient**, is determined by the density and mass of the electrons.

The total specific heat (per volume) of a metal consists of the contribution of the electrons and the lattice and is approximated at high or low temperatures by

$$c_V^{ges} = \gamma T + \begin{cases} 3n_A k_B & \text{for} \quad T > \Theta , \\ \beta T^3 & \text{for} \quad T \ll \Theta . \end{cases} \tag{8.36}$$

To avoid any confusion, we have denoted the number density of atoms with n_A.

At *high temperatures* ($T > \Theta$) the contribution of the lattice dominates, and can be approximated by the Dulong-Petit law.[4] The electrons do not contribute significantly to the heat capacity in this temperature range. At *low temperatures*, the lattice contributes the term βT^3 to the specific heat. The constant β is given by equation (6.94) if we convert that expression to give specific heat per volume. At this point, we should also refer to

4 The Dulong-Petit law $C_V = 3N_A k_B = 3R_m$ refers to one mole, the expression in the above equation refers to volume.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 6.32, which shows the temperature variation of the specific heat of some metals at relatively high temperatures.

These two contributions to the heat capacity are illustrated in Figure 8.8a showing the low temperature data of copper. Clearly the measured specific heat agrees very well with the sum of the electron and lattice contributions, which are equal at about 4 K. Above this temperature the phonons increasingly dominate, as their number increases much faster with temperature than the number of excited electrons. To separate the two contributions, it is best to plot C_V/T as a function of T^2. The value of γ can then be taken directly from the axis intercept, with the slope giving the value of β. Such a plot is shown in Figure 8.8b again demonstrating the good agreement with (8.36).

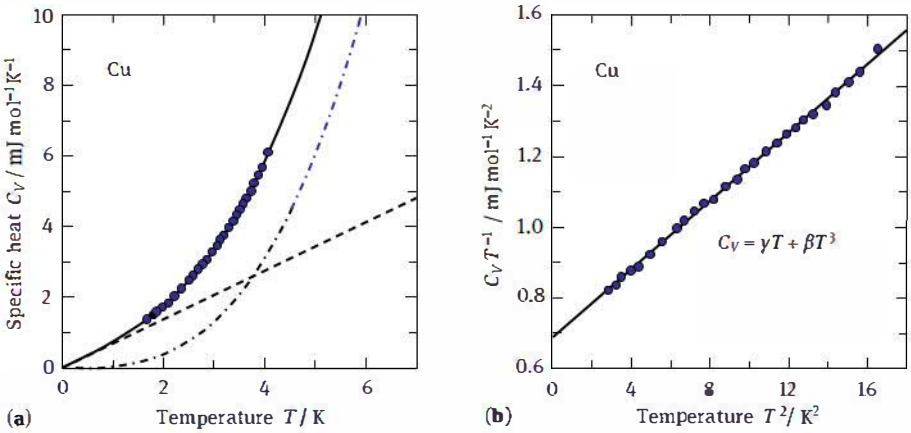


Fig. 8.8: a) The temperature dependence of the specific heat of copper at low temperatures. The specific heat is made up of the linear contribution of the electrons (dashed line) and the T^3 contribution of phonons (dashed-dotted line). **b)** The specific heat of copper plotted as C_V/T as a function of T^2. (After J.A. Rayne, Austral. J. Phys. **9**, 191 (1956).)

How well, in practice, does the approximation of the free Fermi gas model describe the specific heat? The values in Table 8.2 show that for simple metals, the agreement is relatively good as the ratio of experimental to theoretical values of the Sommerfeld coefficient $\gamma_{exp}/\gamma_{theo}$ is close to one, but the quality of the approximation varies for different metals. According to (8.35), γ is proportional to the mass and the concentration of electrons. Since both quantities can be determined with very high accuracy, the reason for the varying quality of the agreement cannot be the poor quality of the input data. The reason for the differences is that we have been assuming the electrons to be free particles. In reality, the conduction electrons interact with the periodic potential of the lattice, distort the crystal in their environment and interact with other electrons. All of these effects can be accounted for by introducing an **effective mass** m^*. In the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

present case, we refer to it as the *thermal effective mass* m^*_{th}, defined by the simple relationship $m^*_{th}/m = \gamma_{exp}/\gamma_{theo}$.

Tab. 8.2: The Sommerfeld coefficient γ_{exp} for various metals and a comparison of the experimental data with the values derived from the free-electron model. (From different sources.)

Element	γ_{exp}	γ_{theo}	m^*_{th}/m	Element	γ_{exp}	γ_{theo}	m^*_{th}/m
Ag	0.64	0.64	1.00	Cu	0.69	0.50	1.37
Al	1.35	0.91	1.48	Ga	0.60	1.02	0.59
Au	0.69	0.64	1.08	In	1.66	1.26	1.31
Ba	2.70	1.95	1.38	K	2.08	1.75	1.19
Be	0.17	0.49	0.35	Li	1.65	0.75	2.19
Ca	2.73	1.52	1.80	Mg	1.26	1.00	1.26
Cd	0.69	0.95	0.73	Na	1.38	1.3	1.22
Cs	3.97	2.73	1.46	Pb	2.99	1.50	1.99

For many transition metals, the ratio of measured to calculated specific heat is much higher than the values given in Table 8.2. For example for Nickel we find the ratio $m^*_{th}/m \approx 15$. However, the reason for this is not from the interaction of the quasi-free *s*-electrons, but is due to the partially filled *d*-shells of these metals where the approximation of the free electron gas does not apply at all, because *d*-wave functions that point in preferred directions, are involved in covalent bonds and do not fill the crystal isotropically. In many metals, *d*-electrons cause a high density of states at the Fermi energy and therefore contribute particularly strongly to the specific heat. Figure 8.9 shows the electronic density of state for nickel, which illustrates this point. For nickel at the Fermi energy the density of states is dominated by the *d*-electrons, whereas in our description only the *s*-electrons were taken into account. The somewhat more complicated conditions that prevail in many metals will be considered in more detail in connection with the discussion of the *band structure* in the second half of this chapter.

Such deviations from the simple free-electron picture can be extreme. In the so-called heavy fermion systems, such as $CeAl_3$ or $CeCu_2Si_2$, the strong interactions between the electrons results in a specific heat so large at low temperatures that it can only be described assuming effective electron masses of up to $m^*_{th} \approx 1000m$. A discussion of these metals with their very surprising low temperature properties would go beyond the scope of this book.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

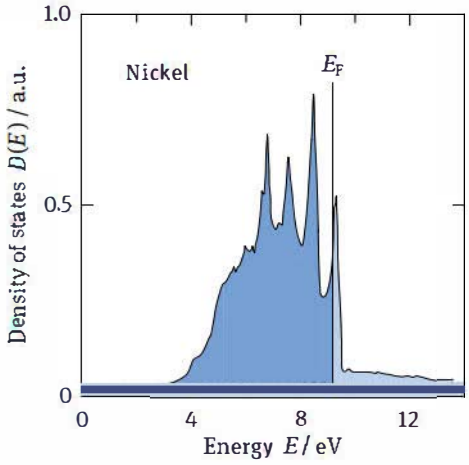


Fig. 8.9: Density of states of nickel. The contribution of the *d*-electrons overlaps that of the *s*-band, whose imagined contribution is shown as dashed line. The occupied states are highlighted by darker blue. (After J. Callaway, C.S. Wang, Phys. Rev. B 7, 1096 (1973).)

8.3 Collective Phenomena in the Electron Gas

Up to now we have assumed that electrons are practically independent of one another. This is astonishing at first, considering that in copper, for example, the average distance between two conduction electrons is only about 2.56 Å. The resulting energy of Coulomb repulsion of 5.6 eV is greater than the average kinetic energy of the Fermi gas of 4.2 V. There are three principal reasons why the approach pursued so far has nevertheless produced good results. We only mention the first two here in passing, as we will deal with them more fully later. First, owing to the Pauli principle, electron-electron scattering is largely suppressed. Secondly, in experiments, it is not the free electrons that are observed, but rather *quasiparticles*, which, despite their interaction with one another, behave almost like free electrons. The third and perhaps most important reason is that the electrostatic interaction between two electrons is largely screened by the presence of others. We now will discuss this in a little more detail.

8.3.1 Screened Coulomb Potential

The electron gas reacts to local electrical charge variations, such as those caused by charged point defects. Depending on the sign of the charge, electrons are either attracted or repelled, thus building up a space charge and causing a screening (in other words a diminution) of the electric field of the defect. Screening also plays an important role in undisturbed solids, because it alters the potential of the atomic cores and reduces the interaction between the conduction electrons.

To make a simplified model of the screening effect, we imagine that an additional point charge e is introduced into the solid at location $\mathbf{r}_0$, which gives rise to a perturbation potential $\delta\varphi(\mathbf{r})$. The additional charge shifts the local energy zero point by an

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

amount $e\,\delta\varphi(\mathbf{r})$. As shown in Figure 8.10, in the vicinity of the additional positive charge the chemical potential is maintained constant by an influx of electrons from the rest of the solid. Similarly, an additional negative charge causes an increase in potential and thus an outflow of electrons.

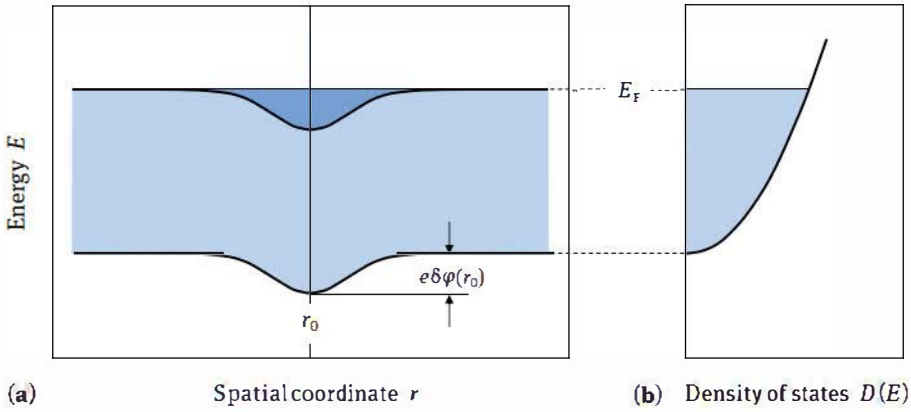


Fig. 8.10: Schematic illustration of the local change in the density of states induced by an additional positive charge. The occupied states are highlighted in light blue. **a)** The energies of the occupied states in the vicinity of a positive charge. The additional (in this case positive) charge reduces the energy of the distribution locally, producing a dip which is then filled with more electrons from the rest of the solid (shown in dark blue) so that the Fermi energy remains constant throughout the sample. **b)** The density of states at a distance far from the additional positive charge.

In the vicinity of the perturbation, the electron concentration is increased by

$$\delta n(\mathbf{r}) = |e| D(E_F) \delta\varphi(\mathbf{r}) \tag{8.37}$$

and shifted with respect to the equilibrium value. Somewhat away from the additional point charge, the Poisson equation links δn and $\delta\varphi$ as

$$\nabla^2(\delta\varphi) = \frac{e}{\varepsilon_0}\delta n = \frac{e^2}{\varepsilon_0} D(E_F)\delta\varphi \,. \tag{8.38}$$

The differential equation has a spherically symmetrical solution of the form

$$\delta\varphi(r) = -\frac{e}{4\pi\varepsilon_0}\,\frac{\mathrm{e}^{-r/r_{TF}}}{r}\,, \tag{8.39}$$

where r_{TF} denotes the **Thomas-Fermi screening length**[5]

$$r_{TF} = \sqrt{\frac{\varepsilon_0}{e^2 D(E_F)}}\,. \tag{8.40}$$

5 Llewellyn Hilleth Thomas, *1903 London, †1992 Rayleigh, U.S.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The screening causes an exponential factor to appear in the equation for the potential function in addition to the typical $1/r$-dependence of the Coulomb potential. The result can also be directly applied to the screening of negative charges. Figure 8.11 compares the pure Coulomb potential with the **screened Coulomb potential**, which has the same form as the *Yukawa potential* used in nuclear physics and was first used by *P. Debye* and *E. Hückel*[6] to describe the electrostatic interaction of ions in electrolytes.

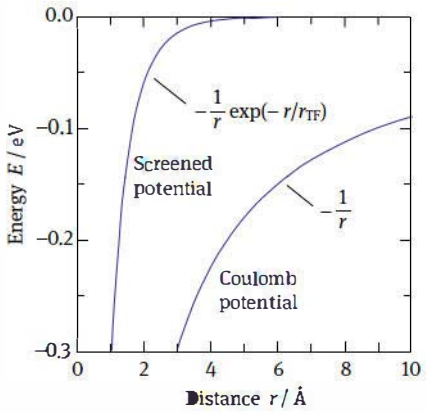


Fig. 8.11: Comparison between pure and screened Coulomb potential. The value $r_{TF} = 1$ Å was chosen for the screening length. As can be seen, with this value of r_{TF}, the effect of the local charge has almost reduced to zero already at a distance of 4 Å.

To get a feeling for the effectiveness of the screening, we use the electron density for copper $n_{Cu} = 8.5 \times 10^{28}\ m^{-3}$ and obtain the very small value $r_{TF} \approx 0.55$ Å, confirming that in metals this screening is extremely effective owing to the high electron density.

Electrostatic screening plays an important role not only in the more in-depth treatment of the electron-electron interaction, but also in many other areas of solid state physics. For example, it is of considerable importance in the quantitative description of defect properties, because screening greatly reduces the range of the defect fields. Thus on adding an additional charge, electrons beyond the screening length r_{TF} are hardly affected by its field.

However, it should be noted here that a problem occurs with real metals, where the screening length and the critical radius of the pseudopotentials (cf. Section 2.5) are comparable meaning that the simple assumptions of the Thomas-Fermi approximation are no longer valid for small distances. To improve the approximation, the decreasing effectiveness of the screening for $r \lesssim r_{TF}$ must therefore be taken into account. Among other things, this leads to oscillations in the space-charge density for screening. Depending on the context, these are known as Friedel or Ruderman-Kittel[7,8] oscillations.

6 Erich Armand Arthur Joseph Hückel, *1896 Berlin, †1980 Marburg
7 Charles **Kittel**, *1916 New York, †2019 Berkeley
8 Malvin **Ruderman**, *1927 New York

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

8.3.2 Metal-Insulator Transition

The concept of screening can also be used to classify materials as either metals or insulators. For example, we can ask ourselves the question of whether a hypothetical solid of hydrogen atoms would be a metal or an insulator. The answer is that this depends on the lattice constant. If the lattice constant is small, the solid should have a metallic character, whereas if the constant is large, it should be an insulator. Under high pressure, for example inside the planet Jupiter, it is therefore possible that hydrogen is present in metallic form.

As already mentioned, the fields of regular atomic cores are also screened by the presence of valence electrons. This reduces the range of the nuclear potential. The higher the electron density, the stronger is the effect. The screening is accompanied by a stronger spatial localization of the core electrons, which simultaneously increases the kinetic energy of the electrons in the atomic core due to the uncertainty relation. In this way, the states are energetically raised until finally, at a very high electron density, the energies of the states of the outer core electrons are so high that they are no longer bound but can move freely. In this case we are dealing with free electrons, and thus we have a metal.

We will now take a closer look at the value of this critical electron concentration. To do this we need to solve the Schrödinger equation for an electron in a screened potential of the form (8.39). The numerical solution shows that bound eigenstates exist only for screening lengths $r_{\mathrm{TF}} > 0.84\,a_0$, where a_0 is the Bohr radius. If we find a bound state, then the electrons condense on the atomic cores and the sample is an insulator.

To estimate the critical electron density at a given lattice spacing, we rewrite (8.40) and obtain

$$\frac{1}{r_{\mathrm{TF}}^2} = \frac{3ne^2}{2\varepsilon_0 E_{\mathrm{F}}} = \frac{4(3\pi^2)^{1/3}}{\pi}\,\frac{n^{1/3}}{a_0} \approx \frac{4n^{1/3}}{a_0}\,. \tag{8.41}$$

Given the above condition that $r_{\mathrm{TF}} > 0.84\,a_0$, a substance can only be an insulator if the following condition for the electron density and the Bohr radius holds:

$$n < \frac{0.045}{a_0^3}\,. \tag{8.42}$$

A crystal with a simple-cubic lattice and an electron density of $n = 1/a^3$ should therefore be an insulator so long as $a > 2.8\,a_0$.

The transition from one substance class to another can be achieved by changing external parameters such as the pressure or magnetic field. This phenomenon is known as **Mott's metal-insulator transition**[9]. In oxides of transition metals, as well as in glasses, and amorphous or liquid semiconductors, electron densities can be varied by changing the material composition. In these materials, abrupt changes in conductivity are observed at certain concentrations arising from the screening effect discussed here.

9 Nevill Francis Mott, *1905 Leeds, †1996 Milton Keynes, Nobel Prize 1977

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The concentration of free electrons in doped semiconductors (see Section 10.2) can be changed in a particularly simple way. If the donor or acceptor concentration is increased, a transition from a semiconducting to a metallically conductive phase occurs. Figure 8.12 shows the electrical conductivity of silicon crystals with various densities of free electrons. The concentration was changed both by doping and also by uniaxial pressure. In the vicinity of the transition, a reduction in the electron density causes a drastic decrease in electrical conductivity until finally, at a density of $n = 3.74 \times 10^{24}\ \mathrm{m}^{-3}$, the transition from metal to insulator takes place.

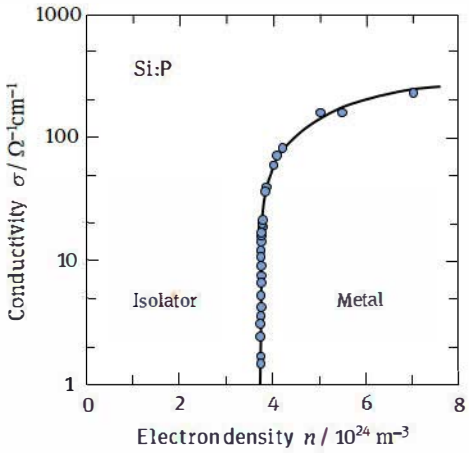


Fig. 8.12: Electrical conductivity of heavily doped silicon samples. The metal-insulator transition occurs at an electron density of $n = 3.74 \times 10^{24}\ \mathrm{m}^{-3}$. (After H.F. Hess et al., Phys. Rev. B **25**, 5578 (1982).)

However, in the example presented here, although the effect is evident, quantitative comparison of theory and experiment is not easy. We need to take into account that the Bohr radius in (8.42) must be replaced by a much larger *effective* Bohr radius, since the Coulomb potential is also screened in the non-metallic phase. We will discuss this aspect of screening of charged impurities in semiconductors and their description with the help of dielectric constants in Section 10.2. Furthermore, the electronic structure of silicon and the fact that the dopant atoms are statistically distributed on lattice sites must also be taken into account.

8.4 Electrons in a Periodic Potential

The free-electron model is impressive in its simplicity yet is able to explain a number of physical properties. However, it also has clearly discernible limitations. For example, we might expect that whenever the electron shells of an element are not completely filled, the electrons of these shells would be able to move relatively freely and the element in question would thus have metallic character. This is obviously not the case with diamond, which is a good insulator, despite the fact that the outer electron shells

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

of the carbon atoms are only half-filled. Furthermore, very drastic deviations occur when considering the Hall effect, which we look at in more detail in Section 9.3. For example, while in the case of sodium we find, as expected, that *one* electron per atom is free to move, whereas in the case of beryllium there seem to be 0.4 *positive* charge carriers per atom instead of the two electrons we would expect from its position in the periodic table. This fact makes it seem that it is not electrons but positively charged particles that are responsible for the current transport in this metal! Obviously, we have too drastically modify the description of the electronic properties in this case. What we have been missing here is that the electrons in metals are not really free but move in a periodic potential.

8.4.1 The Bloch Function

To take this into account we begin, as in the discussion of the free electron gas, by describing the behavior of the electrons using the periodic lattice potential in the *one-electron approximation*. As we will see, the Schrödinger equation of the chosen electron can be split into a set of linear equations which can be solved approximately. The crucial point here is that the potential $\tilde{V}(\mathbf{r})$ has the translational symmetry of the lattice. This means that $\tilde{V}(\mathbf{r}) = \tilde{V}(\mathbf{r} + \mathbf{R})$ applies if $\mathbf{R}$ is any lattice vector. The potential can then be expanded like the scattering density in Section 4.3 into a Fourier series of reciprocal lattice vectors $\mathbf{G}$:

$$\tilde{V}(\mathbf{r}) = \sum_{\mathbf{G}} \tilde{V}_{\mathbf{G}} \, e^{i\mathbf{G}\cdot\mathbf{r}} \,. \tag{8.43}$$

The Fourier coefficients $\tilde{V}_{\mathbf{G}}$ are characteristic of the crystal under consideration.

We take the obvious step of expanding the wave function $\psi(\mathbf{r})$ of the chosen electron in plane waves, because in the limiting case of free electrons this representation has already been proven to work. We therefore again make an educated guess at the wave function:

$$\psi(\mathbf{r}) = \sum_{\mathbf{k}} c_{\mathbf{k}} \, e^{i\mathbf{k}\cdot\mathbf{r}} \,, \tag{8.44}$$

where the coefficients $c_{\mathbf{k}}$ will be determined during the calculation.

We took a similar approach in Section 6.2 when discussing the lattice vibrations, when we transformed the displacements of the atoms into normal modes. For the electrons, however, there is no restriction that the wave vector must lie in the first Brillouin zone, since their wavelength can be smaller than the lattice constant. However, as with the free electron gas, the wave vector $\mathbf{k}$ is subject to the given boundary conditions. We now insert into the Schrödinger equation

$$H\psi(\mathbf{r}) = \left[-\frac{\hbar^2}{2m}\Delta + \tilde{V}(\mathbf{r}) \right] \psi(\mathbf{r}) = E\,\psi(\mathbf{r}) \tag{8.45}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the power series expansion and our guess at the wave function and find:

$$\sum_{\mathbf{k}} \frac{\hbar^2 k^2}{2m} c_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} + \sum_{\mathbf{k}',\mathbf{G}} c_{\mathbf{k}'} \widetilde{V}_{\mathbf{G}} e^{i(\mathbf{k}'+\mathbf{G})\cdot\mathbf{r}} = E \sum_{\mathbf{k}} c_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} . \tag{8.46}$$

In the second sum, we have denoted the wave vector not by $\mathbf{k}$ but by $\mathbf{k}'$ in order to get $(\mathbf{k}' + \mathbf{G}) \rightarrow \mathbf{k}$ after the renaming of the summation indices to arrive at a simpler expression. This renaming is allowed because the summation is performed over *all* values of the wave vectors $\mathbf{k}$ and over all reciprocal lattice vectors $\mathbf{G}$. The renaming does not change the value of the elements in the sum, but only the order in which they are added. After designating the summation indices, we can simplify and write:

$$\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \left[\left(\frac{\hbar^2 k^2}{2m} - E \right) c_{\mathbf{k}} + \sum_{\mathbf{G}} \widetilde{V}_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} \right] = 0 . \tag{8.47}$$

This equation must be valid for all position vectors $\mathbf{r}$, but this is only possible if the expression in the square brackets goes to zero separately for each wave vector $\mathbf{k}$. Thus we obtain the important result

$$\left(\frac{\hbar^2 k^2}{2m} - E \right) c_{\mathbf{k}} + \sum_{\mathbf{G}} \widetilde{V}_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} = 0 . \tag{8.48}$$

This set of algebraic equations is the representation of the **Schrödinger equation** in $\mathbf{k}$-space for an electron moving in a periodic potential. For each wave vector $\mathbf{k}$ there is a system of equations with the wave function $\psi_{\mathbf{k}}(\mathbf{r})$ and the corresponding eigenvalue $E_{\mathbf{k}}$ as solutions. The number of possible solutions is limited by the boundary conditions. There are exactly as many solutions as there are different allowed wave vectors.

An important result is that in (8.48) the only coefficients that occur are those with indices that differ by a reciprocal lattice vector from the $\mathbf{k}$, i.e. only the coefficients $c_{\mathbf{k}-\mathbf{G}}$, $c_{\mathbf{k}-\mathbf{G}'}$, $c_{\mathbf{k}-\mathbf{G}''}$, ... appear. This means that $\psi_{\mathbf{k}}(\mathbf{r})$ is made up of plane waves whose wave vectors also are differing by a reciprocal lattice vector $\mathbf{G}$. This simplifies the expansion (8.44) of the wave function to:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}} . \tag{8.49}$$

Since the reciprocal lattice reflects the Fourier expansion of the scattering density of the real lattice, generally only a view coefficients provide significant contributions. The wave function can therefore be represented by the superposition of relatively few terms.

Before we go into details of the calculation of the coefficients $c_{\mathbf{k}-\mathbf{G}}$, let us briefly consider the structure of the wave function and point out some of its important properties. We rearrange the wave function (8.49) to take the form:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \left(\sum_{\mathbf{G}} c_{\mathbf{k}-\mathbf{G}} e^{-i\mathbf{G}\cdot\mathbf{r}} \right) e^{i\mathbf{k}\cdot\mathbf{r}} , \tag{8.50}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

to make the structure of the solution clearer. The expression in brackets is the Fourier expansion of a lattice periodic function. The solutions of the Schrödinger equation for a periodic potential are therefore *plane waves*, multiplied by a *modulation factor* with the *periodicity of the lattice* which we call $u_{\mathbf{k}}(\mathbf{r})$. Thus the wave function of electrons in a periodic potential can be described by the **Bloch function**

$$\psi_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r})\, e^{i\mathbf{k}\cdot\mathbf{r}} \tag{8.51}$$

where the lattice periodicity appears in the relationship

$$u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) \,. \tag{8.52}$$

A simple example of a Bloch function is shown in Figure 8.13.

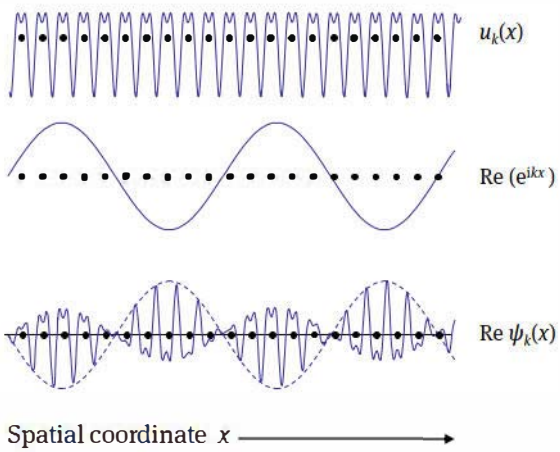


Fig. 8.13: Illustration of a one-dimensional Bloch function. Shown are the lattice periodic function $u_k(x)$ (top), the real part of the phase factor e^{ikx} (middle) and (bottom) the real part of the Bloch function $\psi_k(x)$. The points indicate the position of the atomic cores.

The two equations (8.51) and (8.52) are a special case of a theorem formulated by *F. Bloch*[10] for crystals: it states that for any arbitrary wave function satisfying the Schrödinger equation there exists a wave vector $\mathbf{k}$ such that a translation by a lattice vector $\mathbf{R}$ is equivalent to the multiplication by the phase factor $e^{i\mathbf{k}\cdot\mathbf{R}}$. By inserting this, it is easy to see that the solution fulfills this condition:

$$\psi_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{\mathbf{k}}(\mathbf{r} + \mathbf{R})\, e^{i\mathbf{k}\cdot\mathbf{r}} e^{i\mathbf{k}\cdot\mathbf{R}} = \psi_{\mathbf{k}}(\mathbf{r})\, e^{i\mathbf{k}\cdot\mathbf{R}} \,. \tag{8.53}$$

In the simple case of free electrons, $u(\mathbf{r})$ is a constant. Furthermore, the elastic lattice waves we discussed in Chapter 6 are also Bloch waves in this sense.

Bloch functions $\psi_{\mathbf{k}}(\mathbf{r})$ have two important properties, which we should point out here. Bloch functions with wave vectors differing by a reciprocal lattice vector are

10 Felix Bloch, *1905 Zurich, †1983 Zollikon (Zurich), Nobel Prize 1952

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

identical, which follows directly from (8.50): if we add the reciprocal lattice vector $\mathbf{G}'$ to the wave vector $\mathbf{k}$, we obtain

$$\psi_{\mathbf{k}+\mathbf{G}'}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{k}+\mathbf{G}'-\mathbf{G}}\, e^{i(\mathbf{k}+\mathbf{G}'-\mathbf{G})\cdot\mathbf{r}} . \tag{8.54}$$

By renaming the summation indices $(\mathbf{G} - \mathbf{G}') \rightarrow \mathbf{G}''$ it follows that:

$$\psi_{\mathbf{k}+\mathbf{G}'}(\mathbf{r}) = \underbrace{\sum_{\mathbf{G}''} c_{\mathbf{k}-\mathbf{G}''}\, e^{-i\mathbf{G}''\cdot\mathbf{r}}}_{u_{\mathbf{k}}(\mathbf{r})}\, e^{i\mathbf{k}\cdot\mathbf{r}} . \tag{8.55}$$

Furthermore, if we write $\mathbf{G}$ instead of $\mathbf{G}'$, we get

$$\psi_{\mathbf{k}+\mathbf{G}}(\mathbf{r}) = \psi_{\mathbf{k}}(\mathbf{r}) . \tag{8.56}$$

The eigenvalues $E_{\mathbf{k}}$ also repeat periodically in reciprocal space! For eigenvalues with wave vectors differing by a reciprocal lattice vector, the Schrödinger equations are:

$$H\psi_{\mathbf{k}} = E_{\mathbf{k}}\psi_{\mathbf{k}} \qquad \text{and} \qquad H\psi_{\mathbf{k}+\mathbf{G}} = E_{\mathbf{k}+\mathbf{G}}\,\psi_{\mathbf{k}+\mathbf{G}} . \tag{8.57}$$

With equation (8.56) we get

$$H\psi_{\mathbf{k}} = E_{\mathbf{k}+\mathbf{G}}\,\psi_{\mathbf{k}} \qquad \text{and} \qquad E_{\mathbf{k}} = E_{\mathbf{k}+\mathbf{G}} . \tag{8.58}$$

There are also other ways to describe electrons in solids. For example, the wave function can also be represented by means of **Wannier functions**[11], which are assigned to the individual lattice atoms, so that the function can only assume large values there. Nevertheless, in the further discussion here we will only make use of Bloch functions to describe the electrons.

The wave functions of electrons and their eigenvalues repeat themselves periodically in $\mathbf{k}$-space. Therefore, it is sufficient to *restrict ourselves to the solutions in the first Brillouin zone*, because the solutions in this part of the $\mathbf{k}$-space already contain all the information on the whole system. This procedure is called **reduction to the first Brillouin zone** and the corresponding representation is called the **reduced zone scheme**. This will be discussed in more detail in the following sections.

8.4.2 Quasi-Free Electrons

We will now familiarize ourselves with the Schrödinger equation (8.48) for electrons in a periodic potential through the use of a few simple examples. As an extreme starting simplification, we begin by assuming that the amplitude of the periodic potential is

11 Gregory Hugh Wannier, *1911 Basel, †1983 Portland

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

so small that we may set $\widetilde{V}_{\mathbf{G}} \approx 0$, while we assume that the symmetry of the lattice should still force periodic solutions. We then speak of an **empty lattice**. With these assumptions, the periodicity of the eigenvalues (8.58) gives the simple solution:

$$E_{\mathbf{k}} = \frac{\hbar^2 k^2}{2m} = E_{\mathbf{k}+\mathbf{G}} = \frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|^2 \,. \tag{8.59}$$

These are clearly parabolas, shifted in **k**-space by **G** relative to each other.

We first investigate in more detail a one-dimensional empty lattice with lattice constant a. In this simple case the reciprocal lattice vectors are given by multiples of $g = 2\pi/a$. As shown in Figure 8.14a, the solution (8.59) consists of a periodic arrangement of parabolas. As noted above, the periodicity of the solution allows a representation of all energy eigenvalues in the range $-\pi/a < k \leq \pi/a$. In the reduction to the first Brillouin zone, the parts of the parabolas lying outside the first Brillouin zone are shifted by a reciprocal lattice vector such that they are transposed into the first zone. Thus, we obtain a representation of the energy eigenvalues $E_{\mathbf{k}}$ as shown in Figure 8.14b. Note that different eigenvalues $E_{\mathbf{k}}$ belong to the same wave vector **k** in this representation.

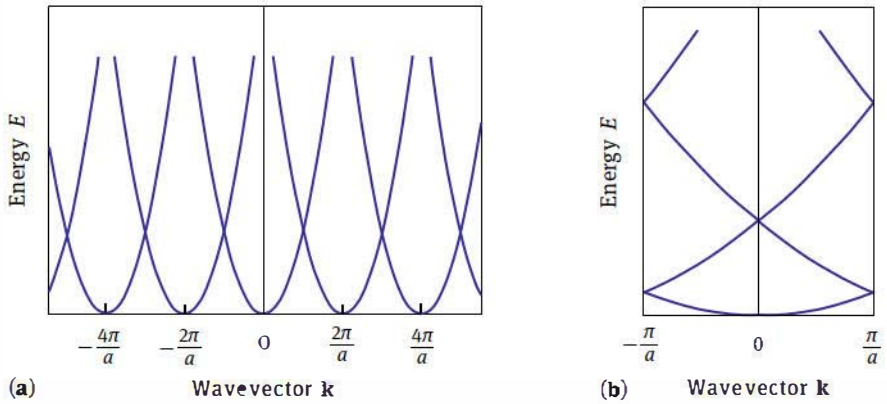


Fig. 8.14: a) A representation of the solution of the Schrödinger equation for a one-dimensional *empty lattice*. The energy eigenvalues are repeated in k-space with period $g = 2\pi/a$. **b)** Reduction of the energy eigenvalues to the first Brillouin zone. All allowed energy eigenvalues can be transposed to lie in the range $-\pi/a < k \leq \pi/a$ by addition of a suitable reciprocal lattice vector. Note that the energy scale in the two figures is different.

In three-dimensional space the situation is somewhat more complicated. For a simple-cubic lattice, (8.59) can be expressed in the form:

$$E_{\mathbf{k}} = \frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|^2 = \frac{\hbar^2}{2m}\left[(k_x+G_x)^2 + (k_y+G_y)^2 + (k_z+G_z)^2\right] . \tag{8.60}$$

As one can easily see from this equation, the reduction to the first Brillouin zone increases the degeneracy of the dispersion curves, i.e. equal parts of the $E_{\mathbf{k}}$-curves

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

belong to different reciprocal lattice vectors. In addition, this type of representation results in a large number of curves at higher energies.

Figure 8.15 shows the $E_{\mathbf{k}}$-curves for specific directions of a face-centered cubic lattice in the reduced representation. As in the discussion of the phonon dispersion curves in Section 6.3, one also starts this representation of the electron dispersion at the Γ-point, goes to the boundary of the Brillouin zone at point X, then moves along the zone boundary to points W and L and then returns to the starting point Γ. This path is easily understandable with the help of Figure 4.8. Starting from the Γ-point, the path leads via the K-point to point X′, as shown in Figure 6.19. In this figure, the energy scale is normalized by the Fermi energy E_F, assuming that there are three electrons per atom.

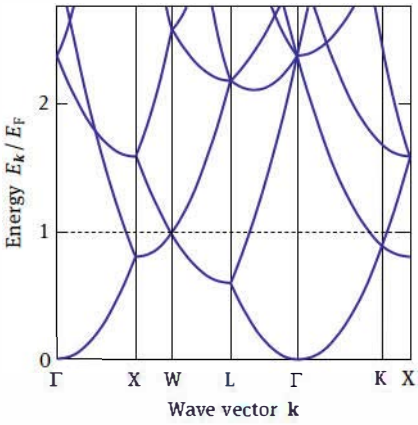


Fig. 8.15: The energy-dispersion curves of free electrons in an empty face-centered cubic lattice after reduction to the first Brillouin zone. The inset on the right indicates the path in the zone. The energy scale is normalized to the Fermi energy E_F for a metal with three free electrons per atom.

Figure 8.16 shows the energy dispersion curves of aluminum derived from a much fuller calculation. We note that the energy scale differs from that in Figure 8.15. The similarity between the two figures is surprisingly large, because we still find all the branches when calculating energy eigenvalues in the empty lattice approximation. It seems that the approximation $\widetilde{V}_{\mathbf{G}} \approx 0$ for aluminum is not too far from the real situation. On closer inspection, however, we can see qualitative differences: on the one hand, degeneracies are eliminated. For example, the energy eigenvalues of the first and second Brillouin zone, which were degenerate in the empty lattice approximation, are clearly split. On the other hand, **energy gaps** occur at the boundaries of the Brillouin zone, which are caused by the finite strength of the lattice potential. The areas in between, which are accessible to the electrons, are known as **energy bands**.

Energy gaps are important and we introduce the concept using a very simple example: if an electron with wave vector $\mathbf{k}$ moves along a chain of atoms, it is scattered at each atom. If the wave vector of the electron satisfies the diffraction condition $\mathbf{K} = \mathbf{G}$, the electron experiences a Bragg reflection, in which the partial waves are constructively superimposed. Since the reflected wave runs in the opposite direction, the relation $\mathbf{K} = 2\mathbf{k}$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

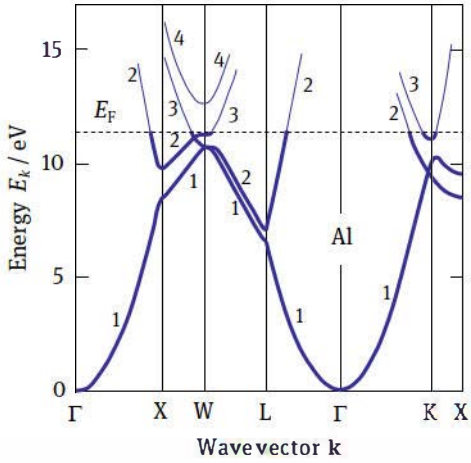


Fig. 8.16: The calculated energy dispersion curves for aluminum. The numbers labelling the curves indicate the order of the Brillouin zones in which the marked parts of the dispersion curve occur before reduction. The Fermi energy E_F is also indicated. (After B. Segall, Phys. Rev.**124**, 1797 (1961).)

applies to the scattering vector. Bragg reflection therefore occurs at the wave number

$$k = \pm\frac{g}{2} = \pm\frac{\pi}{a} \tag{8.61}$$

where $g = 2\pi/a$ again stands for the smallest reciprocal lattice vector. Incoming and reflected waves overlap, resulting in a standing wave. In the stationary state, both partial waves have the same amplitude, resulting in the two possibilities for the resulting wave function:

$$\begin{aligned}\psi_s &\propto \left(e^{igx/2} + e^{-igx/2}\right) \propto \cos(gx/2)\,,\\ \psi_a &\propto \left(e^{igx/2} - e^{-igx/2}\right) \propto \sin(gx/2)\,.\end{aligned} \tag{8.62}$$

This means there are spatially modulated charge densities, which can be expressed by

$$\rho_s = e|\psi_s|^2 \propto \cos^2\left(\frac{\pi x}{a}\right) \qquad \text{and} \qquad \rho_a = e|\psi_a|^2 \propto \sin^2\left(\frac{\pi x}{a}\right)\,. \tag{8.63}$$

On the other hand, the charge density of a running wave is constant:

$$\rho = e|\psi|^2 \propto e^{-ikx}\, e^{ikx} = \text{const.} \tag{8.64}$$

Figure 8.17 shows the spatial variation of an arbitrarily chosen pseudopotential and the electron densities (8.63). For the solution ψ_s the charge density at the nuclear sites is a maximum while it is a minimum for ψ_a. (In analogy with atomic physics, the symmetric solution ψ_s is often called s-type, the antisymmetric solution ψ_a p-type.) The high electron density near the nuclei leads to a lowering of the electron energy and a corresponding increase when the electron density at the nuclei is a minimum. This leads to a lowering of the electron energy in the symmetric case and to an increase in the antisymmetric case. Since the Schrödinger equation thus has two solutions with different energies at the zone boundary, there must clearly be an energy gap.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

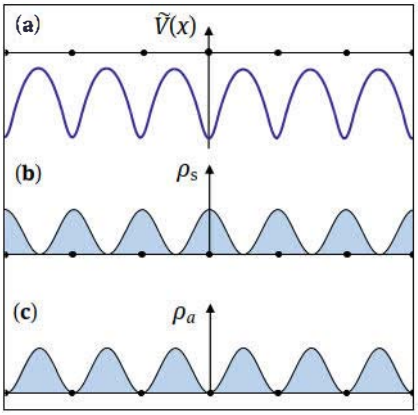


Fig. 8.17: The charge density distribution of electrons with wave number π/a. The positions of the atomic cores are marked by dots. **a)** Qualitative spatial variation of the potential energy. **b)** Charge density ρ_s resulting from the even wave function and **c)** charge density ρ_a associated with the odd wave function.

Now let us consider in more detail the formal solution of the Schrödinger equation at the boundary of the Brillouin zone. We want to know what quantity is responsible for the energy gaps and what form the wave function takes in the vicinity of the zone boundary. As can be seen in Figure 8.14a, the one-dimensional empty lattice has two parabolas which intersect at $k_x = g/2 = \pi/a$, so that there are two degenerate solutions resulting from the Schrödinger equation with the coefficients $c_{\mathbf{k}}$ and $c_{\mathbf{k}-\mathbf{g}}$, respectively. If the periodic perturbation is weak, the main contribution to the wave function should stem primarily from these two solutions. Contributions from "more distant" parabolas, i.e. contributions from higher order Brillouin zones, should not play any role. To show this, we write the Schrödinger equation (8.48) for the wave vector $(\mathbf{k} - \mathbf{G})$:

$$\left(E - \frac{\hbar^2}{2m}|\mathbf{k}-\mathbf{G}|^2\right) c_{\mathbf{k}-\mathbf{G}} = \sum_{\mathbf{G}'} \tilde{V}_{\mathbf{G}'} c_{\mathbf{k}-\mathbf{G}-\mathbf{G}'} \tag{8.65}$$

and solve the equation for the coefficient $c_{\mathbf{k}-\mathbf{G}}$:

$$c_{\mathbf{k}-\mathbf{G}} = \frac{\sum_{\mathbf{G}'} \tilde{V}_{\mathbf{G}'}\, c_{\mathbf{k}-\mathbf{G}-\mathbf{G}'}}{E - \frac{\hbar^2}{2m}|\mathbf{k}-\mathbf{G}|^2} \,. \tag{8.66}$$

If the potential $\tilde{V}(\mathbf{r})$ represents only a small perturbation, the energy eigenvalue E we are looking for is close to the energy $\hbar^2 k^2/2m$ of the free electrons. The largest deviation comes when the denominator in equation (8.66) goes to zero, i.e. when

$$\mathbf{k}^2 \approx |\mathbf{k}-\mathbf{G}|^2 \,. \tag{8.67}$$

This condition is fulfilled at $\mathbf{G} = \mathbf{g}$ and at $\mathbf{G} = 0$, that is, at the edge of the Brillouin zone at π/a and at the origin. This means that only the two coefficients $c_{\mathbf{k}-\mathbf{g}}$ and $c_{\mathbf{k}}$ assume a significantly high value. The remaining coefficients are relatively small and can be neglected.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Since in the further discussion only the two coefficients $c_{\mathbf{k}}$ and $c_{\mathbf{k}-\mathbf{g}}$ are included in the calculation, this approach is called *two-component approximation*. The system of equations (8.65) thus consists only of the following two equations:

$$\left(\frac{\hbar^2\mathbf{k}^2}{2m} - E\right)c_{\mathbf{k}} + \widetilde{V}_{\mathbf{g}}c_{\mathbf{k}-\mathbf{g}} = 0\,,$$

$$\left(\frac{\hbar^2|\mathbf{k}-\mathbf{g}|^2}{2m} - E\right)c_{\mathbf{k}-\mathbf{g}} + \widetilde{V}_{-\mathbf{g}}c_{\mathbf{k}} = 0\,. \tag{8.68}$$

In the following we assume that the lattice has inversion symmetry, then $\widetilde{V}_{\mathbf{g}} = \widetilde{V}_{-\mathbf{g}}$. We also note that in the expansion of the periodic potential, in addition to the first Fourier coefficients $\widetilde{V}_{\mathbf{g}}$ and $\widetilde{V}_{-\mathbf{g}}$, the coefficient $\widetilde{V}_0$ should also appear. But since the latter describes only a constant potential, we can set $\widetilde{V}_0 = 0$. We use the energy eigenvalues of the empty lattice, to which we assign the index ●, as abbreviations as follows:

$$E^0_{\mathbf{k}} = \frac{\hbar^2k^2}{2m} \qquad \text{and} \qquad E^0_{\mathbf{k}-\mathbf{g}} = \frac{\hbar^2|\mathbf{k}-\mathbf{g}|^2}{2m}\,. \tag{8.69}$$

Thus the eigenvalues of equations (8.68) can be expressed in the form

$$E_{\mathrm{s,a}} = \frac{1}{2}\left(E^0_{\mathbf{k}-\mathbf{g}} + E^0_{\mathbf{k}}\right) \mp \sqrt{\frac{1}{4}\left(E^0_{\mathbf{k}-\mathbf{g}} - E^0_{\mathbf{k}}\right)^2 + \widetilde{V}^2_{\mathbf{g}}}\,, \tag{8.70}$$

where the negative sign is assigned to E_{s} and the positive sign to E_{a}.

Since for $\mathbf{k} = \mathbf{g}/2$ the energy eigenvalues of the empty lattice are equal, $E^0_{\mathbf{k}-\mathbf{g}} = E^{●}_{\mathbf{k}}$, we can simplify (8.70) at the boundary of the Brillouin zone to obtain:

$$E_{\mathrm{s,a}} = E^{●}_{\mathbf{k}} \pm |\widetilde{V}_{\mathbf{g}}|\,. \tag{8.71}$$

Thus the energy difference, the width of the energy gap, is given by

$$E_{\mathrm{a}} - E_{\mathrm{s}} = \delta E = 2|\widetilde{V}_{\mathbf{g}}|\,. \tag{8.72}$$

The energy dispersion curves at the boundary of the Brillouin zone are shown schematically in Figure 8.18.

In this simple approximation, the energy splitting at the boundary of the Brillouin zone is determined by the first Fourier component of the potential $\widetilde{V}(\mathbf{r})$. The eigenvalue E_{a} has the higher energy and therefore belongs to the upper energy band. We should note here that the calculation above is only valid for the boundary of the Brillouin zone at $k = \pi/a$. At $k = -\pi/a$ we must deal instead with the coefficients $c_{\mathbf{k}}$ and $c_{\mathbf{k}+\mathbf{g}}$, which naturally lead to an equally large energy gap.

Now we consider how the wave functions $\psi = c_{\mathbf{k}}\exp(\mathrm{i}\mathbf{k}\cdot\mathbf{r}) + c_{\mathbf{k}-\mathbf{g}}\exp[\mathrm{i}(\mathbf{k}-\mathbf{g})\cdot\mathbf{r}]$ of the two energy bands differ near the zone boundary. For this we determine the coefficients $c_{\mathbf{k}}$ and $c_{\mathbf{k}-\mathbf{g}}$ at $\mathbf{k} = \mathbf{g}/2$ by inserting the two eigenvalues E_{s} and E_{a} into equation (8.68), finding:

$$\frac{c_{\mathbf{k}-\mathbf{g}}}{c_{\mathbf{k}}} = \frac{E - \hbar^2k^2/2m}{\widetilde{V}_{\mathbf{g}}} \tag{8.73}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

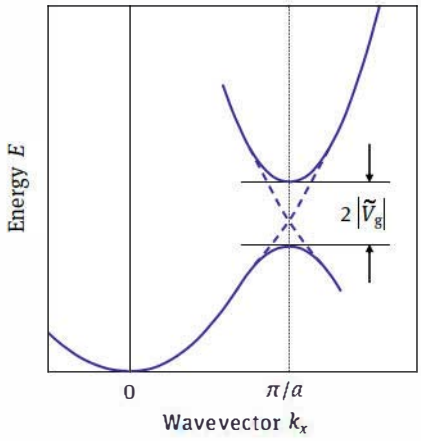


Fig. 8.18: Dispersion curves for quasi-free electrons in a one-dimensional sample. At $k_x = \pi/a$ the periodic potential causes the splitting of $|2\tilde{V}_g|$. The dashed lines show the solutions for the empty lattice.

and thus

$$\left.\frac{c_{\mathbf{k}-\mathbf{g}}}{c_{\mathbf{k}}}\right|_{\mathbf{k}=\mathbf{g}/2} = \frac{E_k^0 \pm |\tilde{V}_\mathbf{g}| - E_k^0}{\tilde{V}_\mathbf{g}} = \pm 1 \,. \tag{8.74}$$

The last equation yields the ratio of the coefficients at the boundary of the Brillouin zone, which is either +1 or -1. At the zone boundary the wave function is either a symmetrical or antisymmetrical superposition of the wave functions of the uncoupled systems. Therefore, which solution has the lower energy depends on the sign of $\tilde{V}_\mathbf{g}$. If $\tilde{V}_\mathbf{g}$ is negative, the ratio is positive for the first band with the energy eigenvalue E_s. Thus the total wave function of the lower band is represented by the sum of the two wave functions of the uncoupled system, i.e. it is symmetrical. At the boundary of the Brillouin zone, both solutions contribute equally to the total wave function.

As we move away from this point, the ratio changes very quickly. The contribution of the coefficients $c_\mathbf{k}$ and $c_{\mathbf{k}-\mathbf{g}}$ to the wave function for the first band is shown schematically in Figure 8.19. In the second band the roles of the coefficients are simply reversed.

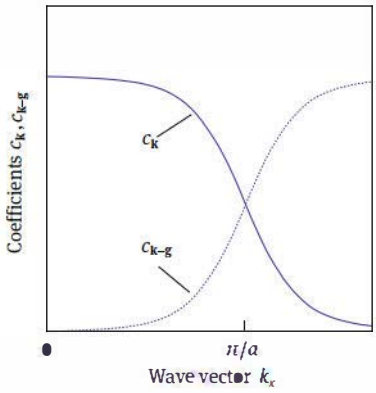


Fig. 8.19: Schematic representation of the contributions of the coefficients $c_\mathbf{k}$ and $c_{\mathbf{k}-\mathbf{g}}$ to the wave function of the first band.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

What is the energy dispersion curve near the zone boundary? This question can be answered with a expansion of the energy eigenvalues (8.70) around the point $k = g/2$, but we will omit this simple but somewhat lengthy calculation here. As can be easily guessed, an expression of the form $E_{s,a} \propto \pm|\delta\mathbf{k}|^2$ results, where $\delta\mathbf{k} = (\mathbf{k} - \mathbf{g}/2)$ expresses the deviation of the wave vector from its value at the boundary of the Brillouin zone, where, as expected, the dispersion curves $E_\mathbf{k}$ are parabolic.

Figure 8.20 shows schematically the energy dispersion curve $E_\mathbf{k}$ for a one-dimensional lattice presented in three different ways. The same dispersion curve is shown in the **extended**, **reduced** and **periodic** zone scheme. We have seen that taking into account the periodicity of the potential in the Schrödinger equation leads naturally to energy bands and to **forbidden regions**, the so-called **band gaps.** This treatment of the electrons is particularly well suited for the description of metals. Starting with free electrons, the electronic properties can be even better described by considering an increasing number of Fourier coefficients of the series expansion of the potential (8.43). However, in the following section we will take a different approach, which is more suitable for crystalline insulators and also for amorphous solids.

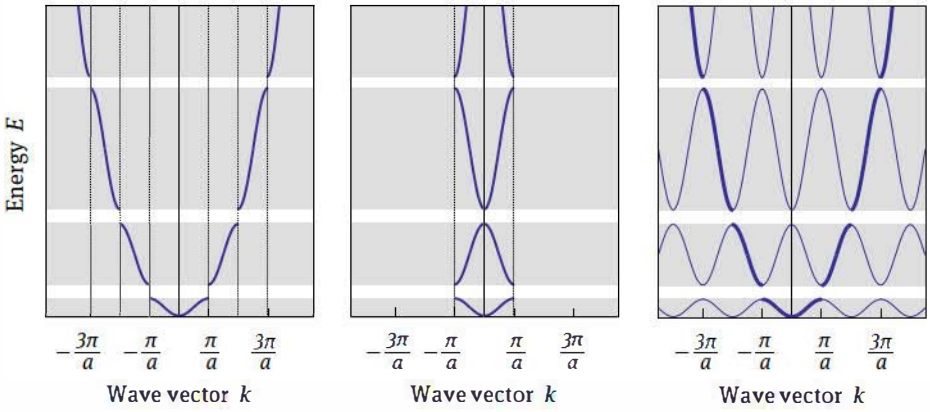


Fig. 8.20: The energy dispersion curve $E_\mathbf{k}$ for a one-dimensional lattice in the *extended*, *reduced* and *periodic* zone scheme. The allowed energy ranges are tinted grey, while the forbidden ranges are left white. The sections of the curve that correspond to the energy parabola of a free electron are drawn thicker in the periodic zone scheme.

8.4.3 Tightly Bound Electrons

Up to now we have been assuming that the electrons can move freely in the solid. Now we will consider the problem from the opposite point of view and assume that the electrons are mainly located near the atomic cores. While the approximation just

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

discussed above is particularly suitable for metals, this second approach is more applicable to molecular and ionic crystals. For simplicity, we will only deal with crystals with a monatomic basis here.

We start by considering isolated atoms, where the electrons are firmly bound to the atomic cores. We further assume that we have already solved the Schrödinger equation for isolated atoms:

$$H_A\bar{\psi}_i = E_i\bar{\psi}_i \tag{8.75}$$

and that the wave functions $\bar{\psi}_i$ and the eigenvalues E_i are therefore known. Here H_A is the Hamiltonian operator of free atoms and the index i denotes the different energy levels. To simplify the notation, we omit the index i for now and reintroduce it when necessary.

In crystalline solids, the wave functions of the electrons overlap, so we have a regular arrangement of weakly interacting atoms. As in the previous section, we limit our considerations again to the *one-electron approximation*. We first establish the Hamiltonian operator H of an electron moving in the potential of all atoms of the crystal. The potential can be split into two parts, first $\bar{V}_A$ that for the free atom and the second the perturbation potential H_S by the neighboring atoms, assuming that the perturbation of the neighboring atoms is small compared to $\bar{V}_A$. Therefore, for the Hamiltonian operator we can write:

$$H = H_A + H_S = -\frac{\hbar^2}{2m}\Delta + \bar{V}_A(\mathbf{r} - \mathbf{R}_m) + H_S(\mathbf{r} - \mathbf{R}_m)\,, \tag{8.76}$$

where the position of the reference atom is defined by the lattice vector $\mathbf{R}_m$.

The perturbation potential H_S is the sum of the atomic potentials of *all the other* atoms, i.e. those with lattice vectors $\mathbf{R}_n \neq \mathbf{R}_m$:

$$H_S(\mathbf{r} - \mathbf{R}_m) = \sum_{n\neq m} \bar{V}_A(\mathbf{r} - \mathbf{R}_n)\,. \tag{8.77}$$

This potential is shown schematically in Figure 8.21. In contrast to the behavior of free electrons in metals, for which the description using pseudo potentials is particularly suitable, tightly bound electrons move in the modified Coulomb potential of the atoms, as indicated in the figure.

If the Hamiltonian H and the wave function $\psi_\mathbf{k}$ were known, the energy eigenvalue $E_\mathbf{k}$ of the electron under consideration could be calculated exactly:

$$E_\mathbf{k} = \frac{\int \psi_\mathbf{k}^* H \psi_\mathbf{k}\,\mathrm{d}V}{\int \psi_\mathbf{k}^* \psi_\mathbf{k}\,\mathrm{d}V}\,. \tag{8.78}$$

However, in the present case the actual wave function $\psi_\mathbf{k}$ is unknown. To arrive at an approximate solution, we replace the actual wave function by a linear superposition $\Phi_\mathbf{k}$ of the atomic eigenfunctions $\bar{\psi}(\mathbf{r} - \mathbf{R}_n)$ and write

$$\psi_\mathbf{k} \approx \Phi_\mathbf{k} = \sum_m a_m\bar{\psi}(\mathbf{r} - \mathbf{R}_m)\,. \tag{8.79}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

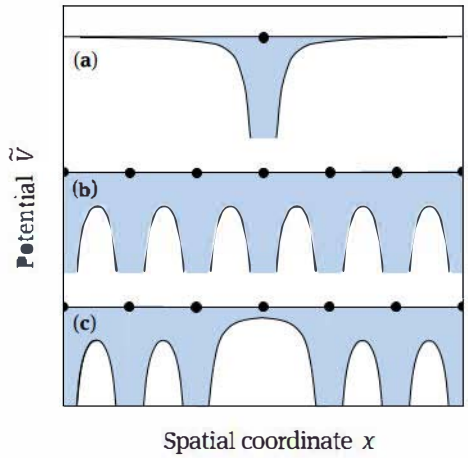


Fig. 8.21: Schematic illustration of the potential in the tight-binding model: **a)** The potential for a single atom. **b)** The superposition of the individual atomic potentials to create the lattice potential, **c)** The resulting perturbation potential, i.e. the total potential minus the potential of the reference atom. The positions of the atomic nuclei are indicated by dots.

According to Bloch's theorem, the solution of the Schrödinger equation in a periodic potential contains a lattice periodic component and a phase factor of the form $\exp(i\mathbf{k}\cdot\mathbf{R})$, where $\mathbf{R}$ is a lattice vector. The lattice periodicity of the function $\Phi_\mathbf{k}$ follows automatically from the periodic arrangement of the atoms on their lattice sites as we sum the individual contributions. The normalization and phase factors, however, must be included in the coefficient a_m. For these factors we choose the expression $a_m = N^{-1/2} e^{i\mathbf{k}\cdot\mathbf{R}_m}$, where the phase is defined by $\exp(i\mathbf{k}\cdot\mathbf{R}_m)$ and the normalization by the number N of atoms. This leads to our guessed wave function:

$$\Phi_\mathbf{k} = \frac{1}{\sqrt{N}} \sum_m \tilde{\psi}(\mathbf{r} - \mathbf{R}_m)\, e^{i\mathbf{k}\cdot\mathbf{R}_m} . \tag{8.80}$$

We should note here that the approach using a lattice-periodic function is not absolutely necessary, but it simplifies the following calculation of the energy eigenvalues.

Inserting $\Phi_\mathbf{k}$ into (8.78) we get for the eigenvalue:

$$E_\mathbf{k} \approx \frac{1}{N} \sum_{m,n} e^{i\mathbf{k}\cdot(\mathbf{R}_m - \mathbf{R}_n)} \int \tilde{\psi}^*(\mathbf{r} - \mathbf{R}_n) \left[H_A + H_S(\mathbf{r} - \mathbf{R}_n)\right] \tilde{\psi}(\mathbf{r} - \mathbf{R}_m)\, dV. \tag{8.81}$$

Here we have already taken advantage of the fact that the overlap of the wave functions is small, so we can set $\int \Phi_\mathbf{k}^* \Phi_\mathbf{k}\, dV \approx 1$ in the denominator of equation (8.78). When evaluating the integral, the following aspects are important for the individual terms: The integral $\int \tilde{\psi}^* H_A \tilde{\psi}\, dV$ reflects the eigenvalue E of the isolated atoms, which we assumed to be known. We split the integral $\int \tilde{\psi}^* H_S \tilde{\psi}\, dV$, in which the perturbation H_S by the neighbors is taken into account, into two parts. With the abbreviation α we refer to the shift of the energy eigenvalue of the atom due to the potential change caused by the neighbors:

$$\alpha = -\int \tilde{\psi}^*(\mathbf{r} - \mathbf{R}_n)\, H_S(\mathbf{r} - \mathbf{R}_m)\, \tilde{\psi}(\mathbf{r} - \mathbf{R}_n)\, dV . \tag{8.82}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The abbreviation β represents the energy change caused by the overlap of the wave function of the reference atom with the wave functions of the other atoms:

$$\beta_n = -\int \bar{\psi}^*(\mathbf{r}-\mathbf{R}_n)\, H_S(\mathbf{r}-\mathbf{R}_m)\, \bar{\psi}(\mathbf{r}-\mathbf{R}_m)\, dV \,. \tag{8.83}$$

Since the procedure is the same for all eigenvalues E_i, we reintroduce the index i, characterizing the different energy levels, and write:

$$E_{\mathbf{k},i} \approx \frac{1}{N}\sum_{m,n} e^{i\mathbf{k}\cdot(\mathbf{R}_m-\mathbf{R}_n)}(E_i - \alpha_i - \beta_{i,n}) \,. \tag{8.84}$$

When calculating the energy eigenvalues, we take into account that the quantities E_i and α_i only refer to the reference atom. Since the overlap of the wave function is not important for these two quantities, we may set $R_m = R_n$. Neither E_i nor α_i and $\beta_{i,n}$ depend on the reference atom m. Thus the summation over m simply results in the factor N, which just cancels out with the normalization. We therefore write:

$$E_{\mathbf{k},i} \approx E_i - \alpha_i - \sum_n \beta_{i,n}\, e^{i\mathbf{k}\cdot(\mathbf{R}_n-\mathbf{R}_n)} \,. \tag{8.85}$$

In the calculation of $\beta_{i,n}$ the overlap of the wave function of the reference atom with the wave functions of the neighboring atoms is explicitly taken into account. If the electrons are tightly bound and thus strongly localized, we do not need to sum over the contributions beyond those of the nearest neighbors. The value of β_i is then determined by the number of neighboring atoms, the strength of the perturbation potential, and above all by the overlap of the wave functions.

The overlap of the wave functions between neighboring atoms, which we have repeatedly mentioned, depends on the type of the bond and the crystal structure. Therefore, the constants β_i differ, even when only the nearest neighbors are being considered. A particularly simple situation can be found in cubic crystals. Furthermore, if the bonding is based on s-wave functions, β is not direction-dependent and has the same value for all neighbors.

As a simple example we consider a simple-cubic lattice, where there are six nearest neighbors with coordinates $(\mathbf{R}_m - \mathbf{R}_n) = (\pm a, 0, 0)$, $(0, \pm a, 0)$ and $(0, 0, \pm a)$. The energy eigenvalues then follow from (8.85)

$$E_{\mathbf{k},i} \approx E_i - \alpha_i - 2\beta_i \left[\cos(k_x a) + \cos(k_y a) + \cos(k_z a)\right] \,. \tag{8.86}$$

Similar expressions can be found for body-centered and face-centered cubic lattices, where the number of nearest neighbors is correspondingly larger.

From (8.86) it is clear that the interaction between neighboring atoms leads to an energetically lowered band of width $12\beta_i$ which replaces the discrete energy levels E_i of isolated atoms. The transition from discrete atomic levels at infinite atomic separation to bands at normal atomic spacings for solids is shown in Figure 8.22 for the case discussed here. Some interesting points should be mentioned in this context: since

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the perturbation potential is negative and the integral (8.82) contains only the wave function of the considered atom, the quantity α_i is positive. It describes the average decrease of the energy levels due to the change of the potential caused by the neighboring atoms at the location of the reference atom. As a result, α_i increases with decreasing interatomic distance. Since the value of β_i is mainly determined by the overlap of the wave functions, the width of the bands depends on it. Because electrons are more localized at lower energy levels, they "see" less of their surroundings and the band width is correspondingly smaller. In contrast, higher bands originating from the outer electron shells can even overlap. Whether a maximum or minimum of the dispersion curves occurs at the Γ point depends on the sign of β_i, which in turn is determined by the sign of the overlapping wave functions ψ_m and ψ_n.

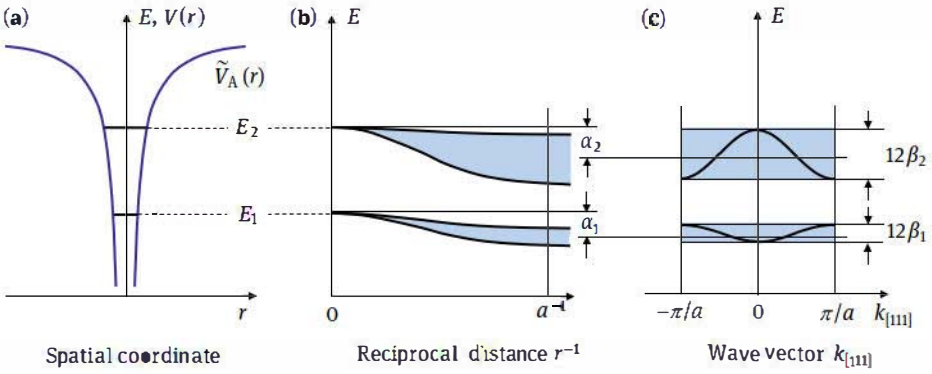


Fig. 8.22: Illustration of the model of strongly bound electrons. **a)** Position of the levels E_1 and E_2 in the potential $\tilde{V}_A$ of an isolated atom, **b)** lowering and broadening of the levels as a function of the reciprocal atomic distance a^{-1}, **c)** dispersion curve E_k along the [111] direction shown in the reduced zone diagram.

The dispersion relation (8.86) can be expanded at the Γ point for small wave numbers. Since the magnitude of the wave vector is given by $k^2 = k_x^2 + k_y^2 + k_z^2$ for the energy eigenvalues we find:

$$E_{\mathbf{k},i} \approx E_i - \alpha_i - 6\beta_i + \beta_i a^2 k^2 \,. \tag{8.87}$$

Disregarding the zero point energy, the resulting dispersion relation corresponds to that for the free electron gas. At small wave numbers, electrons in cubic crystals behave like in an isotropic medium. If one compares (8.87) with the dispersion relation $E = \hbar^2 k^2/2m$ for free electrons, we can see that we assign the electrons in this crystal the **effective mass**

$$m_i^* = \frac{\hbar^2}{2\beta_i a^2} \,. \tag{8.88}$$

It is proportional to $1/\beta_i$ and therefore increases with decreasing bandwidth. Its sign depends on the sign of β_i and can be positive or negative! We will explain the physical

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

meaning of this statement in Section 9.1. Here we only point out that a corresponding expansion is also possible at the maximum of the dispersion curve. In this case, we find the sign of the effective mass opposite to that given by (8.88).

The calculation of the band structure using the strongly-bound electron model shows that the appearance of energy bands is not a result of the periodicity of the lattice. Bloch functions were only used for mathematical simplicity in the assumed wave functions Φ_k. Since only those atoms in the immediate vicinity of the reference atom have to be taken into account, the calculation can be extended to amorphous materials leading to similar results. This fact should be emphasized, because it explains why electronic energy bands also occur in amorphous solids. The bands are not a consequence of the periodic arrangement of the atoms but of the interactions between them.

As examples for this approximation, we consider the results for the ionic crystal KCl and briefly touch on the somewhat more complicated situation of the tetrahedrally-bound semiconductors. In the case of ionic crystals, we would expect the width of the bands to be relatively small compared to their spacing, since the ions of these crystals have closed electron shells and the wave functions hardly overlap. As Figure 8.23 shows for the case of potassium chloride, this is confirmed by the calculation. Only at small ion distances does the band broadening become noticeable. At the equilibrium distance R_0 of the ions, only the uppermost level is significantly broadened. An interesting point, which we will discuss in more detail, is that the s-bands are completely occupied by two, and the p-bands by six electrons per atom.

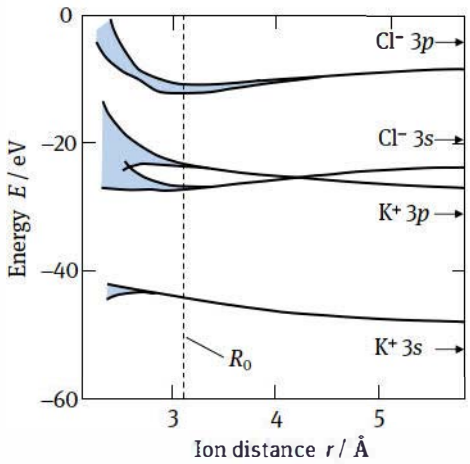


Fig. 8.23: The dependence on the ion separation for the four highest occupied energy bands of KCl. The energies of the eigenstates of the free ions are marked by arrows. The equilibrium distance $R_0 = 3.15$ Å is indicated by the dashed line. (After L.P. Howland, Phys. Rev. **109**, 1927 (1958).)

On the other hand, with covalent bonding, the overlap of the wave functions is relatively large. This leads to pronounced bands, as shown for the example of the tetrahedrally coordinated diamond in Figure 8.24. As discussed in Section 2.4, when the atoms from

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the s- and p-states approach each other, sp^3-hybrid orbitals with binding and antibinding states are formed. At the equilibrium distance R_0, the four electrons available per atom sit in the *valence band*, which is completely filled. The *conduction band*, which is separated from the valence band by an energy gap E_g of 5.5 eV, is empty.

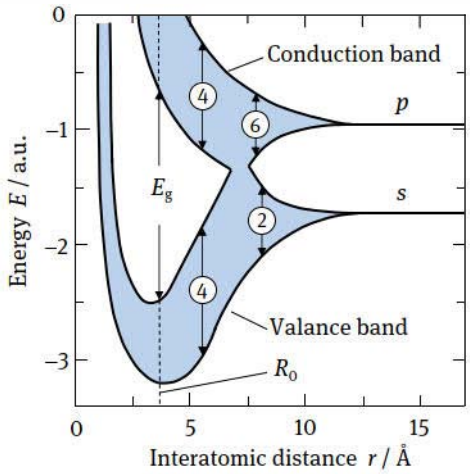


Fig. 8.24: The dependence of the band structure on the interatomic distance in diamond. The numbers in circles indicate the number of occupable states per atom. The equilibrium distance and energy gap take the values $R_0 = 3.57$ Å and $E_g = 5.5$ eV, respectively. (After A.H. Wilson, *The Theory of Metals*, Cambridge Univ. Press, 1965.)

Of course, the approximations we are discussing here have different applicabilities to different substances. If there is a very strong overlap of the wave functions, as in the case of metals, the description using quasi-free electrons in a periodic potential is a good approximation. Conversely, for substances with covalent bonds or for ionic crystals, i.e. whenever the valence electrons are strongly localized, the approximation using strongly bound electrons is much more suitable. In practice, the real situation often lies between these two limiting cases. In order to take this fact into account, improved methods have been developed by which the band structure of real solids can be reliably calculated.

8.5 Energy Bands

8.5.1 Metals and Insulators

In light of the band model, we will now briefly address the question of the parameters that determine whether a solid is metallic or insulating. The answer to this question depends crucially on whether the existing bands are completely or partially filled with electrons. As we show in the next chapter, full bands do not contribute to electrical conductivity, which is why such solids are insulators. Although this is incomprehensible in the framework of the free electron gas approximation, we will use this knowledge

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

in anticipation to discuss qualitatively some typical features of band structures. We assume the validity of Figure 8.20, in which a one-dimensional crystal was considered. Although this strong simplification is not justified in many cases, the basic principles become clear with this following.

The interaction of the electrons with each other or with the lattice neither creates nor destroys states. For example, for a system of N interacting atoms, a band of N states is found for every possible electronic state of the isolated atoms. This means that, taking into account the spin orientation from the s- and p-states of the isolated atoms, bands with $2N$ and $6N$ states are formed. We first consider alkali or noble metals with a monatomic base and *one* conduction electron per atom. The lower **valence bands** are fully occupied. In the **conduction band**, which is formed by s-electrons, there is space for $2N$ electrons. However, since there is only one electron per atom, the conduction band is half filled. This is typical of simple *metals* and is shown schematically in Figure 8.25a. Here the energy dispersion curves are plotted, with the occupied states indicated by thicker lines, and the position of the Fermi energy is also indicated. If the dependence of the energy on the wave vector is not important, the bands are often represented symbolically by boxes as shown by the right-hand picture. The energy is plotted on the ordinate.

If the number of electrons per atom is even, as assumed in Figure 8.25b, the lower bands are full and the upper band empty. Such solids are *insulators*. The situation is

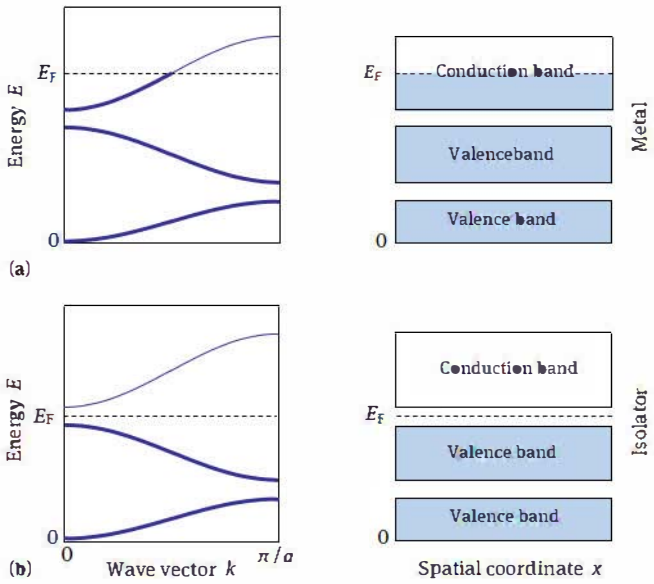


Fig. 8.25: Occupation of the bands of monatomic-based crystals. **a)** The states of the two valence bands are fully occupied, the conduction band of the *metal* is only half filled. **b)** The states of the two valence bands are fully occupied, the conduction band of the *insulator* is empty.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

similar in the case of the semiconductors silicon or germanium. Since there are two tetravalent atoms in the elementary cell, eight valence electrons are available per unit cell in these materials. The dependence of the band broadening on the interatomic distance in the case of diamond in Figure 8.24 shows that the valence band is completely filled with 2×4 electrons (or with four electrons per atom), i.e. the conduction band is empty. At absolute zero, these materials are therefore also insulators. Their semiconducting properties, which we discuss in detail in Chapter 9, are based on the small size of the gap.

However, this picture is not necessarily correct in two- or three-dimensional systems, because according to the above argumentation there should be no bivalent metals. However, in most cases, as shown in Figure 8.26, the dispersion curves are different in different crystal directions. Thus, bands which, if only viewed in a single direction, are separated by an energy gap, but when all directions are taken into account in fact may overlap. This overlap results in partially filled bands and thus in solids with metallic character, such as magnesium. If the overlap is small, only a few electrons are available for current transport. These materials are known as **semi-metals**, examples being arsenic, antimony and bismuth.

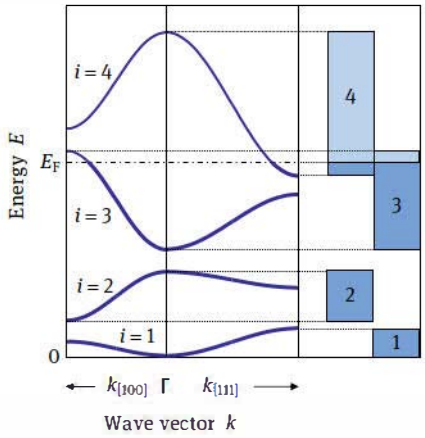


Fig. 8.26: The significance of the directional dependence of the dispersion curves. The two upper bands overlap. Therefore electrons from band $i = 3$ cross into band $i = 4$. If the overlap is small, the material is called a *semi-metal*.

8.5.2 Brillouin Zones and Fermi Surfaces

In the discussion so far, we have ignored the fact that the shape of the Brillouin zone has a considerable influence on the Fermi surface. Taking as an example a two-dimensional square lattice, the first three Brillouin zones of which are shown in Figure 8.27, the relationship between structure and electronic properties is particularly clear.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

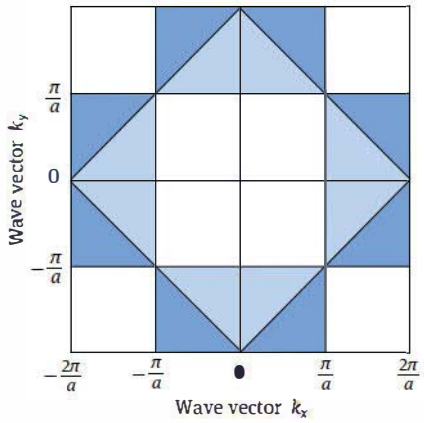


Fig. 8.27: The first three Brillouin zones of a square lattice. The large white square in the center represents the first Brillouin zone with boundaries $k_x = \pm\pi/a$ and $k_y = \pm\pi/a$. The second Brillouin zone is indicated in light blue, the third in dark blue.

Now imagie introducing electrons into the first Brillouin zone. In the free-electrons model, the Fermi surface is the surface of a sphere with radius $k_F = (3\pi^2 n)^{1/3}$. If the Fermi sphere lies completely within the first Brillouin zone, as shown in Figure 8.28a, then the Fermi surface near the zone boundary is of secondary importance. However, if $k_F > \pi/a$, as shown in Figure 8.28b, then the Fermi sphere extends into the second Brillouin zone, which now contains some of the occupied states, which are thus in the next higher band.

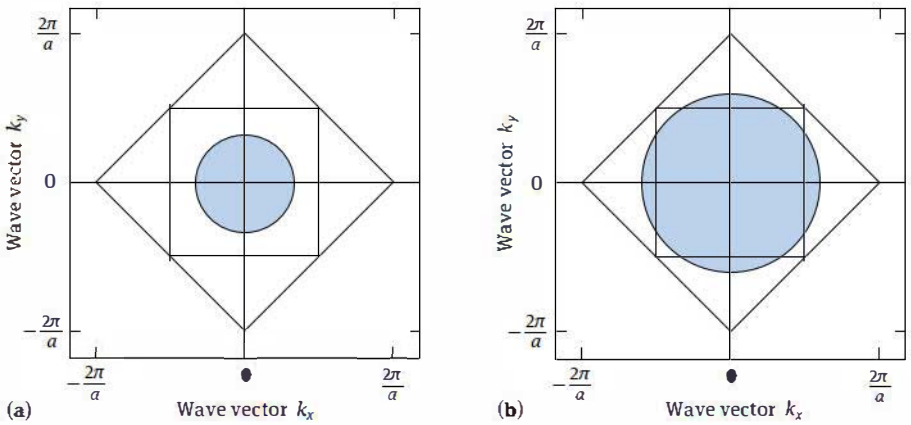


Fig. 8.28: The Brillouin zones and Fermi sphere (colored blue) of a square lattice in the free-electron model. **a)** The Fermi sphere lies completely within the first Brillouin zone, since $k_F < \pi/a$. **b)** The Fermi sphere extends into the second Brillouin zone.

The reduction to the first Brillouin zone (as discussed earlier in the chapter) leads to Figure 8.29. The first Brillouin zone is not completely filled with electrons because there are still free spaces in its corners. In other words the first band is only partially filled,

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

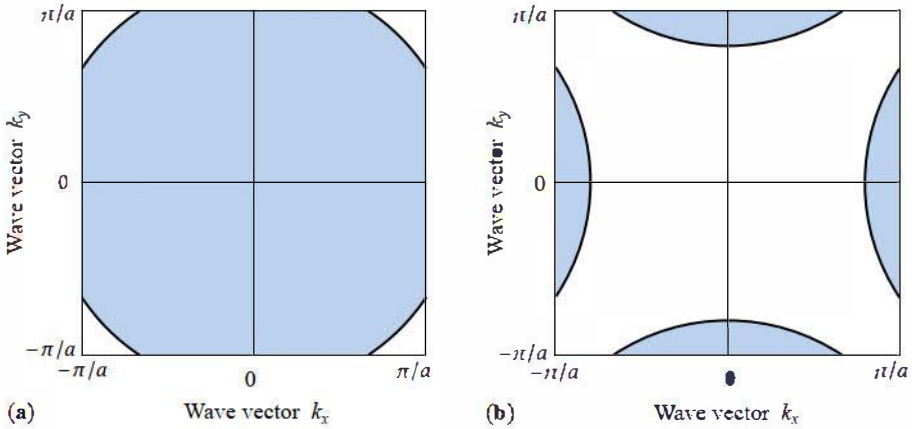


Fig. 8.29: Illustration of the Fermi sphere with $k_F > \pi/a$ in the reduced zone scheme. Note that the blue-tinted electron states in the two pictures are in different bands. **a)** In the corners of the first Brillouin zone, states are still unoccupied. **b)** By folding back the states of the second Brillouin zone, "electron pockets" have been created.

and the electrons in the second Brillouin zone form a partially occupied second band. Since the wave functions of electrons in different bands are orthogonal to each other, the electrons do not "see" each other (in the one-electron approximation), even if they have the same wave vectors after refolding.

In real metals, the periodic potential of the lattice causes an energy gap at the boundary of the Brillouin zone. Figure 8.30 shows lines of constant energy for the first three and a part of the fourth Brillouin zone of a square lattice. Since the solutions of the Schrödinger equation are standing waves at the zone boundary, the group velocity $\partial\omega/\partial k = (1/\hbar)\nabla_k E$ goes to zero there. The gradient $\nabla_k E$ therefore runs parallel to the

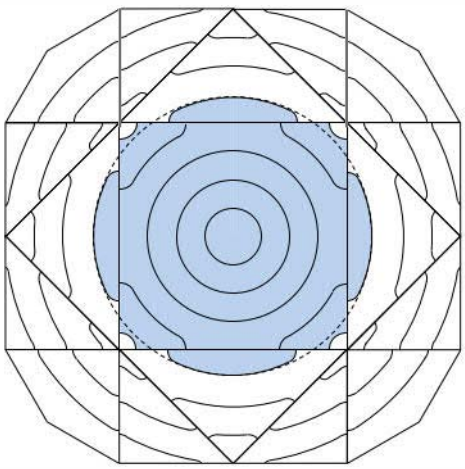


Fig. 8.30: Lines of constant energy taking into account the interaction of the electrons with the periodic lattice. The first three Brillouin zones and a part of the fourth are shown in the extended zone scheme. The occupation of the states for an arbitrarily chosen electron density is indicated by the blue coloration.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

boundaries of the Brillouin zone, the lines of constant energy are perpendicular to the boundary. Thus the Fermi sphere is deformed near the zone boundary. This effect is naturally directly related to the bending of the energy-dispersion curves at the boundary of the Brillouin zone. For an arbitrarily assumed electron density, approximately corresponding to that shown in Figure 8.29b, the occupancy of the states is indicated by blue coloration. Thus the Fermi surface has discontinuities at the Brillouin zone boundaries, as a consequence of the existing energy gaps.

The bending of the Fermi surface near the zone boundary allows additional electrons to be introduced into the first Brillouin zone. Since the electrons there have a lower energy than in the next higher zone, the total energy of the solid is lowered. The dashed circle and the blue area have equal surface areas, since the interaction of the electrons with the lattice does not change the number of occupied states nor their density in **k**-space.

For illustration, Figure 8.31 shows lines of constant energy for the second Brillouin zone in the extended, reduced and periodic zone scheme. From the right-hand illustration it is clear that extrema occur in the second Brillouin zone with the grey areas having an energy maximum at their centers and the blue areas a minimum.

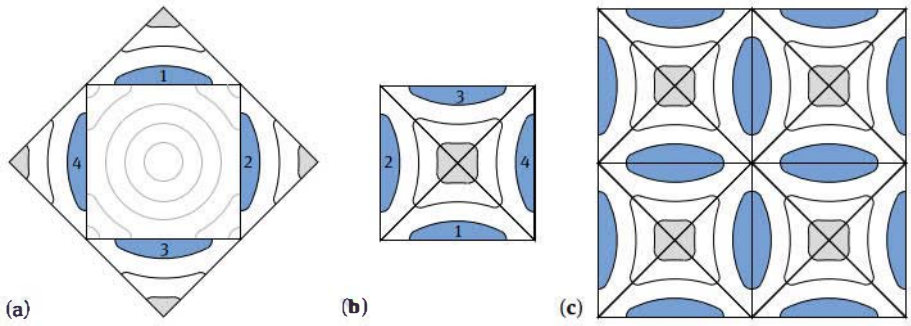


Fig. 8.31: Representations of the lines of constant energy for the second Brillouin zone in the extended **(a)**, reduced **(b)** and periodic **(c)** zone scheme. From **(c)** we see that there is an energy maximum at the centers of the grey areas and in the blue areas an energy minimum.

The knowledge we have obtained in the discussion of two-dimensional structures can be directly applied to real, three-dimensional metals. As an example we choose aluminum, whose band structure can be described with the help of quasi-free electrons and was already shown in Figure 8.16. Aluminum is trivalent and has a face-centered cubic structure. The states of the first Brillouin zone are completely filled, those of the next two partially so. The Fermi surface of the second zone is shown in Figure 8.32a in the reduced zone scheme. In this representation the Fermi surface is closed, but it should be noted that the states in the zone outside the surface are occupied but those inside are empty. The states in the third zone (Figures 8.32b and 8.32c) form a system

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

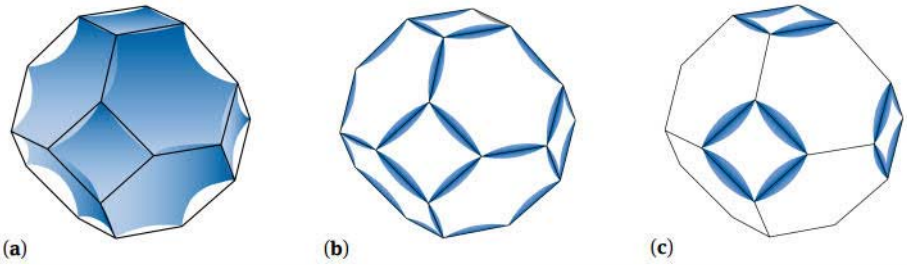


Fig. 8.32: Fermi surfaces of a trivalent metal with a weak periodic potential, shown in the reduced zone scheme. **a)** Second Brillouin zone. States in the volume area outside the blue-tinted surface are occupied but those inside are empty. **b)** Third Brillouin zone in the quasi-free electron model. **c)** Third Brillouin zone for weak interactions, as in the case of aluminum. (After R. Lück, dissertation, TH Stuttgart (1965).)

of interconnected "tubes" in the free-electron model. However, the interaction with the lattice causes these tubes to transform into ring-like structures, and their parts are no longer connected.

8.5.3 Density of States

For many purposes, especially in the case of metals, it is sufficient to know the density of states $D(E)$ instead of the total information contained in the band structure. According to equation (8.7) the density of states is given by:

$$D(E)\,\mathrm{d}E = \frac{2}{(2\pi)^3}\int_{E}^{E+\mathrm{d}E} \mathrm{d}^3k = \frac{2}{(2\pi)^3\hbar}\,\mathrm{d}E \int_{E=\text{const.}} \frac{\mathrm{d}S_E}{v_g}\,, \tag{8.89}$$

where $\mathrm{d}S_E$ stands for an element of the energy surface $E_{\mathbf{k}}$. As with the density of phonon states, the most important contributions come from regions in $\mathbf{k}$-space where the energy surfaces are horizontal. At these places one finds Van-Hove singularities in the density of states. The correlation between the maxima of the density of states and extrema in the band structure is shown in Figure 8.33. This figure displays the $E_{\mathbf{k}}$-curves of copper for directions of high symmetry and the corresponding density of states. The different contributions of *s*- and *d*-electrons to the density of states already mentioned in Section 8.1 in the discussion of Figure 8.9 are clearly visible. In copper, the *s*-band is approximately 12 eV wide and overlaps with a series of high, relatively narrow maxima originating from the *d*-bands. The *s*-band is only partially occupied, so that copper has the typical metallic properties.

The most important method for the measurement of band structures and densities of states is **photoemission spectroscopy** also known as **photoelectron**

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

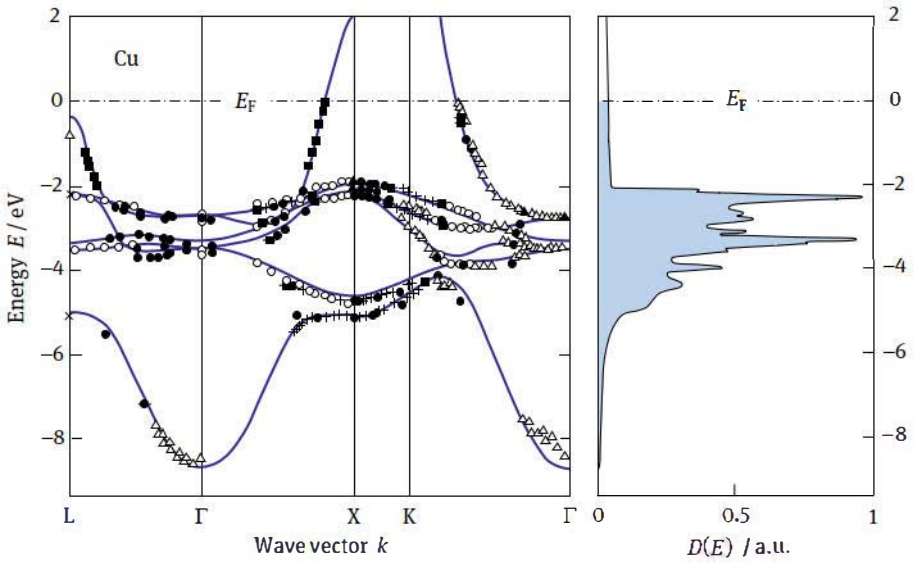


Fig. 8.33: Energy dispersion curves E_k (After R. Courths, S. Hüfner, Phys. Rep. **112**, 53 (1984).) and density of states (After H. Eckhardt et al., J. Phys. **F14**, 97 (1984).) for copper. The experimental data were provided by various authors, the E_k curves and the density of states were calculated by H. Eckhardt et al.

spectroscopy developed by K. Siegbahn[12]. Here the solid is irradiated with ultraviolet light or X-rays and the numbers and energies of the emerging photoelectrons is measured. Depending on the radiation source used, the technique is known as UPS (**UV** **P**hotoemission **S**pectroscopy) or XPS (**X**-ray **P**hotoemission **S**pectroscopy). The basic setup is shown in Figure 8.34. Gas-discharge lamps with photon energies between 20eV and 40eV are used as light sources. In the X-ray range, in addition to classical X-ray tubes, synchrotron radiation is playing an increasingly important role. Electrostatic analyzers for electrons are often used for energy discrimination. By changing the voltage between the cylindrical deflection plates, the energy of the electrons which can pass through the exit slit can be tuned. In the figure a 127°-angle cylindrical sector analyzer is drawn.

The emerging photoelectrons come from depths between 5 Å and 50 Å, depending on their kinetic energy. Consequently, the measurement results depend strongly on the surface condition, so that photoemission spectroscopy has also become an extremely important tool in surface physics. Information on the bulk of the sample can be obtained so long as the surface does not differ significantly from the bulk of the solid in terms of composition and structure. In all cases, a well-defined state of the surface is required. Consequently such measurements must be carried out in ultra-high vacuum.

12 Kai Manne Börje Siegbahn, *1918 Lund, †2007 Ängelholm, Nobel Prize 1981

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

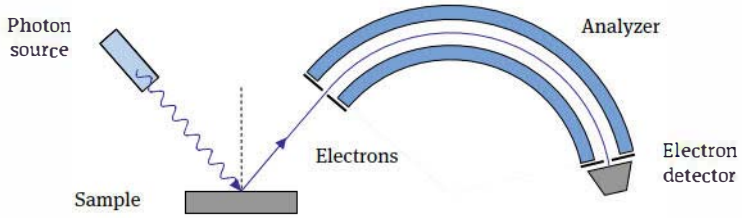


Fig. 8.34: Scheme of the measurement setup for determining the electronic density of states by photoemission spectroscopy.

As shown in Figure 8.35, the internal photoelectric effect excites electrons from the occupied to the empty states above the vacuum level E_{Vac}. If $\hbar\omega$ is the energy of the incident photons, Φ is the work function, E_{kin} the kinetic energy of the emitted electrons and E_{b} their binding energy with respect to the Fermi energy, then the energy balance is:

$$\hbar\omega = \Phi + E_{\mathrm{kin}} + E_{\mathrm{b}} \,. \tag{8.90}$$

The number of emitted electrons is measured as a function of their kinetic energy. Since the probability of excitation by a photon within a band generally depends only weakly on the energy of the state, the number of electrons recorded reflects the density of states of the solid. The highest kinetic energy $E_{\mathrm{kin}}^{\mathrm{max}}$ is carried by those electrons which have been lifted from states at the Fermi energy into the vacuum. Superimposed on the spectrum is a background of electrons that have already undergone inelastic scattering before leaving the solid. If the measurement is angle resolved (known as ARPES from

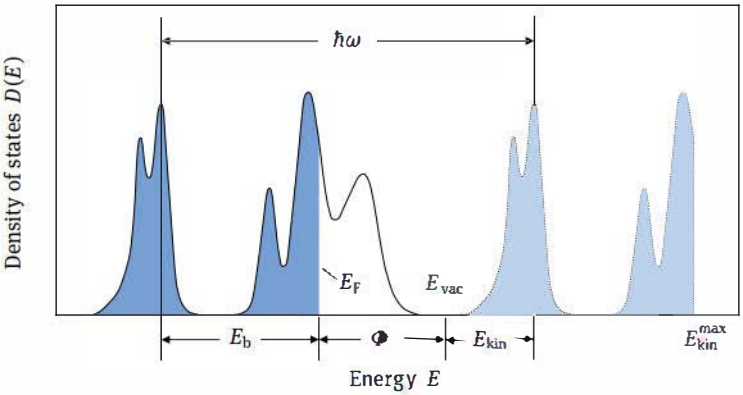


Fig. 8.35: The determination of the electronic density of states by photoemission spectroscopy. The density of states of two separate bands is shown with the occupied part indicated in medium blue. The electrons from the occupied states are excited into continuous states above the vacuum level E_{Vac} by the incident photons, creating a copy of the initial filled occupied states indicated by light blue. $E_{\mathrm{kin}}^{\mathrm{max}}$ is the maximum kinetic energy of the excited electrons.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Angle-**r**esolved **p**hoto**e**mission **s**pectroscopy), information on the wave vector of the electrons can be obtained and the electronic band structure can be determined directly.

8.5.4 Graphene and Nanotubes

Finally, in this chapter we briefly discuss the question of the differences between "classical" solids and the two-dimensional layers of graphene and nanotubes. We begin by considering the band structures of these materials, which can be calculated with the tight-binding model. The unusual behavior of graphene is of course a consequence of its unusual structure, which we have already briefly discussed in Chapter 3. Figure 8.36a shows the hexagonal graphene lattice again. The primitive unit cell, which contains two carbon atoms, is highlighted in grey. The non-primitive reciprocal lattice (Figure 8.36b) is hexagonal, where, as we will see, the K-points play a very important role.

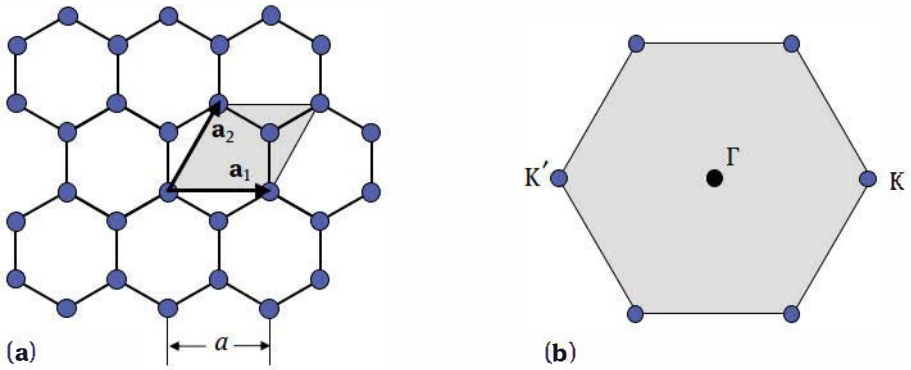


Fig. 8.36: The graphene lattice. **a)** The honeycomb structure of the real lattice has a two-atomic primitive unit cell with lattice constant $a = 2.46$ Å. **b)** The hexagonal first Brillouin zone. The K-points are of special importance.

According to the *Mermin-Wagner theorem*[13,14] two-dimensional crystals with long-range order should not exist, because they should be destroyed by fluctuations. According to this theorem, the crystals should curl up or clump together. In fact, graphene layers are corrugated, such that the fluctuations are suppressed by the anharmonic coupling of stretched and compressed regions. Graphene can also be stabilized by the interaction with the substrate when grown on the surface of three-dimensional crystals.

Although the structure does not seem to show any surprising properties, the energy dispersion curves for graphene differ fundamentally from those discussed so far. The

13 N. David Mermin, *1935 New Haven
14 Herbert Wagner, *1935

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

calculation of the band structure leads to the surprising result that at the Γ-point the valence band has a minimum and the conduction band a maximum. The shape of the dispersion curves at the K-point plays a crucial role in determining the electrical properties, because in its vicinity the bands have the shape of a double cone, as shown in Figure 8.37. The Fermi energy in pure samples coincides with the point of intersection of the two cones. At absolute zero, therefore, the conduction band is empty, while the valence band is completely occupied. The unusual shape of the dispersion curve and the fact that the valence and conduction bands meet at the K-point has considerable consequences for the properties of this unusual solid.

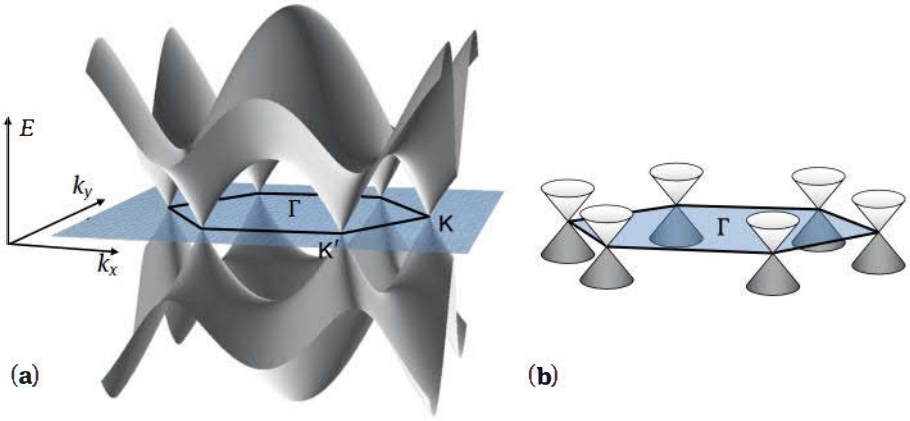


Fig. 8.37: a) Band structure of graphene (After M. Orlita and M. Potemski, Semicond. Sci. Technol. **25** 063001 (2010).). **b)** In the vicinity of the K-points, the bands have the shape of a double cone. At $T = 0$, the lower (valence) band is completely filled with electrons while the upper (conduction) band is empty.

In the vicinity of the K-point, as shown in Figure 8.37, the bands can be described by the relation:

$$E_{\mathbf{k}} = \hbar\,|\mathbf{q}|\,v_{\mathrm{F}} \tag{8.91}$$

The vector $\mathbf{q}$ is defined by $\mathbf{q} = (\mathbf{k} - \mathbf{K})$, where $\mathbf{K}$ is determined by the position of the K-point. The Fermi velocity v_{F} is given by:

$$v_{\mathrm{F}} = \frac{\sqrt{3}\gamma_0 a}{2\hbar} \tag{8.92}$$

Here γ_0 reflects the overlap of the wave functions of neighboring atoms. With $\gamma_0 = 3.2$ eV and $a = 2.46$ Å, the value of the Fermi velocity is $v_{\mathrm{F}} = 1.0 \times 10^6$ m/s.

As we will see in the next chapter, in normal solids free electrons move with different velocities. In graphene, however, the electrons on the double cone *all* move at the *same velocity*. We have therefore used the abbreviation for the Fermi velocity v_{F} in both equations (8.91) and (8.92) for the velocity. Surprisingly, the electron mass does

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

not appear in the latter expression. This means that in graphene the electrons behave formally like photons, i.e. like massless relativistic particles. But we will not go into this often used analogy further here and only mention that in the vicinity of the K-point, which in graphene is often called the *Dirac point*, the Schrödinger equation has the same form as the *Dirac-Weyl equation*[15] for massless neutrinos.

As already mentioned, in graphene the Fermi energy coincides with the intersection of the two cones. Graphene can therefore be regarded as a semiconductor with a vanishing band gap or as a semi-metal. By the application of electric fields or by doping (see Section 10.2), the level of occupation can be shifted, as shown schematically in Figure 8.38. It then lies above or below the intersection of the two cones. Accordingly, charge transport is then effected either by electrons or by "holes", holes being what we call unoccupied states in the valence band, the properties of which will be discussed in detail in the following chapters.

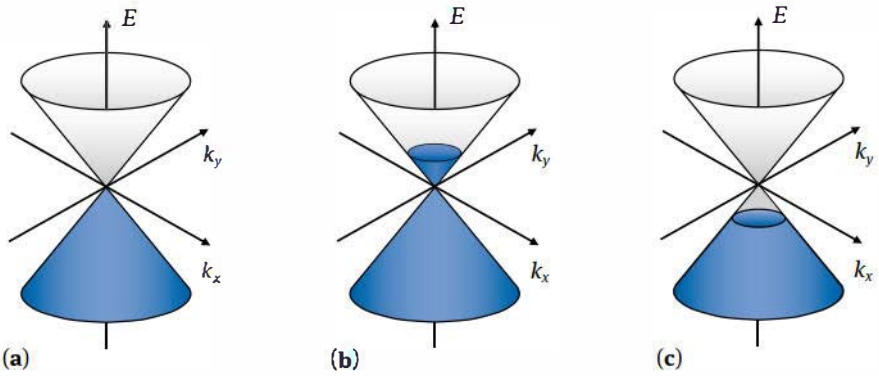


Fig. 8.38: Occupation of the bands of graphene at the K-point. The states occupied by electrons are shown in blue. **a)** Pure graphene. The Fermi energy lies at the intersection of the double cones. **b)** By applying an electrostatic voltage, states in the conduction band can be occupied. **c)** If the polarity is reversed, the valence band is partially depleted in the vicinity of the intersection.

We will now continue by asking the question how does the unusual band structure change when graphene is "rolled" into a nanotube? The answer is surprising, because the properties depend crucially on how the tube was rolled, i.e. whether it is an armchair, zigzag or chiral structure. (see Section 3.3.6)

Due to the rotational symmetry, the component $k_\perp$ of the wave vector oriented perpendicular to the tube axis can only take on discrete values. If U is the circumference of the tube, the condition

$$k_\perp = \frac{2\pi p}{U} \tag{8.93}$$

15 Hermann Klaus Hugo Weyl, *1885 Elmshorn, †1955 Zürich

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

must be fulfilled where p is an integer. The maximum value of p is determined by the number of unit cells around the circumference of the tube. The component $k_\parallel$ of the wave vector parallel to the cylinder axis is quasi-continuous due to the comparatively large tube length. This means that the allowed wave vectors in reciprocal space form straight lines parallel to the tube axis, separated by the distance $2\pi/U$ from each other. The two-dimensional band structure of graphene thus reduces to a set of discrete parallel lines, as shown in Figure 8.39.

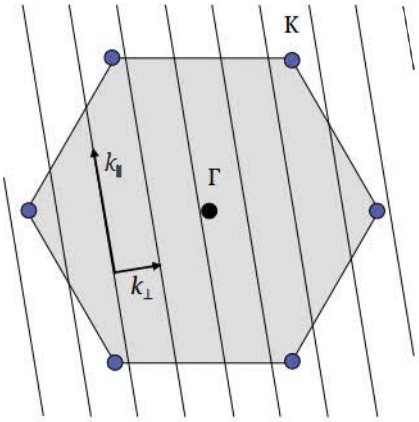


Fig. 8.39: Lines of allowed k-vectors for a nanotube in the Brillouin zone of graphene. The K-point is not touched here, so this nanotube will be, as we will see, semi-conductive.

As can be easily seen, in tubes with the armchair structure (see Figure 3.34), one of the straight lines runs through the K-point. At Figure 8.40 the corresponding dispersion curve and the corresponding density of states is sketched. The great similarity to

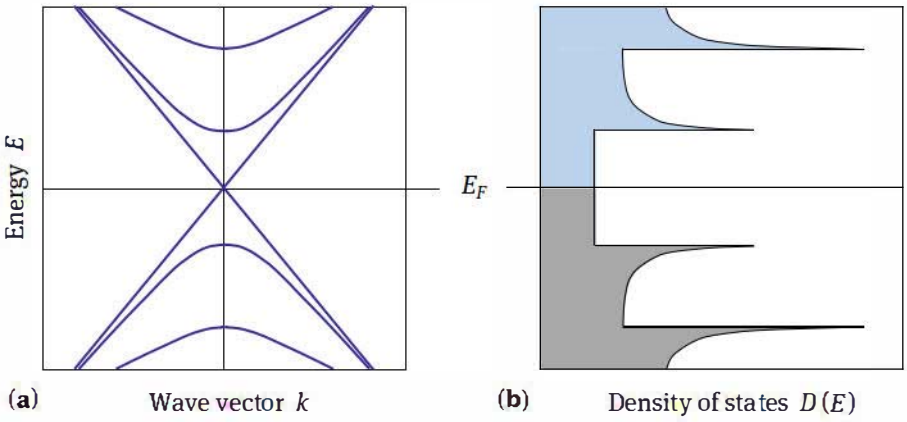


Fig. 8.40: **a)** The band structure of a carbon nanotube with armchair structure in the vicinity of the K-point. **b)** The corresponding density of states of this tube, which has metallic properties.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

graphene is obvious. However, there is a crucial difference: the density of states does not disappear at the Fermi energy, but has a finite value there. Nanotubes with an armchair structure therefore have metallic properties.

Let us now look at the dispersion curve and the density of states of tubes with a zigzag structure (see Figure 3.34), which can be seen in Figure 8.41. Obviously there is a small gap between the fully occupied valence band and the empty conduction band. The distance between the two bands is about 300 meV for a tube of radius 1.5 nm. As we will see in Chapter 10, this band gap is typical of semiconducting materials.

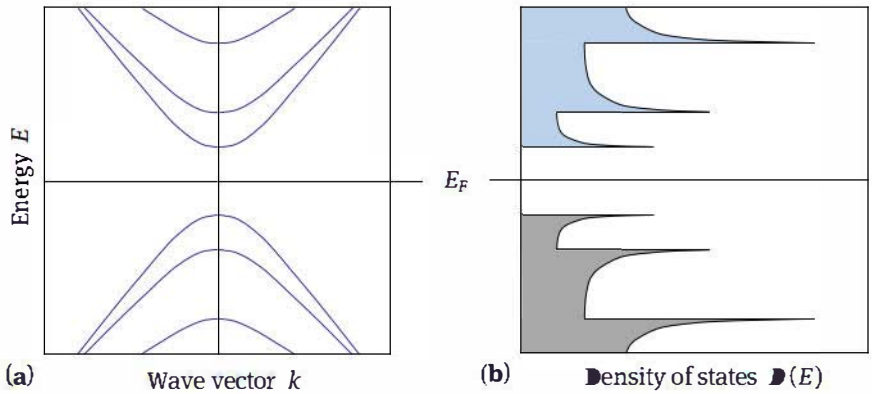


Fig. 8.41: a) The band structure in the vicinity of the K-point of a carbon nanotube with zigzag structure. **b)** The corresponding density of states. Such a tube has semiconducting properties.

It should also be noted that, depending on the "roll-off angle", nanotubes with a chiral structure have metallic or semiconducting properties. Even tubes with a zigzag structure can have metallic properties if their diameter is just large enough so that one of the straight lines in the reciprocal space cuts through the K-point. Overall, about one third of nanotubes have metallic properties, the rest have semiconducting properties.

With the help of a scanning tunneling microscope, which we described in Section 4.6, the density of states of nanotubes can be investigated. If the distance between the sample and the metal tip of the microscope is kept constant and the applied voltage U is varied, the derivative of the measured tunnel current I with respect to the voltage, i.e. dI/dU, is proportional to the density of states, as we will see in Section 11.2. The data shown in Figure 8.42 were measured on a tube with a diameter of 1.3 nm. The Van-Hove singularities which were mentioned in Section 8.1, are clearly visible. The relatively large width of the peaks is due to the interaction of the nanotube under investigation with the gold substrate.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

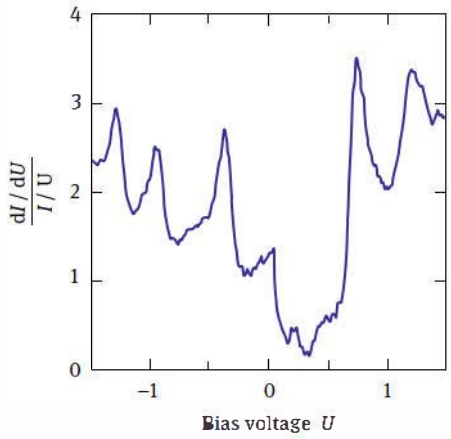


Fig. 8.42: Van-Hove singularities in the density of states of nanotubes measured with a scanning tunneling microscope. (After J.W.G. Wildöer et al., Nature **391**, 59 (1998).)

8.6 Exercises and Problems

1. Fermi Distribution. At $T = 0$ all electron states in metals are occupied up to the Fermi energy. What is the average electron velocity compared to the Fermi velocity?

2. Specific Heat of Potassium. For potassium (body-centered cubic lattice, lattice constant $a = 5.23$ Å, other material parameters are as given in Chapters 6 and 8) compare the contribution of the phonons with that of the electrons to the specific heat at room temperature. Over what temperature range does the electronic contribution dominate?

3. Free Electron Gas. Consider the two simple metals sodium (bcc) and copper (fcc) with lattice constants $a = 4.21$ Å and $a = 3.61$ Å, respectively. In both cases the Fermi surface can be described as spherical to a good approximation.
(a) Calculate the electron density and the Fermi wave vector for the two metals.
(b) Are all the electrons in the first Brillouin zone? Why is the approximation justified to treat the electrons as a free gas?

4. Pressure of the Conduction Electrons. The motion of the conduction electrons causes a pressure which can be calculated via the volume dependence $p = -(\partial U/\partial V)_{T,N}$ of the internal energy U. How large is this pressure in the case of gold at $T \approx 0$?

5. Fermi Surface and Brillouin Zones. Identical atoms with five electrons each are arranged on a two-dimensional square lattice with lattice constant a. Construct the first five Brillouin zones. How does the shape of the Fermi surfaces qualitatively change if a weak periodic potential is present? (Assume freely moving electrons in the lattice plane with no vertical motion. Take account of the spin degeneracy).

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

6. Liquid ^{3}He. Due to its high zero-point energy, ^{3}He is liquid even at $T = 0$ where it has a density of only $\varrho = 0.08\,\mathrm{g/cm^3}$. ^{3}He behaves like a Fermi gas and many of its properties can be quantitatively described using this model. Use this approximation to determine the Fermi energy, the Fermi velocity and the Fermi temperature. Calculate the specific heat for $T \ll T_\mathrm{F}$ and compare this at $T = 50\,\mathrm{mK}$ with that of copper. The effective mass $m^*_{^3\mathrm{He}}$ of a ^{3}He atom is about 2.8 times that of free ^{3}He atom.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

9 Electronic Transport Properties

The knowledge of the electronic density of states, the application of Fermi statistics and the inclusion of the interaction of the electrons with the periodic lattice potential has allowed us to explain a number of fundamental static solid state properties. We now turn to phenomena in which the *motion* of the electrons plays a crucial role. As we will see, the periodic modulation of the lattice potential and the associated band structure has rather amazing consequences for electron transport. Two of these properties have already been mentioned in the previous chapter without further explanation: fully occupied bands make no contribution to the electrical conductivity and the *effective mass* of the electrons can be positive or negative. In this chapter, we develop the concept of the effective mass (see also Section 8.4) and the concept of positively charged *holes*. Both of these concepts will be of great importance in later chapters. Subsequently, we will discuss charge and heat transport within metals. Particularly interesting phenomena are observed when the sample is located in a magnetic field. Since electrons in magnetic fields move on surfaces of constant energy, such experiments can be used to determine the shape of Fermi surfaces. Furthermore, magnetic fields restrict the motion of the electrons causing the quantization of their orbits. Consequences of this effect become especially clear in connection with the quantum Hall effect, which we discuss at the end of the chapter.

9.1 Equation of Motion and Effective Mass

9.1.1 Electrons as Wave Packets

When deriving the density of states, we treated electrons as waves. The question arises: to what extent are classical equations (such as Newton's second law) applicable when considering wave-like electrons in solids? It is known from quantum mechanics that particles cannot be precisely localized simultaneously in position and momentum space. Since the inequality $\delta k\, \delta r > 1$ must always be fulfilled, the position vector $\mathbf{r}$ and the wave vector $\mathbf{k}$ cannot be simultaneously and accurately defined. However, a limited *localization* of the wave function with respect to position and momentum can be achieved by the superposition of states with different wave vectors. For example, free electrons can be represented as wave packets by the superposition of plane waves. The wave function $\psi(\mathbf{r}, t)$ then takes the form:

$$\psi(\mathbf{r}, t) = \sum_{\mathbf{k}} g(\mathbf{k})\, \mathrm{e}^{\mathrm{i}\left(\mathbf{k}\cdot\mathbf{r} - \hbar k^2 t/2m\right)} \,. \tag{9.1}$$

If the coefficients $g(\mathbf{k})$ are suitably distributed (e.g. Gaussian) within an interval δk, this leads to the desired spatial localisation. Wave packets are therefore suitable for the representation of electron motion in free space.

https://doi.org/10.1515/9783110666502-009

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In the same way, localized electrons in a solid can be described by using Bloch waves rather than plane waves. The ensuing wave packets formed in this way allow a simple and vivid **semiclassical description** of the electron motion. Since the wave vector should be relatively well defined, the wave packets must be constructed such that their uncertainty in **k**-space is much smaller than the extent of the Brillouin zone. Therefore, the semiclassical description is only applicable if the spatial extension δx of the wave packet is much larger than the lattice spacing a, i.e. the electron extends over several unit cells. At the same time, the wavelength λ of an externally applied field must be large compared to the width of the wave packet so that the packet reacts as a compact particle. The characteristic dimensions are shown schematically in Figure 9.1, although we should note that the actual differences between the length scales are much larger than can be shown in such a figure.

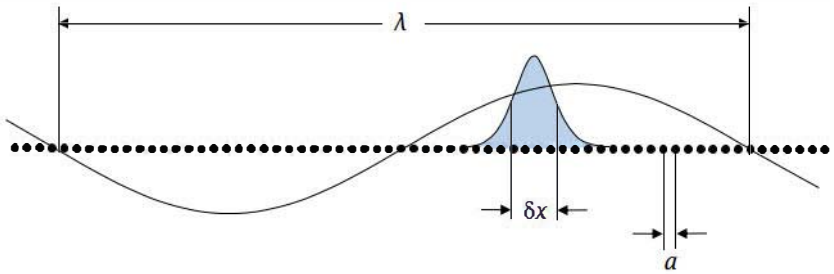


Fig. 9.1: Typical relative length scales in the semiclassical approximation. The constraints that the lattice constant $a \ll$ width δx of the wave packet, and $\delta x \ll$ the wavelength λ of the applied external field must apply. The proportions are not to scale.

With these conditions, the time dependence of the position vector **r** and the wave vector **k** of an electron can be described as a wave packet even in the presence of an external electric field $\mathcal{E}$ and/or a magnetic field **B**. The velocity of the wave packet center-of-mass is given by the group velocity, which follows directly from the dispersion relation $E_n(\mathbf{k})$:[1]

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{v}_n(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}} E_n(\mathbf{k}) = \frac{1}{\hbar}\frac{\partial E_n(\mathbf{k})}{\partial \mathbf{k}} , \tag{9.2}$$

where n is the index of the band under consideration. The velocity is unambiguously linked to the energy surface $E_n(\mathbf{k})$, and no further parameters are needed in the equation. For free electrons with a parabolic dispersion relation $E = \hbar^2\mathbf{k}^2/2m$ the group velocity is $\mathbf{v}_g = \hbar\mathbf{k}/m$.

1 The notation $E(\mathbf{k})$ used here is intended to indicate that these are not the eigenvalues $E_{\mathbf{k}}$ of the Bloch functions, but that they refer to the wave-packet energy.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

If a force **F** acts on an electron, its wave vector changes and thus also the *quasi momentum* $\hbar\mathbf{k}$ according to the equation

$$\hbar\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = \mathbf{F} = -e\left[\mathcal{E}(\mathbf{r},t) + \mathbf{v}_n(\mathbf{k}) \times \mathbf{B}(\mathbf{r},t)\right]. \tag{9.3}$$

This is the **semiclassical equation** of motion, which is important for the description of the motion of electrons in solids. For free charged particles, this equation is well known. However, the derivation for the case of electrons interacting with the lattice is complicated and can be found in theoretical solid state physics textbooks. As with the Bloch functions, the wave vector is determined only up to a reciprocal lattice vector **G**, which is why $\hbar\mathbf{k}$ is called *quasi momentum* instead of momentum. We further assume that the applied fields are small enough to prevent *interband transitions* from occurring. In this case, the band index n does not change and we can omit it from the discussion. If transitions do occur between the bands at very high field strengths, this is referred to as *electrical* or *magnetic breakdown*. We will treat breakdown in semiconductors in-depth in Chapter 10.

From the definition (9.2) follows with (9.3) for the derivative of the group velocity:

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{1}{\hbar}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial E(\mathbf{k})}{\partial \mathbf{k}}\right) = \frac{1}{\hbar}\frac{\partial^2 E(\mathbf{k})}{\partial \mathbf{k}\partial \mathbf{k}}\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = \frac{1}{\hbar^2}\frac{\partial^2 E(\mathbf{k})}{\partial \mathbf{k}\partial \mathbf{k}}\mathbf{F}. \tag{9.4}$$

Thus we get for the cartesian components $\dot{v}_i$:

$$\frac{\mathrm{d}v_i}{\mathrm{d}t} = \frac{1}{\hbar^2}\sum_{j=1}^{3}\frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j}F_j = \sum_{j=1}^{3}\left(\frac{1}{m^*}\right)_{ij}F_j \tag{9.5}$$

with

$$\left(\frac{1}{m^*}\right)_{ij} = \frac{1}{\hbar^2}\frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j}. \tag{9.6}$$

Thus by using the effective mass tensor $[\mathbf{m}^*]$, the connection to the classical equation of motion $\mathbf{F} = m\dot{\mathbf{v}}$ becomes evident. The effective mass is determined by the reciprocal of the *curvature* of the energy surface. It contains the interaction of the electrons with the atomic core potentials that they experience when moving in solids. This is why we speak of a **dynamic mass**. It is remarkable that the knowledge of the energy dispersion $E(\mathbf{k})$ is sufficient to describe the electron motion, and no further details of the crystal potential are needed. The concept developed here to describe the electron motion in the crystal is called the **effective mass approximation**. From here on, we will always use the effective mass m^* instead of the free electron mass.

The usual expression $\hbar\mathbf{k} = m\mathbf{v}$ for the momentum now has the form $\hbar\mathbf{k} = [\mathbf{m}^*]\mathbf{v}$. If a force acts on a free electron, it is accelerated in the direction of the force, and the magnitude of the acceleration is independent of its wave number. In contrast, the acceleration of electrons in solids does not necessarily occur in the direction of the applied force, and the magnitude of the acceleration usually depends on the wave

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

number. In the following, we will mostly omit the tensor labeling to keep the notation simple, but we will continue to keep the tensor character of the effective mass in mind.

As can be seen directly from the definition (9.5), the tensors of the reciprocal effective mass $(1/m^*)_{ij}$ and the effective mass m^*_{ij} are symmetrical. Therefore, they can be transformed using the principal axis theorem which reduces the number of independent components to a maximum of three. In particularly simple cases, all components have the same value, and the effective mass $m^*(k) = \hbar^2/[\mathrm{d}^2E(k)/\mathrm{d}k^2]$ is a scalar quantity. The solid in question then possesses isotropic electrical properties, even if m^* still depends on the wave number. An effective mass independent of the wave number appears in parabolic bands. This simple case can be found in the vicinity of band extrema at high-symmety points since at these points the energy surfaces can be well approximated by paraboloids.

The principle relationship between the effective mass and the wave vector is shown in Figure 9.2. Starting from $k = 0$, the curvature of the band $E(k)$ decreases with the wave vector at first, and thus the effective mass increases. At the inflection points of the dispersion curve, m^* becomes infinitely large and then negative. The effective mass of the electrons is therefore positive at the band minima and negative at the maxima. Near the extrema, the effective mass is approximately constant.

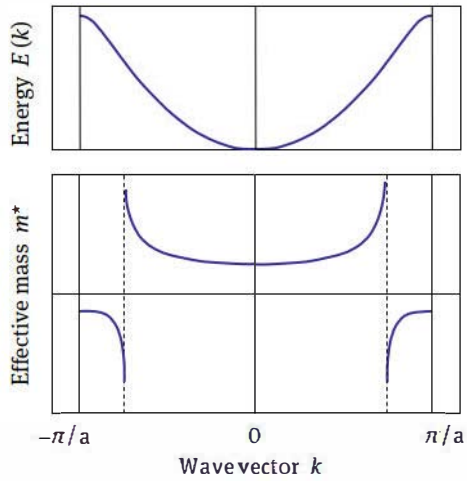


Fig. 9.2: The relationship between the dispersion function $E(k)$ and the effective mass m^* in a one-dimensional representation.

It is not surprising that the effective mass depends on the motion of the electrons. Similar observations have been made for classical systems. To illustrate this dependence, we imagine that a constant force acts on a sphere submerged in a liquid. When the force is applied at time $t = 0$, friction is not important, and the acceleration of the ball is determined by the inertial mass. After some time, the velocity of the ball becomes constant, because the friction and the external force are in balance. If the interaction of the ball with the liquid is not taken into account in the description and instead the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

concept of effective mass is applied, the effective mass increases with velocity until the ball finally has to be assigned an infinite mass because no acceleration occurs despite a force acting on it.

The energy dependence of the effective electron mass leads to surprising conclusions. If a constant electric field $\mathcal{E}$ is applied to an ideal crystal, a constant force $\mathbf{F} = -e\mathcal{E}$ acts on the electrons. According to equation (9.3), this force causes a uniform motion of the electrons in $\mathbf{k}$-space. Since the wave function and the energy eigenvalue of an electron in a crystal is periodic in $\mathbf{k}$-space, this results in a periodic change of velocity in real space. This means that the electrons perform an oscillatory motion as shown in Figure 9.3. In the ideal case of collision-less electron motion, there is no direct current conduction in an infinitely extended crystal! In such a perfect crystal at $T = 0$ the electrons would simply perform **Bloch oscillations**. Another way of looking at this is that when an electron approaches the boundary of the Brillouin zone, it experiences a Bragg reflection and moves in the opposite direction, although the force still acts in the original direction.

The period of this oscillation is easy to determine: As mentioned above, when an electric field is applied, the electrons move with constant velocity in $\mathbf{k}$-space, namely with $|\dot{k}| = e\mathcal{E}/\hbar$ through the Brillouin zone over the range $2\pi/a$. The period of oscillation T_B is therefore

$$T_B = \frac{2\pi}{a} \Big/ \frac{e\mathcal{E}}{\hbar} = \frac{h}{ae\mathcal{E}} \,. \tag{9.7}$$

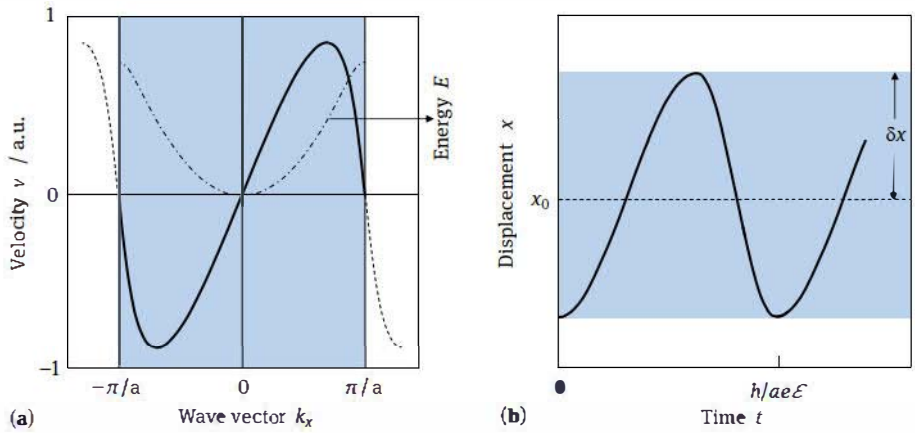


Fig. 9.3: Electron motion under the influence of a constant electric field $-\mathcal{E}_x$. **a)** The velocity (solid curve) and energy (dashed-dotted curve) as a function of the wave number of the electrons: At the boundary of the Brillouin zone, the electrons "jump" in the reduced zone scheme from π/a to $-\pi/a$. **b)** The time dependence of the spatial displacement, an oscillatory motion around the mean value x_0. As the initial condition $k(t = 0) = 0$ was chosen.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Choosing the values $\mathcal{E} = 1\,\mathrm{kV/m}$ and $a = 2\,\text{Å}$, we find the numerical values $T_B \approx 20\,\mathrm{ns}$ for the period or equivalently $\nu_B \approx 50\,\mathrm{MHz}$ for the oscillation frequency. To estimate the amplitude δx of the displacement, we can take the Fermi velocity as the mean velocity $\bar{v}$. With $\bar{v} \approx 10^6$ m/s the amplitude is approximately $\delta x \approx \bar{v} T_B/4 \approx 5\,\mathrm{mm}$.

If such Bloch oscillations did actually occur, then bulk-metallic samples would not conduct direct currents. Obviously we have taken the idealization too far, because in real materials the electrons do not pass through the Brillouin zone unhindered, but are scattered by impurities, phonons and other electrons. Since these scattering processes occur frequently, they drastically change the electron behavior. Given that the typical average scattering times (see Section 9.2) are on the order of 10^{-14} s, electrons only travel about 10 nm before scattering. The occurrence of Bloch oscillations would not only require defect-free crystals and very low temperatures, but also a negligible electron-electron interaction (cf. Section 9.2). Nevertheless, Bloch oscillations have been observed in *semiconductor heterostructures*, as we will discuss in Section 10.5. In such heterostructures, very narrow bands known as *minibands* with effective lattice constants in the range of 100 Å can be created. The expected oscillations have a much higher frequency due to large lattice constants and can be detected by ultra-fast optical spectroscopy. These oscillations can also be used to generate THz radiation.

9.1.2 Electron Motion in Bands

Although such Bloch oscillations in normal metals do not actually occur, the unusual behavior of electrons in a periodic potential has significant consequences for a number of properties, especially for charge transport.

The contribution of the individual electrons to the current density $\mathbf{j}$ depends on the electron velocity. This in turn is determined by their wave vectors according to (9.2). For the current density we write:

$$\mathbf{j} = -\frac{e}{V} \sum_{\mathbf{k}} \mathbf{v}(\mathbf{k}) \,. \tag{9.8}$$

To proceed, we convert from the sum $\sum_{\mathbf{k}}$ format to the integral $\int \varrho_k \, \mathrm{d}^3k$. We have already derived the density of states in momentum space $\varrho_k = 2V/(2\pi)^3$ in Section 8.1. Since only the occupied states contribute to the charge transport, we obtain the expression:

$$\mathbf{j} = \frac{-e}{4\pi^3} \int \mathbf{v}(\mathbf{k}) f(E,T) \, \mathrm{d}^3k \,, \tag{9.9}$$

where $f(E,T)$ is the Fermi-Dirac distribution, indicating the probability with which a state is occupied.

To obtain the simplest possible expressions, we consider the current density at absolute zero, where the Fermi-Dirac distribution is simply a step function separating occupied from unoccupied states. The integration over all occupied states then only

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

extends up to the Fermi energy, resulting in current density:

$$\mathbf{j} = \frac{-e}{4\pi^3} \int\limits_{\text{occupied}} \mathbf{v}(\mathbf{k})\, \mathrm{d}^3k = \frac{-e}{4\pi^3\hbar} \int\limits_{\text{occupied}} \nabla_{\mathbf{k}} E(\mathbf{k})\, \mathrm{d}^3k\,. \tag{9.10}$$

For the following treatment it is important that the current goes to zero in the absence of an external field. This can easily be shown for crystals with inversion symmetry. Since the reciprocal lattice has the point symmetry of the real lattice, then $E(\mathbf{k}) = E(-\mathbf{k})$. Thus the velocity of an electron with wave vector $-\mathbf{k}$ is:

$$\mathbf{v}(-\mathbf{k}) = \frac{1}{\hbar}\nabla_{-\mathbf{k}} E(-\mathbf{k}) = -\frac{1}{\hbar}\nabla_{\mathbf{k}} E(\mathbf{k}) = -\mathbf{v}(\mathbf{k})\,. \tag{9.11}$$

In other words, electrons with opposite wave vectors move in opposite directions. Without applied field for each electron with wave vector $+\mathbf{k}$ there is an electron with wave vector $-\mathbf{k}$, and thus for each electron with velocity $\mathbf{v}(\mathbf{k})$ there is also an electron with $\mathbf{v}(-\mathbf{k}) = -\mathbf{v}(\mathbf{k})$. The integral (9.10) is therefore zero, i.e. $\mathbf{j} \equiv 0$. If an electric field is applied, the argument is still valid if the band is full. Although each electron gains an additional quasi momentum, the occupation of the band remains the same. As can be seen in Figure 9.3, the electrons jump to the other side of the Brillouin zone as soon as they reach the zone boundary at $\pm\pi/a$. Thus the spatial and time averages of the electron velocity disappear when averaged over the fully occupied band.

The relation (9.11) can be extended to structures without inversion symmetry, if we note that for every electron whose spin is in the positive direction with energy $E(\mathbf{k})_\uparrow$, there is also an electron with $E(-\mathbf{k})_\downarrow$, so that an analogous approach is possible. Since the up-spin current flow in the subbands is compensated for by the down-spin flow, the general rule is that full bands do not contribute to the current flow. This is why insulators exist despite the fact that the electrons in the bands can move.

The integration in equation (9.10) can be carried out in various ways, since we can take into account not only the occupied but also the unoccupied states:

$$\mathbf{j} = \frac{-e}{4\pi^3} \int\limits_{\text{occupied}} \mathbf{v}(\mathbf{k})\, \mathrm{d}^3k = \frac{-e}{4\pi^3} \Bigg[\underbrace{\int\limits_{\text{BZ}} \mathbf{v}(\mathbf{k})\, \mathrm{d}^3k}_{\equiv 0} - \int\limits_{\text{empty}} \mathbf{v}(\mathbf{k})\, \mathrm{d}^3k \Bigg] = \frac{+e}{4\pi^3} \int\limits_{\text{empty}} \mathbf{v}(\mathbf{k})\, \mathrm{d}^3k\,. \tag{9.12}$$

In this equation we have taken the integral over all occupied states and split it into an integral over the entire Brillouin zone (BZ) and subtract the contribution of the empty states. As we have just discussed, when the band is full the integration over the Brillouin zone makes no contribution to the current. In the new approach of equation (9.12) the current appears to be transported by positive charge carriers in the empty states, i.e. those not occupied by electrons. These fictitious charge carriers are called **holes**, or **defect electrons**. The introduction of this concept may seem a little artificial at first. However, we will see that it offers many advantages, for example when the band is *almost* filled. If we wish to calculate the electrical conductivity in this case, we only need the energy dispersion near the band maximum. There, the dispersion curve can

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

usually be described by a paraboloid to a very good approximation which considerably simplifies the calculation.

If the band is not full, the application of an electric field causes a change in the velocity distribution within the band. Without scattering processes the electrons would execute Bloch oscillations as described above. The scattering processes, however, just result in a slight shift in the electron distribution in **k**-space, as we will see in the next section. Because the electric field singles out one direction, the occupation of the states in a partially filled band is no longer inversion-symmetric, so $\mathbf{j} \neq 0$. In the next section we will perform the integration and find a suitable approximation for the conductivity.

9.1.3 Electrons and Holes

Holes closely resemble positive charge carriers. Nevertheless, we should not simply treat them like positively charged particles. In the following, we will briefly explain their most important properties. In order to distinguish between electrons and holes, we mark the quantities occurring in this subsection with the index "n" or "p" for negative and positive depending on whether they are associated with electrons or holes.

With a full band, the sum over all wave vectors vanishes, i.e., $\sum \mathbf{k} = 0$. Now let us assume that, as shown in Figure 9.4a, *one* electron with wave vector $\mathbf{k}_n$ is excited from a valence band (which is full) into a conduction band (which is empty). Optical absorption can cause such a transition, an effect which we will discuss in Section 10.1 in the context of the optical properties of semiconductors.

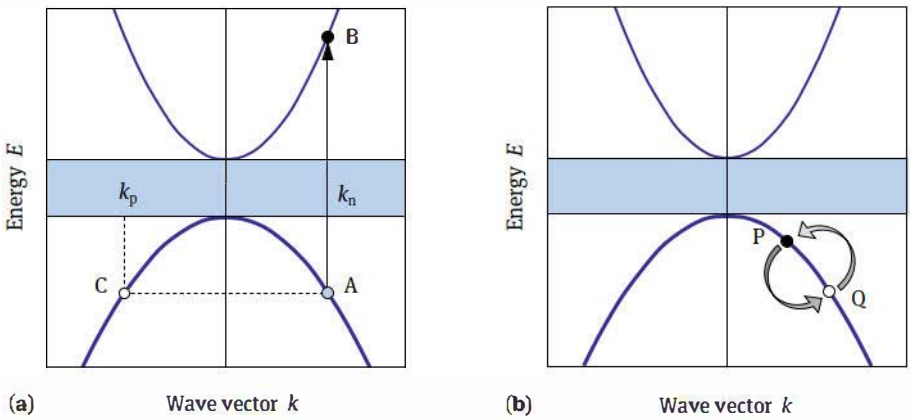


Fig. 9.4: On the concept of holes. **a)** An electron at A is excited, e.g. by optical absorption, from the full valence band into the empty conduction band, at B. The electron in the conduction band is represented by a black circle and the missing electron in the valance band by a light blue circle. The resulting hole, however, has the wave vector $\mathbf{k}_p = -\mathbf{k}_n$, i.e. that of the electron occupying the state of the open circle C. **b)** If an electron jumps (down) from P to an empty state Q, the corresponding hole is raised. The energy of the system falls, i.e. the energy difference between the two states is released.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Since an electron is missing in the valance band, we now have $\sum \mathbf{k} = -\mathbf{k}_n$. We can therefore assign to the missing electron, i.e. to the hole, the wave vector[2]

$$\mathbf{k}_p = -\mathbf{k}_n \,. \tag{9.13}$$

Some caution needs to be taken when interpreting diagram 9.4a. While, the electron is transferred from point A to point B, the corresponding hole does *not* have the wave vector $\mathbf{k}_n$ of the missing electron, but rather that of its opposite on the dispersion curve $-\mathbf{k}_n$, i.e. the electron at point C.

As shown in Figure 9.4b, a hole deeper in the valence band, for example at Q, can jump to the upper edge of the band at P, which is equivalent to the electron from P jumping to the empty state Q. During this process energy is released. Conversely, additional energy must be provided in order to move a hole from the upper edge of the valence band downwards, because an electron from deeper in the band must be raised to the edge of the valence band requiring extra energy. On the basis of these considerations it is understandable that the relation between the electron energy $E_n(\mathbf{k})$ and the hole energy $E_p(\mathbf{k})$ is:

$$E_p(\mathbf{k}) = -E_n(\mathbf{k}) \,. \tag{9.14}$$

The relation between the effective masses m^* of the two types of charge carriers follows directly from the definition (9.5) and the inversion symmetry of the bands. The masses have opposite signs, since:

$$\left[\frac{1}{m^*}\right]_p = \frac{1}{\hbar^2}\left[\frac{\partial^2 E(\mathbf{k})}{\partial \mathbf{k}\partial \mathbf{k}}\right]_p = \frac{1}{\hbar^2}\left[\frac{-\partial^2 E(\mathbf{k})}{(-\partial \mathbf{k})(-\partial \mathbf{k})}\right]_n = -\left[\frac{1}{m^*}\right]_n \quad \text{and} \quad m_p^* = -m_n^* \,. \tag{9.15}$$

Under the influence of an electric field, all electrons in $\mathbf{k}$-space move at the same velocity. The unoccupied state, which we describe here as a hole, naturally follows the motion of the electrons. This can easily be shown, because when calculating the velocity of the holes, a negative sign appears both for the gradient and for the hole energy. This means that for the velocity of the hole we have:

$$\nabla_{\mathbf{k}} E_p(\mathbf{k}) = \nabla_{\mathbf{k}} E_n(\mathbf{k}) \qquad \text{and} \qquad \mathbf{v}_p(\mathbf{k}_p) = \mathbf{v}_n(\mathbf{k}_n) \,. \tag{9.16}$$

The equations of motion for electrons and holes thus take the following form:

$$\hbar \dot{\mathbf{k}}_n = -e(\mathcal{E} + \mathbf{v}_n \times \mathbf{B}) \,, \tag{9.17}$$

$$\hbar \dot{\mathbf{k}}_p = +e(\mathcal{E} + \mathbf{v}_p \times \mathbf{B}) \,. \tag{9.18}$$

Thus looked at from outside, the hole appears to act like a particle with positive charge!

One important conclusion should be pointed out here: during charge transport, electrons and holes move in opposite directions, since the electrons are located at the lower edge of the conduction band, whereas holes are located at the upper edge of the valence band. As a result of their opposite charge, the contributions of electrons and of holes to the charge transport is additive.

2 An analogous consideration also applies to the spin that the holes appear to carry.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

9.2 Transport Properties

The Drude Model. A milestone in the understanding of the transport properties of metals came with the theory of electrical conductivity developed by *P. Drude*[3] in 1900. The theory was able to explain the linear relationship between the current and the electric field, which is the basis of Ohm's law.[4] One of the successes of Drude's theory was the derivation of the *Wiedemann-Franz law*. However, no explanation could be found for the measured small values of the specific heat and the paramagnetic susceptibility of the conduction electrons, since the Pauli principle which plays a crucial role in these effects had not yet been formulated.

Drude's theory was based on the assumption that the motion of electrons can be described with the help of kinetic gas theory. The electrons are treated as free particles moving with the thermal velocity $\mathbf{v}_{th}$ and constantly scattering from the atomic cores. The theory involves two important quantities, the **drift velocity** $\mathbf{v}_d$ and the **mean scattering** or **relaxation time** τ, both of which are included in the classical equation of motion for the electron:

$$m\frac{d\mathbf{v}}{dt} = -e\,\mathcal{E} - m\frac{\mathbf{v}_d}{\tau}\,. \tag{9.19}$$

The term $m\mathbf{v}_d/\tau$ takes the usual form for a friction or damping force and thus takes into account the damping effect of the scattering processes. The drift velocity $\mathbf{v}_d = (\mathbf{v} - \mathbf{v}_{th})$ reflects the additional velocity caused by the field. The relaxation time τ is the characteristic time with which $\mathbf{v}_d$ exponentially relaxes towards the equilibrium value $\mathbf{v}_d = 0$ after the electric field has been switched off.

In the stationary case $\dot{\mathbf{v}} = 0$, the drift velocity is given by

$$\mathbf{v}_d = -\frac{e\tau}{m}\mathcal{E} = -\mu\,\mathcal{E}\,. \tag{9.20}$$

Here we introduce the concept of the **mobility** $\mu = e\tau/m$. If n is the number density of the electrons, the current density is then:

$$\mathbf{j} = -en\mathbf{v}_d = \frac{ne^2\tau}{m}\mathcal{E} = ne\mu\,\mathcal{E} \tag{9.21}$$

and thus we have for the electrical conductivity:

$$\sigma = \frac{j}{\mathcal{E}} = \frac{ne^2\tau}{m} = ne\mu\,. \tag{9.22}$$

This allows Ohm's law to be traced back to two specific parameters of the material concerned, the electron density and the mean relaxation time. Taking typical conductivityvalues for metals, we find the relaxation time τ to be of the order of 10^{-14} s. Since

3 Paul Karl Ludwig Drude, *1863 Braunschweig, †1906 Berlin
4 Georg Simon Ohm, *1789 Erlangen, †1854 Munich

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Drude assumed that the electrons move at the thermal velocity of about 10^5 m/s, this results in a mean free path $\ell = v\tau$ of about 10 Å, a length comparable to the interatomic distance.

Note that in this derivation *all* conduction electrons are accelerated and *scattered*. However, this approach is not compatible with the fact that electrons have to follow the Fermi-Dirac distribution. As we will see, the correct quantum-mechanical calculation surprisingly leads to the same result for the conductivity.

9.2.1 Sommerfeld Theory

A much improved theory was later developed by *A. Sommerfeld*, which we have already used in discussing the specific heat of metals. As we will see, this theory explains both the electrical conductivity and the Wiedemann-Franz law. In this theory, electrons are regarded as quasi-free particles satisfying the Schrödinger equation and subject to the Pauli principle. The simplifying assumptions of the Sommerfeld theory have only limited applicability to metals with complicated band structures, so that here we mainly discuss simple and noble metals. Since the conduction bands of such metals are about half filled, their Fermi surfaces can be described to a good approximation as *Fermi spheres*.

When an external force $\mathbf{F}$ or an electric field $\mathcal{E}$ is applied, the wave vector changes according to (9.3) as follows:

$$\hbar \frac{d\mathbf{k}}{dt} = \mathbf{F} = -e\mathcal{E} \,. \tag{9.23}$$

The effect of the electric field on the Fermi sphere in $\mathbf{k}$-space is shown in Figure 9.5: with no external field, the center of the Fermi sphere and the origin of $\mathbf{k}$-space coincide at the Γ-point. However, if we apply an electric field, each wave vector and thus also the Fermi sphere as a whole starts moving by an amount (9.23)

$$\delta\mathbf{k} = \frac{-e\,\mathcal{E}\,\delta t}{\hbar} \,, \tag{9.24}$$

where δt is the elapsed time since the electric field was applied. During the time δt, the whole pattern of wave vectors moves steadily in the negative field direction (because the electron has a negative charge). The consequent displacement of the Fermi sphere is illustrated in Figure 9.5b.

Scattering processes, which we will discuss in more detail in the next section, cause the transport of electrons from the front to the rear of the Fermi sphere. Shortly after the field is applied, a dynamic equilibrium is reached between the displacement of the Fermi sphere and the rearrangements due to the scattering processes and the sphere comes to a standstill. We should emphasize that the Fermi sphere shifts only slightly, although the wave vectors of the electrons are subject to constant change according to (9.3). Since the displacement of the Fermi sphere causes only small changes in energy,

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

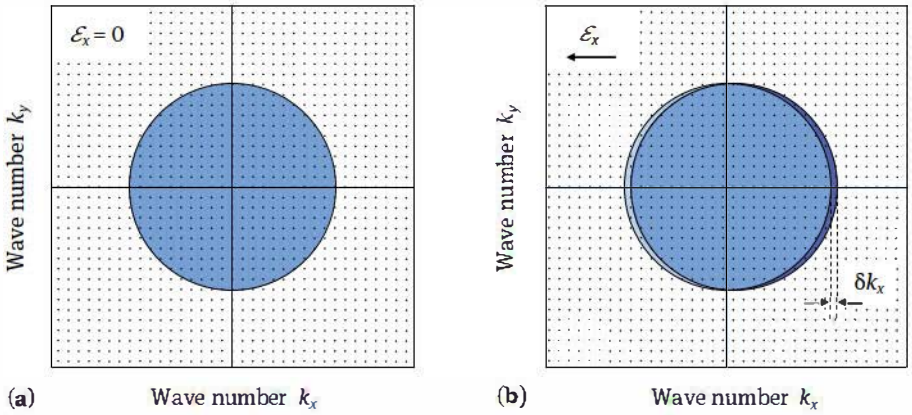


Fig. 9.5: The Displacement of the Fermi sphere in an electric field. The lattice points symbolize the allowed wave vectors. **a)** The Fermi sphere without applied electric field. The coordinate origin and the center of the sphere coincide. **b)** Under the influence of an electric field $-\mathcal{E}_x$ all wave vectors and the electron distribution are shifted to the right by δk_x. Electrons are transported from the dark-tinted area at the front side of the distribution to the light-tinted area at the rear side by scattering processes.

and scattering can only occur into empty final states, only electrons with wave vectors close to the Fermi surface can participate in this rearrangement process.

While in the absence of a field, the current flow in the positive and negative hbox$x$-direction cancels out exactly, this is no longer the case in the presence of a field. The resulting current is caused by the fast electrons of the dark tinted area or by the empty states in the lighter area. In contrast to the classical approach, where all electrons contribute to the current flow with a small drift velocity $\mathbf{v}_d$, in Sommerfeld's model only the few fast electrons at the Fermi surface are involved. Given the earlier discussion, we need to address the question of how to introduce the *mean scattering time*, i.e. the time which passes between applying the field and reaching the stationary displacement. We first discuss this question and then go into the various scattering mechanisms.

9.2.2 Boltzmann Equation

With no external electric field, the occupancy of the electronic states is determined by the Fermi-Dirac distribution function $f_0(\mathbf{k})$, where we denote the equilibrium value by the index 0. If a field is applied, the electrons are accelerated and after a short time scattering processes lead to the establishment of a new stationary non-equilibrium value $f(\mathbf{k}, \mathbf{r}, t)$.

Basically, there are three causes for changes in the distribution function: diffusion arising from fluctuations in the spatial electron density, the action of external fields and scattering processes. The temporal and spatial development of the distribution function

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

can be described by the **Boltzmann equation** (or Boltzmann transport equation). To derive this equation, we consider the changes of the distribution function over a short time interval dt. According to *Liouville's theorem*[5] of classical mechanics, in the absence of scattering processes the density in phase space is maintained. For the distribution function we can therefore write:

$$f(\mathbf{k} + d\mathbf{k}, \mathbf{r} + d\mathbf{r}, t + dt) = f(\mathbf{k}, \mathbf{r}, t) \,. \tag{9.25}$$

This equation can be rewritten for small changes of the variables as:

$$f(\mathbf{k} + d\mathbf{k}, \mathbf{r} + d\mathbf{r}, t + dt) - f(\mathbf{k}, \mathbf{r}, t) = \frac{\partial f}{\partial \mathbf{k}} \cdot d\mathbf{k} + \frac{\partial f}{\partial \mathbf{r}} \cdot d\mathbf{r} + \frac{\partial f}{\partial t}\, dt = 0 \,. \tag{9.26}$$

If scattering occurs, a "correction term", must be added containing the total contribution of all the scattering processes. If we differentiate the above equation with respect to time we obtain the *Boltzmann equation*:

$$\dot{\mathbf{k}} \cdot \frac{\partial f}{\partial \mathbf{k}} + \dot{\mathbf{r}} \cdot \frac{\partial f}{\partial \mathbf{r}} + \frac{\partial f}{\partial t} = \left.\frac{\partial f}{\partial t}\right|_{\text{coll}} \,. \tag{9.27}$$

The first two terms on the left-hand side are the *field* and the *diffusion term*, whereas we find the *scattering* or *collision term* on the right side. When discussing the electrical conductivity, the diffusion term plays no role because a homogeneous electric field does not change the spatially homogeneous electron distribution. Therefore, we will omit this term in the following.

For the collision term we use the simple **relaxation time approximation**

$$\left.\frac{\partial f(\mathbf{k})}{\partial t}\right|_{\text{coll}} = -\frac{f(\mathbf{k}) - f_0(\mathbf{k})}{\tau(\mathbf{k})} \,. \tag{9.28}$$

The meaning of the relaxation time $\tau(\mathbf{k})$ or the *scattering rate* τ^{-1} was discussed in Section 7.2 when treading phonon scattering. In the present case, a non-equilibrium distribution $f(\mathbf{k}, \mathbf{r}, t)$ relaxes towards equilibrium by the scattering of electrons with wave vectors $\mathbf{k}$ into states with wave vectors $\mathbf{k}'$. As with phonon-phonon scattering, we have to take into account the fact that scattering processes are reversible, i.e. scattering from $\mathbf{k}'$ to $\mathbf{k}$ can also occur. The transition probabilities are the same for both scattering directions, but the net processes do not cancel each other out because the occupation of the respective initial and final states for the forward and backward direction is different owing to the different distribution functions $f(\mathbf{k}, \mathbf{r}, t)$ and $f_0(\mathbf{k}, \mathbf{r}, t)$.

The importance of the relaxation time in the context of the displacement of the Fermi sphere can be illustrated by a simple example: let us assume that under the influence of an electric field, a stationary non-equilibrium distribution $\overline{f}(\mathbf{k})$ has been

5 Joseph Liouville, *1809 Saint-Omer, †1882 Paris

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

established. If the field is then abruptly switched off, the distribution $f(\mathbf{k})$ evolves from the initial distribution $\overline{f}(\mathbf{k})$ back to the equilibrium distribution $f_0(\mathbf{k})$. With the relaxation time approximation (9.28) and with no external field (9.27) takes the simple form:

$$\frac{\partial f(\mathbf{k})}{\partial t} = -\frac{f(\mathbf{k}) - f_0(\mathbf{k})}{\tau(\mathbf{k})} \,. \tag{9.29}$$

With the initial condition $f(\mathbf{k}, t = 0) = \overline{f}(\mathbf{k})$ we find:

$$f(\mathbf{k}) - f_0(\mathbf{k}) = \left[\, \overline{f}(\mathbf{k}) - f_0(\mathbf{k}) \,\right] \mathrm{e}^{-t/\tau} \,. \tag{9.30}$$

After switching off the field, the deviation from the equilibrium distribution decays exponentially with the characteristic time constant τ. Correspondingly, when the electric field is switched on, the subsequent displacement of the Fermi sphere takes place with the same time constant. Thus τ is the time over which the distribution function adjusts to the new conditions in response to a sudden change of the external field, in other words, the time in which a new stationary state is reached. The connection with the Drude model becomes apparent here.

9.2.3 Electric Charge Transport

Following the detailed preliminary discussion, we now turn to the direct electrical current conductivity [σ] of metals. Since this is given by [σ] = $\mathbf{j}/\mathcal{E}$, we calculate the current density $\mathbf{j}$, generated by a constant electric field $\mathcal{E}$ with the help of the Boltzmann equation. We perform the calculation for an isotropic gas of free electrons, to avoid needing to take into account any tensor character of the conductivity. In this simple case: $\sigma = \sigma_{xx} = j_x/\mathcal{E}_x$. We disregard transient switch-on processes and consider only the stationary state, where the term $\partial f/\partial t$, which describes the explicit time dependence, vanishes in equation (9.27). In this simple case the distribution function depends neither on position nor time, so that the Boltzmann equation with $\dot{\mathbf{k}} = -e\,\mathcal{E}/\hbar$, equation (9.23), takes the simple form:

$$-\frac{e}{\hbar}\,\mathcal{E} \cdot \nabla_{\mathbf{k}} f(\mathbf{k}) = -\frac{f(\mathbf{k}) - f_0(\mathbf{k})}{\tau(\mathbf{k})} \,. \tag{9.31}$$

Solving for $f(\mathbf{k})$, we find:

$$f(\mathbf{k}) = f_0(\mathbf{k}) + \frac{e\tau(\mathbf{k})}{\hbar}\,\mathcal{E} \cdot \nabla_{\mathbf{k}} f(\mathbf{k}) \,. \tag{9.32}$$

For *small deviations* of the distribution function from the equilibrium value, we can simply solve this equation by replacing the gradient of the actual distribution function by the gradient of the equilibrium distribution. We then obtain the **linearized Boltzmann equation** for electric charge transport:

$$f(\mathbf{k}) \approx f_0(\mathbf{k}) + \frac{e\tau(\mathbf{k})}{\hbar}\,\mathcal{E} \cdot \nabla_{\mathbf{k}} f_0(\mathbf{k}) \,. \tag{9.33}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The current density is given by equation (9.9), which we repeat here:

$$\mathbf{j} = -\frac{e}{4\pi^3}\int \mathbf{v}(\mathbf{k}) f(\mathbf{k})\, \mathrm{d}^3k\,. \tag{9.34}$$

The free-electron model assumes that metals behave isotropically with respect to their electronic properties. Thus, $j_y = j_z = 0$, if the electric field is applied in the x-direction. In polycrystalline materials, this assumption is quite well fulfilled. Now we insert $f(\mathbf{k})$ from the linearized Boltzmann equation (9.33) into (9.34). The equilibrium distribution $f_0(\mathbf{k})$ does not contribute to the current flow, since the mean value over all velocities $\mathbf{v}(\mathbf{k})$ clearly vanishes. The term containing the gradient provides the current flow of interest. Since $\partial f_0(\mathbf{k})/\partial k_x = [\partial f_0(\mathbf{k})/\partial E]\,\hbar v_x$, we can write for the current density:

$$j_x = -\frac{e^2\mathcal{E}_x}{4\pi^3}\int v_x^2\,\tau(\mathbf{k})\,\frac{\partial f_0(\mathbf{k})}{\partial E}\,\mathrm{d}^3k\,. \tag{9.35}$$

When evaluating this integral, we take into account the fact that the Fermi-Dirac function has a very steep drop at the Fermi energy, but elsewhere is approximately constant. The derivative of this function is therefore only significantly different from zero in the immediate vicinity of the Fermi energy. Since the derivative takes on very large values there, we replace it with a delta function, i.e., we set $[\partial f_0(\mathbf{k})/\partial E] \approx -\delta(E - E_\mathrm{F})$. We have already dealt with the volume element d^3k in detail in Sections 6.4 and 8.1 in deriving the density of states of phonons and electrons. We use (6.79) and obtain for the electrical conductivity

$$\begin{aligned}\sigma = \frac{j_x}{\mathcal{E}_x} &\approx \frac{e^2}{4\pi^3\hbar}\int \frac{v_x^2(\mathbf{k})}{v(\mathbf{k})}\,\tau(\mathbf{k})\,\delta(E - E_\mathrm{F})\,\mathrm{d}E\,\mathrm{d}S_\mathrm{E} \\ &\approx \frac{e^2}{4\pi^3\hbar}\int \frac{v_x^2(\mathbf{k}_\mathrm{F})}{v(\mathbf{k}_\mathrm{F})}\,\tau(\mathbf{k}_\mathrm{F})\,\mathrm{d}S_\mathrm{F}\,.\end{aligned} \tag{9.36}$$

As in (6.79), in the first line $\mathrm{d}S_\mathrm{E}$ is an element of the constant-energy surface in $\mathbf{k}$-space. The second line is obtained after integration over energy, where $\mathrm{d}S_\mathrm{F}$ is a surface element on the Fermi sphere.

Since the velocity distribution of the electrons within our simple model is isotropic, we may simply set $v_x^2 = v^2/3$ (which also applies to the Fermi velocity). The relaxation time τ also shows no directional dependence, so that we can take both quantities out of the integral. This yields the simple expression $\int \mathrm{d}S_\mathrm{F} = 4\pi k_\mathrm{F}^2$ for the surface integral. Noting that $m^* v_\mathrm{F} = \hbar k_\mathrm{F}$, we obtain for the electrical conductivity:

$$\sigma_{xx} = \sigma = \frac{j_x}{\mathcal{E}_x} = \frac{e^2}{3\pi^2}\,\frac{\tau(E_\mathrm{F})}{m^*}\,k_\mathrm{F}^3\,. \tag{9.37}$$

Also noting that from equation (8.26) the magnitude of the Fermi wave vector is given by $k_\mathrm{F}^3 = 3\pi^2 n$, the expression is further simplified to:

$$\sigma = \frac{n e^2}{m^*}\,\tau(E_\mathrm{F})\,. \tag{9.38}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

This equation corresponds to the classical formula (9.22) of *P. Drude*, which is based on all electrons participating equally in the charge transport. However, as we have just seen above, only electrons at the Fermi surface contribute to the conductivity, because only there is the gradient of the distribution function clearly different from zero. The main difference here is that we take the relaxation time $\tau(E_F)$ to be that of the electrons at the Fermi surface, whereas Drude's theory includes an unspecified mean scattering time τ, which can be interpreted as the scattering of all electrons with neighboring atoms and among each other.

Here are some typical numerical values: in high-purity copper the average scattering time τ is approximately 2×10^{-14} s at room temperature and 6×10^{-11} s at 4 K. Since only electrons at the Fermi surface can be scattered, and these are moving with the Fermi velocity $v_F = 1.6 \times 10^6$ m/s, the mean free path $l = v_F \tau$ can be derived from the scattering time resulting in $l \approx 30$ nm and $l \approx 0.1$ mm, respectively. It is also interesting to estimate the relative displacement $\delta k/k_F$ of the Fermi sphere. Taking the mean scattering time of copper at room temperature and an electric field strength of $\mathcal{E} = 100$ V/m in (9.24), we obtain $\delta k \approx 3 \times 10^3$ m^{-1} and thus $\delta k/k_F \approx 10^{-7}$. The Fermi sphere thus shifts only slightly.

We should point out here that the Drude model also gives the correct expression for the alternating current conductivity. If we look for a stationary solution of equation (9.19) for the alternating field $\mathcal{E} = \mathcal{E}_0 \exp(-i\omega t)$, we find with (9.21) a frequency-dependent alternating current conductivity:

$$\sigma_{ws} = \frac{ne^2\tau}{m^*} \frac{1}{1 - i\omega\tau} . \tag{9.39}$$

The inertia of the electrons gives rise to a phase shift between current and voltage, with the factor $(1 - i\omega\tau)$ in the expression being typical of so-called *relaxation processes*. We will discuss these in detail in Section 13.3.

9.2.4 The Scattering of Conduction Electrons

Apart from the electron-electron interaction, the effectiveness of which we will discuss later, conduction electrons in crystals should be able to move freely with the effective mass. This applies both to the free-electron gas model, where the electrons move in a constant potential, and to electrons in the periodic potential of a crystal. As we will see, electrons are always scattered when there are deviations from the constant or periodic potential caused by defects or lattice vibrations. In the following, we will make some qualitative observations regarding these processes.

Electron-Defect Scattering. We first address the interaction with static lattice defects such as point defects, impurity atoms or dislocations. According to simple scattering theory, which was discussed in Sections 4.2 and 7.3, waves can be elastically scattered

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

at static scattering centers. In such processes, the direction of motion of the electrons changes, but not the magnitude of their wave vectors. Labelling the electron momentum before and after the scattering process by $\hbar\mathbf{k}$ and $\hbar\mathbf{k}'$, respectively, wave vector conservation can be described by $\mathbf{k} = \mathbf{k}' + \mathbf{K}$ where $\mathbf{K}$ is any wave vector whose magnitude satisfies $K \leq 2k$. The momentum transfer $\hbar\mathbf{K}$ is absorbed by the lattice as a whole.

Figure 9.6 shows such an elastic scattering process schematically. In addition to the Fermi sphere, the Brillouin zone of a simple cubic lattice with the proportions corresponding approximately to those of a simple metal is shown. The Fermi sphere is displaced by the electric field by the amount δk_x, its surface is "softened" by the finite temperature. Both effects are greatly exaggerated in this figure.

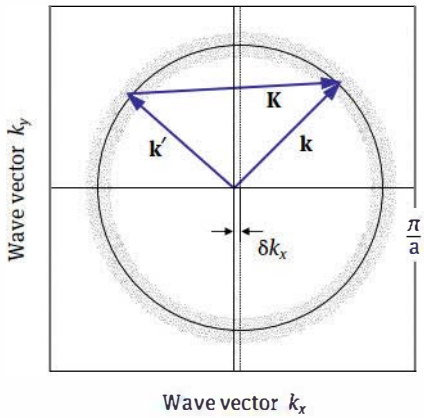


Fig. 9.6: Electron-defect scattering. The wave vectors $\mathbf{k}$ and $\mathbf{k}'$ of the electron before and after the scattering process both lie in the softened region of the Fermi sphere. The lattice as a whole absorbs the momentum $\hbar\mathbf{K}$. The center of the Fermi sphere is shifted to the right by δk_x owing to the applied electric field.

The electrons do not experience any change of energy during elastic scattering at defects. Since the final state must be unoccupied, the process can only take place within the thermally softened region of the Fermi sphere.

Electon-Phonon Scattering. In contrast to scattering by defects, electron-phonon scattering is inelastic and can be treated analogously to neutron scattering as discussed in Section 6.3. Since $E_F \gg \hbar\omega_D$, the electron energy hardly changes with the generation or absorption of a phonon. Thus to good approximation we can assume that for the electron wave numbers $|\mathbf{k}| \approx |\mathbf{k}'|$. Consequently, only those electrons in the vicinity of the Fermi surface participate in the scattering process. For the wave vector the conservation law $\mathbf{k} = (\mathbf{k}' \pm \mathbf{q} + \mathbf{G})$ applies, where $\mathbf{G}$ is a reciprocal lattice vector. The sign of $\mathbf{q}$ depends on whether a phonon is generated or absorbed. As with phonon-phonon scattering, two processes can be distinguished here: the normal process without the involvement of a reciprocal lattice vector, and the umklapp process, where $\mathbf{G} \neq 0$.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Normal Process: Normal processes have different effects on the electrical conductivity depending on the sample temperature. As shown schematically in Figure 9.7, the mean scattering angle ϑ of the electrons depends on temperature, since the mean momentum transfer is determined by the wave number and thus by the frequency of the dominant phonons. At low temperatures, each scattering process causes only a relatively small angular change. Therefore, many scattering events have to occur before an electron from the front of the Fermi sphere can reach the rear. The time needed for the transition from the front side to the rear gives an effective mean scattering time. However, with increasing temperature, the scattering angle increases and the normal processes become more and more effective until finally one single scattering event is sufficient. In Figures 9.7 and 9.8 we have not shown the spherical free-electron Fermi surface, but rather a surface deviating from this shape, as this actually occurs in a similar form in many metals.

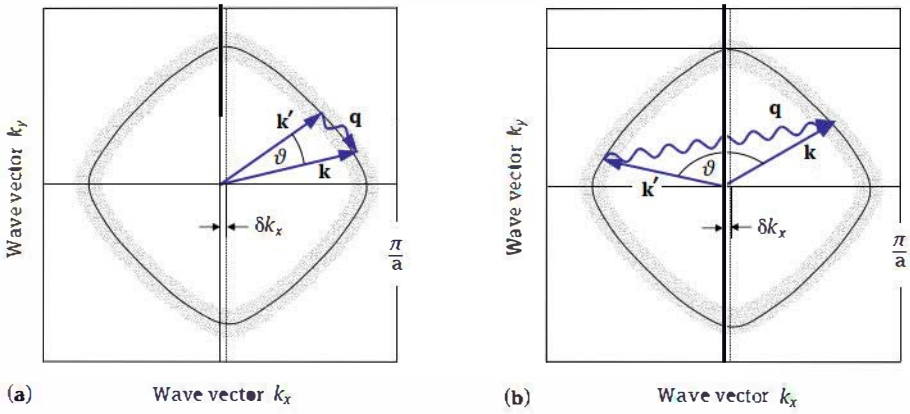


Fig. 9.7: Normal processes. **a)** A scattering process at low temperatures with small momentum transfer. **b)** A scattering process at high temperatures from the front to the back of the "Fermi sphere" by a single scattering event. In both cases, the electric field shifts the center of the Fermi surfaces to the right with respect to the coordinate-system origin by δk_x.

Umklapp processes: An umklapp process occurs when the final wave vector of an electron, after scattering with a phonon, falls outside the first Brillouin zone, as shown schematically in Figure 9.8. Similarly to the phonon umklapp process, a minimum momentum of the phonon involved is required. However, since the Fermi surface is close to the boundary of the Brillouin zone in many cases, the momentum or energy of the phonon can be relatively small depending on the shape of the Fermi surface. As can be seen in Figure 9.8, the addition of a reciprocal lattice vector brings the electron back to the Fermi surface in the first zone but at the rear of the distribution. If the Fermi surface is suitably shaped, umklapp processes can lead to large changes in the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

direction of the electron momentum even though the momentum of the participating phonon can be so small that a normal process would only cause a small angular change. Consequently, these processes provide a particularly effective mechanism for reaching and maintaining the stationary equilibrium during electric charge transport.

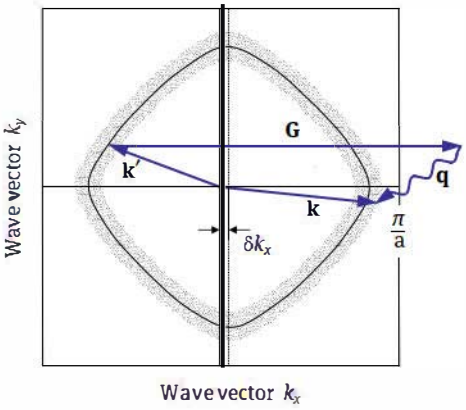


Fig. 9.8: An umklapp process. The scattered electron generates a phonon and is brought to the rear of the Fermi surface with the help of a reciprocal lattice vector. The center of the Fermi sphere is shifted with respect to the origin of the coordinate system by δk_x.

Electron-Electron Scattering. Up to now we have always ignored the interaction *between* the electrons and have treated them in the context of the *one-electron approximation*. However, if instead we go over to a many-body description, then interactions between electrons can occur and can be taken into account. Such interactions lead to a scattering out of or into the states of the one-particle approximation. Given the strength of the Coulomb potential and the high electron density in metals of about one electron per unit cell, one might expect that the rate caused by this scattering mechanism might well be higher than that of the mechanisms discussed so far. However, there is a strong suppression of electron-electron scattering for two reasons. First, the shielding effects which we discussed in the previous chapter cause a strong reduction of the range of the Coulomb potential, and secondly, there is the effect of the Pauli principle whose consequences for electron-electron scattering we will now discuss.

To show the influence of the Pauli exclusion principle, we choose the simplest and most common scattering process: the interaction between two electrons with energies E_1 and E_2. For simplicity we look at the scattering process at $T = 0$ and assume that only the electron with energy E_1 is excited, thus having energy $E_1 > E_F$. The energy of the second electron is $E_2 < E_F$. Since the scattering can only take place into unoccupied states, the energies E_1' and E_2' of the two electrons after the scattering process must both be greater than the Fermi energy, i.e. $E_1', E_2' > E_F$ must apply. In addition, energy conservation requires that:

$$E_1 + E_2 = E_1' + E_2' \,. \tag{9.40}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

This results in the restriction that E_2 and E_1' must be in the vicinity of the Fermi surface, i.e. within a spherical shell of thickness $|E_1 - E_F|$. This means that of all of the possible states, only a fraction $9(E_1 - E_F)^2/E_F^2$ is available as final states. The restriction for E_2' is already contained in (9.40) and does not represent a further limitation. Let us now consider finite temperatures where there is a softening of the Fermi sphere of the order of $k_B T$. The above argument can be applied to this case, where $(E_1 - E_F)$ is replaced by the range of the thermal softening. It can therefore be expected that the scattering rate $1/\tau_E$ for electron-electron processes is reduced approximately by the factor $(k_B T/E_F)^2$ compared to the scattering rate which would occur without the Pauli principle.

In our analysis thus far, we have not considered shielding effects. To take these into account, we have to replace the Rutherford scattering cross section, which is based on the unshielded Coulomb potential, by the scattering cross section which is based on the shielded potential according to *Thomas* and *Fermi*. This means that in the expression for the scattering rate, a factor $r_{TF}^2 \propto E_F$ is added. If one includes this effect, the following relation for the scattering rate results:

$$\frac{1}{\tau_E} \approx \frac{1}{\hbar} \frac{(k_B T)^2}{E_F} \,. \tag{9.41}$$

If we use typical numerical values, we find $\tau_E \approx 10^{-12}$ s for the relaxation time at room temperature, i.e. a value that is approximately four orders of magnitude larger than the experimentally-observed mean scattering time. This means that the one-electron approximation describes the conditions very well and that electron-electron scattering may be neglected in most cases. However, it should be possible to observe the influence of this scattering in very pure, largely defect-free metals at low temperatures, because in this case the other two scattering mechanisms (electron-defect and electron-phonon) hardly come into play. An example where this effect *can* be observed is described in the next section.

The above discussion makes it clear that the independent-electron model provides a good approximation for electrons at the Fermi edge. Surprisingly, it is also true for materials where the electron-electron interaction is relatively strong. The reason for this is that in experiments it is not just the properties of the electrons alone that are observed, but those of the electrons including their interaction. If the interaction noticeably changes the properties of the free electrons, then we no longer speak of a Fermi gas, but of a **Fermi liquid**. In this concept, introduced by *L.D. Landau*[6], the various electron interactions are taken into account by replacing the „bare" electrons by what is known as a **quasiparticles**. A quasiparticle is an electron accompanied by a local distortion in the surrounding electron gas representing the effects of the interactions, and has an effective mass different from that of a free electron. These

6 Lev Davidovich Landau, *1908 Baku, †1968 Moscow, Nobel Prize 1962

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

interactions can cause quantitative changes of the transport processes. The theory of the Fermi liquid has been applied very successfully in other areas of physics, but we will not pursue it further here.

9.2.5 Temperature Dependence of the Electrical Conductivity

We first consider the temperature dependence of the electrical conductivity of copper. As can be seen in Figure 9.9, it is constant at low temperature, then decreasing steeply with increasing temperature and finally showing a more moderate $1/T$ decrease at room temperature. The conductivity of the alloy *Manganin*[7] is also shown which depends only very weakly on the temperature over the entire temperature range and is always much smaller than that of copper. From the equation for the electrical conductivity (9.38), we see that the mean scattering time $\tau(k_F)$ is the only quantity that can have a temperature dependence. Thus, in order to understand the temperature dependence of the conductivity, we must therefore take a closer look at the temperature dependence of the scattering processes already discussed.

As already explained, the relaxation time τ is determined by the scattering of the conduction electrons from defects, from lattice vibrations and from other electrons. Since the three scattering mechanisms act independently of each other, the effective scattering rate is the sum of the individual scattering rates:

$$\frac{1}{\tau} = \frac{1}{\tau_D} + \frac{1}{\tau_G} + \frac{1}{\tau_{EE}} \,. \tag{9.42}$$

Initially, neglecting the last term, i.e. the contribution of the electron-electron interaction, we divide the electrical resistivity $\varrho = 1/\sigma$ into two parts, that based on the defect

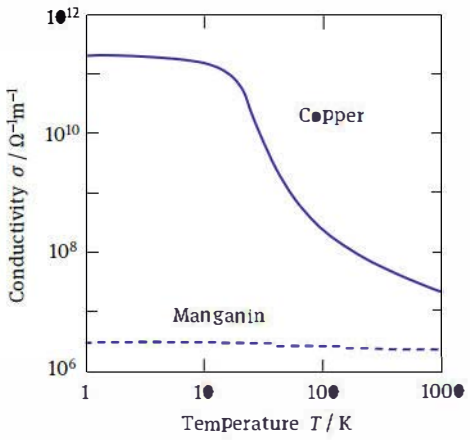


Fig. 9.9: Temperature dependence of the electrical conductivity of high-purity copper (solid line) and that of the alloy manganin (dashed line).

7 Manganin consists of 86% copper, 12% manganese and 2% nickel.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

scattering, the other that caused by the thermal motion of the lattice:

$$\varrho = \frac{m^*}{ne^2\tau} = \varrho_D + \varrho_G = \frac{m^*}{ne^2\tau_D} + \frac{m^*}{ne^2\tau_G(T)} \,. \tag{9.43}$$

This empirical law is known as **Matthiessen's rule**[8].

Confirmating Matthiessen's rule, Figure 9.10 shows the temperature dependence of the electrical resistance of three sodium samples. As T approaches zero, a temperature-independent **residual resistance** is observed. Since phonon scattering dies out at low temperatures, the resistance can only be caused by defects. The resistivity ratio $\varrho\,(300\,\mathrm{K})/\varrho\,(4.2\,\mathrm{K})$ is a measure of the concentration of the scattering defects and thus also of the purity and quality of the sample. In the literature, the residual resistivity ratio is often used for this purpose and is referred to by the abbreviation **RRR** standing for the **R**esidual **R**esistivity **R**atio. In elementary metals, a ratio of 1000 : 1 can easily be achieved. In contrast, the temperature variation in strongly alloyed and amorphous metals is comparatively weak, so that $\varrho\,(300\,\mathrm{K})/\varrho\,(4.2\,\mathrm{K}) \approx 1$. In the latter case structural and substitutional disorder causes such strong defect scattering that any electron-phonon scattering is hardly noticeable. Typical resistivity values for amorphous (and also liquid) metals are in the range of 1 to 5 μΩ m. As an example, Figure 9.9 shows the conductivity of the alloy manganin, which has a resistivity of $\varrho \approx 0.43\,\mu\Omega\,\mathrm{m}$.

At high temperatures, i.e. at $T > \Theta$, in pure metals the electon-phonon scattering dominates, and the previously used relation $l^{-1} = n\sigma_{\mathrm{cross}}$ applies, for the inverse mean free path of the scattered electrons.[9] Since the scattered electrons have energies near the

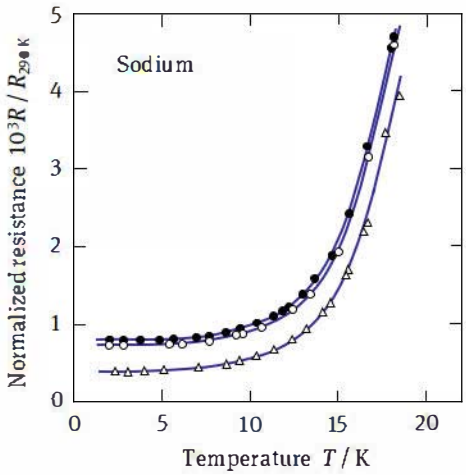


Fig. 9.10: The temperature dependence of the electrical resistance of three sodium samples with different residual resistances. (After D.K.C. MacDonald, K. Mendelssohn, Proc. Roy. Soc. London **202**, 103 (1950).)

8 Augustus Matthiesen, *1831 London, †1870 London

9 To avoid confusion with the conductivity σ, the index "cross" is added to the denotation of the scattering cross section, i.e. we write σ_{cross} here.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Fermi energy and are mainly interacting with phonons with the Debye frequency, which are excited in large numbers at high temperatures. The scattering cross section σ_{cross} is approximately constant for this process. The inverse mean free path increases with the phonon density n proportionally to the temperature (see Section 6.4) and thus $l^{-1} \propto T$. Since the mean scattering time is directly related to the mean free path via $l = \tau v_F$, the electrical conductivity at high temperatures $T > \Theta$ is inversely proportional to the temperature, i.e.

$$\sigma \propto \tau_G \propto \frac{1}{T} . \tag{9.44}$$

We now look at the temperature dependence of the electrical conductivity of a pure metal in the medium-temperature range, i.e. between 10 K and 100 K. Starting from room temperature, as just discussed, the conductivity initially varies as $1/T$ as the temperature decreases, but the rate becomes increasingly steeper with falling temperature. There are several reasons for this. First, the number of phonons and thus the number of scattering partners decreases. At the same time, the scattering cross section for electron-phonon scattering also decreases. This is due to the fact that, as in the case of phonon-phonon scattering (cf. Section 7.2), the scattering cross section σ_{cross} is proportional to the frequency of the phonons involved, but for $T < \Theta$ this becomes progressively smaller as the temperature decreases. Finally, with decreasing phonon frequency, the momentum change associated with each scattering event also becomes smaller, so that the angular change which can be caused by a single process is also smaller (see Figure 9.7). Thus, to transfer the electron wave vector from the front to the rear of the Fermi surface requires an increasing number of scattering processes. Formally, this dependence can be expressed by the relationship $\tau_G^{-1} \propto n \eta \sigma_{cross}$ where n is the phonon density and η reflects the change in angle caused by each scattering process and thus expresses the effectiveness of the scattering events. Again σ_{cross} is the scattering cross section. All three quantities decrease increasingly rapidly with decreasing temperature.

Since the derivation of the so-called **Bloch-Grüneisen law**, taking into account the three effects discussed above, requires a somewhat more extensive effort, we simply state the result here:

$$\varrho_G = A \left(\frac{T}{\Theta} \right)^5 \int_0^{\Theta/T} \frac{x^5 dx}{(e^x - 1)(1 - e^{-x})} , \tag{9.45}$$

where Θ is the Debye temperature. This universal expression should be valid for all pure metals where electron-phonon scattering is dominant. On evaluating the integral for $T \to 0$, we find $\varrho_G \propto (T/\Theta)^5$ and for high temperatures $\varrho_G \propto (T/\Theta)$ in agreement with the experimental observations.

Figure 9.11 shows the resistance R/R_Θ for a number of metals, normalized to the resistance at the Debye temperature as a function of T/Θ. The observed dependence is consistent with the theoretical prediction, shown by the solid line in the figure. The transition from the T^5-dependence to the linear dependence is also clearly visible.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

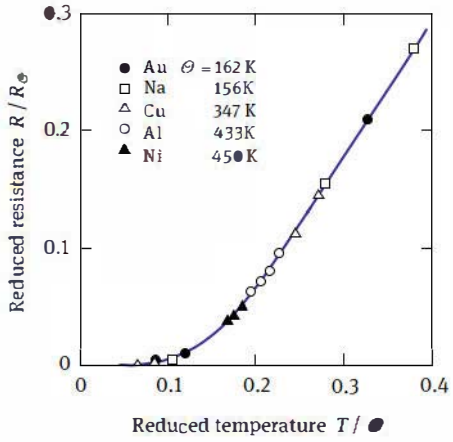


Fig. 9.11: Reduced resistance of various metals as a function of the reduced temperature T/Θ. (After D.K.C. MacDonald, *Handbook of Physics XIV*, S. Flügge, ed., Springer, 1956.)

We can summarize the results for the three temperature ranges as follows:

$$\varrho \propto \begin{cases} \text{const.} & \text{for} \quad T \ll \Theta\,, \\ T^5 & \text{for} \quad T < \Theta\,, \\ T & \text{for} \quad T \gg \Theta\,. \end{cases} \tag{9.46}$$

Finally we return to electron-electron scattering. As discussed above, in the majority of cases electron-defect scattering dominates at low temperatures. However, in very pure metals, especially in alkali metals with relatively small Fermi energies, the electron-electron scattering can be observed. Figure 9.12 shows measurements of the electrical resistivity of potassium by various authors. Starting at high temperature, electron-phonon scattering is first observed. With decreasing temperature the curve becomes steeper, and in agreement with the Bloch-Grüneisen law, is roughly proportional to T^5. Below 2 K the conductivity is determined by electron-electron scattering and is therefore proportional to temperature squared.

9.2.6 One-dimensional Conductors

A question that is not only interesting from a theoretical point of view, but is also gaining technological importance due to increasing miniaturization, is that of the electrical conductivity of one-dimensional systems. To treat this problem, we consider a thin wire of length L, connecting two metal contacts between which a voltage U is applied. For simplicity, we consider a wire with a rectangular cross-section, which is extended in the x-direction, but sufficiently narrow in the y- and z-direction. For the

10 The data were provided by various authors, as described in J. Bass et al, Rev. Mod. Phys. **62**, 645 (1990).

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

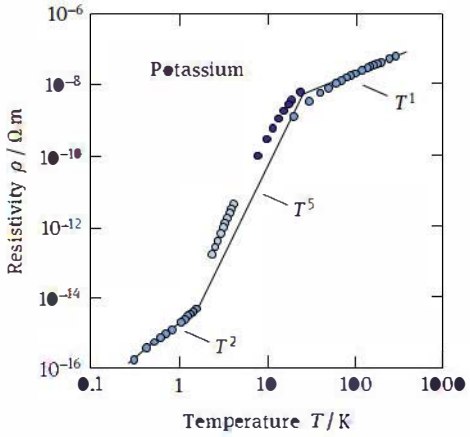


Fig. 9.12: Electrical resistivity of high purity potassium. The three ranges with different slopes described in the text are clearly visible. At the lowest temperatures electron-electron scattering dominates. (After M.P. Marder, *Condensed Matter Physics*, John Wiley & Sons, 2000.[10])

sample to have one-dimensional properties, the characteristic dimensions of width and thickness must be comparable to or smaller than the Fermi wavelength λ_F. Another important assumption is that the mean free path of the electrons is large compared to the wire length. As a result there can be no scattering processes in the wire, and the motion of the electrons becomes *ballistic*.

As indicated in Figure 9.13, the voltage U between the connecting electrodes causes a change in the chemical potential given by $\Delta\mu = eU$. We should note that during ballistic electron propagation the voltage cannot drop along the wire, so that any voltage jump must take place at the connecting electrodes. Another interesting fact is that the chemical potential for forward and backward moving electrons is different in the wire. This does not contradict the definition of μ, because the chemical potential is only constant in a system in equilibrium. This is not the case here, since the electrons do not exchange energy either with each other or with the lattice due to the lack of scattering. We should also note that when describing transport phenomena in one-dimensional systems, the bands are usually referred to as *channels*.

When calculating the current transport, we first assume that the cross-section of the wire only allows for *one* transverse energy eigenvalue $E_{i,j}$ (see Section 8.1). The charge transport then takes place in only one channel. We can derive the corresponding expressions using the arguments of the discussion of one-dimensional heat transport in Section 7.3. In analogy with (7.30) and (7.31) we can write for the electric charge transport:

$$I = \frac{1}{L}\sum_{\mathbf{k}} e v_{\mathbf{k}} = \frac{1}{L}\int \varrho_{\mathbf{k}}^{(1)}\, e\, v(\mathbf{k})[f(E + eU/2) - f(E - eU/2)]\,\mathrm{d}k\,. \tag{9.47}$$

In this expression we write the electron velocity as $v = \hbar^{-1}(\partial E/\partial k)$ and the density of states as $\varrho_{\mathbf{k}}^{(1)} = L/\pi$, and integrate over power E rather than k. From the position dependence of the chemical potential in sketch 9.13 we can see that the square bracket can be replaced approximately by eU, if the transition to the connecting electrodes takes

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

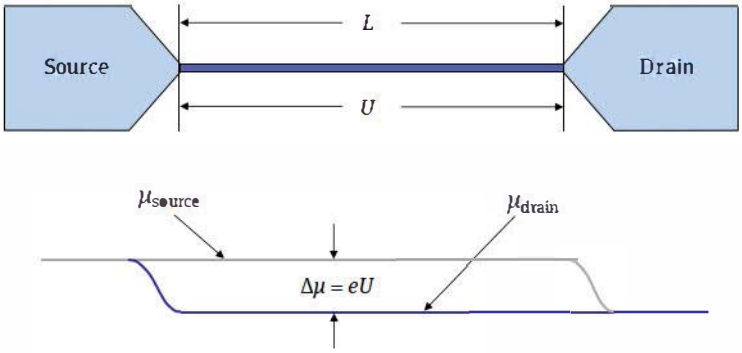


Fig. 9.13: Schematic representation of a one-dimensional conductor. The position dependence of the Fermi energies of the electrons of the source (grey curve) and the drain (blue curve) are also shown.

place over a distance small compared with the length of the wire. Thus the expression for the current can be simplified to:

$$I = \int_0^\infty \frac{2e}{h} \left[f(E + eU/2) - f(E - eU/2) \right] \mathrm{d}E = \frac{2e^2}{h} U \,. \tag{9.48}$$

We thus find for the electrical conductance G and the resistance R:

$$G_Q = \frac{I}{U} = \frac{2e^2}{h} \quad \text{and} \quad R_Q = \frac{h}{2e^2} = 12.906404\,\mathrm{k\Omega} \,. \tag{9.49}$$

A perfectly transmissive one-dimensional conductor has a finite conductance, which can be expressed in terms of the fundamental constants and is known as the conductance quantum $\mathbf{G_Q}$, while the inverse quantity represents the **resistance quantum** $\mathbf{R_Q}$. We should also note here that occasionally the *spin degeneracy*, which halves the conductance or doubles the resistance, is also taken into account.

The quantization of the conductance can be demonstrated in what is in principle a surprisingly simple experiment requiring no great deal of technical effort. Two wires arranged perpendicular to each other are brought into contact and the current flowing through the contact subject to a voltage is monitored. Now, the two wires are excited into mechanical oscillation leading to a break in the contact. If the current is measured during the time that the contact is actually being broken, a time dependence is observed as shown in Figure 9.14 for an experiment with two gold wires. There are clearly visible steps in the current and thus also in the conductance. Obviously, the conductance changes by one quantum or multiple quanta of the conductance.

How can we understand the occurrence of these jumps? It could be shown that as the two wires are separating, instead of suddenly breaking, the midsection forms a narrowing bridge by plastic deformation. In Section 8.1 we derived how the density of states of one-dimensional systems is composed of the contributions of the sub-bands

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

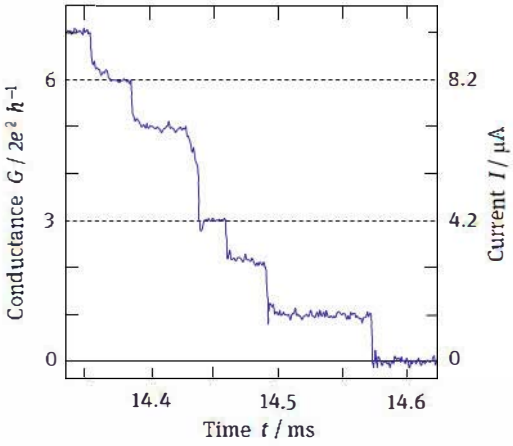


Fig. 9.14: Conductance and current as a function of time when a contact between two oscillating gold wires is broken under ultrahigh vacuum. The x-axis shows the elapsed time since the beginning of the separation process. (After J.L. Costa-Krämer et al., Surf. Sci. **342**, L1144 (1995).)

(see Figure 8.5b). The density of states of the subbands diverges at the respective transverse energy $E_{i,j}$, which in this case depends on the dimension of the wire cross-section. If the cross-section decreases as in the experiment described, the values of $E_{i,j}$ increase and the singularities in the density of states move to the right in Figure 8.5b. Each time a transverse eigenvalue crosses the Fermi energy, a channel disappears and the conductance decreases by one quantum. If the level in question is degenerate, the jump can also be a multiple of G_Q.

Quantized Conduction of Point Contacts. Fabricating metallic "wires" of suitable sizes is not easy since the Fermi wavelength of metals is very small, about $\lambda_F \approx 5$ Å. Conditions are much more favorable for *semiconductor heterostructures* (see Section 10.4), where a two-dimensional electron gas with much longer Fermi wavelength can be generated. We briefly describe such an experiment below.

The first proof of conductance quantization was demonstrated in a *quantum dot contact*. This was in an AlGaAs/GaAs heterostructure, as shown in Figure 9.15a, the precise design being irrelevant here. Owing to the low electron density, the Fermi wavelength was $\lambda_F \approx 42$ nm. The crucial factor was that a two-dimensional electron gas (shown in dark blue in the figure) could be generated at the interface between the two semiconductors. An electrode, separating the source and the drain, was deposited on the surface but with a 250 nm wide and 1 μm long wedge-shaped gap. By applying a voltage of about $U_g \approx -0.6$ V, the electron gas under the electrode was repelled. The current flow from the source to the drain therefore occurred exclusively through the narrow gap. At the measuring temperature of $T = 0.6$ K, the mean free path of the electrons was 8.5 μm, i.e. much larger than the characteristic dimensions of the gap. The electrons thus passed through the constriction without scattering.

If the potential of the electrode was lowered below −0.6 V, the electron gas could be further suppressed in the vicinity of the electrode, so that the effective gap width b could

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

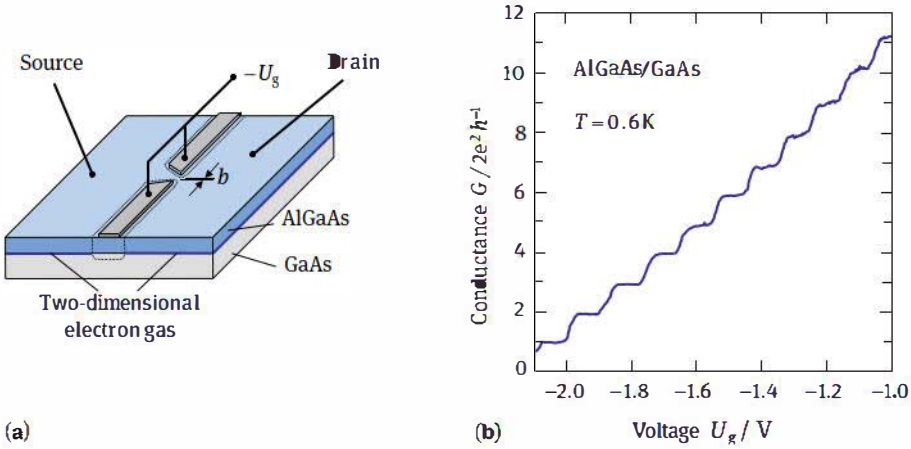


Fig. 9.15: Quantum dot contact. **a)** Schematic representation of a quantum dot contact. The two-dimensional electron gas between the GaAs and AlGaAs layer is shown in dark blue. Under the electrode the electron gas is suppressed by the negative potential. The effective gap width b is determined by the electrode voltage U_g. **b)** The conductance of such a quantum dot contact at 0.6 K as a function of the voltage U_g at the electrode. (After B.J. van Wees et al., Phys. Rev. **60**, 848 (1988).)

be progressively reduced. At a voltage below $U_g = -2.2$ V the gap became completely closed. If, as in Figure 8.15b, the measured current from the source to the drain is plotted as a function of the voltage U_g, we see plateaus in the current which reflect the quantization of the conductance. Each time a further conductivity channel is closed due to the increasing narrowing of the gap, the conductance decreases by one unit of the conductance quantum $G_Q = 2e^2/h$.

9.2.7 Luttinger Liquid

We have already discussed the one-dimensional conductors that are created when thin gold wires are further thinned by plastic deformation and are finally broken. What was surprising at the time was the observation that the conductance decreased stepwise as a function of time during this process leading to the conclusion that conductance is clearly quantized. We will now go one step further and briefly discuss strictly one-dimensional conductors. Such systems should have properties differing significantly from those of classical solids. In our discussion, we first consider a metal containing just *one* conduction electron per atom, so that, taking into account the spin degeneracy, the conduction band is half-filled. A description of such an unusual solid can be found in the theory of the so-called Luttinger liquid developed by *S. Tomonaga*[11] and

11 Shin-Itiro Tomonaga, *1906 Tokyo, †1979 Tokyo, Nobel Prize 1965

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

J. M. Luttinger[12]. However, a detailed treatment would be well beyond the scope of this introductory textbook, and we will just pick out a few interesting aspects in order to familiarise ourselves a little with the particularities of Luttinger liquids.

Of course, electrons in one-dimensional samples are severely restricted in how they can move, because they cannot avoid collisions or interactions with other particles. In contrast to the motion of electrons in two- or three-dimensional solids, motion in one-dimensional samples is always strictly correlated. Therefore, any disturbances will propagate as wave-like excitations along the sample. The forced displacement of a single electron therefore causes a density and a spin-density wave in the electron gas. We will look at both phenomena in more detail for conventional solids. Density waves in the electron gas are called plasmons, which we will discuss in detail in Chapter 13. Disturbances in the spin alignment cause spin waves, which we will look at in Chapter 12.

Surprisingly, in one-dimensional samples, spin and charge move independently of each other, resulting in a so-called *spin-charge separation*. This effect can be explained by using a simple example. As shown in Figure 9.16, we consider a one-dimensional antiferromagnet whose conduction band is half filled. The properties of antiferromagnets are discussed in detail in Chapter 12. Here, the only important thing is that the electron spins of neighboring atoms are oppositely oriented. In our thought experiment, we now illuminate the sample with light, and an electron is removed from the band via the photoelectric effect and thus a hole is created. This process is shown schematically in the central part of Figure 9.16. The other electrons are initially not affected by this event. As show at the bottom, the hole can now jump to an adjacent site, or in other words, an adjacent electron can take the place of the hole. This process has great formal similarity with the vacancy diffusion we discussed in Section 5.1. However, when a neighboring electron jumps onto the vacant site, its spin, due to the antiferromagnetic order, will be opposite to the spin originally present at this space and thus the magnetic order will be disturbed. If the hole moves on, there is no further disorder in the spin alignment, as can easily be seen. The absorbed photon has thus created two "defects": a hole in the electron occupation and a localized spin defect.

There is another interesting aspect. Surprisingly, the distribution function of the electron momentum, even at absolute zero, does not correspond with the distribution we know from the Fermi gas or the Fermi liquid. Even at $T = 0$, the distribution does not have an abrupt step at the Fermi edge, but instead there is a power law of the form $\mathrm{sgn}(k - k_F)|k - k_F|^{\alpha}$. The electronic density of states thus has the form $D(E) \propto E^{\alpha}$ in the vicinity of the Fermi energy. The numerical value of α is determined by the character and strength of the interaction. If the interaction is negligible, $\alpha = 0$.

The modified density of states and the correlation effects naturally exert a profound influence on the physical properties of Luttinger fluids. We can demonstrate this via

12 Joaquin Mazdak Luttinger, *1923 New York, †1997 New York

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

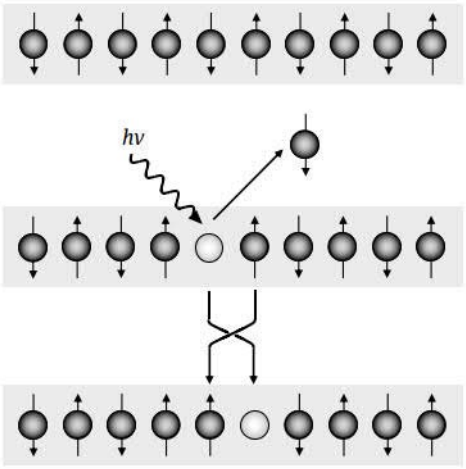


Fig. 9.16: Spin-charge separation in a half-filled one-dimensional band. Starting with the undisturbed state at the top, as we move down the figure, an electron is first removed from the chain by photon absorption and a hole is created. Finally, if an adjacent electron jumps onto the free site, the hole propagates further and simultaneously causes magnetic disorder. (After M. Dressel, *Luttinger-Flüssigkeit*, Universität Stuttgart.)

the electrical conductivity of metallic carbon nanotubes. In one such experiment, the electrical properties of nanotubes, weakly coupled to metallic electrodes, were investigated. Were the behavior classical, the current-voltage curve should follow Ohm's law. However, the theory of Luttinger liquids predicts a more complicated dependence. We will not derive the behavior here but merely look at the result. For weak electric fields, i.e. for $eU/k_BT < 1$, we expect the conductance G to obey a power law of the form $G = I/U \propto T^{\alpha}$. For high fields, i.e. for $eU/k_BT > 1$, the relationship $I \propto U^{1+\alpha}$ should apply. If $dI/dU \propto U^{\alpha}$ is plotted, the slope of the current increase can be used to determine the value of α. Dividing the measured data by T^{α} and plotting as a function of temperature should produce *one* common measurement curve. The result of such an investigation on a bundle of single-walled nanotubes is shown in Figure 9.17.

The figure shows that for small values of eU/k_BT the scaled conductance is constant, but for large values the increase follows a power law with exponent $\alpha = 0.46$. Plotted this

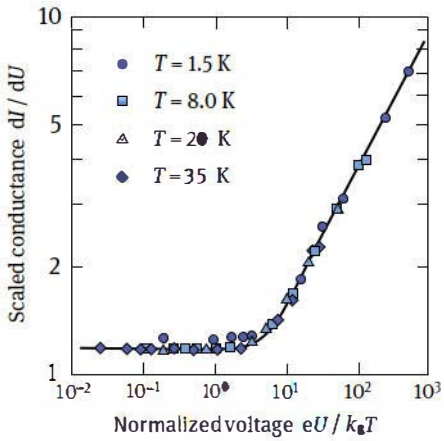


Fig. 9.17: The conductance of a bundle of single-walled nanotubes as a function of voltage measured over a range of temperatures. (After M. Bockrath et al., Nature **397**, 598 (1999).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

way, the measured values recorded at different temperatures all coincide as described above. The measurements therefore agree very well with the theoretical expectations.

Interesting questions arise when several nanotubes are arranged to run parallel. The interaction between the tubes can then be varied by varying their separation. With decreasing separation, the transition from Luttinger to Fermi liquid can be studied. However, we will not pursue this question further, as it goes beyond the scope of this introduction.

Many materials have been investigated showing signs that they could be Luttinger liquids. However, the proof is not so simple. One of the major difficulties is that one-dimensional systems are unstable against small perturbations, so that there is always a tendency for phase transitions. Nevertheless, it is believed that Luttinger liquids can be produced by the constriction of two-dimensional electron gases into a single dimension. The carbon nanotubes already mentioned here are a good example of a Luttinger liquid. In the next chapter, we will look at edge states in the quantum Hall effect, which surprisingly also behave like one-dimensional conductors. Furthermore, the properties of one-dimensional molecule chains in some organic crystals suggest that these can also behave as Luttinger liquids.

9.2.8 Quantum Dots

Having already introduced the idea of the *quantum dot contact* in the demonstration of conductance quantization above, we take a further step in the reduction of dimensionality to consider these zero-dimensional structures in detail. One of the many possible quantum-dot structures based on semiconductor heterostructures (see Section 10.4) is shown in Figure 9.18a. A semi-insulating AlGaAs layer separates the (metallic) *gate*

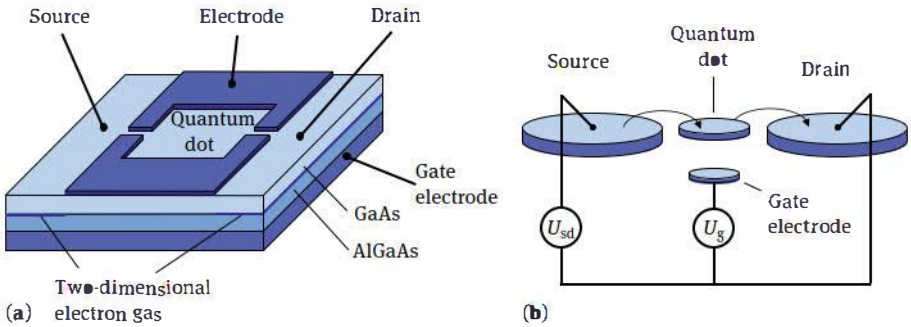


Fig. 9.18: Schematic representation of a quantum dot. **a)** The electrode on the GaAs layer defines the shape of the quantum dot. Tunneling through the potential barriers at the constriction sites allows the electrons to pass through the quantum dot from source to drain. **b)** Simplified representation of a quantum dot. The tunneling of the electrons is indicated by arrows. The source-drain voltage U_{sd} and the gate voltage U_g are shown.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

electrode from the semi-conducting GaAs layer. The positive gate voltage U_g causes the occurrence of a two-dimensional electron gas at the interface between the two materials. The structured electrode *on* the GaAs layer sits at a negative potential with respect to that of the source and drain. This means that the square of electrons at the center of the device, labelled Quantum Dot, are surrounded by an electrostatic potential barrier. If a small voltage is applied between source and drain, electrons can enter the potential well and leave it on the other side by tunnelling, because the potential barrier is somewhat lower at the constrictions created by the gaps in the metalization. The depth of the potential well can be easily changed by adjusting the gate voltage. Figure 9.18b shows the same situation in diagrammatic form.

If we measure the current through the quantum dot when a voltage U_{sd} of a few microvolts is applied, we observe that the conductance $G = I/U_{sd}$ as a function of the gate voltage U_g exhibits sharp maxima which are approximately equally spaced. This can be explained on a purely classical basis: if the potential well is very small, say a square with a 500 nm side, then we must take into account that each electron carries a discrete charge. The charge Q of the quantum dot is therefore not a continuously varying quantity but is given by $Q = Ne$, where N is the number of electrons on the quantum dot. If the gate voltage is now continuously increased, the charge makes jumps as shown schematically in Figure 9.19.

The electrostatic energy E of the quantum dot is given by:

$$E = \frac{Q^2}{2C} - \varphi Q = \frac{(Ne)^2}{2C} - Ne\varphi\,, \tag{9.50}$$

where the first term represents the capacitive *charging energy* of the N electrons, with C the total capacity. The second term represents the potential energy of the quantum dot, where φ is the electrostatic potential. Assuming for simplificity that the main contribution to the capacity C comes from the gate electrode, we can write $C \approx C_g$

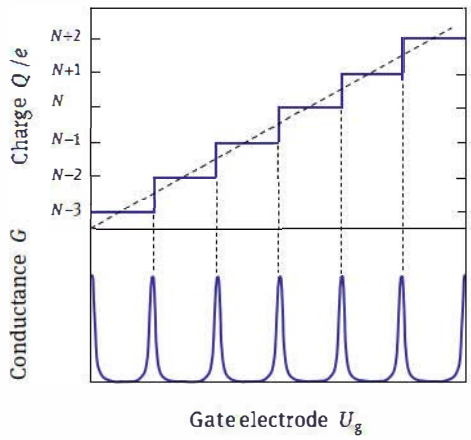


Fig. 9.19: Charge of a quantum dot and origin of the conductance maxima. Upper part: The charge of the quantum point changes by individual jumps, whereas the classical relationship $Q = CU_g$ assumes a continuous change (dotted line). Lower part: schematic representation of the conductance maxima, whose separation is given by e/C.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

and $\varphi \approx U_g$. Equation (9.50) shows that the smaller the quantum dot and associated capacity C, the greater the charging energy.

The gate voltage at which the charge increases by one electron is determined by the condition $E(N+1) = E(N)$. This reappears periodically, whenever $U_g = e(N + 1/2)/C_g$. At this voltage, the states with N or $(N + 1)$ electrons are degenerate, so that the quantum dot can fluctuate between these two states. In contrast, at the gate voltage $U_g = eN/C_g$ according to (9.50), an energy $e^2/2C$ must be supplied (or extracted) if an electron is to be added (or removed). This effect is called the **Coulomb blockade**. The relatively high Coulomb energy required to change the charge state leads to a suppression of the charge fluctuation in the quantum dot. As a result, the peaks in the conductance shown in Figure 9.19 always occur when the two adjacent states with N or with $(N + 1)$ electrons become degenerate as the appropriate voltage is crossed. At this voltage, an electron can tunnel through the first barrier, remains in the quantum point for a short time, and leaves it through the second barrier, and the process can then be continuously repeated. However, at intermediate voltages this cannot happen and current flow is suppressed. With increasing temperature, the steps become rounded and the conductance maxima broaden. A typical conductance oscillation, recorded at 0.1 K, is shown in Figure 9.20.

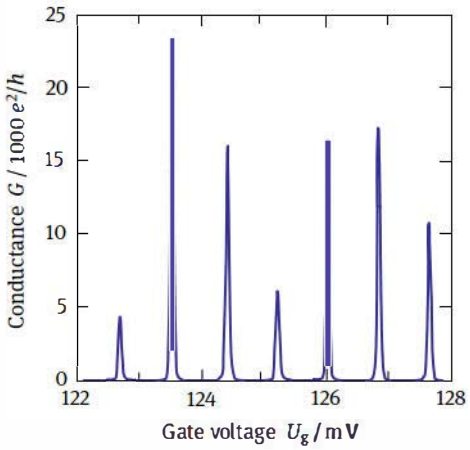


Fig. 9.20: Conductance oscillations of a quantum point in a GaAs/AlGaAs heterostructure as a function of the gate voltage U_g, measured at 0.1 K. The separations between the sharp conductance peaks are nearly equal. (After Control U. Meirav, E.B. Foxman, Semicond. Sci. Technol. **11**, 255 (1996).)

Up to now, only the charge energy has been considered and the eigenstates of the quantum dots have been disregarded. This classical approach is very well suited for the description of **Coulomb oscillations** when the quantum dots are made up of metallic islands. In this case the number of eigenstates of the islands is very large, the states are energetically dense. The Coulomb blockade energy is then exactly $e^2/2C$, and the conductance peaks occur at regular intervals.

For most quantum dots created with the help of heterostructures, the number of trapped electrons is small. Relatively large energy differences comparable to the charge energy occur between the eigenstates of the quantum dots. As follows from

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

elementary quantum mechanics, in the case of a circular potential well with radius r, the separation ΔE of the energy levels between two eigenstates is constant and can be approximated by

$$\Delta E \approx \frac{\hbar^2}{m^* r^2}\,, \tag{9.51}$$

where m^* is the effective mass of the electrons. If the charging energy and level separation are comparable, then in addition to the classical charging effect the splitting of ΔE must also be taken into account. This means that the separations between the conductance maxima are not necessarily the same.

On average, charges remain on the quantum dot for a time $\delta t = RC$, where R is the resistance that occurs when the electrons tunnel. This results in an energy uncertainty

$$\delta E = \frac{h}{\delta t} = \frac{e^2}{C}\frac{h}{e^2}\frac{1}{R}\,. \tag{9.52}$$

The energy uncertainty and the charging energy are comparable if $R \approx h/e^2$. Fluctuations then occur, which broaden the Coulomb charge effect. Of course, the thermal energy must also be small compared to the charging energy if the phenomenon of the Coulomb blockade is to be apparent. This means that defined states of charge only exist if both conditions

$$R \gg \frac{h}{e^2} \qquad \text{and} \qquad \frac{e^2}{C} \gg k_B T \tag{9.53}$$

are fulfilled. However, if the quantum dots under consideration are strongly coupled to source and drain, the charging effects described here are unimportant.

Quantum dots show many interesting effects that we cannot treat in detail. For example, the voltage U_{sd} between the source and the drain has a large influence on the measurement results and allows conclusions to be drawn about the level scheme of the quantum dots. Here we have also completely ignored the effects of electron spin, which strongly influences the properties of quantum dots. Another very interesting field is the behavior of quantum dots in magnetic fields.

9.2.9 Heat Transport in Metals

We now turn to the thermal conductivity of metals. First of all, it is an open question whether the heat transport is primarily due to the motion of electrons or, as in the case of insulators, to the propagation of phonons. Daily experience teaches us that at room temperature it is the electrons that are mainly responsible for heat transport since the thermal conductivity of metals is usually much greater than that of dielectrics. The fact that electrons also provide the dominant contribution at very low temperatures follows from the observation that the specific heat of electrons decreases linearly with decreasing temperature. The specific heat of the lattice, in contrast, decreases as T^3.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Thus at very low temperatures there are obviously many more excited electrons than excited phonons available for energy transport. Apart from the two limiting cases (low and high temperatures), it is difficult to separate the contributions of phonons and electrons in most conductive materials.

Figure 9.21 shows the temperature dependence of the thermal conductivity Λ of high-purity copper. The similarity with the temperature dependence for dielectric crystals is obvious. It is remarkable that although at room temperature metals have a much higher thermal conductivity than dielectric materials, the maximum value is smaller than that for pure dielectric crystals. We can understand this effect if we simply ignore the contribution of the phonons and assume that $\Lambda_{el} \gg \Lambda_{ph}$. To calculate the electronic thermal conductivity Λ_{el} we again use the analogy to the kinetic theory of gases, as we did already for the lattice contribution to the thermal conductivity in Section 7.3. Since the transport of heat, like the transport of charge, is only carried by excited electrons at the Fermi surface, we put into equation (7.21) for the thermal conduction, the specific heat C_V^{cel} taken from equation (8.35), the Fermi velocity v_F and the effective mass m^* and find:

$$\Lambda_{el} = \frac{1}{3} c_V^{el} v\, l = \frac{1}{3} \frac{\pi^2 n k_B^2 T}{m^* v_F^2} v_F\, l = \frac{\pi^2}{3} \frac{n k_B^2 \tau}{m^*} T \,. \tag{9.54}$$

The temperature dependence of electronic thermal conductivity can be divided into three temperature ranges with the help of equation (9.46):

$$\Lambda_{el} \propto l\,T \propto \begin{cases} T & \text{for} \quad T \ll \Theta \,, \\ T^{-4} & \text{for} \quad T < \Theta \,, \\ \text{const.} & \text{for} \quad T \gg \Theta \,. \end{cases} \tag{9.55}$$

At low temperatures ($T \ll \Theta$) the mean free path l of the electrons is limited by the scattering at impurities and is therefore constant. The temperature dependence of the

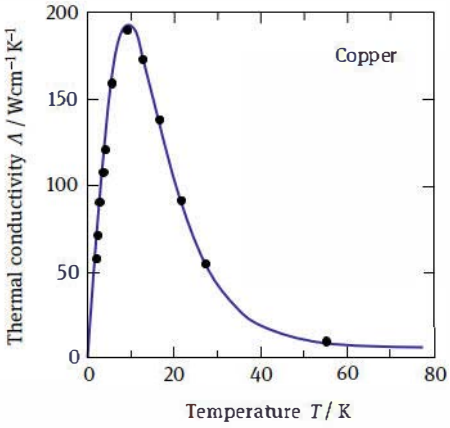


Fig. 9.21: The thermal conductivity of copper. Similarly to dielectric crystals, metals also show a pronounced maximum in the temperature dependence. (After G.K. White, R.J. Tainsh, Phys. Rev. **119**, 1869 (1960).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

thermal conductivity is therefore determined by the specific heat, which increases linearly with temperature. With increasing temperature, as the number and frequency of the dominant phonons increases, electron-phonon scattering also increases, and we reach the regime of the validity of the Bloch-Grüneisen law, where the decrease of the mean free path overcompensates for the increase of the specific heat. Thus, the thermal conductivity decreases proportionally to T^{-4}. At high temperatures, i.e. at $T \gg \Theta$, those phonons at the Debye frequency dominate, with numbers growing linearly with temperature. The mean free path therefore decreases inversely with the temperature, thus compensating for the increase in the specific heat and the thermal conductivity then becomes almost temperature-independent.

We now briefly return to the question of the limited contribution of the phonons to heat conduction. First, the question arises why phonons hardly contribute at all to the heat conduction of pure metals. The reason for this is that in pure metals the phonons are scattered not only by their interaction with one another but primarily by their interaction with the electrons. This scattering mechanism which also limits the electrical conductivity, is very effective at impeding phonon flow. Phonon-phonon scattering, which is dominant in dielectric solids, is therefore usually of no significance in metals. The small contribution of the phonons to heat conduction is therefore generally difficult to observe because it is masked by the heat transport by the electrons. Exceptions are alloys or highly contaminated metals, where the electrons are so strongly scattered at impurity atoms and defects that they contribute very little to thermal conductivity. Their thermal transport is therefore relatively small and largely carried by the phonons.

Wiedemann-Franz Law. Assuming that both electrical and thermal conductivity are caused by the free electrons, we can combine equations (9.54) and (9.38) to find the *Wiedemann-Franz law*[13,14]

$$\frac{\Lambda_{el}}{\sigma} = \frac{\pi^2}{3}\left(\frac{k_B}{e}\right)^2 T = LT \,. \tag{9.56}$$

The ratio of the two transport quantities should thus be determined only by the temperature and the universal **Lorenz number**[15] $L = 2.5 \times 10^{-8}\,\mathrm{W\Omega\,K^{-2}}$.

In Table 9.1 the room temperature values of the electrical and thermal conductivity as well as the Lorenz numbers of various metals are listed. Obviously at room temperature the Lorenz number is close to the expected value and thus the Wiedemann-Franz law is fulfilled to good approximation. This remarkable law is only valid, however, if the mean free path is limited in both cases by the same scattering mechanism.

13 Gustav Heinrich Wiedemann, *1826 Berlin, †1899 Leipzig
14 Rudolph Franz, *1826 Berlin, †1902 Berlin
15 Ludvig Valentin Lorenz, *1829 Helsingør, †1891 Frederiksberg

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Tab. 9.1: Thermal conductivity, electrical conductivity and Lorenz number of some metals at room temperature. (Various sources.)

Element	Λ_{el}/W/mK	$\sigma/(\mu\Omega m)^{-1}$	$L/W\Omega/10^8K^2$	Element	Λ_{el}/W/mK	$\sigma/(\mu\Omega m)^{-1}$	$L/W\Omega/10^8K^2$
Ag	429	68.0	2.15	Li	85	11.7	2.47
Al	237	40.0	2.03	Mg	171	25.4	2.29
Au	317	48.8	2.29	Na	141	23.8	2.02
Cu	401	59.6	2.84	Pb	35	5,2	2,30
Fe	80	11.2	2.44	Sn	67	9.1	2.49
K	102	16.4	2.13	Zn	121	18.4	2.24

Figure 9.22 shows that copper at high temperatures indeed follows the Wiedemann-Franz law and the Lorenz number is close to the expected value. This is not only true for copper, but also for most metals to a good approximation. With decreasing temperature the Lorenz number of copper decreases, passes through a minimum and reaches the theoretical value again at low temperatures. The reason for the temperature dependence is the different weighting of the scattering processes limiting electrical and thermal transport, which we now look at in more detail.

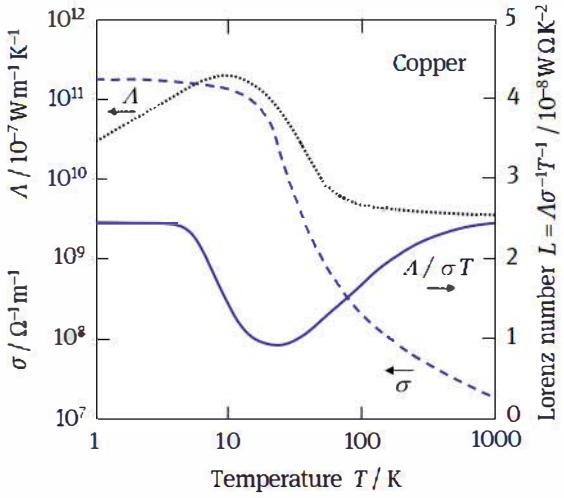


Fig. 9.22: Temperature dependence of the electrical (dashed) and thermal (dotted) conductivity of copper. The blue curve shows the temperature dependence of the extracted Lorenz number. (After J.S. Blakemore, *Solid State Physics*, Saunders, 1974 and G.K. White, R.J. Tainsh, Phys. Rev. **119**,1869 (1960).)

9.2.10 Fermi Function in Stationary Equilibrium

If an electric field is applied to a metal, then according to (9.24) all wave vectors change continuously according to $\delta\mathbf{k} = -e\tau\mathcal{E}/\hbar$ and thus a displacement of the Fermi sphere oc-

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

curs, as already discussed in Figure 9.5. Formally, this also follows from equation (9.33), which shows that $f(\mathbf{k})$ can be understood as the beginning of a Taylor expansion in terms of the wave vector $\mathbf{k}$ about the equilibrium distribution $f_0(\mathbf{k})$. A DC electrical field does not change the shape of the Fermi-Dirac distribution function, but transforms the equilibrium function $f_0(\mathbf{k})$ into a new function $f(\mathbf{k})$, both differing only in their arguments. In the context of the wave vector or the energy of the electrons, this statement can be illustrated by the relationship

$$f(\mathbf{k}) = f_0\left(\mathbf{k} + \frac{e\tau\mathcal{E}}{\hbar}\right) = f_0\Big(E(\mathbf{k}) + e\tau\mathbf{v}(\mathbf{k})\cdot\mathcal{E}\Big) \tag{9.57}$$

Here we have used the fact that the change of the wave vector causes an energy change $\delta E \approx -e\tau\mathbf{v}(\mathbf{k})\cdot\mathcal{E}$.

Expressed in words, this means that those electrons at the front side of the Fermi surface traveling against the field gain energy (because of their negative charge), while those on the rear side are slowed down. Thus, as shown in Figure 9.23, the wave vectors change by the amount $\pm\delta\mathbf{k}$ depending on the direction of motion. In reality, however, this is not a static shift, but a dynamic process in which the electrons are repeatedly scattered, moving along the Fermi surface from the front to the back.

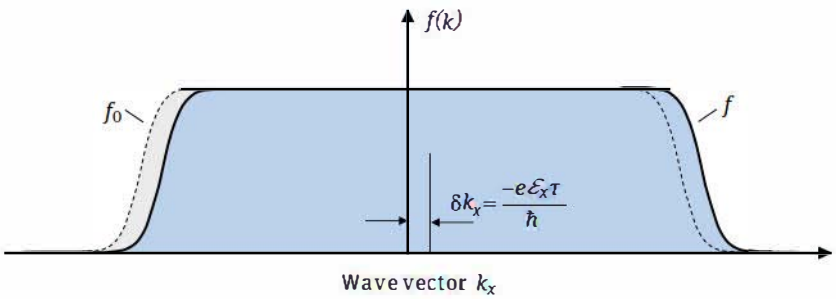


Fig. 9.23: The Fermi-Dirac distribution $f(E)$ in the presence of an electric field $-\mathcal{E}_x$. The equilibrium distribution $f_0(E)$ of the undisplaced Fermi surface is shown as a dashed line. Under the influence of the field the wave vectors are shifted to the right by the amount δk_x. This increases the number of states with a positive wavevector while the number with a negative wave vector decreases.

If heat flows through the sample, the velocity distribution of the electrons depends on the direction in which they move with respect to the heat flow. The electrons traveling from the warm end and moving in the direction of the heat flow have a higher "temperature" and are therefore faster than those moving in the opposite direction. In order to find the transformation of the Fermi-Dirac distribution function analogous to (9.57), we start from (9.27), but now consider the diffusion term, which takes into account the spatial variation of the distribution function. After some calculation steps we find a

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

result analogous to (9.57):

$$f(\mathbf{k}) = f_0\Big(T - \tau \mathbf{v}(\mathbf{k}) \cdot \nabla T\Big) \,. \tag{9.58}$$

The electrons are therefore "warmer" or "colder" by $\delta T = -\tau \mathbf{v}(\mathbf{k}) \cdot \nabla T$ owing to the existing temperature gradient depending on the direction of motion. The Fermi-Dirac distribution of the electrons traveling from the cold end of the sample has a steeper slope, while those coming from the warm end have a flatter slope. This change in the distribution function is sketched for the two cases in Figure 9.24.

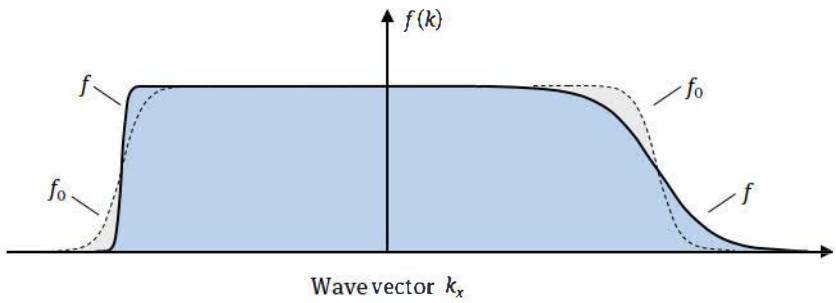


Fig. 9.24: The Fermi-Dirac distribution $f(E)$ in a sample with a temperature gradient. The equilibrium distribution $f_0(E)$ is shown as a dashed line. The distribution function of the electrons from the cold end of the sample is steeper than the thermal- equilibrium profile, while that of the electrons from the warm end is flatter.

Obviously, the requirements for the relaxation processes are different for electrical and thermal transport. For the former it depends on the change of momentum, and for the latter on the change of energy. During heat transport, a stationary state is reached as the "hot" electrons lose their excess energy by scattering and thus return to the Fermi surface. Here the direction of motion plays a minor role, because the "correct" temperature, i.e. the energy and thus the magnitude of the electron momentum, is of primary importance. During the scattering processes, on average the electrons coming from the warm end of the sample give off their excess energy, and the electrons from the cold end of the sample take up energy.

On the other hand, when an electric field is applied, equilibrium is established by the scattering of electrons from the front to the rear of the Fermi surface. Therefore, to be effective scattering processes must primarily change the momentum of the electrons. At high temperatures *one* scattering event is sufficient because the momentum of the dominant phonons is sufficiently large (see Figure 9.7b). These scattering processes are often called *horizontal* processes. At low temperatures several scattering processes are necessary because the scattering events lose their "effectiveness". If the changes in momentum are small, as shown in Figure 9.7a, they are called *vertical* processes.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

9.3 Electrons in Magnetic Fields

So far we have been studying the dynamic behavior of electrons in electric fields and in temperature gradients. Now we turn to the influence of magnetic fields, which give rise to a variety of interesting effects.

In this and the previous chapter it has become clear that the Fermi surfaces are of primary importance in determining the behavior of solids. They are surfaces of constant energy in **k**-space and separate the occupied from the unoccupied states at $T = 0$. Knowledge of their shape is of great importance in the description of many properties. There are several experimental methods for the determination of Fermi surfaces, most of which involve magnetic fields. These include cyclotron resonance, de Haas-van Alphen effect, Shubnikov-de Haas effect, magnetoresistance and ultrasonic attenuation. The Fermi surface can also be determined by photoelectron spectroscopy, measurements of the anomalous skin effect or by optical reflection. We discuss cyclotron resonance and the de Haas-van Alphen effect in more detail here.

9.3.1 Cyclotron Resonance

The measurement of cyclotron resonance is a "classical" experiment for the determination of Fermi surfaces. The sample is placed in a static magnetic field and irradiated with microwaves. On tuning the frequency of the microwave, a distinct absorption maximum can be observed at the cyclotron frequency.

If a magnetic field acts on an electron in a solid, it follows from equations (9.2) and (9.3) that there is a change of wave vector given by:

$$\mathrm{d}\mathbf{k} = -\frac{e}{\hbar^2}\left[\nabla_{\mathbf{k}} E(\mathbf{k}) \times \mathbf{B}\right]\mathrm{d}t \,. \tag{9.59}$$

Since magnetic fields that are constant in time do not cause any change in the energy of the deflected electrons, conservation of energy ensures that the electrons in the magnetic field must move on surfaces of constant energy, i.e. that $\mathrm{d}\mathbf{k}$ must be perpendicular to $\nabla_{\mathbf{k}} E$. The trajectories of the electrons with the Fermi energy are thus restricted to the Fermi surface.

Owing to the symmetry of Fermi surfaces, in many cases the motion will take place around a closed path. As shown in Figure 9.25a for a simple Fermi surface, the electron path is perpendicular to the magnetic field. From (9.59), the change $\mathrm{d}\mathbf{k}$ in the wave vector **k** is proportional to $\nabla_{\mathbf{k}} E(\mathbf{k}) \times \mathbf{B}$. The cross product has the effect that only the component $[\nabla_{\mathbf{k}} E(\mathbf{k})]_\perp$ of the gradient perpendicular to the magnetic field is involved, so that $|\nabla_{\mathbf{k}} E(\mathbf{k}) \times \mathbf{B}| = B\,[\nabla_{\mathbf{k}} E(\mathbf{k})]_\perp$. For simplification, we introduce the new vector $\mathrm{d}\mathbf{k}_\perp$, perpendicular to $\mathrm{d}\mathbf{k}$ and **B**. This allows $[\nabla_{\mathbf{k}} E(\mathbf{k})]_\perp$ to be expressed in a simplified way by $\mathrm{d}E/\mathrm{d}k_\perp$.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

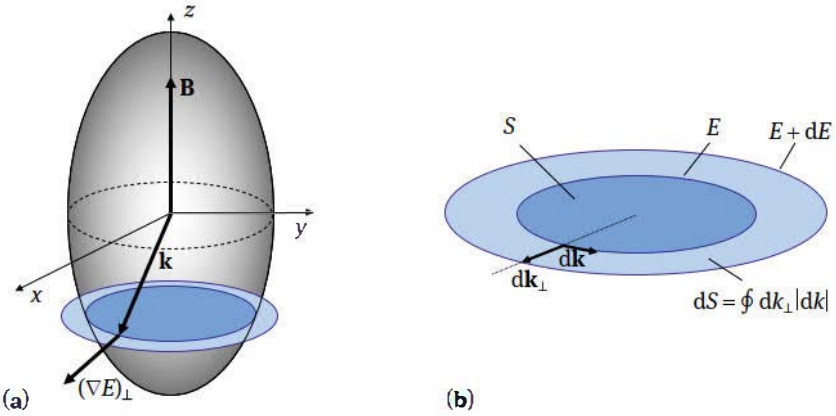


Fig. 9.25: Illustrating the electron motion in a magnetic field. **a)** The orbit of an electron on a Fermi surface having the shape of a rotational ellipsoid. **b)** The orbit plane showing the cross-sectional area S and the definitions of $\mathrm{d}\mathbf{k}$ and $\mathrm{d}\mathbf{k}_\perp$.

We now calculate the time $\tilde{T}$ for an electron to complete one orbit. For this we integrate equation (9.59) and obtain:

$$\tilde{T} = \oint \mathrm{d}t = \frac{\hbar^2}{eB} \oint \frac{|\mathrm{d}\mathbf{k}|}{|\mathrm{d}E/\mathrm{d}k_\perp|} = \frac{\hbar^2}{eB} \oint \frac{\mathrm{d}k_\perp}{\mathrm{d}E} |\mathrm{d}\mathbf{k}| = \frac{\hbar^2}{eB} \frac{\mathrm{d}S}{\mathrm{d}E} \,. \tag{9.60}$$

Here we have used the fact that $\mathrm{d}S = \oint \mathrm{d}k_\perp |\mathrm{d}\mathbf{k}|$ for the cross-sectional area S in $\mathbf{k}$-space, as shown in Figure 9.25b. Thus we obtain the following expression for the circulation time:

$$\tilde{T} = \frac{\hbar^2}{eB} \frac{\mathrm{d}S}{\mathrm{d}E} \,. \tag{9.61}$$

The orbital period is thus determined by the energy dependence of the cross-sectional area in $\mathbf{k}$-space enclosed by the orbit. For the special case of a free electron gas this quantity is $\mathrm{d}S/\mathrm{d}E = 2\pi m/\hbar^2$ independent of the wave vector. This means that in this case all electrons have the same circulation period. The circulation frequency ω_c is then given by the known cyclotron frequency $\omega_c = eB/m$. In general, however, the circulation times differ. To be able to use the expression (9.61) beyond the limiting case of free electrons, we replace the free-electron mass m by the effective mass m^*. It must be noted that this **cyclotron mass** does not always match the *dynamic mass* introduced in Section 9.1. The cyclotron mass is determined by the position of the orbit in question on the Fermi surface and not by an electronic state as defined in equation (9.5). With this generalization for the **cyclotron mass**, the well-known formula for the **cyclotron frequency** remains:

$$\omega_c = \frac{eB}{m^*} \,. \tag{9.62}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

It is worth noting here that the orbit of an electron in real space is somewhat more complicated than that in **k**-space, since the z-component of the wave vector must also be taken into account. Thus electrons will describe spirals in real space.

As already mentioned, the cyclotron resonance is measured by placing the sample in a magnetic field and irradiating it with microwaves of suitable frequency. When the cyclotron and microwave frequencies coincide, the electrons are accelerated absorbing energy from the electric field. As a result, on tuning the microwave frequency, a sharp absorption maximum occurs at the cyclotron frequency.[16] In experiments on metals, a complication arises that electromagnetic waves can hardly penetrate the sample due to the *skin effect*. The penetration depth δ is given by:

$$\delta = \sqrt{\frac{2}{\mu_0 \omega \sigma}}\,, \tag{9.63}$$

where μ_0 is the magnetic permeability constant. The skin depth δ decreases with increasing frequency and at high frequencies becomes smaller than the electron-orbit diameter. If the experiment is performed at a field $B = 1\,\mathrm{T}$, the cyclotron frequency for free electrons is 30 GHz. The penetration depth at this frequency is about 10 nm for pure copper and low temperatures, whereas the orbit radius is about 10 μm. This means that the orbiting electrons only come into contact with the electric field within the penetrtion depth at the surface. There, depending on the phase relation, energy is added or subtracted. Absorption occurs when the electrons are accelerated in phase by

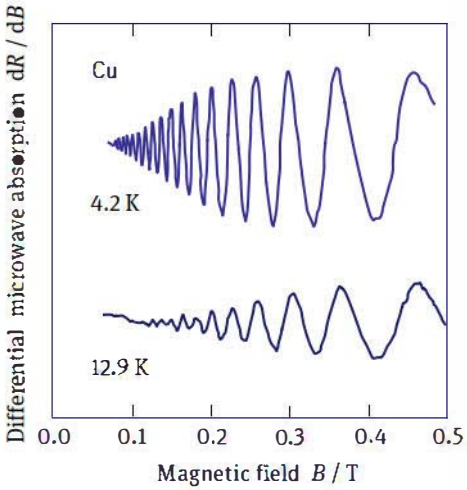


Fig. 9.26: Differential microwave absorption dR/dB of copper at 4.2 K and 12.9 K as a function of the static magnetic field. (After P. Häussler, S.J. Welles, Phys. Rev. **152**, 675 (1966).)

16 As we describe the method above, we perform the experiment at a constant magnetic field and vary the microwave frequency. However, in practice it is usually the magnetic field that is varied and the microwave frequency kept constant, the reason being that it is relatively easy to vary the magnetic field but virtually impossible to vary a microwave frequency continuously.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the electric field as they pass through the interaction zone. As a result, an absorption maximum occurs not only at the cyclotron frequency ω_c, but also when an integral number of periods of the microwave oscillation have elapsed between two cycles of the cyclotron motion. In metals, therefore, a series of maxima is observed, whereas in semiconductors generally only *one* absorption maximum occurs due to the smaller charge carrier density and thus larger penetration depth. The result of a cyclotron resonance experiment on a (110) copper surface at 35 GHz and a static magnetic field in the [100] direction is shown in Figure 9.26. The expected oscillation of the differential absorption dR/dB is clearly visible.[17]

Sharp resonances can only be obtained if the electrons can complete several orbits undisturbed, i.e. if the condition $\omega_c\tau \gg 1$ is fulfilled, where τ is the average electron scattering time. Experiments on cyclotron resonance must therefore be carried out at high frequencies and thus at high magnetic fields. To avoid scattering with phonons and defects and to achieve long scattering times, the experiments must be made at low temperatures using pure materials. The broadening of the maxima as the temperature increases from 4.2 K to 12.9 K, clearly seen in Figure 9.26, shows that the interaction of the circulating electrons with the lattice vibrations leads to a considerable reduction of the mean scattering time even at 10 K.

If the effective mass is not the same for all electrons, their orbital times will differ. Since the electrons do not absorb at one well-defined frequency but at different frequencies, we might not expect any sharp absorption maxima in real metals. However, there are two reasons why the do occur. First, only electrons at the Fermi surface can take up energy and contribute to absorption, since the unoccupied states of higher energy are available only at this location. The electrons inside the Fermi sphere perform orbits, but do not contribute to the measurement signal. However, this restriction alone would not lead to maxima, because the effective mass of the electrons generally depends on their position on the Fermi surface. Secondly, the observed signal comes from a region of the Fermi surface where a large number of electrons circulate at approximately the same frequency. This is the case when the circulation times are stationary with respect to small changes of the in-field component of the wave vector, i.e. when the cross-sectional area S has an extremum. These electrons absorb within a narrow bandwidth at the same microwave frequency or magnetic field. Such paths are called **extremal orbit**. Figure 9.27 shows two such areas. If the magnetic field $\mathbf{B}$ is applied, the electrons in region A fulfill this condition and cause an absorption maximum. In contrast, the electrons in region B circulate at different velocities. Their contribution to the signal contributes to the background and does not cause a maximum. If the magnetic field is tilted in direction $\mathbf{B}'$, the orbits in area A$'$ can be observed, because they exhibit

17 Here we use the term dR/dB, which is common in this field of work, although the abbreviation R is also used for radius or resistance.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

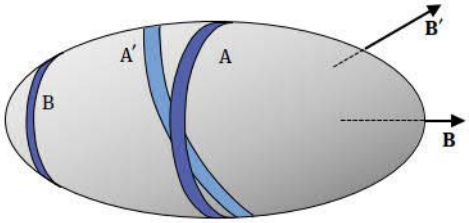


Fig. 9.27: Extremal orbits. The orbits in areas A and A′ are extremal orbits, associated with the magnetic fields $\mathbf{B}$ and $\mathbf{B}'$. The orbits in area B do not fulfill the conditions for extremal orbits and are therefore not visible in cyclotron resonance experiments.

extremal orbits for this field direction. The Fermi surface can thus be scanned in this way by changing the sample orientation.

We refer to Figure 9.28, which shows the Fermi surface of copper in the periodic zone scheme. This illustration indicates various extremal paths, which are less easy to recognize in the reduced zone scheme. Orbits A, B and C are *closed orbits* giving rise to absorption maxima. Trajectory D is an *open path* where no resonance can occur, and is thus not detected in cyclotron resonance experiments. However, open orbits do contribute significantly to *magnetoresistance*, which we will not discuss here. The orbits A and B are called **belly** and **neck orbits**, respectively. In these two cases, the occupied states lie inside the orbital loops, and the direction of circulation corresponds to that of free electrons. The topology of orbit C is different from the first two orbits. As in Figs. 8.29b and 8.32a, the occupied states lie on the *outside* of the orbital loop, thus changing the sign of $\nabla_{\mathbf{k}} E(\mathbf{k})$ in equation (9.59) and thus also the direction of the orbital motion. Orbits A and B are known as **electron orbits** and orbit C is known as a **hole orbit**. In multiply-connected Fermi surfaces, different types of orbits can occur simultaneously if the magnetic field points in the appropriate direction.

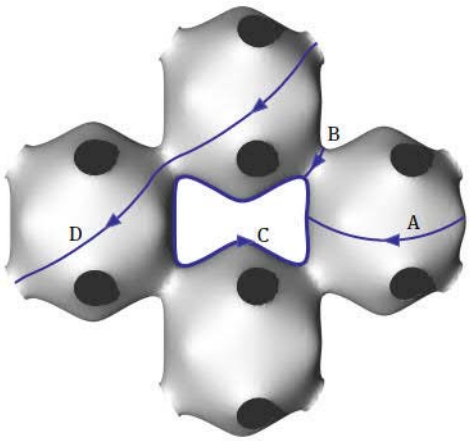


Fig. 9.28: Electron paths in magnetic fields. The Fermi surface of copper in the periodic zone scheme is shown. Trajectory A shows a belly orbit. Trajectory B is a neck orbit, C is a "dog bone" orbit and D is an open trajectory. Trajectories A and B are electron orbits, while trajectory C is a hole orbit. The open trajectory D lies topologically between the two orbit types. The orbits depicted are associated with different orientations of the magnetic field.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

9.3.2 Landau Levels

An extremely interesting phenomenon, which we have so far ignored, is the fact that in magnetic fields electron orbits must be quantized. To study this effect, we consider the motion of quasi-free electrons in a constant magnetic field. To simplify matters, we initially ignore the spin, since its magnetic moment does not influence orbital motion, but only adds a term in the energy calculation. An exception to this is the *quantum Hall effect*, which we will discuss briefly at the end of this chapter. A more detailed discussion of the electron spin will be given in Chapter 12 in connection with the magnetic properties of solids.

In the quantum mechanical description, the magnetic field is taken into account via the vector potential $\mathbf{A}$ in the Hamiltonian operator. Thus the stationary Schrödinger equation takes the following form:

$$\frac{1}{2m^*}(-\mathrm{i}\hbar\nabla + e\mathbf{A})^2\psi = E\psi \,. \tag{9.64}$$

We assume that the magnetic field $\mathbf{B}$ is oriented in z-direction and use the corresponding vector potential $\mathbf{A} = (0, xB, 0)$.[18] Writing the equation in Cartesian coordinates, we obtain:

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial z^2} + \left(\frac{\partial}{\partial y} + \frac{\mathrm{i}eBx}{\hbar}\right)^2\psi + \frac{2m^*E}{\hbar^2}\psi = 0 \,. \tag{9.65}$$

We use here the trial function:

$$\psi = \tilde{\psi}(x)\,\mathrm{e}^{-\mathrm{i}(k_y y + k_z z)} \,, \tag{9.66}$$

which goes back to *L.D. Landau* and maintains plane waves in y- and z-direction. Using this approach we find:

$$\frac{\partial^2\tilde{\psi}}{\partial x^2} + \left[\frac{2m^*E}{\hbar^2} - k_z^2 - \left(\frac{eBx}{\hbar} - k_y\right)^2\right]\tilde{\psi} = 0 \,. \tag{9.67}$$

With the following abbreviations

$$E_\ell = E - \frac{\hbar^2 k_z^2}{2m^*}, \qquad x' = x - \frac{\hbar k_y}{eB} \qquad \text{and} \qquad \omega_\mathrm{c} = \frac{eB}{m^*} \tag{9.68}$$

the equation takes the simple form:

$$\frac{\partial^2\tilde{\psi}}{\partial x'^2} + \frac{2m^*}{\hbar^2}\left[E_\ell - \frac{1}{2}m^*\omega_\mathrm{c}^2 x'^2\right]\tilde{\psi} = 0 \,. \tag{9.69}$$

This is the Schrödinger equation for a linear harmonic oscillator with frequency ω_c, centered at

$$x_0 = \frac{\hbar k_y}{eB} = \frac{1}{\omega_\mathrm{c}}\frac{\hbar k_y}{m^*} \,, \tag{9.70}$$

18 We have chosen the Coulomb gauge here.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

with eigenfunctions $\bar{\psi}(x)$ and eigenvalues $E_\ell = (\ell + 1/2)\,\hbar\omega_c$ well-known from quantum mechanics. Expressing the energy of the electrons with the original variables, we get, with $\ell \geq 0$,

$$E = E_\ell + E(k_z) = \left(\ell + \frac{1}{2}\right)\hbar\omega_c + \frac{\hbar^2 k_z^2}{2m^*} \,. \tag{9.71}$$

The energy of the electrons comprises two parts: first, the electrons have the usual translational energy in the field direction, but secondly, they also have a quantized energy associated with the circular motion. Under the influence of a magnetic field, the parabolic free-electron bands split into sub-bands called **Landau levels** with energy eigenvalues separated by $\delta E = \hbar\omega_c$. The magnetic field causes a quantization of the electron energy in the xy-plane, while nothing changes in the z-direction. Figure 9.29 shows the electrical energy as a function of k_z. If we assume a field of two Tesla, then with $m^* \approx m$ the splitting is $\delta E \approx 0.1\,\text{meV}$ or $\delta E/k_B \approx 1\,\text{K}$. At a typical Fermi temperature $T_F \approx 5 \times 10^4\,\text{K}$, there are a very large number of Landau levels occupied by electrons. The electron spins also contribute an additional splitting, which we will briefly discuss along with the quantum Hall effect, and the effects on the magnetic properties of solids in detail in Section 12.2.

It is remarkable that the spectrum is very similar to that of one-dimensional metals. While in the latter the separation is determined by the transverse eigenvalues and thus by the geometry, here all sub-bands have the same separation $\hbar\omega_c$.

What are the consequences for the motion of the electrons? To illustrate this we use Bohr's correspondence principle, which states that in the limiting case of large quantum numbers, the classical and the quantum-mechanical descriptions become identical. For the observable electrons in the vicinity of the Fermi energy, classical orbits are certainly a very good approximation, since they exhibit large quantum numbers in typical magnetic fields. The electrons move in the xy-plane on a circular orbit

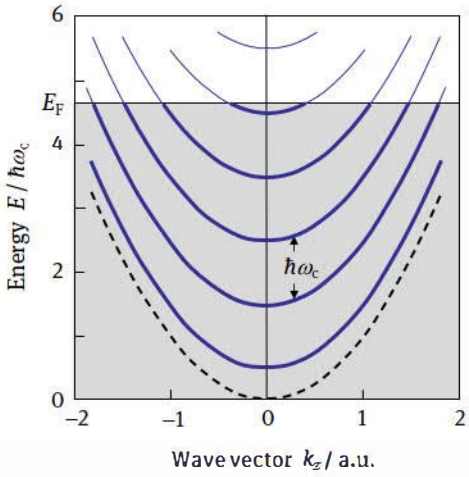


Fig. 9.29: The electron energy in a magnetic field as a function of the wave number k_z. The dashed curve shows the energy of free electrons with no magnetic field. The states of the subbands are occupied up to the Fermi energy.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

whose classical orbit radius r_ℓ is determined by the amplitude of the linear oscillator in equation (9.69). For the radius we therefore find:

$$r_\ell^2 = \left(\ell + \frac{1}{2}\right) \frac{2\hbar}{m^* \omega_C} . \tag{9.72}$$

Clearly the energy values E_ℓ are linked to specific orbital radii. The motion in the z-direction is not affected. It follows directly from (9.72) that the magnetic flux Φ_ℓ enclosed by the orbit is also quantized:[19]

$$\Phi_\ell = \pi r_\ell^2 B = \left(\ell + \frac{1}{2}\right) \frac{h}{e} . \tag{9.73}$$

In **k**-space, the electrons describe circular orbits with quantized radii k_ℓ and enclosing well-defined areas S_ℓ. From the classical orbital velocity, taking equation (9.71) into account, we obtain for magnitude of the wave vector in the xy-plane:

$$k_\ell^2 = \left(\ell + \frac{1}{2}\right) \frac{2m^* \omega_C}{\hbar} . \tag{9.74}$$

Along with the area S_ℓ enclosed in **k**-space:

$$S_\ell = \left(\ell + \frac{1}{2}\right) \frac{2\pi eB}{\hbar} . \tag{9.75}$$

Since the electron orbits in the xy-plane enclose an area $A_\ell = \pi r_\ell^2 = (2\ell + 1)\pi\hbar/eB$, we can write for the areas:

$$A_\ell = \left(\frac{\hbar}{eB}\right)^2 S_\ell . \tag{9.76}$$

Thus as the magnetic field increases, the spirals become tighter in real space, while the circles in momentum space become larger. This relationship between the enclosed areas, A_ℓ and S_ℓ, is not only valid for the spiral tracks of free electrons, but in general applies to all closed orbits.

A closer look at the motion of the electrons shows that the operators for k_x and k_y do not commute and thus their expectation values cannot be determined independently of one another. This means that the magnetic field leads to a breakdown in the wave-vector quantization scheme we have used previously. The quantities k_x and k_y lose their validity as "good" quantum numbers. In the absence of any magnetic field, the allowed wave vectors are evenly distributed in **k**-space. However, in a magnetic field they lie on circles, as shown in Figure 9.30b. However, this representation, where we still use the axes with the old quantum numbers k_x and k_y as coordinates, is not quite correct. The states on the circles cannot really be represented by stationary points, since they rotate around the perimeter with the angular frequency ω_C.

19 Equation (9.73) is only valid for electrons in parabolic bands. In the general case, the factor of 1/2 in the brackets is replaced by a *phase factor* γ.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

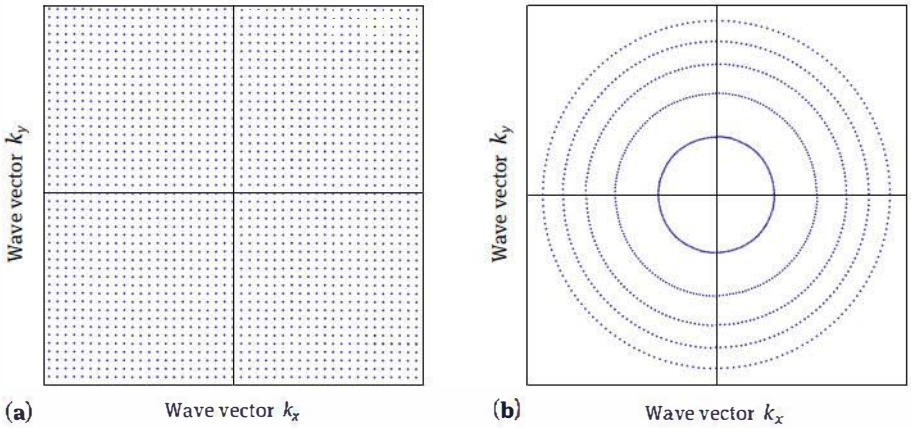


Fig. 9.30: Influence of a magnetic field on the eigenvalues of the wave vectors of free electrons in the xy-plane. The component of the wave vectors in field direction is not influenced. **a)** Allowed wave vectors without magnetic field, **b)** allowed wave vectors in the magnetic field.

If we view the allowed states in three dimensions in **k**-space, we see that they lie on cylindrical surfaces known as **Landau tubes** or **Landau cylinders**. In the general case, the precise profile of the tubes is determined by the energy surface $E(\mathbf{k})$ and hence they can deviate from a cylindrical shape. On applying a field, the states which were allowed with no field "condense" on the nearest tube. All states on a tube are degenerate with respect to k_x and k_y and have the same quantum number ℓ. The wave vector in the direction of the magnetic field, on the other hand, is still determined by the geometric boundary conditions. Figure 9.31 shows the cylinders occupied by electrons, with the surface of the original Fermi sphere forming the spherical boundary.

Let us now consider the re-arrangement of the states in **k**-space in the xy-plane when we apply a magnetic field. We know that the original states in the absence of a magnetic field are evenly distributed in the $k_x k_y$-plane. After the field is applied, the states all lie on circles as shown in Figure 9.30. Now consider the circle with quantum number ℓ. The states which condense on this circle are those original states (in the absence of the magnetic field) lying in the annular ring defined by the condition $k^2_{\ell-1/2} < k^2_\ell \leq k^2_{\ell+1/2}$. The area $\Delta S_\ell = 2\pi m^* \omega_c/\hbar$ of each annulus is independent of the quantum number. If we have a square sample of edge length L, then, with no magnetic field, the density of states in the $k_x k_y$-plane is given by $\varrho^{(2)}_{\mathbf{k}} = 2L^2/4\pi^2$ according to (8.10). In this case, there are $\varrho^{(2)}_{\mathbf{k}} 2\pi m^* \omega_c/\hbar$ states on each ring. Thus each Landau level has the degeneracy:

$$g_e = \frac{L^2}{\pi}\frac{m^*\omega_c}{\hbar} = \frac{2eB}{h}L^2 . \tag{9.77}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

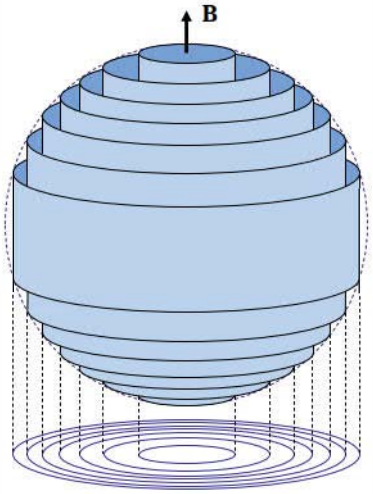


Fig. 9.31: The Landau tubes of a free electron gas. The occupied cylinders are bounded by the Fermi sphere. Also shown is the projection of the cylinders onto the xy-plane.

It should be emphasized here that we have taken into account the spin degeneracy by the added factor of 2 in the expression for $\varrho_{\mathbf{k}}^{(2)}$.[20]

We should note here that to establish the quantization, the electrons must perform several unperturbed cycles for the temporal and spatial phase of the wave function to be defined. As in cyclotron resonance, the condition $\omega_c\tau \gg 1$ must therefore be fulfilled for the formation of well-defined states, with τ again indicating the mean scattering time. Owing to the scattering of electrons by the lattice or by defects, the ideal infinitely thin-walled Landau cylinders become broadened. When an electron is excited, e.g. at cyclotron resonance, a transition from one Landau tube to another with a higher quantum number ℓ takes place. Of course, transitions are only possible near the Fermi surface, since filled initial states and empty final states are only available at this location.

9.3.3 Density of States in Magnetic Fields

The quantization of the electron motion and the degeneracy of the states in magnetic fields has astonishing effects on the density of states and the magnetization of metals. Returning to equation (9.71), which describes the energy eigenvalues of free electrons in a magnetic field, we can already get an idea of the density of states of two- and three-dimensional systems in magnetic fields. The magnetic field leads to a quantization in the xy-plane with energy eigenvalues $(\ell + 1/2)\,\hbar\omega_c$. In two-dimensional systems whose density of states is constant according to (8.11), this results in the complete

20 In the literature, the term $g_e = eBL^2/h$ is also used for the degeneracy. In this case, the spin degeneracy is not included.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

quantization of the system in the new eigenstates. The density of states is, as with harmonic oscillators, a sum of delta functions (see Figure 9.32a), whose weight is given by the degree of degeneracy.

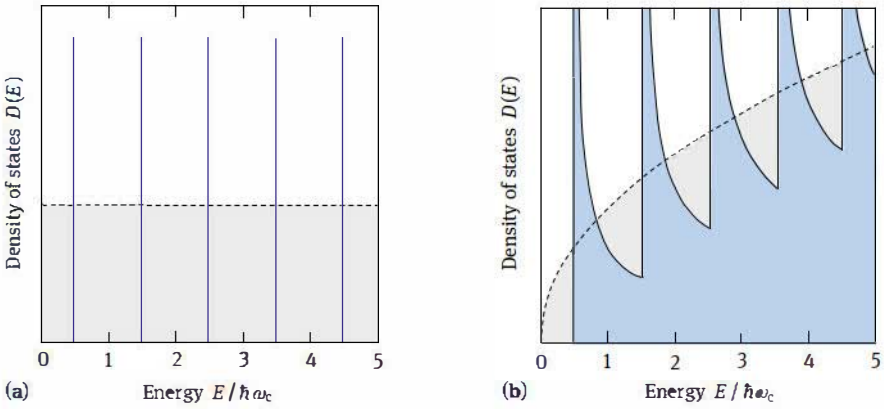


Fig. 9.32: Electronic density of states with (light blue) and without a (grey) magnetic field. For simplicity, the splitting arising from the spins is not included in this illustration. **a)** The density of states for a two-dimensional system. **b)** The density of states for a three-dimensional system.

Of course, in three-dimensional samples the density of states also includes the quantization owing to the electron motion in the z-direction, as shown in Figure 9.32b. This we can see from Figure 9.29, which shows the energy spectrum of the electrons in the subbands, in comparison with Figure 8.4. Obviously this is a superposition of the densities of states as expressed in (8.20) for the one-dimensional case. The spacing of the Van-Hove singularities is given by $\hbar\omega_c$. Since the magnetic field changes the distribution of the states but not their number, the areas under the density of states with an applied field (solid line) and without (dashed line) are equal.

As we will see below, magnetic fields change the internal energy and thus the thermodynamic properties of solids. Let us first look at the situation in *two-dimensional* electron gases, whose density of states with and without an appied magnetic field is shown in Figure 9.33a. From a constant density of states in magnetic fields a series of δ-functions emerge.

As the strength of the applied magnetic field is increased, the spacing between the Landau levels also increases. Thus, the uppermost occupied level, wbich determines the Fermi energy, also shifts to higher energy values. Simultaneously, according to equation (9.77), the degree of degeneracy also increases, so that more and more electrons can be accommodated in the lower levels. The occupation of the uppermost level therefore decreases steadily. At some point, with increasing field this level is completely emptied and the Fermi energy abruptly jumps to the level below.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

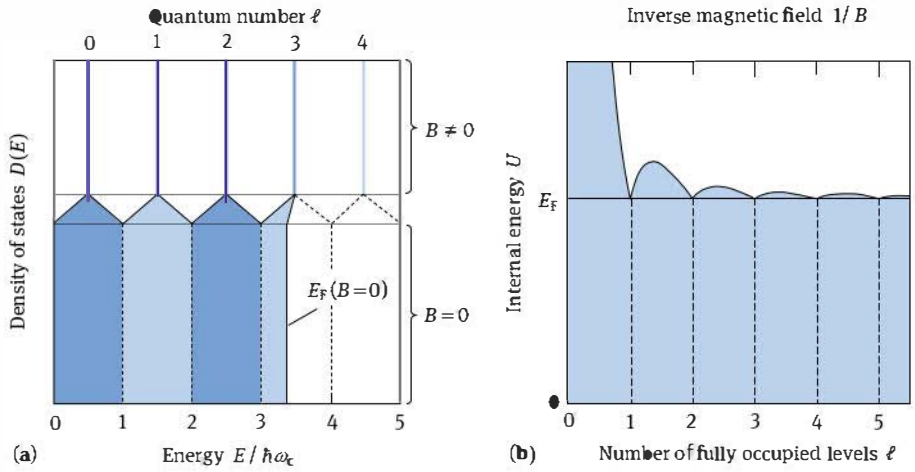


Fig. 9.33: a) The occupation of the states in a two-dimensional electron gas with (top part) and without (bottom part) a magnetic field. In this illustration the states up to $\ell = 2$ are completely occupied, state $\ell = 3$ is only partially occupied, and the states above are empty. The different blue coloration indicates which states coalesce to a common level when a magnetic field B is applied. b) Schematic plot of the internal energy of a two-dimensional electron gas as a function of the inverse magnetic field. The number ℓ of fully occupied Landau levels plotted on the abscissa is proportional to $1/B$.

Whenever $N = \ell g_e$, where N is the number of electrons and ℓ is the number of occupied levels, then (at $T = 0$) the levels are only either completely filled or empty, i.e. there are no partially occupied levels. As can be seen in Figure 9.33a in this case as many states are raised as simultaneously lowered, so that the total internal energy remains unchanged. If, on the other hand, a level is only partially occupied, then more states are raised than lowered and the internal energy is higher than in the magnetic field-free case.

In the case where there are only fully occupied levels, the relation $E_F = \left(\ell + \frac{1}{2}\right)\hbar\omega_c$ applies to two-dimensional systems according to equation (9.71). If we insert the expression for the cyclotron frequency, we find

$$\ell + \frac{1}{2} = \frac{m^* E_F}{e\hbar}\,\frac{1}{B} \qquad \text{and} \qquad \frac{1}{B} = \left(\ell + \frac{1}{2}\right)\frac{e\hbar}{m^* E_F}\,. \tag{9.78}$$

Fully occupied levels therefore occur for equal steps of $1/B$. This is directly connected with the oscillation of the internal energy with the magnetic field, shown schematically in Figure 9.33b.

The question arises whether similar oscillations of the internal energy also occur in three-dimensional systems. A look at Figures 9.31 and 9.32 makes clear, that the same effect can be expected qualitatively in the three-dimensional case. However, a quantitative analysis is much more complex. Of course, there can be no fully occupied levels since the magnetic field does not cause quantization in the in-field direction.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Nevertheless, the passage of the Landau cylinders through the Fermi energy will lead to an oscillation of the internal energy.

One example of the effects resulting from the oscillation of the internal energy with the magnetic field is the variation of specific heat. The result of a measurement on beryllium in magnetic fields around 2 T at a temperature of 1.5 K is shown in Figure 9.34. Relatively small changes in the field already cause a number of distinct maxima and minima.

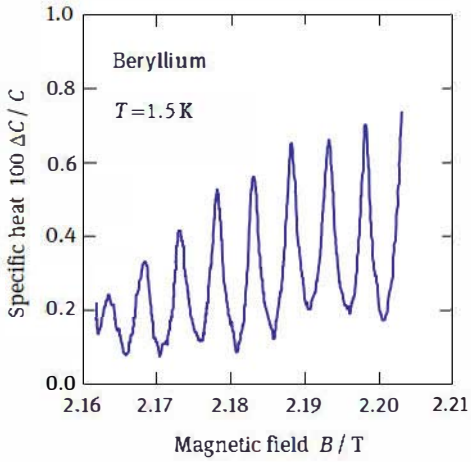


Fig. 9.34: The variation of the specific heat of beryllium at 1.5 K as a function of the applied magnetic field. (After P.F. Sullivan, G. Seidel, Phys. Rev. **173**, 679 (1968).)

9.3.4 De Haas-van Alphen Effect

We now consider in detail the behavior of the magnetization M which at absolute zero is given by $M = -(\partial U/\partial B)$, the magnetic-field dependence of the internal energy. We saw in Figure 9.33b that the internal energy for two-dimensional systems varies with a constant period in $1/B$, and this is also the case for three-dimensional samples. Thus, we would expect a similar behavior for the magnetization. It is remarkable that the derivative of the internal energy can take positive and negative values. If $M > 0$, the contribution of the orbital motion of the electrons is paramagnetic, whereas for $M < 0$ the contribution is diamagnetic. The changes in magnetization with the magnetic field, the **De Haas-van Alphen oscillations**[21,22] are shown schematically in Figure 9.35.

In materials with impurities and defects and also with increasing temperature, the Landau levels become broadened and thus the effects associated with them become

21 Wander Johannes de Haas, *1878 Lisse, †1960 Bilthoven
22 Pieter Marinus van Alphen, *1906 Leiden, †1967 Eindhoven

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

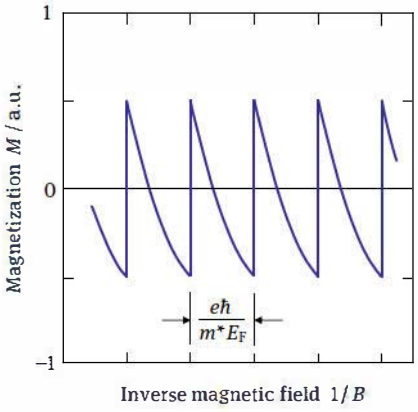


Fig. 9.35: De Haas-van Alphen oscillations. The magnetization M of the conduction electrons is plotted as a function of the inverse magnetic field $1/B$ in units of $e\hbar/m^*E_F$.

blurred. With inceasing temperature the peaks of the de Haas-van Alphen oscillations therefore become rounded, disappearing completely at room temperature. However, the overall effect does not completely disappear. What remains is the **Landau diamagnetism**, which is based on the orbital motion of the free electrons described above. A detailed calculation shows that the resulting susceptibility compensates just one third of the *Pauli paramagnetism*, which we discuss in Section 12.2.

We will now show that the period of the de Haas-van Alphen oscillations allows us to deduce the cross-sectional area of the Fermi surface directly. Differentiating (9.75) leads to

$$\delta\ell = -\frac{\hbar S_\ell}{2\pi e}\frac{\delta B}{B^2} \,. \tag{9.79}$$

Since for our further discussion only the electrons at the Fermi energy are of interest, we consider the cross-sectional area S_F in $\mathbf{k}$-space. With $|\delta\ell| = 1$ we obtain for the period of the de Haas-van Alphen oscillation

$$\delta B = \frac{2\pi e}{\hbar S_F}B^2 \qquad \text{and} \qquad \delta\left(\frac{1}{B}\right) = \frac{2\pi e}{\hbar S_F} \,. \tag{9.80}$$

The period of oscillation measured in the experiment is thus proportional to $1/S_F$. As with cyclotron resonance, essentially only those electrons on extremal orbits contribute to the measurement signal. Figure 9.36a shows a measurement on a gold sample with the field oriented in the [111]-direction. It is clear that the oscillations seen in the measurement have two different periods which can be attributed to the neck orbit H_{111} and the belly orbit B_{111}. From the two periods of the de Haas-van Alphen oscillations, we find that gold has a 1:29 ratio for the cross-sectional areas, with numerical values $S_F = 1.5 \times 10^{15}\,\mathrm{cm}^{-2}$ for the neck orbit and $S_F = 4.3 \times 10^{16}\,\mathrm{cm}^{-2}$ for the belly orbit. If the magnetic field is oriented in the [100] direction, there is no corresponding neck orbit to contribute. Beside the belly orbit B_{100} shown in Figure 9.36b the four-cornered rosette hole orbit appears, whose position can be found with the help of the periodic

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

zone scheme. Making de Haas-van Alphen measurements with different orientation of the sample, allows the shape of the Fermi surface to be reconstructed very well.

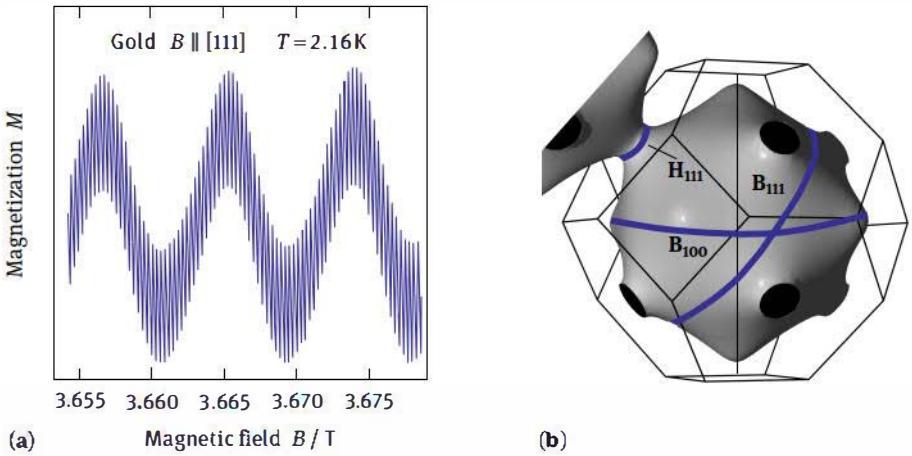


Fig. 9.36: The de Haas-van Alphen effect in gold. **a)** With the magnetic field in the [111] direction, the low-frequency oscillation is caused by the neck orbit H_{111}, the high-frequency oscillation by the belly orbit B_{111}. (With friendly permission of B. Lengeler, RWTH Aachen). **b)** The Fermi surface of gold. The neck orbit H_{111} as well as both belly orbits B_{111} and B_{100} are indicated.

As further examples of Fermi surfaces, Figure 9.37 shows those of lithium and barium. While the surface of lithium differs only slightly from the spherical shape, the Fermi surface of barium has a very complex structure.

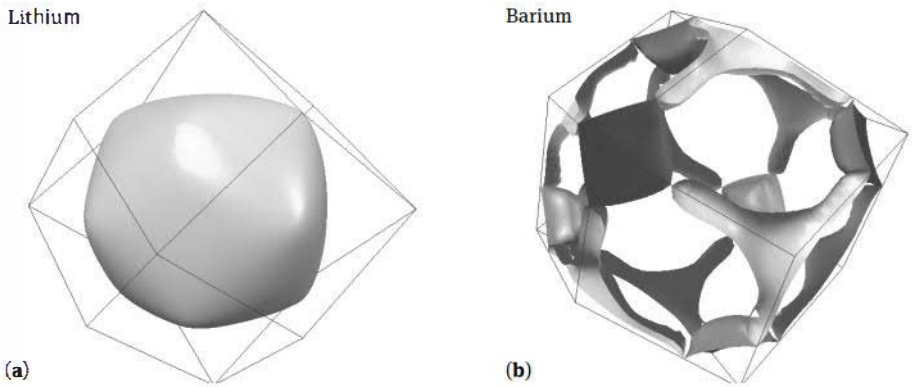


Fig. 9.37: Fermi surfaces and Brillouin zones of **a)** lithium and **b)** barium. (After T.-S. Choyet al., *A database of Fermi surface in virtual reality modeling language (vrml)*. Bull. Am. Phys. Soc., **45**, L36 42, 2000.)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As we have seen, the de Haas-van Alphen effect is based on the dependence of the magnetization on the applied magnetic field. Similar effects are also observed in electrical conductivity. These oscillations are called *Shubnikov-de Haas oscillations.*[23]

9.3.5 Hall Effect

Here we briefly discuss the **classical Hall**[24] **effect**, which in principle allows the charge carrier concentration to be determined. In the Hall effect we measure the electrical transport properties in electrical and magnetic fields simultaneously. A typical arrangement for this is shown in Figure 9.38. Below we derive the equations for the effect, but we should note that the interpretation of the experimental results is not always as simple as one might expect from the simple derivation. However, the resulting equations also form the basis for the discussion of the **quantum Hall effect** observed in two-dimensional electron systems and described at the end of this section.

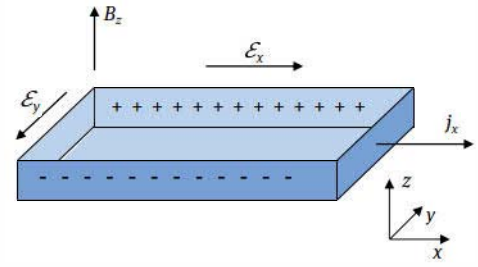


Fig. 9.38: Sample geometry for the discussion of the Hall effect. The current flows in x-direction, the magnetic field is applied in z-direction.

The Classical Hall Effect. We begin our discussion with the equation of motion (9.19), into which we have included the Lorentz force[25], since in Hall measurements we are applying a magnetic field in addition to the electric field:

$$m^* \dot{\mathbf{v}} = -e(\mathcal{E} + \mathbf{v}_\mathrm{d} \times \mathbf{B}) - m^* \frac{\mathbf{v}_\mathrm{d}}{\tau} \,. \tag{9.81}$$

Given that the electrons change their direction of motion with each scattering process, the influence of the Lorentz force on their motion averages out except for the contribution caused by the drift velocity. It is therefore sufficient to deal with the drift velocity alone. We assume that the magnetic field is applied in the z-direction and consider the

23 Lev Vasilyevich Shubnikov, *1901 Saint Petersburg, †1937 (Ukraine)
24 Edwin Herbert Hall, *1855 Gorham, †1938 Cambridge, USA
25 Hendrik Antoon Lorentz, *1853 Arnhem, †1928 Haarlem, Nobel Prize 1902

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

stationary case for which $\dot{\mathbf{v}} = 0$. For the drift velocity we get

$$\begin{aligned} v_{\mathrm{d},x} &= -\frac{e\tau}{m^*}\left(\mathcal{E}_x + v_{\mathrm{d},y}B\right), \\ v_{\mathrm{d},y} &= -\frac{e\tau}{m^*}\left(\mathcal{E}_y - v_{\mathrm{d},x}B\right), \\ v_{\mathrm{d},z} &= -\frac{e\tau}{m^*}\mathcal{E}_z . \end{aligned} \tag{9.82}$$

Now we resolve for $\mathbf{v}_\mathrm{d}$, considering (9.21) and obtain for the current density:

$$\begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = -\frac{\sigma_0}{1+\omega_\mathrm{c}^2\tau^2}\begin{pmatrix} 1 & -\omega_\mathrm{c}\tau & 0 \\ \omega_\mathrm{c}\tau & 1 & 0 \\ 0 & 0 & 1+\omega_\mathrm{c}^2\tau^2 \end{pmatrix}\begin{pmatrix} \mathcal{E}_x \\ \mathcal{E}_y \\ \mathcal{E}_z \end{pmatrix}, \tag{9.83}$$

where $\sigma_0 = ne^2\tau/m^*$ is the conductivity with no magnetic field and ω_c is the cyclotron frequency.

In the following discussion, we consider a flat bar with a rectangular cross-section, a section of which is sketched in Figure 9.38. Since in this geometry, current cannot flow in the z-direction, and there cannot be an electric field in this direction, (9.83) can be simplified to:

$$\begin{pmatrix} j_x \\ j_y \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{xx} \end{pmatrix}\begin{pmatrix} \mathcal{E}_x \\ \mathcal{E}_y \end{pmatrix}, \tag{9.84}$$

where for the sake of clarity we have introduced the following conductivities:

$$\sigma_{xx} = \frac{ne}{B}\frac{\omega_\mathrm{c}\tau}{1+\omega_\mathrm{c}^2\tau^2}, \tag{9.85}$$

$$\sigma_{xy} = -\frac{ne}{B}\frac{\omega_\mathrm{c}^2\tau^2}{1+\omega_\mathrm{c}^2\tau^2}. \tag{9.86}$$

We solve the system of equations with respect to the electric fields, and obtain:

$$\begin{pmatrix} \mathcal{E}_x \\ \mathcal{E}_y \end{pmatrix} = \begin{pmatrix} \varrho_{xx} & \varrho_{xy} \\ -\varrho_{xy} & \varrho_{xx} \end{pmatrix}\begin{pmatrix} j_x \\ j_y \end{pmatrix}. \tag{9.87}$$

The resistivity components that occur are represented by:

$$\varrho_{xx} = \frac{B}{ne}\frac{1}{\omega_\mathrm{c}\tau} = \frac{m^*}{ne^2\tau}, \tag{9.88}$$

$$\varrho_{xy} = \frac{B}{ne}. \tag{9.89}$$

The component ϱ_{xx} corresponds to the usual expression for the resistivity, while ϱ_{xy} is linked to the Hall effect. It should be noted that the quantities ϱ_{yy} and ϱ_{yx} also appear in the general case of electrical current transport in anisotropic materials (9.87). However, for isotropic materials and with the selected direction of the magnetic field $\varrho_{xx} = \varrho_{yy}$ and $\varrho_{yx} = -\varrho_{xy}$.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

With the geometry assumed above, current flows only in the x-direction. With $j_y = 0$ and (9.87) we obtain the following relations:

$$\mathcal{E}_x = \varrho_{xx} j_x \,, \tag{9.90}$$

$$\mathcal{E}_y = -\varrho_{xy} j_x \tag{9.91}$$

and as a consequence

$$\mathcal{E}_y = -\frac{eB\tau}{m^*}\,\mathcal{E}_x \,. \tag{9.92}$$

In a magnetic field, an electric field, called **Hall field**, is built up in the y-direction. The reason being that when the magnetic field in the z-direction is switched on, the electrons are deflected by the Lorentz force and accumulate on the surface, as shown in Figure 9.38. In the stationary state, the resulting electric force of the Hall field just compensates for the Lorentz force.

Since no current flows in the y-direction in the stationary state, we can replace $\mathcal{E}_x$ by equation (9.21) and find for the Hall field

$$\mathcal{E}_y = -\frac{1}{ne} B j_x = R_\mathrm{H} B j_x \,. \tag{9.93}$$

Thus the relationship for the **Hall constant** R_H is

$$R_\mathrm{H} = \frac{\mathcal{E}_y}{j_x B} = -\frac{1}{ne} \,. \tag{9.94}$$

Since $\mathcal{E}_y$, j_x and B are measurable quantities, R_H and thus the electron density n can be determined directly.

However, the interpretation of the experimentally measured values is not as simple as it first appears. It turns out that, contrary to expectation, R_H mostly depends on the magnetic field, the temperature, and the sample preparation. For example, the Hall voltage as a function of the magnetic field can even change its sign in some cases. The reason for this is that usually more than one band contributes to the charge transport.

With sufficiently high magnetic fields, i.e. at $\omega_c \tau \gg 1$, the Hall constant reaches a saturation value. Table 9.2 shows values measured on carefully prepared samples at low temperatures and high magnetic fields. The Hall constant is given in units of electron charge per atom (e/atom). Let us take a brief look at the values for various metals. First of all, it should be noted that in the case of alkali metals, the observed

Tab. 9.2: Hall constant of selected metals. The constants are given in units of electron charge per atom. (The data were taken from N.W. Ashcroft, N.D. Mermin, *Festkörperphysik*, *Oldenbourg*, 2007).

Metal	Li	Na	K	Cu	Au	Be	Mg	In	Al
Valancy	1	1	1	1	1	2	2	3	3
R_H (e^-/Atom)	0.8	1.2	1.1	1.5	1.5	−0.4	−0.8	−0.9	−0.9

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

values correspond quite well to what we would expect. Significant deviations can be found for copper and gold, and for Be, Mg, In and Al where the sign is even negative. The reason for this surprising result is that the current is carried by holes. In the case of aluminum with its three valence electrons, it looks as if approximately one positive charge per atom transports the current.

If the current is transported by various classes of charge carrier, e.g. by electrons in different bands or by electrons and holes, then the derivation carried out here must be modified. This aspect is treaded in the so-called *two-band model*, which we will not go into further here. If both electrons *and* holes contribute to the current, this must be taken into account in the condition $j_y = 0$, which then becomes $j_{y,\mathrm{n}} + j_{y,\mathrm{p}} = 0$, i.e. the *total* current carried separately by the electrons and holes must be zero. If we neglect the contribution of the Hall field $\mathcal{E}_y$ to the current density j_x, since it increases proportionally to B^2, then after a short calculation we obtain the relation:

$$R_\mathrm{H} = \frac{p\mu_\mathrm{p}^2 - n\mu_\mathrm{n}^2}{e(p\mu_\mathrm{p} + n\mu_\mathrm{n})^2}\,, \tag{9.95}$$

where n and p are the densities, μ_n and μ_p the mobilities of the electrons and holes, respectively. Depending on the density and mobility of the charge carriers, the Hall constant can therefore be $R_\mathrm{H} < 0$ or $R_\mathrm{H} > 0$. This relationship will become important in Section 10.2 during the discussion of semiconductors.

9.3.6 Quantum Hall Effect

As we will see in Sections 10.4 and 10.5, electrons can be localized relatively straightforwardly in thin layers to form a two-dimensional electron gas with the help of *semiconductor heterostructures* such as in *field effect transistors*. When Hall effect measurements were first made in such systems surprising new effects were observed.

The electrons can move freely in the layer plane, but cannot move in the z-direction. As we have already seen, applying a magnetic field in the z-direction further restricts the motion of the electrons. Only discrete energy eigenvalues then occur, yielding a density of states as depicted in Fig. 9.32a.

We now take a closer look at this Hall effect in two-dimensional samples. For this purpose, in equation (9.89) we replace the three-dimensional electron density n by the number density $n_{2\mathrm{D}}$ of electrons per unit area. We assume that the magnetic field strength has been chosen so that only the fully occupied Landau levels are present. In this case the N electrons of the sample are distributed on p levels, where according to (9.77) the relation $N = pg_\mathrm{e}/2$ holds, if we take into account that in high magnetic fields the spin degeneracy is lifted. If the sample has an area A, then the two-dimensional electron density is given by $n_{2\mathrm{D}} = N/A$, and we find:

$$\varrho_{xy} = \frac{B}{n_{2\mathrm{D}}e} = \frac{1}{p}\frac{h}{e^2} = \frac{25812\ \Omega}{p}\,, \tag{9.96}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

where the filling factor p can take the values $p = 1,2,3$, etc.. According to (9.89), the transverse resistance ϱ_{xy} should therefore increase linearly with the magnetic field and take on the calculated values for the magnetic fields specified by (9.96).

Since no electron scattering can take place in the fully occupied Landau levels, the mean scattering time τ must be infinite. Thus, according to equations (9.85) and (9.88), both σ_{xx} and ϱ_{xx} are zero! The current is thus not driven through the sample by the longitudinal electric field $\mathcal{E}_x$, which also disappears because $\sigma_{xx} = 0$, but by the Hall field $\mathcal{E}_y$, generated by the applied magnetic field.

However, the experiments carried out by *K. von Klitzing*[26] showed unusual behavior. Figure 9.39 shows as an example the result of such a measurement on a GaAs AlGaAs/GaAs-heterostructure. Although the experiment confirms that ϱ_{xx} and σ_{xx} disappear precisely at the predicted values of the magnetic field, R_x is also observed to disappear over wide ranges of the magnetic field. The long plateaus which the Hall resistance R_y exhibits as a function of the field is also astonishing. Obviously, not only the magnetic fields where exactly all electrons are in fully occupied Landau levels are special.This amazing behavior is known as the **quantum Hall effect.**

In the new *International System of Units* (2019) the quantity ϱ_{xy}, made up of fundamental constants, is now fixed to the value:

$$R_K = \frac{h}{e^2} = \frac{6.62607015 \times 10^{-34}\,\text{Js}}{(1.602176634 \times 10^{-19}\text{C})^2} = 25812.80745\ldots\ \Omega\,. \tag{9.97}$$

This remarkable quantity is usually called the *Von-Klitzing constant*.

How can it be understood that the transverse resistance has steps and the longitudinal resistance does not show the expected *Shubnikov-de Haas oscillations* when the Landau cylinders cross the Fermi level, but disappears over wide ranges of the

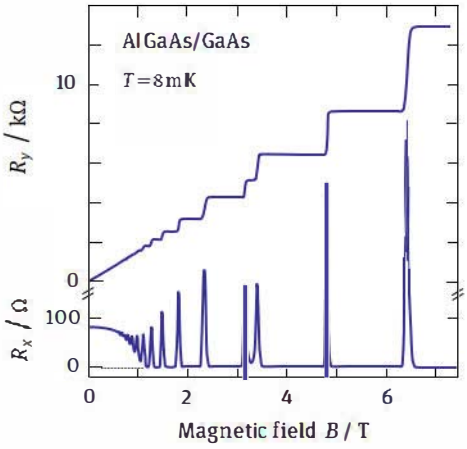


Fig. 9.39: The Hall resistance R_y and longitudinal resistance R_x, measured at very low temperatures on a two-dimensional AlGaAs/GaAs-heterostructure. (After K. von Klitzing, Physica **126B**, 242 (1984).)

26 Klaus von Klitzing, *1943 Schroda, Nobel Prize 1985

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

magnetic field? In a simple phenomenological explanation we assume that in real samples a fraction of the electrons are not really free, but are localized by the effects of impurities. As shown in Figure 9.40, the density of states then no longer consists of a series of δ-functions, but of broadened bell curves. The localized states, which do not contribute to the charge transport, are located on sloping sides of the distribution, and only the states in the centers are delocalized.

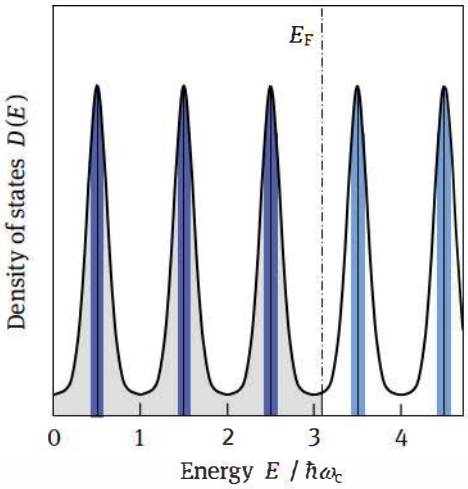


Fig. 9.40: Schematic representation of the density of states of a two-dimensional electron gas in a magnetic field for a sample with disorder. The delocalized states are shown in blue, the localized states in grey. The Fermi energy is shown positioned between the 3rd and 4th Landau levels.

If the density of states consisted only of δ-functions as shown in Figure 9.33, the Fermi level would jump to the next lower level after the emptying of the uppermost occupied Landau level if the applied magnetic field was further increased. However, here the Fermi level is pinned by the localized states between the levels. As the field increases, the Fermi level moves through the localized states, the mean scattering time τ of the electrons in the delocalized states of the fully occupied Landau levels is indeed infinite. Since neither the electrons in the localized states nor the electrons in the fully occupied Landau levels contribute to charge transport, both σ_{xx} and ϱ_{xx} vanish as depicted in Figure 9.39. Since the localized levels have no influence on the Hall resistance, the value of ϱ_{xy} also remains constant in this field region. Thus, this simple phenomenological model provides a clear explanation of the surprising observations.

That said, we will now take the discussion of the quantum Hall effect somewhat further and discuss the influence both of defects and of the edges of the sample on the motion of the electrons. Deep inside the sample, the free electrons move through undisturbed cyclotron paths and thus do not contribute to the charge transport. The corresponding energy states are the previously discussed Landau levels. An important point in understanding the processes in the sample is that at the edge of the sample the potential rises steeply to the vacuum potential. This causes a bending of the Landau levels as shown in Figure 9.41a. The Fermi energy therefore intersects the Landau levels

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

at the edge of the sample, creating one-dimensional conduction channels known as the **edge channels**. At the edge of the sample, the electrons are elastically reflected at the surface and the strong magnetic field suppresses the backscattering so that the electrons "hop" along the edge of the sample as shown in Figure 9.41b. These trajectories are therefore often called *skipping orbits*. The electrons in the edge channels are forced forward by the strong magnetic field, even when scattering by defects occurs. At the opposite sample edge the electrons run in the opposite direction because of the Lorentz force. Each channel contributes to the current and in each the charge transport is quasi-ballistic. In Section 9.2 we have already discussed the current transport in one-dimensional conductors and found the value $G_Q = 2e^2/h$ for the conductance (per channel), taking into account the degeneracy of the spins.

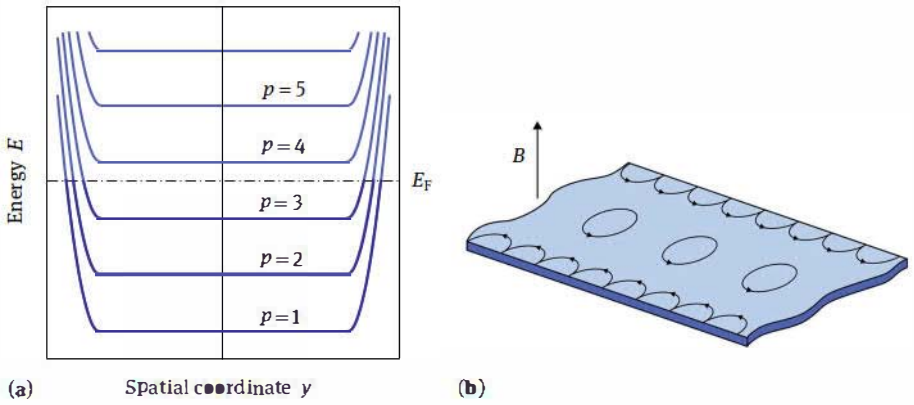


Fig. 9.41: Schematic representation of the edge channels in the two-dimensional electron gas. **(a)** The boundary potential causes a bending of the Landau levels at the boundary of the sample. **(b)** Schematic representation of the trajectory curves. Closed cyclotron orbits occur inside the sample. The electrons are reflected at the edges of the sample and contribute to current transport.

The spatial separation of the forward and return channels has the consequence that inelastic scattering is completely suppressed. For this to happen, an electron would have to be scattered from an edge channel on one side of the sample to an edge channel on the other side. This can be practically ruled out if the Fermi energy is located between the Landau levels, since there are no states inside the sample over which scattering could occur. This means that the electrons entering on one end of the sample have to move along the edge until the next contact.

To explain the quantum Hall effect in the edge channel model, we use the *Landauer-Büttiker formalism*[27,28]. We first assume that the Fermi energy lies between two Landau

27 Rolf Wilhelm Landauer, *1927 Stuttgart, †1999 Briarcliff anor
28 Markus Büttiker, *1950 Wolfwil †2013 Collonge-Bellerive

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

levels. Referring to Figure 9.42, we assume that contact 1 is at potential $\mu_1 = -\mu$, contact 4 is at potential $\mu_4 = 0$. During their passage from contact 1 to contact 4, the electrons moving along the upper edge first reach and enter contact 2. Since no current is drawn there, its chemical potential increases until the same amount of current flows to contact 3. The same argument applies to contact 3, so that the relationship $\mu_1 = \mu_2 = \mu_3 = -\mu$ must hold. The same argument can also be applied to the lower current if we start from contact 4. From this follows $\mu_4 = \mu_5 = \mu_6 = 0$. Thus the lower edge channels do not transport any current.

Per edge channel, the current $I = (\mu_3 - \mu_4)G_Q/e = -\mu G_Q/e$ flows through the sample. If p channels are present, the relationship for the total current is

$$I = -p\frac{G_Q}{e}\mu = -p\frac{2e}{h}\mu\,. \tag{9.98}$$

For the Hall resistance the following applies

$$R_{35} = \frac{(\mu_3 - \mu_5)}{eI} = \frac{1}{p}\frac{h}{2e^2} \tag{9.99}$$

and for the longitudinal resistance $(\mu_5 - \mu_6)/eI = 0$. The current therefore flows without resistance.

We should also point out a few interesting aspects. In our derivation, we have used the value $G_Q = 2e^2/h$ for the conductance quantum, which takes into account the spin degeneracy of the Landau levels which explains the factor of two in the expressions above. This simplification is no longer justified for high magnetic fields,

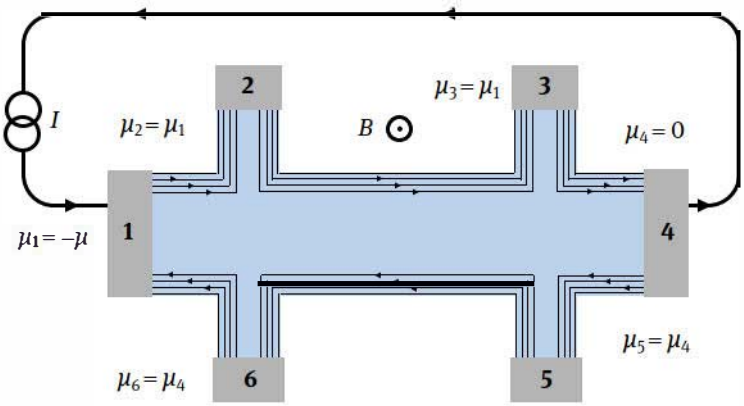


Fig. 9.42: Quantum Hall resistance in the edge channel description. The sample is light blue, the contacts are shown in grey. The current I flows through the sample with the six contacts. The arrows indicate the direction of motion of the electrons and *not* the conventional current direction. The contacts 1 and 4 are at the potentials $\mu_1 = -\mu$ and $\mu_4 = 0$ respectively. The sketch is based on the assumption that the Fermi energy lies between the third and fourth Landau level, i.e. that there are three edge channels ($p = 3$).

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

because then the additional splitting of the Landau levels (and the halving of the conductance quantum) must also be taken into account. In addition, to simplify the mathematical expressions, we have assumed perfect transmission between the contacts and additionally that the electrons are not reflected at the contacts. Of course these aspects would have to be considered in a more detailed treatment.

Up to now, we have assumed that the Fermi energy is maintained when the magnetic field varies *between* two Landau levels. The small number of edge states cannot be responsible for this. In real samples, as mentioned above, impurities and crystal defects cause a spatial potential variation. Therefore, the energy dispersion shows, as schematically depicted in Figure 9.43, local fluctuations. This broadens the δ-function shaped Landau levels, as already shown in Figure 9.40.

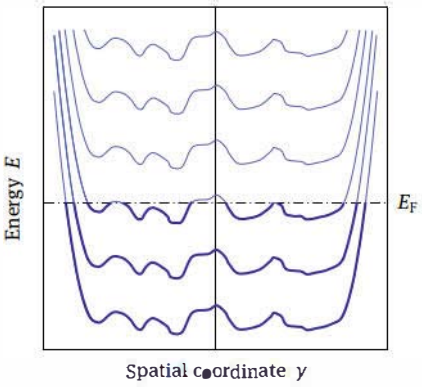


Fig. 9.43: The influence of crystal defects on the energy dependence of the Landau levels. The disorder causes a spatial fluctuation in the energy dispersion curves. Occupied states are indicated by a darker blue line.

These fluctuations have the consequence that linear channels are also created within the sample, but forming closed loops (see Figure 9.44a). Since they are spatially localized, they do not contribute to the charge transport. As long as the electrons in the edge channels cannot be scattered into these states, charge transport continues to flow without resistance. If the magnetic field is increased, the Landau levels are raised in energy and the uppermost occupied Landau level approaches the Fermi energy. Thus the isolated loops expand and new ones are created. When the field is further increased (Figure 9.44b), the two inner channels move away from the edge and inelastic scattering of electrons becomes possible (Figure 9.44c). Finally, if the Fermi energy is at the Landau level, the innermost channel breaks up into localized states (Figure 9.44d). These now provide extended states within the sample via which the electrons can be scattered from one edge to the other, and the longitudinal resistance R_x is at a maximum. The Hall resistance ϱ_{xy} changes. In general, if the Fermi energy and the uppermost Landau level coincide, the longitudinal resistance is at a maximum, but a minimum occurs if the Fermi energy lies between the levels. However, it also shows that, surprisingly, the quantum Hall effect is less pronounced in nearly perfect samples with only a few

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

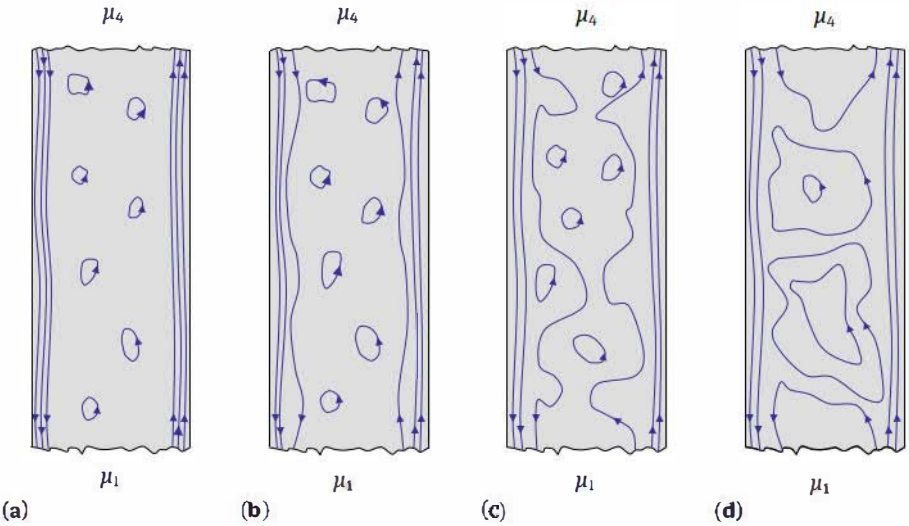


Fig. 9.44: Electron states at the Fermi energy. **a)** In addition to the three edge channels, localized loops of states can be seen in the middle. **b)** As the magnetic field increases, the localized areas of states increase, and the innermost edge channel also becomes distorted and moves away from the sample edge. **c)** Inelastic scattering between the electrons of the two innermost edge channels can take place. **d)** Fermi energy and Landau level coincide, the innermost edge channel is broken up into localized states. The longitudinal resistance is at a maximum.

imperfections. Nevertheless, if the concentration of the impurities becomes too high, the scattering times become too short and the whole quantum Hall effect disappears.

It is remarkable that the Hall resistance (9.97) can be expressed in terms of fundamental constants alone. It is closely related to *Sommerfeld's fine structure constant* α, which is:

$$\alpha = \frac{1}{2\varepsilon_0 c}\frac{e^2}{h} \approx \frac{1}{137.036} \tag{9.100}$$

and can therefore be used to determine the Hall resistance constant.

With the help of a highly accurate resistance bridge, traditional resistance standards can be compared with the quantized Hall resistance and thus be calibrated absolutely. These resistance standards serve as a further step as transfer standards for the calibration of customer standards.

At very low temperatures and even higher magnetic fields, a new phenomenon, the **fractional quantum Hall effect** is observed, in which fractional quantum numbers are found. Despite its striking similarity to the integer quantum Hall effect, its physical origin is different. This is immediately obvious from the fact that values of $p < 1$ also occur. In this effect, only the lowest Landau level is populated, but it is not fully occupied at these high fields. To describe this phenomenon, the one-particle approximation we have used so far is no longer sufficient. We have to consider strongly interacting

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

electrons, which move in a correlated manner as a *quantum liquid*. There are then effective charges $e^* = p\,e$, which can be smaller than the elementary charge e. However, a treatment of these effects goes beyond the scope of this book.

9.3.7 Quantum Hall Effect in Graphene

We return here once again to graphene, which, as already pointed out, exhibits a number of remarkable properties. These include the quantum Hall effect, which differs from that in "classical" solids. In metals without a magnetic field, every electronic state in momentum space is normally doubly occupied owing to spin degeneracy. If the conduction band has several "valleys", the degree of degeneracy can be even greater. This also applies to graphene where the states are fourfold degenerate, so that in the equation for the conductivity σ_{xy} the factor e^2/h must be replaced by $4e^2/h$. However, this correction alone is not enough. As we have seen in Section 8.5, the ground state of the electrons as well as the holes lies exactly at $E =$ ●. Consequently, the first quantum Hall plateau already appears at a filling half the level known from the classical quantum Hall effect. For the Hall conductivity we thus obtain the expression:

$$\sigma_{xy} = \pm\frac{4e^2}{h}\left(p + \frac{1}{2}\right) . \tag{9.101}$$

The plateaus occur at a conductivity that is increased or decreased by $\sigma_{xy} = \pm 2e^2/h$ compared to the classical value. Formally, the quantum Hall effect in graphene can be described with the help of the Dirac-Weyl equation, which we referred to in Section 8.5.

The result of a measurement of the quantum Hall effect in graphene is shown in Figure 9.45 where both the Hall resistance R_y and the longitudinal resistance R_x are shown. The experiment was performed at 4 K in a magnetic field of 14 T, with the

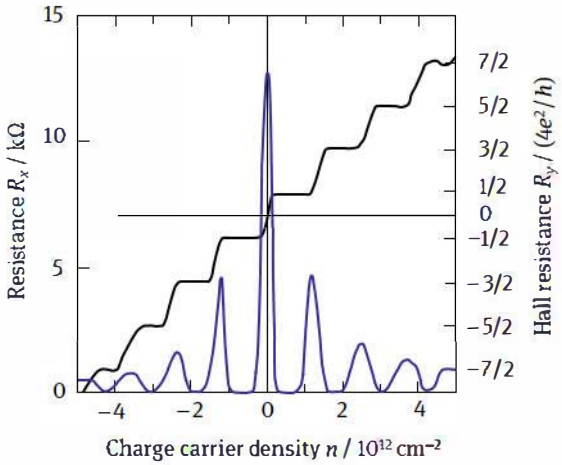


Fig. 9.45: Quantum Hall effect of a single graphene layer as a function of the charge carrier concentration. The Hall resistance R_y and longitudinal resistance R_x were measured at 4 K in a magnetic field of 14 T. (After K.S. Novoselov et al., Nature **438**, 197 (2005).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

graphene layer as the active element of a field effect transistor (see Section 10.5). With this switching element, if a positive gate voltage is applied, electrons are injected into the graphene layer, whereas holes are injected if the voltage is negative. It is remarkable that this experimental arrangement not only allows the carrier density to be varied over a wide range, but also allows for a change of sign such that the behavior of both electrons *and* holes to be investigated.

9.4 Exercises and Problems

1. Fermi Sphere. A current of 5 A flows through a gold wire of 5 mm diameter. Calculate the displacement of the Fermi sphere and compare it with the radius of the sphere. What is the electron drift velocity?

2. Electrical Parameters of Potassium. The electrical properties of potassium at room temperature are to be investigated in more detail by means of a few measured variables. Calculate the following quantities: the Fermi wave vector, the Fermi energy and relaxation time, the mobility, the mean free path, the velocities of electrons in a field of 50 mV/m and the thermal conductivity at room temperature. In addition to the data that appear in the current text, you may assume the following parameters for potassium, density, 0.862 g/cm^3 and resistivity, 61 nΩm.

3. Strongly Bound Electrons. Consider a crystal with a simple cubic lattice and a diatomic basis (lattice constant $a = 4$ Å). The conduction and valence bands can be described by

$$E_{\mathbf{k},i} = E_i^0 - 2\beta_i[\cos k_x a + \cos k_y a + \cos k_z a]$$

with the parameters $E_V^0 = 0$ eV, $E_L^0 = 12$ eV, $\beta_V = -0.8$ eV and $\beta_L = 1.0$ eV.

By absorbing a photon, an electron with $k = 10^8$ m^{-1} is lifted from the filled valence band into the empty conduction band. The wave vector change of the electron can be neglected. (Why?)

(a) Sketch the dispersion relations for the electrons in the conduction and valence bands in the x-direction and indicate which states are occupied at $T = 0$.
(b) What are the values of the effective masses m_p^*/m and m_n^*/m of the generated hole and the excited electron after the photon absorption?
(c) What are the velocities v_s and v_n in the vicinity of the Γ-point?
(d) What acceleration $\dot{v}_p$ and $\dot{v}_n$ would be experienced by the charge carriers in an electric field of 1 V/m in the x-direction?

4. Scattering at Point Defects. The electrical resistivity of sodium (body-centered cubic lattice) is limited at room temperature by electron-phonon scattering and therefore increases linearly with temperature. Electrons are also scattered at vacancies,

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

whose number increases exponentially with temperature. Their scattering cross section is determined to a first approximation by their geometrical size. Estimate at which temperature the two scattering mechanisms are equally effective.

In addition to the material parameters in Chapter 8 you may use the following numerical values: resistivity at room temperature $\rho_{300} = 47.5\,\text{n}\Omega\text{m}$, vibration entropy $S_\text{L}/k_\text{B} = 5.8$, formation energy $E_\text{L} = 0.42\,\text{eV}$ and lattice constant $a = 4.23\,\text{Å}$.
Note: Sodium melts at 371 K.

5. Scattering at a Wire Surface. We are considering a gold nanowire with a diameter of $D = 50\,\text{nm}$. The electrons which flow through this wire are not only scattered by the phonons but also at the wire surface. Estimate at what temperature the two contributions to the resistance would be equal. The resistivity of gold at room temperature is $\varrho_\text{Au} = 20.5\,\text{n}\Omega\text{m}$.

6. Free Electrons in the Magnetic Field. A potassium crystal (body-centered cubic lattice with lattice constant $a = 5.32\,\text{Å}$) is placed in a magnetic field $B = 0.8\,\text{T}$.
(a) Calculate the Hall constant.
(b) How many Landau tubes are occupied at $T = 0\,\text{K}$?
(c) What is the radius of the extremal orbits in real space?
(d) What is the minimum average scattering time τ of the electrons required for de Haas-van Alphen oscillations to be easily measurable?

6. Electrical and Thermal Conductivity. Figure 9.46 shows the resistivity of three materials between 1.5 K and room temperature. The three samples tested were two high-purity gold wires, (one untreated and the other annealed) and a further wire of $Au_{50}Pd_{50}$ alloy.

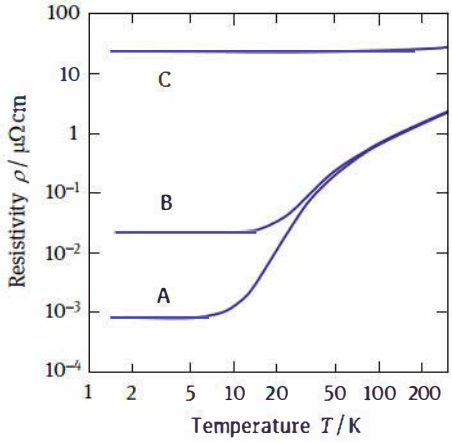


Fig. 9.46: The resistivities of three samples over the temperature range from 1.5 K to room temperature.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

(a) Assign the materials mentioned to the measurement curves A to C.
(b) Calculate the residual resistance ratios and the mean free paths of the electrons in the gold samples at temperatures below 5 K and at room temperature.
(c) In the tempered gold wire which scattering mechanism dominates at room temperature and which at temperatures below 5 K? Which one dominates in the $Au_{50}Pd_{50}$ alloy?
(d) Estimate the thermal conductivities of the three samples at a temperature of 1 K.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

10 Semiconductors

The term **semiconductor** is given to those materials with electrical conductivities lying between those of metals and insulators. Good metals have resisitivities between 10^{-7} Ωm and 10^{-8} Ωm, whereas the resistivity of good insulators is above 10^{12} Ωm. For semiconductors, the resistivity can vary over a wide range, from approximately 10^{-4} Ωm to 10^{7} Ωm.

Semiconducting elements with a simple crystal structure are mostly found in the fourth main group of the periodic table. Among these, silicon and germanium are the most interesting from a technical perspective. Carbon in the form of diamond belongs rather to the class of insulators, while lead to the metals. In contrast, fullerenes show semiconducting properties. Tin has both a semiconducting and a metallic modification. Other semiconducting elements are red phosphorus, boron, selenium and tellurium with covalent bonding and a relatively complex crystal structure. Furthermore, a number of elements also become semiconducting under high pressure. In addition, there are many compound materials which are semiconductors. Some III-V compound semiconductors such as GaAs, GaP, InP, InSb or InAs have a zincblende structure. The bond between the atoms is partly covalent and partly ionic. There are also several II-VI compounds which show semiconducting properties such as ZnO, ZnS, CdS or CdSe, and IV-IV semiconductors such as SiC or SiGe.

We should also mention here the organic semiconductors, whose conductivity is based on their conjugated double bonds between the carbon atoms. Semiconducting properties of organic materials were investigated for the first time in polyacetylene (polyethyne). Finally, we should also mention the magnetic semiconductors, which include EuS.

At absolute zero, semiconductors behave as insulators. Since the lower bands are completely filled with electrons and the higher bands are completely empty, no charge transport can occur. At finite temperatures, valence electrons can be thermally excited into the empty conduction band making the material conductive. For technical applications, the intentional generation of defects by *doping* is of crucial importance, because this can drastically change the electrical properties of the starting material to adapt them for technical applications.

In this chapter, we will first consider pure, so-called *intrinsic semiconductors*, and in doing so we will discuss the distinction between *direct* and *indirect* semiconductors. Then we look at the effects of doping semiconductors, which leads to *impurity conduction*, and then we deal with amorphous semiconductors. Following that, we will look at the p-n junction, the properties of *heterostructures* and finally discuss the mode of operation of some semiconductor-based *electronic devices*.

https://doi.org/10.1515/9783110666502-010

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

10.1 Intrinsic Crystalline Semiconductors

10.1.1 Band Structure, Band Gap and Optical Absorption

In semiconductors, the highest fully occupied band is called the **valence band** and the first empty band above it is called the **conduction band**. We write E_V for the energy of the upper edge of the valence band and E_C for the lower edge of the conduction band. The empty **band gap** $E_g = (E_C - E_V)$ between these two energies is crucial for enabling many electronic properties. Its magnitude depends slightly on temperature. Starting from low temperatures, the band gap first increases quadratically and then linearly with increasing temperature. The total change up to room temperature is about 10%, arising from thermal expansion and the effects of the electron-phonon interaction.

Table 10.1 lists the band gaps at room temperature and at absolute zero for a number of semiconductors. In addition, whether a **direct** or **indirect band gap** is present is indicated. If the valence band maximum and the lowest conduction band minimum both lie at the Γ-point, then we speak of a **direct semiconductor** or a direct band gap. If the extrema in **k**-space are not at the same **k**-vector, then the crystal is an **indirect semiconductor**. The difference between the two types of semiconductors will be discussed in more detail below.

Tab. 10.1: The size of the band gap E_g, and whether it is direct or indirect, for several semiconductors. (The majority of the data taken from W. Martienssen, *Springer Handbook of Condensed Matter and Material Data*, W. Martienssens, H. Warlimont, eds., Springer, 2005).

	E_g(300K)/eV	E_g(0K)/eV	Type
Diamond	5.43	5.45	indirect
Si	1.12	1.17	indirect
Ge	0.66	0.74	indirect
AlAs	2.15	2.23	indirect
GaP	2.27	2.35	indirect
GaAs	1.42	1.52	direct
InSb	0.18	0.24	direct
InAs	0.35	0.42	direct
InP	1.34	1.42	direct
ZnO	3.20	3.44	direct
CdS	2.48	2.58	direct
CdTe	1.48	1.61	direct

To introduce the subject we begin with the band structures of the two technically important and scientifically interesting semiconductors *gallium arsenide* and *indium*

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

antimonide, shown in Figure 10.1. These plots are the results of band structure calculations adapted to spectroscopic measurements.

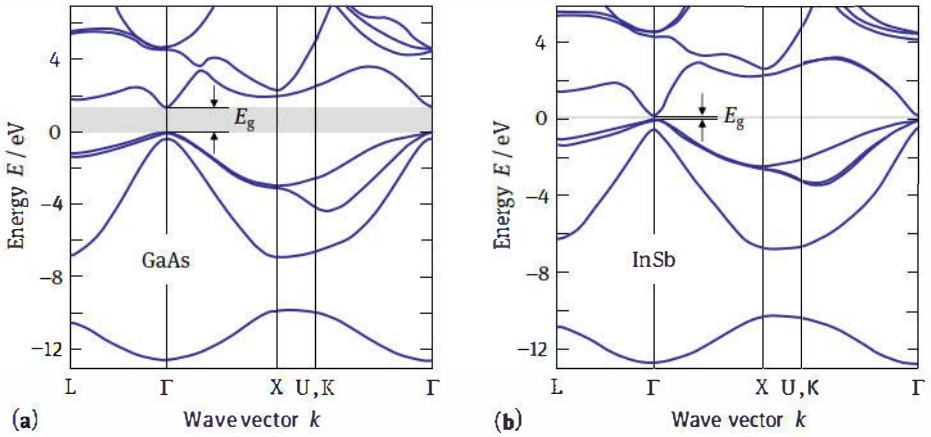


Fig. 10.1: Band structure of **a)** gallium arsenide and **b)** indium antimonide. The energy gaps E_g are marked in grey. Valence band maximum and lowest conduction band minimum are opposite each other at the Γ-point. The materials in both cases are direct semiconductors. (After J.R. Chelikowsky, M.L. Cohen, Phys. Rev. B **14**, 556 (1976).)

From the figure we can see that both gallium arsenide and indium antimonide are direct semiconductors, because the valence band maximum and the lowest conduction band minimum are located at the same point in **k**-space, i.e. at the Γ-point.

The band gap can be determined in a particularly simple way with the aid of optical absorption. As already mentioned in Section 9.1, when a photon is absorbed, an electron is lifted into the conduction band, leaving a hole in the valence band. We will talk about such **interband transitions** again in a more general context in Chapter 13. In Figure 10.2a two such absorption transitions are shown by arrows. The transition follows vertically in the band diagram, since the photon momentum $\hbar k_\gamma$ is negligibly small compared to typical values of the electron momentum. Since the band gap must be overcome, a photon can only be absorbed if its energy exceeds the minimum value $\hbar\omega_\gamma \geqq (E_C - E_V) = E_g$. Many semiconductors are therefore transparent in the near infrared. Above the threshold energy, which is given by the band gap, the optical absorption rises steeply with increasing frequency. An example is shown in Figure 10.2b. There the absorption coefficient of indium antimonide is plotted logarithmically as a function of the photon energy. The steep increase at the absorption edge is characteristic of direct semiconductors.

Now let us look at indirect semiconductors. Figure 10.3 shows the band structure of the two best known semiconductors silicon and germanium. In both cases the valence band maximum is found at the Γ-point, but the conduction band minimum directly

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

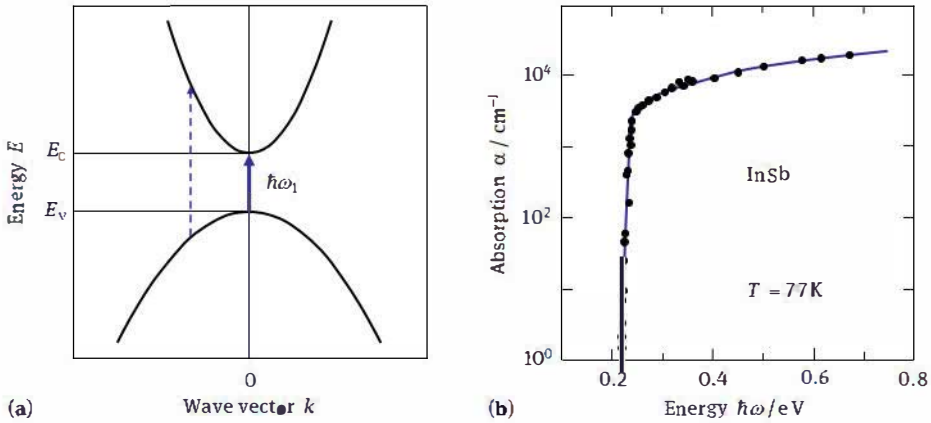


Fig. 10.2: Optical absorption at the direct band gap. **a)** Schematic representation of the absorption process. The thick blue arrow marks the transition having the lowest possible energy. At higher photon energies (dashed blue arrow) electrons at lower energies are excited. **b)** Optical absorption coefficient of indium antimonide on a logarithmic scale as a function of the energy of the irradiating photons. (After Ch. Kittel, *Introduction to Solid State Physics*, Wiley, 2005.)

above it does not have the smallest energy gap to the valence band. As can be seen in the figure, there is a lower conduction band minimum for silicon near the X-point, and for germanium at the L-point.

The optical-absorption mechanism for indirect semiconductors is somewhat more complicated than for direct semiconductors. As indicated in Figure 10.4a, absorption

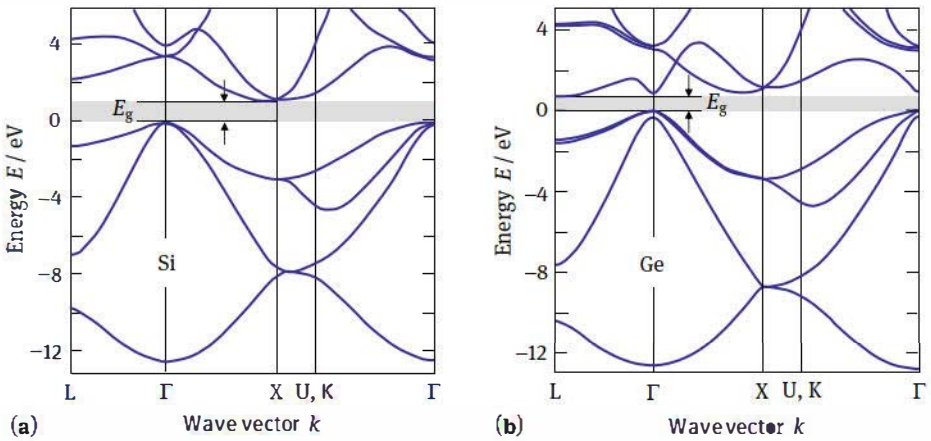


Fig. 10.3: Band structure of **a)** silicon and **b)** germanium. The energy gaps E_g are marked in grey. In indirect semiconductors, the valence band maximum is also located at the Γ-point, but the lowest conduction band minimum is near the X point for silicon and at the L point for germanium. (After J.R. Chelikowsky, M.L. Cohen, Phys. Rev. B **14**, 556 (1976).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

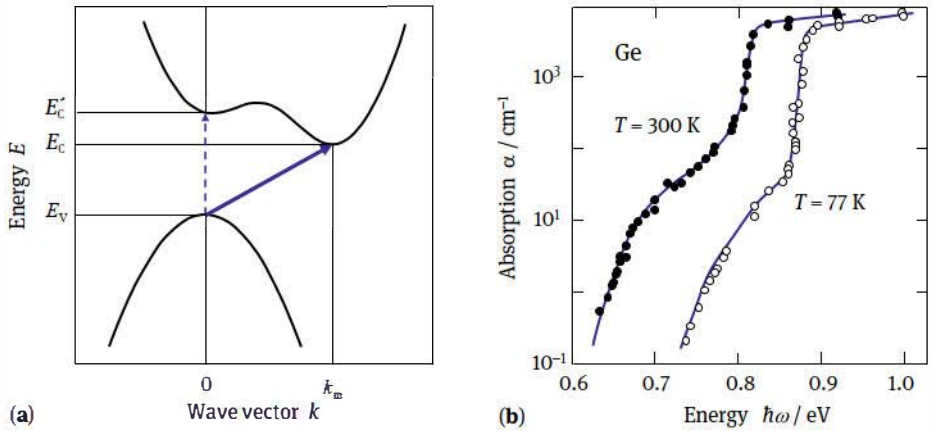


Fig. 10.4: Optical absorption at indirect band gaps. **a)** Schematic representation of the process. The energy of the conduction band minimum at the Γ-point is called E_C'. The transition with the smallest possible energy (solid blue arrow) requires the cooperation of a phonon. The direct transition with the smallest energy is shown by the dashed line. **b)** Absorption coefficient of germanium on a logarithmic scale as a function of the energy of the irradiating photons. With increasing temperature the weaker indirect absorption precedes the direct absorption. (After W.C. Dash, R. Newman, Phys. Rev. **99**, 1151 (1955).)

begins already at a photon energy smaller than the band gap energy ($E_C' - E_V$) at the Γ-point. If the lowest lying minimum of the conduction band is at $\mathbf{k}_m$, no direct transition is possible owing to the small photon momentum. For energy and quasi-momentum conservation, the participation of a phonon is necessary. With ω_q and $\mathbf{q}$ to denote the phonon frequency and wave vector, the conditions for generating an electron-hole pair with the smallest possible photon energy can be expressed as:

$$\begin{aligned} \hbar\omega_\gamma \pm \hbar\omega_q &= E_g \,, \\ \hbar\mathbf{k}_\gamma \pm \hbar\mathbf{q} &= \hbar\mathbf{k}_m \,. \end{aligned} \tag{10.1}$$

Since $\hbar\omega_q \ll E_g$ and $|\mathbf{k}_\gamma| \ll |\mathbf{k}_m|$ the photon provides the energy and the phonon provides the required momentum. The probability of this process taking place is much smaller than that of the direct process, because there has to be simultaneously both a photon and a phonon for the electron to interact with and in consquence the associated absorption is much weaker. The experimental curves for the optical absorption of germanium in Figure 10.4b support this conclusion. The absorption coefficient is relatively small near the minimum band gap energy and increases with the photon energy. The increase in the absorption curve is noticeably stronger, as soon as the direct process indicated by the dashed blue arrow in Figure 10.4a starts to be possible. The data displayed here also show that the position of the absorption edge depends on the temperature, since the band gap increases with decreasing temperature.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

10.1.2 The Effective Mass of Electrons and Holes

The electrical properties of semiconductors are mainly determined by the electrons in the conduction band minimum and the holes in the valence band maximum. We therefore need to know the band structure in these energy ranges a little more precisely. In Section 9.1 we found that the effective mass of electrons and holes is determined by the band curvature. In the vicinity of the band extrema, the curvature is approximately constant as thus also is the dynamic effective mass, which in this case corresponds very well to the cyclotron mass (cf. Section 9.3). Measurements of the cyclotron resonance are therefore suitable for determining the dynamic effective mass. Since the penetration depth of microwaves into semiconductors is relatively large meanin that the circulating electrons are constantly exposed to the microwave field, just *one* resonance is usually observed for a certain effective mass.

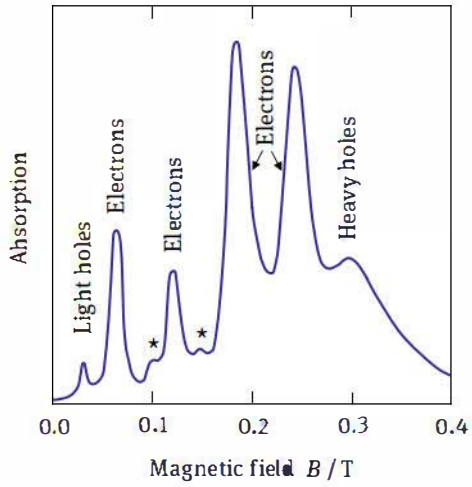


Fig. 10.5: Cyclotron resonance measured on germanium at 4 K and 23 GHz. With the selected sample orientation, all kinds of charge carrier appear as a function of the applied magnetic field. The maxima labeled *electrons* stem from electrons with different extremal orbits in the conduction band. The starred maxima (*) are harmonics of *heavy holes*. (After R.N. Dexter et al., Phys. Rev. **104**, 637 (1956).)

Cyclotron resonance experiments are performed at microwave frequencies. From the relation $\omega_c = eB/m$ a resonance frequency of 28 GHz is calculated for a field of one Tesla. Figure 10.5 shows a measurement on germanium at 23 GHz. If the magnetic field is tuned, a series of resonances occur which can be assigned to electrons and holes in different bands. We will discuss the names and meanings of the individual maxima in the following. In the experiment shown here, the germanium crystal was oriented in the magnetic field in such a way that all the different effective masses could be observed by just tuning simply the field.

In order to provide good signals for measurement, the electrons must cycle their orbits several times without scattering i.e. the condition $\omega_c \tau \gg 1$ must be fulfilled. Since the scattering times τ are about 10^{-13} s at room temperature, meaningful experiments are only possible on very pure samples at low temperatures. Under these conditions,

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the small number of available charge carriers is problematic. Fortunately, the number of charge carriers can be artificially increased by irradiating the semiconductor by light with energy greater than the band gap.

The masses of the various electrons and holes can be determined by selecting the appropriate magnetic field strength and direction. In the case of semiconductors with a direct band gap such as GaAs or InSb (see Figure 10.1), the properties of the electrons are crucial for the properties of the conduction band at the Γ-point. Since the energy at this location hardly depends on the direction, electrons only have *one effective mass* m_n^*[1] and the dispersion relation has the simple form:

$$E_n = E_C + \frac{\hbar^2 k^2}{2m_n^*} . \tag{10.2}$$

Table 10.2 lists the electron effective masses for various semiconductors. It is remarkable that these are usually much smaller than the mass of free electrons, and therefore the interactions lead to an apparent reduction in mass.

Tab. 10.2: Effective mass of the electrons at the conduction band minimum. The indices "ℓ" and "t" stand for *longitudinal* and *transverse*. (The majority of the data is from the homepage of the Ioffe Institute, Saint Petersburg).

	m_n^*/m	$m_{n,\ell}^*/m$	$m_{n,t}^*/m$
C	—	1.4	0.36
Si	—	0.98	0.19
Ge	—	1.59	0.082
GaP	—	1.12	0.22
GaAs	0.063	—	—
GaSb	0.041	—	—
InP	0.073	—	—
InAs	0.023	—	—
InSb	0.014	—	—

For semiconductors with indirect band gaps, such as germanium or silicon, the band structure is more complicated. As already mentioned, in germanium the conduction band minima are located at the L-points, i.e. along ⟨111⟩-direction. In silicon, the minima along the ⟨100⟩-direction are near the X-points. The $E(\mathbf{k})$-surface located near the minimum is not isotropic, but has the shape of a rotational ellipsoid. For illustration, Figure 10.6 shows areas of constant energy for the conduction band of germanium.

1 As in Section 9.1, we use the indices n and p to denote the effective masses of electrons and holes, respectively,

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The size of the eight semi-ellipsoids compared to the Brillouin zone depends on the arbitrarily chosen energy scale. In the periodic zone scheme, the energy areas are complete ellipsoids in which the electrons circulate at the cyclotron resonance frequency. Depending on the magnetic field direction, different extremal orbits contribute to the signal. The shape of the surfaces of constant energy can therefore be measured by tilting the samples.

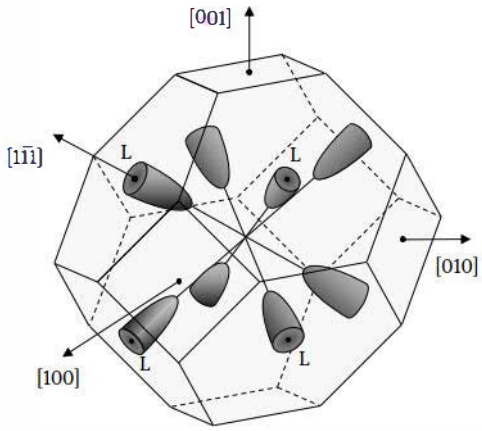


Fig. 10.6: Areas of constant energy for the conduction electrons in germanium. The conduction band minima occur along the ⟨111⟩-directions at the eight L-points.

The effective mass is defined by two quantities, namely the **longitudinal mass** $m_{n,\ell}$ and the **transverse mass** $m_{n,t}$. The surface of constant energy can therefore be described near the minima by the equation:

$$E_n = E_C + \hbar^2 \left[\frac{k_1^2 + k_2^2}{2m_{n,t}^*} + \frac{k_3^2}{2m_{n,\ell}^*} \right], \tag{10.3}$$

where the origin of the coordinate system lies at the respective minimum. For silicon, the direction of k_3 is identical with the ⟨100⟩-, and for germanium with the ⟨111⟩-direction. Depending on the direction, the curvatures of the energy surfaces differ considerably and thus also the effective masses. For silicon and germanium they are listed in Table 10.2. The four maxima in Figure 10.5 attributed to the electrons, originate from extremal orbits on different ellipsoids, since the sample was not aligned along a preferred axis.

Now we come to the holes in the valence band. On a closer look Figure 10.1 shows that the semiconductors depicted have maxima associated with two valence bands which coincide at the Γ-point. Owing to their different curvature, the holes have different effective masses and are called **heavy** and **light** holes. As Figure 10.5 shows, both types of holes can be observed in cyclotron-resonance measurements. A further band is observed which is due to the spin orbit coupling lowered by the energy Δ. The charge carriers of this band are called **split-off holes** and have an effective mass m_Δ^*. The

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

corresponding band is shown in Figure 10.1 just below the valence band maximum. To a rough approximation, the valence bands of almost all semiconductors can be taken as spherical at the Γ-point, so that *one* effective mass per band is sufficient for characterization. Table 10.3 lists the effective masses of the various hole types for a number of semiconductors along with the energies Δ of the spin-orbit splitting.

Tab. 10.3: Effective mass of the holes at the valence band maximum and the spin-orbit splitting Δ of some semiconductors. The indices "s" and "l" denote *heavy* and *light* respectively. (The majority of the data is from the homepage of the Ioffe-Institute, St. Petersburg.)

	$m^*_{p,l}/m$	$m^*_{p,s}/m$	m^*_{Δ}/m	Δ (eV)
C	0.7	2.12	1.06	0.006
Si	0.16	0.49	0.24	0.044
Ge	0.043	0.33	0.084	0.29
GaP	0.14	0.79	0.25	0.08
GaAs	0.082	0.51	0.15	0.34
GaSb	0.05	0.4	0.14	0.80
InP	0.089	0.58	0.17	0.11
InAs	0.026	0.41	0.16	0.41
InSb	0.015	0.43	0.19	0.82

10.1.3 Charge Carrier Density

In semiconductors, both electrons and holes contribute to the current transport. For the conductivity of intrinsic semiconductors we therefore write as an extension of (9.22)

$$\sigma = e(n\mu_n + p\mu_p)\,, \tag{10.4}$$

where n and p are the densities of the free electrons and holes, respectively. The current transport by holes is dominated by the contribution of the lightest holes. Because the charge carrier densities are extremely dependent on temperature, the conductivity also changes very rapidly with temperature. In the following, we first discuss the charge carrier densities and then the mobility μ_n and μ_p.

First we consider the **intrinsic conductivity** of semiconducting crystals, which can be observed in very pure samples. Electrons are excited into the conduction band by thermal excitation and creating holes in the valence band at the same time. When determining the carrier density, we do not need to distinguish between direct and indirect semiconductors, since in thermal equilibrium the dynamics of the excitation is not important.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

We obtain the electron density n in the conduction band by integrating the product of density of states and the occupation probability. Since the Fermi function $f(E,T)$ decreases rapidly with increasing energy, only the states at the lower edge of the conduction band are important. To simplify the calculation we are thus justified in using infinity for the upper integration limit rather than the energy of the upper band edge. We also apply the same argument to the calculation of the hole density p. Thus we can write:

$$n = \int_{E_C}^{\infty} D_C(E)\, f(E,T)\, \mathrm{d}E \qquad \text{and} \qquad p = \int_{-\infty}^{E_V} D_V(E)\, [1 - f(E,T)]\, \mathrm{d}E\,. \tag{10.5}$$

The occupation probability of the holes is just written $[1 - f(E,T)]$, since they are unoccupied electron states. Since, in the vicinity of the band extrema the dispersion curves are parabolic, we can use the density of states of the free electron gas for the densities of state D_C and D_V of the two bands, if we replace the mass of the free electrons in (8.9) by the effective masses m_n^* and m_p^*:

$$D_C(E) = \frac{1}{2\pi^2}\left(\frac{2m_n^*}{\hbar^2}\right)^{3/2}\sqrt{E - E_C} \qquad \text{for} \qquad E > E_C\,, \tag{10.6}$$

$$D_V(E) = \frac{1}{2\pi^2}\left(\frac{2m_p^*}{\hbar^2}\right)^{3/2}\sqrt{E_V - E} \qquad \text{for} \qquad E < E_V\,. \tag{10.7}$$

The two bands are separated from each other by an energy gap, in the range $E_V < E < E_C$ where there are no states.

As we will see in a moment, the Fermi energy lies approximately in the middle of the energy gap in the case of intrinsic conductance. This means that its separation from the band edges, i.e. $|E_C - E_F|$ and $|E_V - E_F|$, is always much greater than the thermal energy $k_B T$. When integrating the two equations (10.5), we can therefore approximate the Fermi energy by an exponential function:

$$f(E,T) = \frac{1}{\mathrm{e}^{(E-E_F)/k_BT} + 1} \approx \mathrm{e}^{-(E-E_F)/k_BT} \qquad \text{for} \qquad E > E_F\,,$$

$$1 - f(E,T) = \frac{1}{\mathrm{e}^{(E_F-E)/k_BT} + 1} \approx \mathrm{e}^{-(E_F-E)/k_BT} \qquad \text{for} \qquad E < E_F\,. \tag{10.8}$$

This means that we can directly apply Boltzmann statistics to the charge carriers in semiconductors. Electrons in the conduction band and holes in the valence band thus move largely like the atoms of classical gases and can be described with the help of kinetic gas theory. In reservation, it must of course be said that this statement is only valid so long as the Fermi level is sufficiently far from the band edge. As we will see, this **approximation of non-degeneracy** is valid for intrinsic semiconductors and for those with low impurity densities. With heavily-doped materials, the Fermi energy moves nearer to the band edge or even lies within the band, and the simple approximation breaks down and we have a **degenerate semiconductor**.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

It should be pointed out that in the literature different nomenclatures for Fermi energy and chemical potential are used. Often only the chemical potential introduced in Section 8.1 at $T = 0$ is called Fermi energy. In order to avoid confusion with the mobility μ, we will use the term **Fermi level** and – not quite consistently – the abbreviation E_F or $E_F(T)$ to denote the position of the chemical potential $\mu(T)$ at finite temperatures.

Replacing the Fermi distributions with the corresponding Boltzmann distributions allows the charge carrier densities to be calculated in a simple way. For the electrons we find for example:

$$n = \int D_C(E)\, f(E)\, dE = \frac{1}{2\pi^2}\left(\frac{2m_n^*}{\hbar^2}\right)^{3/2} e^{E_F/k_B T} \int_{E_C}^{\infty} \sqrt{E - E_C}\, e^{-E/k_B T}\, dE\,. \tag{10.9}$$

The integral can readily be solved analytically and for the two charge carrier types we obtain:

$$n = 2\left(\frac{m_n^* k_B T}{2\pi\hbar^2}\right)^{3/2} e^{-(E_C - E_F)/k_B T} = \mathcal{N}_C e^{-(E_C - E_F)/k_B T}\,, \tag{10.10}$$

$$p = 2\left(\frac{m_p^* k_B T}{2\pi\hbar^2}\right)^{3/2} e^{(E_V - E_F)/k_B T} = \mathcal{N}_V e^{(E_V - E_F)/k_B T}\,. \tag{10.11}$$

The **effective densities of states** $\mathcal{N}_C$ and $\mathcal{N}_V$ introduced here, compared to the exponential factor, are only "weakly" dependent on temperature. Thus we achieve a very far-reaching simplification: in this approximation it looks as if we are no longer dealing with two broad bands but just two discrete levels with energies E_C and E_V. We will use this simplification very often later in this chapter.

The charge carrier densities n and p are both determined, among other things, by the position of the Fermi level. However, the Fermi energy E_F does not appear in the product $n\,p$ of the two carrier densities. Taking into account that the band gap is determined by the difference $E_g = (E_C - E_V)$, from (10.10) and (10.11) we obtain the relation:

$$np = \mathcal{N}_C \mathcal{N}_V\, e^{-E_g/k_B T} = 4\left(\frac{k_B T}{2\pi\hbar^2}\right)^3 \left(m_n^* m_p^*\right)^{3/2} e^{-E_g/k_B T}\,. \tag{10.12}$$

The product of the charge carrier densities is completely characterized by the masses of the charge carriers and the energy gap and has a characteristic value at fixed temperature for each non-degenerate semiconductor. In the thermodynamics context, the relation $n\,p$ = const. is often called the **law of mass action**. In equilibrium, new charge carriers are constantly generated by thermal excitation, and after a short time are annihilated again by recombination. During the time interval between generation and the recombination, they diffuse in the sample and cover a distance known as the **diffusion length**.

So far we have made no assumptions which would restrict our treatment simply to the case of intrinsic conductivity. Therefore we will also use the relationships derived here for doped semiconductors further on in this chapter. Beside (10.10) and (10.11) especially the law of mass action will also play an important role.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In the rest of this section, we limit ourselves to considering only intrinsic semiconductors in which the source of the conduction electrons is the valence band alone. Therefore the important relation $n_i = p_i$ applies, where we use the index "i" to characterize intrinsic quantities. For intrinsic semiconductors, it follows from (10.12) that

$$n_i = p_i = \sqrt{\mathcal{N}_C \mathcal{N}_V}\, e^{-E_g/2k_B T} . \tag{10.13}$$

Typical numerical values for the energy gap and intrinsic charge carrier density for some important semiconductors are given in Table 10.4. In Si and GaAs the density at room temperature is very low. It is therefore extremely difficult or even impossible to observe intrinsic conduction in these materials at room temperature, because impurities generally give rise to carrier densities higher than the intrinsic values.

Tab. 10.4: Band gap and calculated charge carrier density for some important semiconductors at 300 K.

	E_g/ eV	n_i/m^{-3}
Germanium	0.66	2.4×10^{19}
Silicon	1.12	1.1×10^{16}
Gallium arsenide	1.42	1.8×10^{12}

The position of the Fermi level and its temperature dependence can be generally determined from the requirement for charge neutrality. With $n_i = p_i$, for intrinsic conduction it follows from (10.10) and (10.11) that:

$$E_F = \frac{E_C + E_V}{2} + \frac{k_B T}{2} \ln\left(\frac{\mathcal{N}_V}{\mathcal{N}_C}\right) = \frac{E_C + E_V}{2} + \frac{3}{4} k_B T \ln\left(\frac{m_p^*}{m_n^*}\right) . \tag{10.14}$$

For $T = 0$ the Fermi level is in the middle of the energy gap. If the effective masses of the electrons and the holes are the same, i.e. if the valence band and the conduction band have the same curvature, the position does not change with increasing temperature. If the masses differ, the Fermi level shifts, but the temperature dependence is small compared to the magnitude of the band gap.

Figure 10.7 clearly summarizes the most important results. The number of excited electrons and holes initially increases at the band edges due to the parabolic dependence of the density of states. At higher excitation energies, it then drops again due to the decreasing probability of occupation. Since intrinsic semiconductors have the same number of electrons and holes, the darker blue area in the conduction band and the lighter blue area in the valence band are of the same size. As shown in the right-hand picture, for this to be possible with different densities of states in the two bands, the Fermi level is shifted towards the band with the lower density of states.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

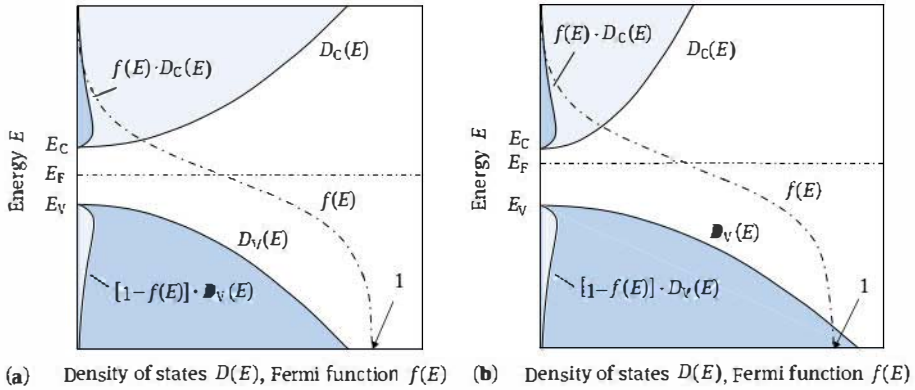


Fig. 10.7: Densities of state $D(E)$ and Fermi functions $f(E)$ in the valence and conduction bands for $T > 0$. The electrons are shown in dark blue, the holes in light blue. **a)** Semiconductors with the same density of states in the valence and conduction bands, i.e. $\mathcal{N}_C = \mathcal{N}_V$, **b)** Semiconductors with $\mathcal{N}_V > \mathcal{N}_C$, i.e. with the larger number of states at the valence band edge.

10.2 Doped Crystalline Semiconductors

10.2.1 Doping

In real crystals there are always impurities which cannot be avoided during crystal growth. Usually these impurities are "electrically active", i.e. they contribute additional charge carriers in the bands. For example, in the purest and best grown silicon or germanium crystals, there are about 10^{18} charge carriers per cubic meter and in GaAs crystals about 10^{22} charge carriers per cubic meter instead of the values given in Table 10.4. This means that at room temperature, intrinsic properties can only be observed in germanium. The electrical conductivity in silicon is too small for application purposes because of the small charge carrier density. If we insert the numerical values into the equation for the conductivity of silicon, one finds the small value $\sigma_i \approx n_i e \mu_n \approx 4 \times 10^{-4}\,(\Omega\mathrm{m})^{-1}$.

As simple examples we take the tetravalent elements silicon and germanium. As discussed in Section 2.4, the atoms in these crystals exhibit an sp^3-hybrid bond. If we now introduce pentavalent P, As or Sb atoms, the neighboring silicon atoms try to force a quadruple coordination, but since the foreign atoms introduced have an s^2p^3-configuration, they have one more electron than is required for the tetrahedral bond. This electron cannot participate in the bond, but remains (weakly) bound to the positive atomic core at low temperatures due to Coulomb interaction, creating a local cloud of net negative charge. Figure 10.8a illustrates the insertion of an arsenic atom. As we will see, the binding energy of the excess electron is so low that it can already at room temperature separate from the atomic core and contribute to the electrical conductivity. Pentavalent impurity atoms in crystals with quadruple bonds are therefore known as **donors**, because they "donate" electrons at room temperature.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

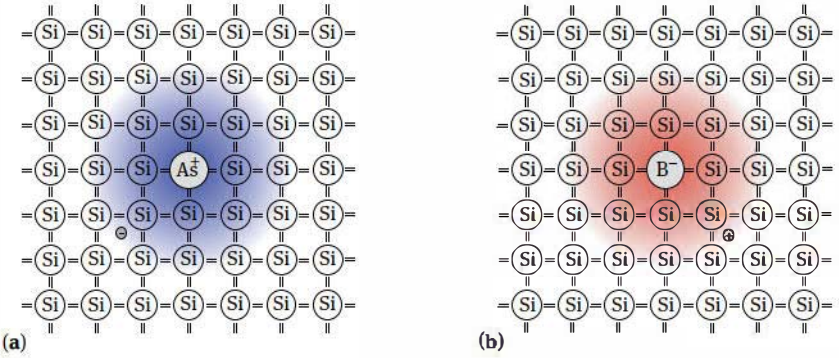


Fig. 10.8: Dopant atoms in silicon. **a)** *Donor:* the extra electron not contributing to the bonds of the donor atom loosely orbits the positively charged body of the arsenic atom producing a local negative charge. **b)** *Acceptor:* similarly, the deficit electron needed for the acceptor atom bonds is collectively provided by the neighboring atoms producing a net local positive charge, loosely coupled to the boron atom.

At first glance, trivalent impurity atoms, such as B, Al, Ga or In, behave quite differently. Although the trivalent impurity atoms are also incorporated in regular lattice sites, they lack one electron to complete the fourth bond. Nevertheless, the cubic symmetry remains intact. The missing electron is collectively subtracted from the neighboring silicon atoms leaving, in this case, a surrounding localized cloud of net positive charge. To the outside, it looks as if a positive charge is circling around the negatively charged atomic core of the impurity atom. A schematic representation of this situation is shown in Figure 10.8b for the case of a boron atom. At room temperature, the net positive charge can be detached from the impurity atom and transferred to the lattice or, in other words, an electron equivalent is absorbed permanently by the impurity, leaving a net positive charge to contribute to the conduction. Trivalent impurity atoms therefore act as **acceptors**.

The question now arises as to which new properties the weakly bound electrons of the donor atoms have. Let us first look at the optical absorption by the donors. In the simplest approximation, the motion of the additional electron around the positively charged atomic core of the donor atom can be described with the help of the *Bohr model of the hydrogen atom*. In this approximation, the energy eigenvalues E_ν are given by

$$E_\nu = -\frac{1}{2}\frac{m^* e^4}{(4\pi\varepsilon_r\varepsilon_0)^2\hbar^2}\cdot\frac{1}{\nu^2}\,, \tag{10.15}$$

where ν is the principal quantum number.[2] This equation differs in two ways from the well-known expression for the energy levels of the hydrogen atom. In the numerator, instead of the mass m of the free electron, the *effective mass* m^* is used and in the

2 To avoid confusion with particle number density, ν here represents the principal quantum number.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

denominator the dielectric constant ε_r appears, taking into account the shielding effect of the Coulomb potential from the neighboring atoms. These quantities have the effect that the energy levels are reduced by a factor of 100 to 1000 in comparison to those of the free hydrogen atom. Furthermore, it is remarkable, that the effective Bohr radius $a_0 = 4\pi\varepsilon_r\varepsilon_0\hbar^2/m^*e^2$ is approximately 50 times greater than that of a hydrogen atom. Thus within the electron orbit there are about 1000 lattice atoms!

Figure 10.9 shows the position of the donor and acceptor levels in the semiconductor band gap for the two types of doped semiconductors. Materials which are dominated by donors are known as **n-type** materials, as they have extra negative carriers, and acceptor dominated materials are known as **n-type** materials, as they have extra positive carriers. Using $\nu = 1$, we obtain the ionization energy, denoted by E_d and E_a in the figure. For donors in silicon, $\varepsilon_{r,Si} = 11.7$ yielding a value $E_d \approx 30$ meV. The base of the conduction band corresponds to the vacuum level for the hydrogen atom, i.e. the ground state E_D of the donor is 30 meV below E_C. We have taken the effective mass used in the estimate to be $m_n^* \approx 0.3\,m$. Treating germanium in the same way, since $\varepsilon_{r,Ge} = 16.6$ and $m_n^* \approx 0.15m$, the ionization energy with $E_d \approx 9$meV is even smaller. The donor and acceptor atoms are not themselves mobile and do not contribute directly to conductivity. At low temperatures the impurities are in the ground state, i.e. they are neutral. At room temperature they are positively or negatively charged due to the low ionization energy.

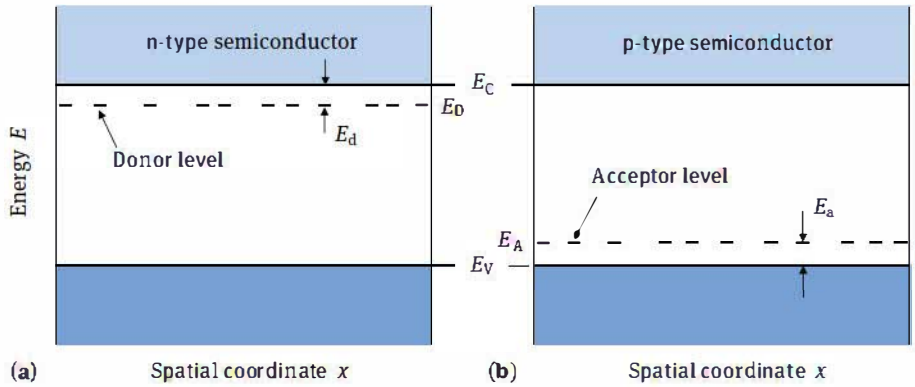


Fig. 10.9: Donator and acceptor levels in semiconductors. **a)** The ground state E_D of the donators with ionization energy E_d lies directly below the conduction band edge E_C. **b)** The ground state E_A of the acceptors with ionization energy of the holes E_a is just above the valence band edge E_V.

In the hydrogen model, the energy of the levels is independent of the nature of the donor or acceptor atoms. Table 10.5 shows the experimental values of the ionization energies of some impurities for silicon and germanium. Considering the rough approximation used in the calculation, the agreement of the experimental values with those of the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Tab. 10.5: Ionization energies of donors in silicon and germanium in meV. (The majority of the data taken from the homepage of the Ioffe-Institute, St. Petersburg).

	Donator	P	As	Sb	Bi	Acceptor	B	Al	Ga	In
Silicon		45	54	43	69		45	72	74	157
Germanium		13	14	10	13		11	11	11	12

simple calculation is surprisingly good. Obviously, the shielding can be well taken into account simply by adjusting the dielectric constant. The main reason for this is the large effective radius of the "hydrogen-like atoms".

The ionization energy of the donors is small, corresponding approximately to the thermal energy at room temperature. Therefore, as already mentioned, the donors are ionized at this temperature. This also applies to acceptors which we will not further discuss in detail here. In analogy with hydrogen, we would expect further energy eigenstates with the corresponding higher values of the quantum number ν. Figure 10.10 demonstrates that this expectation is justified in principle. It shows the infrared spectrum of antimony-doped germanium, where a number of absorption lines are visible. The measurement was made at low temperature so that the electrons not participating in the bonding of the antimony atoms remain in the ground state. However, the observed spectrum is not structured as simply as would be expected from (10.15). Since the donor electrons move in the crystal field which has cubic symmetry rather than in a spherically symmetric Coulomb field, this is not surprising. Above the ionization energy of 9.6 meV – for germanium $\varepsilon_r^2 m/m^* \approx 1500$ – a transition to continuous states, that is states of the *conduction band*, takes place. Similar spectra are also observed when other impurity atoms are incorporated in germanium or other semiconducting host crystals.

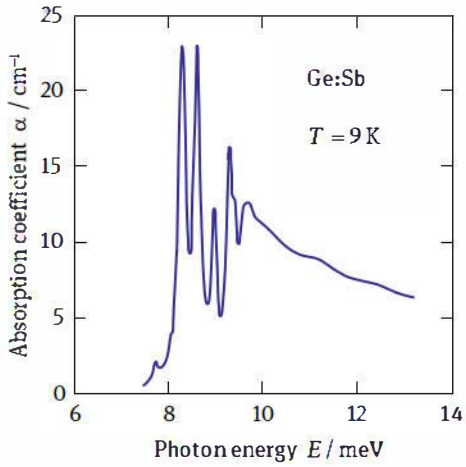


Fig. 10.10: Infrared absorption of germanium with a doping of 7×10^{26} m^{-3} antimony atoms. (After J.H. Reuszer, P. Fisher, Phys. Rev. **135**, A1125 (1964).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The theoretical description of the electronic eigenstates of impurity atoms can be significantly improved by taking into account a number of solid-state effects. Above all, the actual form of the local electric field must be taken into account, since this causes a noticeable splitting of the ground state and can imply selection rules for the optical transitions differing from those of the hydrogen atom. If a quantitative description of the energy levels is of importance in the context of different doping elements and host lattices, the details of the covalent bond in question must also be taken into account, as well as the influence of the effective masses and their directional dependence. In this way quantitative agreement with the experimental data is much improved, but such subtleties are not important for the treatment here.

Figure 10.11 shows the values of the ionization energies of a number of different dopant elements within the band gap of germanium. Obviously, the position of these levels for elements not coming from the third or fifth main groups cannot be described in the simple hydrogen model. In particular, elements such as zinc, chromium or copper can have several charge states with different energies. Gold, for example, can form four different levels in germanium and can act as both donor or acceptor.

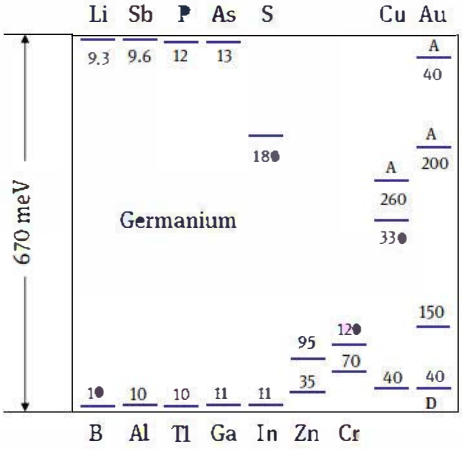


Fig. 10.11: Ionization energies of impurity atoms in germanium. The numbers indicate the separation from the nearest band edge. The letters A and D denote *acceptor* and *donor*, respectively, when the nature of the state concerned is not directly obvious from the position. (After S.M. Sze, *Physics of Semiconductor Devices*, Wiley, 1981.)

10.2.2 Charge Carrier Density and Fermi Level

We now calculate the charge carrier densities in doped semiconductors, but limit ourselves to *non-degenerate semiconductors*, i.e. semiconductors whose impurity density is relatively low. The interaction between the impurities is neglected, since it only becomes important at high concentrations.

In real semiconductors there are always unwanted donors and acceptors due to uncontrollable contamination. As a result, the energetically higher-lying donors trans-

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

fer electrons to the lower-lying acceptors, but this electron transfer is not associated with the generation of free charge carriers. The effect of the two types of impurities cancel out, an effect that is called **compensation** and is used in the production of high-resistance semiconductors. In this process, the often known but unavoidable impurities can be compensated by well-defined additions of doping material.

The energy levels of band states hardly depends on the occupation of the states. It is independent of whether a state is empty, occupied by one electron or two electrons (with opposite spin). This is understandable because the electrons in these states are delocalized, i.e. smeared all over the crystal. Given this smearing, they do not "see" each other; the one-electron approximation is therefore valid. Electrons at impurities are spatially localized. If we leave aside structurally complicated impurities with configurations of different energies, each impurity state can in principle be unoccupied, or occupied by one or two electrons as in the band states. In the case of double occupation, however, the Coulomb repulsion between the spatially localized electrons considerably increases the energy of the impurity. Double occupied states are therefore very often located in the conduction band and are normally irrelevant, so that only single charged and neutral states have to be considered.

When calculating the number of free charge carriers, we divide the density n_D or n_A of donors and acceptors into a neutral and a charged part:

$$n_D = n_D^0 + n_D^+ \qquad \text{and} \qquad n_A = n_A^0 + n_A^- . \tag{10.16}$$

The occupation of the ground state, i.e. the probability that an impurity atom is not ionized, can be expressed by means of the Fermi-Dirac distribution:

$$\frac{n_D^0}{n_D} = \left[\frac{1}{g_D} e^{(E_D - E_F)/k_B T} + 1 \right]^{-1} \quad \text{and} \quad \frac{n_A^0}{n_A} = \left[g_A e^{(E_F - E_A)/k_B T} + 1 \right]^{-1} . \tag{10.17}$$

The weighting factors g_D and g_A take into account the degeneracy of the impurity levels, which slightly modifies the usual expression for the Fermi statistics. For simple donors, which we are looking at here, an electron can be added spin up or down. This "choice" gives the state a double statistical weight and leads to the weighting factor $g_D = 2$. In the case of acceptors the situation is somewhat more complicated. There, the degeneracy of the valence band must also be taken into account for the common semiconductors. This also gives the ground state a higher weight and thus lowers g_A. Although the degree of degeneracy is important in numerical calculations, we will omit it in the further discussion to simplify the formulas, because we are only interested in the primary behavior.

Free charge carriers with densities n and p are generated by excitation of electrons from the valence band and especially by ionization of the impurities. In the previous section we already derived the corresponding expressions (10.10) and (10.11) for the band states. Since no assumption that would cause a restriction to intrinsic charge carriers was made, the equations apply to doped semiconductors as well. We can

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

therefore also use them here. In addition we have

$$n + n_A^- = p + n_D^+ , \tag{10.18}$$

which reflects the condition of charge neutrality for doped semiconductors.

Despite these highly simplified assumptions, the situation is relatively complex, as there are four different charge densities in temperature-dependent equilibrium. In the following we will consider the simple and realistic case that one type of impurity dominates. We discuss an *n-type semiconductor*, in which many donors and a few acceptors are present. For the sake of clarity, we repeat the relevant equations adapted for this specific case:

$$n = N_C e^{-(E_C - E_F)/k_B T} , \tag{10.19}$$

$$n_D = n_D^\bullet + n_D^+ , \tag{10.20}$$

$$\frac{n_D^\bullet}{n_D} = \frac{1}{e^{(E_D - E_F)/k_B T} + 1} , \tag{10.21}$$

$$n + n_A^- = n_D^+ + p . \tag{10.22}$$

Equation (10.19) is of course only valid for non-degenerate semiconductors which fulfill the condition $(E_C - E_F) \gg k_B T$. The doping should be so high that the contribution of the dopant atoms to the concentration of the free charge carriers dominates, i.e. **extrinsic conduction** prevails. Then $n_D^+ \gg p$, i.e. the density n_i of the excited electrons from the valence band can be neglected compared with the contribution of the impurities. A further simplification follows from the assumption $n_D \gg n_A$ made above. In this case all acceptors capture an electron which originally belonged to a donor. As there are practically no neutral acceptors anymore we can neglect $n_A^\bullet$ and replace n_A^- with n_A. Thus (10.22) simplifies to $(n + n_A) = n_D^+$.

Figure 10.12 illustrates the situation described here. In the compensated n-type semiconductor the valence band and acceptor levels are fully occupied by electrons.

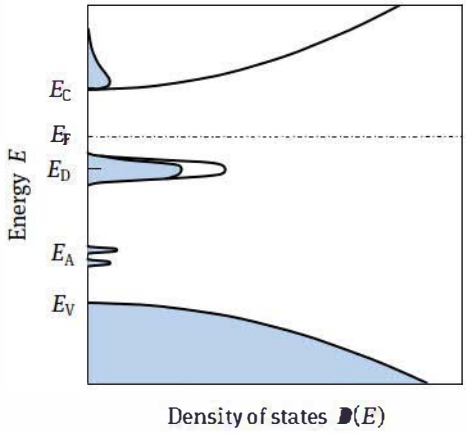


Fig. 10.12: Occupation of the states in a compensated n-type semiconductor. States occupied by electrons are shown in blue. The two levels marked E_A represent the densities of two different impurities acting as acceptors.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Part of the donors are ionized by transferring electrons to the acceptors and most importantly into the conduction band.

With the simplifications described above, the concentration n of the conduction electrons can be determined by using (10.20) – (10.22):

$$n = n_D^+ - n_A = (n_D - n_D^0) - n_A = n_D \left[1 - \frac{1}{e^{(E_D - E_F)/k_B T} + 1} \right] - n_A \,. \tag{10.23}$$

With (10.19) we eliminate E_F and get the final expression with $E_d = (E_C - E_D)$, which we will use to discuss the temperature dependence of n:

$$\frac{n(n_A + n)}{n_D - n_A - n} = \mathcal{N}_C \, e^{-E_d/k_B T} \,. \tag{10.24}$$

The resulting temperature dependence of electron density n is shown qualitatively in Figure 10.13. If the electron density is known, the position of the Fermi level can be directly derived from (10.19) and is also shown.

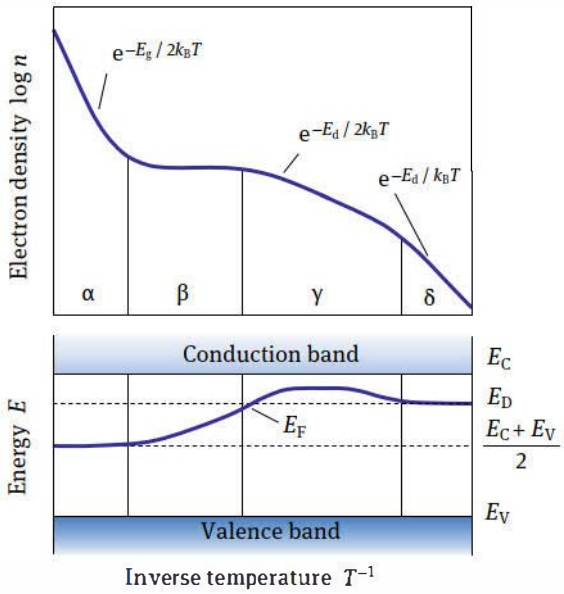


Fig. 10.13: The electron density in the conduction band of an n-type semiconductor (top) and position of the Fermi level (bottom) as a function of the reciprocal temperature. The four temperature ranges α – δ are discussed in the text.

Intrinsic Regime. At *high temperatures* (range α), intrinsic conduction is dominant and most charge carriers are excited from the valence band $p \gg n_D$. The Fermi level lies approximately in the middle of the band gap and the concentration of free charge carriers increases strongly with temperature. Equations (10.10) and (10.14) which we derived for intrinsic semiconductors apply.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Saturation Regime. At *room temperature* (range β) the intrinsic conduction becomes unimportant because the available thermal energy is not sufficient to excite electrons from the valence band into the conduction band. We have $k_B T \approx E_d$, so we can set $\exp(-E_d/k_B T) \approx 1$ as a good approximation. Since the acceptors are not important, equation (10.24) takes the form $n^2 \approx \mathcal{N}_C(n_D - n)$. With $n \ll \mathcal{N}_C$ the simple relation $(n_D - n) \approx 0$ follows. The density of the free charge carriers is now determined by the number of impurities and is independent of temperature. If we neglect n_A, the following applies

$$n \approx n_D = \text{const.}\,, \tag{10.25}$$

$$E_F \approx E_C - k_B T \ln\left(\frac{\mathcal{N}_C}{n_D}\right)\,. \tag{10.26}$$

The Fermi level moves continuously upwards with decreasing temperature. Since all dopant atoms are ionized, it is called *saturation regime*.

Freeze-out Regime. At *low temperatures* (range γ), the electron density decreases rapidly with decreasing temperature because the available thermal energy can ionize fewer and fewer donors. This phenomenon is called *freeze-out*. In this situation the inequality $n_A \ll n \ll n_D$ is valid. This results in the expressions

$$n \approx \sqrt{n_D \mathcal{N}_C}\, e^{-E_d/2k_B T}\,, \tag{10.27}$$

$$E_F \approx E_C - \frac{E_d}{2} - \frac{k_B T}{2} \ln\left(\frac{\mathcal{N}_C}{n_D}\right)\,. \tag{10.28}$$

In this range the Fermi level lies about halfway between the conduction band and the donor level.

Compensation Regime. At *very low temperatures* (range δ), i.e. for $k_B T \ll E_d$, the inequality $n \ll n_A \ll n_D$ holds. Thus (10.24) and (10.19) can be approximated by

$$n \approx \frac{\mathcal{N}_C n_D}{n_A}\, e^{-E_d/k_B T}\,, \tag{10.29}$$

$$E_F \approx E_C - E_d + k_B T \ln\left(\frac{n_D}{n_A}\right)\,. \tag{10.30}$$

At $T \to 0$ the position of the Fermi level is determined by the donors, which are partially charged due to the compensation effect. Therefore the Fermi level initially sits at the donor energy. As the donors also transfer electrons to the conduction band with rising temperature, E_F then gradually increases, while at the same time the charge carrier density n increases exponentially. Compared with low temperatures (range γ), the exponent in the exponential factor in the expression for the density of the conduction electrons increases by a factor of two.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

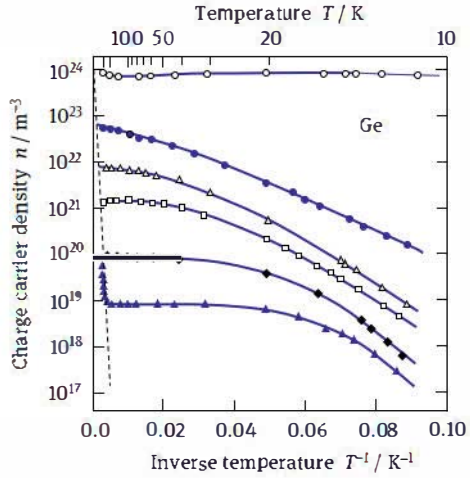


Fig. 10.14: The density of charge carriers in n-type germanium as a function of the inverse temperature, measured using the Hall effect. The range of intrinsic conduction is shown as a dashed line. Following the curves upward, the arsenic doping density increases in steps of approximately ten covering the range from 10^{19} m^{-3} to 10^{24} m^{-3}. (After E.M. Conwell, Proc. I.R.E., 1327 (1952).)

Figure 10.14 shows the experimentally determined densities of free electrons in n-doped germanium. The donor concentration of the samples can be read in this figure directly from the charge carrier density at higher temperatures (range β in Figure 10.13). The ranges of the saturation and ionization regimes are clearly visible. The intrinsic conduction was only investigated for the lowest doped sample. At first glance, the sample with the highest donor concentration ($n_D = 10^{24}$ m^{-3}) show a surprising behavior, since the charge carrier density, contrary to expectations, is largely independent of temperature. At this high impurity concentration, however, the mean distance between the donors is small so that their wave functions overlap. Thus, the donor electrons are no longer localized even at low temperatures. We have already seen this effect in Section 8.3, when we looked at the metal-insulator transition. As an example, in Figure 8.11 the results for silicon was shown, which was highly doped with phosphorus. For arsenic-doped germanium, the metal-insulator transition occurs at a concentration of 3.5×10^{23} m^{-3} which means that the most highly doped sample in the figure actually has metallic properties, since the Fermi level is already in the conduction band, ensuring that the carrier concentration is independent of temperature.

The easiest way of determining the density of free charge carriers is by using the Hall effect. In semiconductors both types of charge carriers contribute. Equation (9.95) applies, which we repeat here:

$$R_H = \frac{p\mu_p^2 - n\mu_n^2}{e(p\mu_p + n\mu_n)^2} \,. \tag{10.31}$$

The sign of R_H is determined by the charge carriers making the main contribution to the charge transport. If one type of charge carrier is dominant, its density can be determined directly from the Hall voltage. In this case the constant R_H is independent of the mobility and given by equation (9.94). Compared to metals, the Hall constant of semiconductors is relatively large due to their much smaller carrier density.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 10.15 shows Hall measurements on n- and p-doped InSb samples. In the saturation regime, one charge carrier species dominates. Thus, the simplification $R_H \approx -1/en \approx -1/en_D$ and $R_H \approx 1/ep \approx 1/en_A$ holds. Since the charge carrier density is constant, the mobility has no influence on the Hall voltage. At high temperatures, the Hall constant decreases exponentially with increasing temperature given that the density of free charge carriers increases exponentially with intrinsic conduction. Due to their higher mobility (see Table 10.6), it is usually the electrons and not the holes that contribute mainly to conductivity and the Hall constant in the case of intrinsic conduction. This leads to a remarkable temperature dependence in the p-doped samples: at the transition from the p-conduction to the intrinsic conductivity the Hall constant disappears according to (10.31) and then changes its sign.

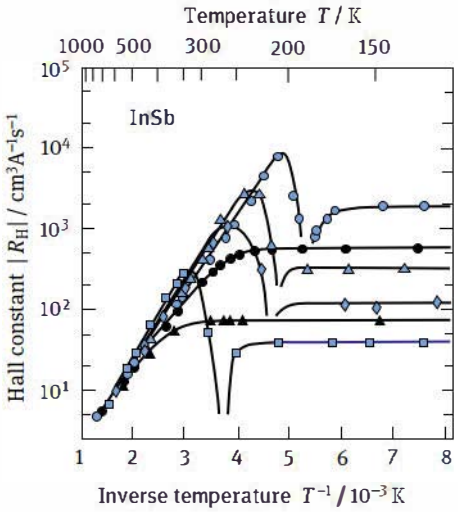


Fig. 10.15: Hall constant $|R_H|$ of indium antimonide as a function of the inverse temperature. The measured data for the n-doped samples are shown in black and those for the p-doped samples in blue. (After O. Madelung, H. Weiss, Z. Naturf. **9a**, 527 (1954).)

10.2.3 Mobility and Electrical Conductivity

To understand fully conductivity in semiconductors, we are lacking a discussion of the mobility $\mu = e\tau/m^*$, for which some typical values are listed in Table 10.6. The mobility is determined by the effective mass and above all by the scattering time. The electrons in semiconductors are mainly scattered by defects and phonons. Positive and negative charge carriers behave in a similar way with respect to scattering times, since the scattering of holes can ultimately be attributed to the scattering of electrons. The electron-electron scattering is of no importance, because of the low electron density. There are other scattering mechanisms in play, but we will only mention here scattering by optical phonons, since the other mechanisms do not play a major role in most cases.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Tab. 10.6: Mobility μ_n of electrons and μ_p of holes in semiconductors at room temperature. (The data were taken from various sources).

Material	μ_n (cm^2/Vs)	μ_p (cm^2/Vs)	Material	μ_n (cm^2/Vs)	μ_p (cm^2/Vs)
C	1 800	1 400	InSb	77000	850
Si	1 400	450	InP	5 400	200
Ge	3 900	1 900	AlSb	900	400
GaAs	8 500	400	PbS	550	600
GaSb	5 000	1 000	PbSe	1020	930
InAs	40000	500	PbTe	2 500	1 000

In crystals with ionic bonding, such as GaAs, scattering occurs by longitudinal optical phonons because the electrons couple to the local electric fields of these phonons. The mobility of the electrons in aluminum-doped Mg_2Ge is shown in Figure 10.16. Sample (1) was not intentionally doped, but owing to impurities has an effective dopant density of $(n_D - n_A) = 1.3 \times 10^{22}$ m^{-3}. The aluminum concentration of sample (2) is 4.2×10^{22} m^{-3} and of the most heavily doped sample (3) 8.2×10^{23} m^{-3}. With the exception of sample (3) which shows metal-like behavior due to the high donor density, starting at low temperatures we find a sharp increase in mobility with temperature, a maximum in the range of 50 K to 100 K and then a steep drop.

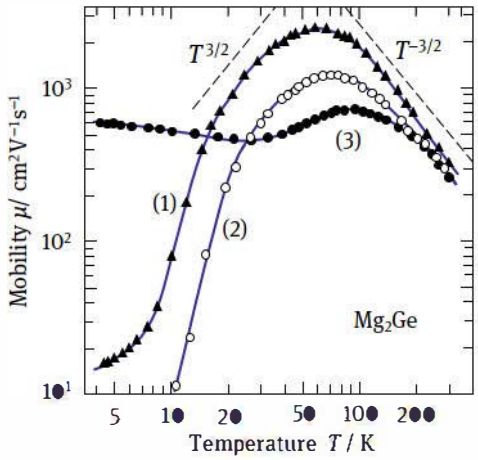


Fig. 10.16: The mobility of electrons in aluminum-doped Mg_2Ge, measured using the Hall effect. The samples, labelled (1), (2) and (3), contained 1.3×10^{22}, 4.2×10^{22} and 8.2×10^{23} impurities per m^3, respectively. Sample (3) shows metal-like behavior. (After P.W. Li et al., Phys. Rev. B 6, 442 (1972).)

We start by discussing the temperature range below the mobility maximum, where scattering by dopant atoms dominates. Noting that the impurities are charged, we need

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

to treat the scattering process as *Rutherford scattering*.[3] According to the well-known formula from nuclear physics, the total scattering cross section is $\sigma_{st} \propto v^{-4}$, which results in the expression $l^{-1} = n_{st}\sigma_{st} \propto n_{st}/\bar{v}^4$ for the inverse mean free path. The mean velocity $\bar{v} = (3k_BT/m^*)^{1/2}$ can be expressed with the help of the kinetic gas theory, since the charge carriers in non-degenerate semiconductors obey Boltzmann statistics. Assuming for simplicity that the number of charged impurities is independent of temperature, we obtain:

$$\frac{1}{\tau} = \frac{\bar{v}}{l} \propto \frac{n_{st}\bar{v}}{\bar{v}^4} \propto \frac{n_{st}}{T^{3/2}} \qquad \text{and} \qquad \mu = \frac{e\tau}{m^*} \propto \frac{T^{3/2}}{n_{st}} . \tag{10.32}$$

The actual number of charged scattering centers depends on the temperature and the details of the compensation effect, which are not known for the measurements shown in Figure 10.16. The assumption of a constant number of scattering centers is therefore questionable. In fact, the observed temperature dependence of the mobility considerably deviates from the prediction (10.32). However, we will not pursue this aspect, but simply use the data for estimating the scattering cross section of aluminum ions. If we take the mobility of sample (2) at 30 K, we find a value of $\mu \approx 500\ \text{cm}^2/\text{Vs}$. With (7.11) and (9.22), this gives a value of $\sigma_{st} \approx 10^{-11}\ \text{cm}^2$ for $\sigma_{st} = e/(n_{st}m^*\mu\bar{v})$. The ions thus have a scattering cross-section diameter of about 300 Å, giving rise to extraordinarily strong scattering.[4]

Above the mobility maximum, scattering processes with acoustic phonons dominate. To derive the temperature dependence of the mobility in this range, we again use the relation between the mean free path and the scattering cross section. At room temperature phonons with Debye frequency ω_D dominate, so that we can assume a temperature-independent scattering cross section σ_{st}. According to equation (6.100) the phonon density $n_{ph} \propto T$. Taking into account $\bar{v} \propto T^{1/2}$ we therefore obtain

$$\frac{1}{\tau} = \frac{\bar{v}}{l} = \bar{v}n_{ph}\sigma_{st} \propto \bar{v}n_{ph} \propto T^{3/2} \qquad \text{and} \qquad \mu \propto T^{-3/2} . \tag{10.33}$$

A look at Figure 10.16 shows good agreement with the experimental results. The special metal-like character of the charge transport in the heavily doped samples (3) is very clearly visible at low temperatures where the mobility has an almost temperature-independent high value.

Finally in this section we return to the conductivity, whose temperature curve is shown in Figure 10.17 for n-doped germanium. Starting from low temperatures, the conductivity increases exponentially with increasing carrier concentration. The temperature dependence of the mobility and the effective density of states plays a minor role. In the saturation regime, where the carrier concentration is approximately

3 Ernest Rutherford, *1871 Brightwater, New Zealand, †1937 Cambridge, Nobel Prize 1908

4 In metals, charged point defects scatter much less effectively because the Coulomb field is very strongly shielded by the high concentration of free electrons (see Section 8.3).

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

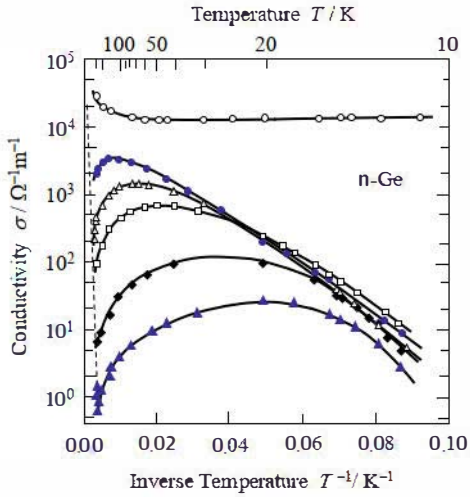


Fig. 10.17: Electrical conductivity of n-germanium. The samples are identical to those used in the determination of the charge carrier density of Figure 10.14. The sample with the highest conductivity has a metal-like character. (After E.M. Conwell, Proc. I.R.E., 1327 (1952).)

constant, the conductivity decreases again with increasing temperature due to the decrease in mobility. At high temperatures, intrinsic conductivity sets in as clearly visible for the lowest doped sample.

A special case is the sample (●) with the high doping level of $n_D = 10^{24}\,m^{-3}$, whose conductivity changes only slightly. Figure 10.14 shows that the concentration of the charge carriers in this sample is largely independent of temperature. We have already discussed the cause of this above.

Ohm's law (9.21) is only valid as long as the mobility is not affected by the magnitude of the electric field. In fact, for technically important semiconductors, the mobility begins to decrease at field strengths above $10^5\,V/m$. The drift velocity $v_D = \mu\mathcal{E}$ approaches a limiting value of about $10^5\,m/s$. The reason for this is the electron-phonon coupling. At high field strengths, the energy gain of the charge carriers in the electric field is sufficient to generate optical phonons. Since this process is extremely effective owing to the high density of states of the optical phonons, it limits the drift velocity of electrons and holes. With direct semiconductors, further effects also contribute but we will not go into them here in detail. However, this limitation of the drift velocity is very relevant in state-of-the-art devices, since field strengths above $10^7\,V/m$ occur due to their usually small dimensions.

10.3 Amorphous Semiconductors

In our previous discussion we have repeatedly stated that crystalline and amorphous solids have many similar properties, but also differ greatly in some aspects. The last remark also applies to the electrical properties of amorphous semiconductors, which we now discuss.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As we have seen, bands evolve from atomic levels, with the band structure being primarily determined by the short-range order. Since this differs only slightly between crystalline and amorphous phases, bands and band gaps are present in both cases. An important structural difference is that in the amorphous phase, small variations in bond length and bond angle occur. The band gap, which depends on the distance to the nearest neighbors, is therefore "smeared", which results in the exponential tails of the density of state. Since these states occur infrequently, they are also spaced far apart, and their wave functions do not overlap, so that electrons in these states are localized. This fact can be taken into account in the simplest approximation by assuming, as suggested by *N.F. Mott* and indicated in Fig. 10.18, that there are *localized* and *delocalized* states, which can be separated from each other by a **mobility edge**.

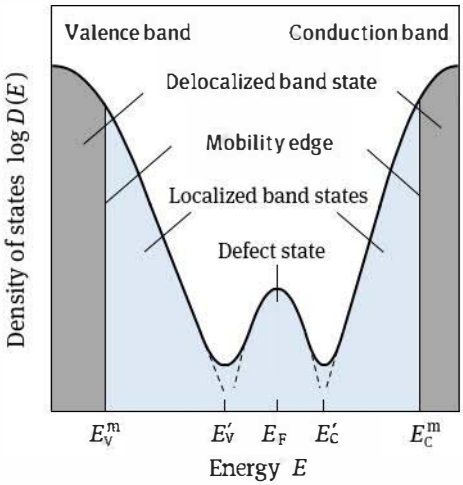


Fig. 10.18: Schematic representation of the density of states of amorphous semiconductors. Delocalized states are indicated in grey and localized states in light blue. The quantities E_V^m and E_C^m mark the mobility edges, and E_V' and E_C' the edges of the localized band states.

The structural disorder has another important consequence, the existence of defect states *in* the band gap. A very important defect of this kind, the unsaturated chemical bond, has already been discussed in Section 5.3. Since these are neither binding nor anti-binding states, they are located in energy approximately at the center of the energy gap. Owing to the variation in their environments, the energy of these defects has a broad distribution. As we will see, they strongly influence the location of the Fermi level and thus the electrical conductivity.

At this point, it is useful to add a brief remark on the *optical absorption* since the absorption is sensitive to the energy gap. Here it is of great importance that owing to the irregular arrangement of the atoms, quasi-momentum conservation is lifted in amorphous semiconductors. Therefore, optical transitions from the valence band to the conduction band do not necessarily occur vertically in momentum space, i.e. while maintaining the wave number of the excited electron, even if no phonons are involved in the absorption process (cf. Section 10.1). Since transitions between localized band

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

states also occur, absorption already starts at energies which are smaller than the energy gap of crystalline semiconductors of the same composition. The various optical properties of amorphous and crystalline silicon is shown in Figure 10.19, which depicts the imaginary part ε'' of the dielectric function.[5] In the amorphous phase, the structure of the absorption curve for the crystalline sample is broadened and the absorption clearly starts at lower photon energies. One result is that thin layers of amorphous silicon are much less transparent than similar layers of crystalline materials. This phenomenon is exploited in the construction of solar cells (see Section 10.5), because when amorphous rather then crystalline silicon is used, the sunlight is already absorbed at longer wavelengths and by much thinner layers.

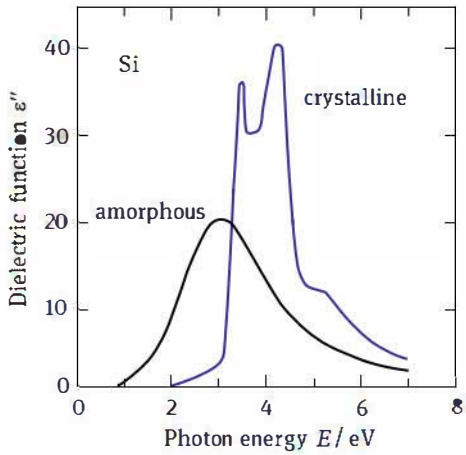


Fig. 10.19: Imaginary part ε'' of the dielectric function of amorphous and crystalline silicon as a function of the energy of the incident photons. (After J. Stuke, *Proc. 10th Int. Conf. Physics of Semiconductors*, S.P. Keller et al., eds., US Atomic Energy Comm. Washington, (1970).)

10.3.1 Electrical Conductivity

In this Section we consider the intrinsic conductivity of *undoped* amorphous semiconductors, for example amorphous silicon. Since electrons are the main contributors to the conductivity in these materials, we can leave holes out of the discussion. Furthermore, we assume that the Fermi level is independent of temperature and is situated at the peak of the defect density of states. The reasons for this assumption will be explained at the end of the section. First of all, we must take into account that electrons in amorphous semiconductors are not described by Bloch waves, since these require a periodic lattice. Electrons are therefore already relatively strongly scattered in ideal,

5 ε'' is linked to the optical extinction coefficient κ via the relation $\varepsilon'' = 2\kappa n'$, where n' is the refractive index. As explained in Section 13.2, the quantity ε'' reflects the density of states of the bands.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

defect-free, amorphous networks. As a result, the mobility of charge carriers in amorphous materials in comparison with that of crystalline materials is greatly reduced. Amorphous semiconductors therefore have a comparatively low intrinsic conductivity at high temperatures.

Depending on the temperature, various mechanisms contribute to charge transport. At *high temperatures* charge transport takes place in the delocalized band states sitting above the mobility edge E_C^b. The transport mechanism is not fundamentally different from the intrinsic conduction of crystalline semiconductors. Accordingly, we would expect an exponential relation between the conductivity and the temperature of the form

$$\sigma = \sigma_0 e^{-(E_C^b - E_F)/k_B T} . \tag{10.34}$$

As with crystalline semiconductors, the exponential factor reflects the temperature dependence of the charge-carrier concentration. The influence of temperature on the mobility of the delocalized electrons is comparatively small and is therefore neglected in the above equation. The mobility can be determined from the experimental value of the constant σ_0, if the density of states at the mobility edge is known. Typical values fall in the range of $1 - 10\,\mathrm{cm}^2/\mathrm{Vs}$ and are thus smaller by a factor of 1000 than in crystals.

At *room temperature* only the localized band states are occupied. Charge transport now takes place by electrons jumping from one localized state to another. This transport mechanism is called the **hopping conductivity**. During this process electrons absorb thermal energy and tunnel to localized states in the neighborhood This process is explained in more detail in Figure 10.20, showing schematically the "potential landscape" at the band edge.

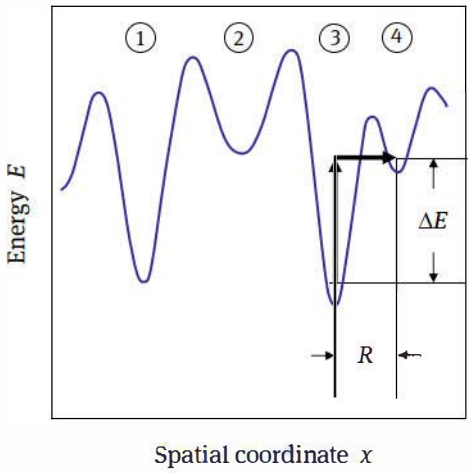


Fig. 10.20: Hopping in amorphous semiconductors. The electron in well 3 is thermally excited and tunnels to the nearest potential minimum 4. Here R is the distance of the nearest state and ΔE is the activation energy for the marked jump.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

A jump from one potential well to the next can be broken down into two steps: first, the electron absorbs thermal energy ΔE from the phonon bath, which is required to overcome the energy difference between the two states. The probability for this is given by the Boltzmann factor. This is followed by the tunnelling process, where the overlap of the wave functions in the two wells is the determining factor. The probability that a jump occurs can be expressed by the jump rate ν, which can be approximated by

$$\nu = \nu_0 \, e^{-\Delta E/k_B T} \, e^{-2R/\alpha} . \tag{10.35}$$

We already know the first two factors on the right-hand side from the discussion of thermally activated diffusion of atoms in Section 5.1. The *attempt frequency* ν_0, with which the electron runs against the potential barrier, is $10^{13}\ s^{-1}$. This is comparable to that of atomic diffusion. In fact, it is a typical lattice frequency, because the motion of the electron in the potential well is generated by the motion of the neighboring atoms. The factor $\exp(-2R/\alpha)$ describes the overlap of the wave functions between the neighboring states separated by the distance R and is therefore a measure of the tunnel probability. The wave function of the localized electrons can be approximately represented by $\psi \propto \exp(-r/\alpha)$, where α is the *localization length* of the electrons.[6] Since the depth of the potential wells varies from case to case, the individual activation energies are replaced by the mean value $\overline{\Delta E}$. The tunnelling probability does not depend on the temperature and can therefore be added to the pre-factor in the expression for the hopping conductivity. Formally, this expression now has the appearance of the conductivity (10.34) for the delocalized states. Instead of E_C^m, however, the smaller activation energy $(E_C' + \overline{\Delta E})$ is involved, and the pre-factor σ_0 containing the mobility, is also greatly reduced.

If we use typical values for the occurring variables, we find values for the mobility of around $0.01\ cm^2/Vs$. The conductivity is thus significantly lower than that for charge transport in delocalized states. In amorphous semiconductors not only does the charge carrier concentration decrease strongly with temperature, but also the mobility, since the transport mechanisms differ in the different temperature ranges.

At *low temperatures*, the charge transport takes place in the immediate vicinity of the Fermi level, i.e. in the tails of the localized band states or, in other words, in the localized defect states. At low temperatures, however, the energy difference between spatially adjacent states can be much greater than the thermal energy and can therefore hardly be surmounted. For the electrons it is then more advantageous to tunnel to states at a greater distance, but with energies closer to those of the initial defect, than to its immediate neighbors. This kind of charge transport is called **variable range hopping**.

To illustrate this, refer back to Figure 10.20 and the "potential landscape" portrayed there to describe the defect states. At low temperatures, the electron in well 3 would most likely jump to well 1 and not to well 4.

6 The expression (10.35) is formally identical to (5.4), since $\nu_0 \exp(-2R/\alpha)$ can be reduced to a temperature-independent constant.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The probability that a jump from a state with energy E_i to a state with energy E_j will take place is given by (10.35), where ΔE stands for $\Delta E = (E_j - E_i)$ and R is the distance between the two states. For a jump to occur, at least *one* state with the energy difference ΔE must be present in the volume $(4\pi/3)R^3$. Since the number of these states in this volume is given by $(4\pi/3)R^3 D(E_F)\Delta E$, we can replace ΔE in (10.35) by $\Delta E = [(4\pi/3)R^3 D(E_F)]^{-1}$ to obtain:

$$\nu = \nu_0 \exp\left\{-2R/\alpha - [4\pi/3R^3 D(E_F)k_B T]^{-1}\right\} . \tag{10.36}$$

We find the most likely jumps by looking for the maximum rate ν with respect to distance. Assuming that the density of states at the Fermi edge is approximately constant, $\partial\nu/\partial R = 0$, results in the condition $2/\alpha = 9/4\pi R^4 D(E_F)k_B T$. We insert this result into (10.36) and obtain for the conductivity, which is proportional to the hopping rate, the relation:

$$\sigma = \sigma_0 \exp\left[-\left(\frac{T_0}{T}\right)^{1/4}\right] \qquad \text{with} \qquad T_0 \approx \frac{2.06^4}{\alpha^3 D(E_F)k_B} , \tag{10.37}$$

where T_0 is a constant containing the important parameters, the density of state at the Fermi energy, and the localization length. The power in the exponent depends on the shape of the defect density of states at the Fermi level. Only if the density of states at the Fermi level is approximately constant and the sample under investigation is three-dimensional do we obtain the numerical value 1/4 for the exponent. Figure 10.21 shows the resistance $R_{Si} \propto 1/\sigma$ of an amorphous silicon film produced by sputtering. The resistance of the sample shows the expected temperature dependence.

With this we have given a brief insight into the mechanisms of charge transport in amorphous semiconductors. Except at high and low temperatures, any experimental separation of the various transport mechanisms proves to be extremely difficult, since the variations are not sharp, but rather smooth. A further complication is that the properties of the samples not only depend on their composition, but in many cases on the sample preparation which also has a strong influence on the experimental result.

10.3.2 Defect States

The discussion of electrical conductivity and optical absorption of amorphous semiconductors makes clear that defects already occur in pure materials, i.e. in the absence of doping. As we already mentioned in Section 5.3, the nature and properties of these defects depend on the coordination number of the material under consideration. Defects in tetrahedral semiconductors such as a-Si behave differently from those in chalcogenide glasses such as a-As_2S_3 or in a-Se. In pure amorphous silicon, for example, about $10^{25} - 10^{26}$ m^{-3} dangling bonds can be found. These can be easily detected as localized, unpaired spins using electron spin resonance. However, in chalcogenide glasses the number of unpaired electrons is below the detection limit for this measurement technique.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

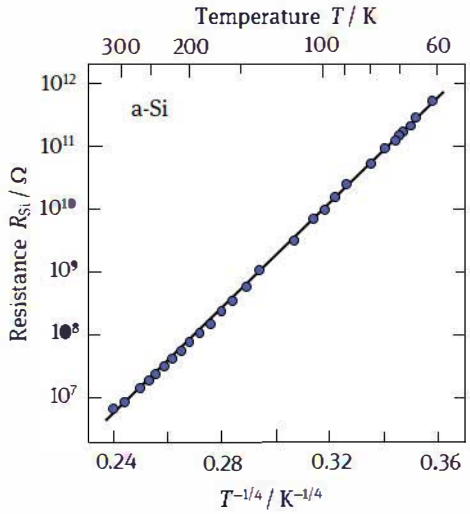


Fig. 10.21: Resistance of a 2.16 μm thick amorphous silicon film as a function of $T^{-1/4}$. The film was produced by sputtering at room temperature. (After J.J. Hauser, Phys. Rev. B 8, 3817 (1973).)

Open, unsaturated bonds are the typical defects in amorphous semiconductors. As long as the environment of the defects can be considered rigid, the states are expected to lie at the middle of the energy gap, since neither a binding nor an anti-binding orbital exists. By *relaxation*, i.e. by rearrangement of the neighboring atoms, the energy of these states is lowered. Uncharged defects have the peculiarity that they can act both as donors and as acceptors, since they can give off the non-binding electron or take up an additional electron. Under "normal" circumstances, i.e. with the usual impurities in crystalline semiconductors, the energy of the state is strongly increased in the second case, since the two electrons on the defect are in the same place and repel each other. The energy difference between the neutral and the negatively charged state is called the **correlation** or **Hubbard[7] energy** U and is typically of the order of one electron volt in crystalline semiconductors. When discussing the impurities in crystalline semiconductors we have therefore assumed that the doubly charged states are located in the conduction band and need not be considered further. In amorphous solids, local structural changes are relatively easy to achieve. The structure around a charged defect is capable of considerable changes which can thus significantly reduce the energy of the defect. Defects that are occupied by two electrons therefore play an important role in amorphous semiconductors.

As far as the correlation energy U is concerned, there are two possibilities. Either the energy of the doubly charged state lies above that of the neutral or positively charged state or it lies below it. In the former case, the correlation energy is positive, as in a crystal, although much smaller, and in the latter case it is negative. In the simple concept discussed here, as illustrated schematically in Figure 10.22, there are two

7 John Hubbard, *1931 London, †1980 San José

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

maxima in the density of states within the energy gap. The neutral defect states S^0 with an unpaired electron sit at energy E_S. These cause the strong electron spin resonance signal. If such a defect gives up its unpaired electron, the state S^+ can be created with no significant shift in energy. For simplicity, we assume that S^0 and S^+ have the same energies. The emitted electron is picked up by another defect whereby a defect S^- is being formed. Whereas in our simple picture the energy of S^+ is still at E_S, the energy of S^- is at $E_S + |U|$ or $E_S - |U|$, depending on whether the correlation energy is positive or negative. In the folowing discussion, we must now distinguish between the two cases, since different consequences result.

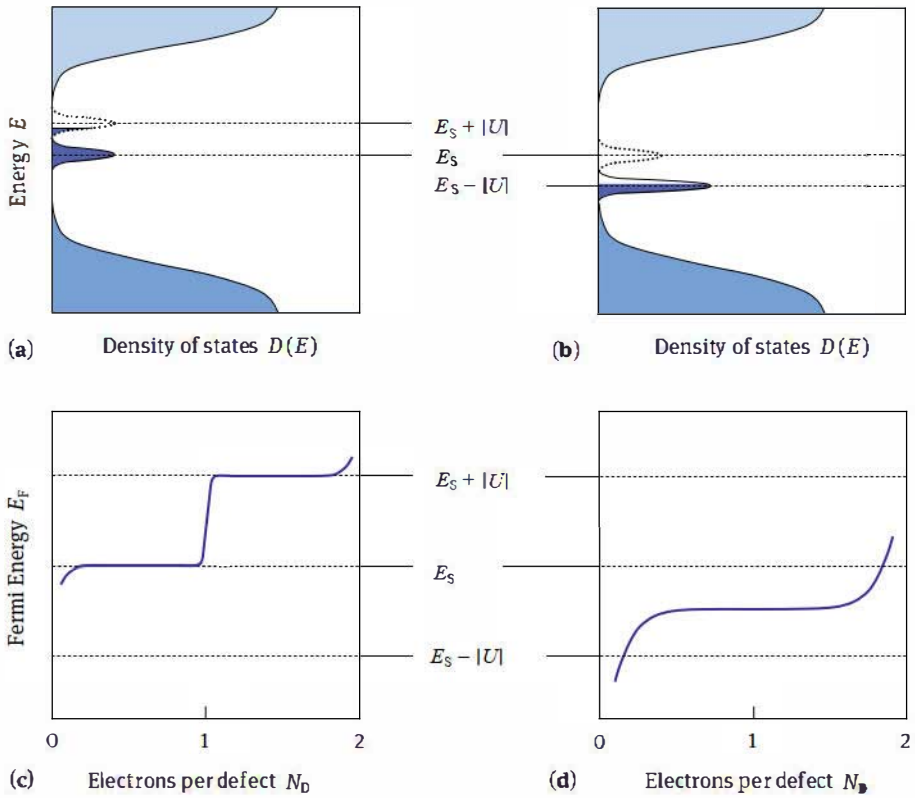


Fig. 10.22: The occupation of defect states and their influence on the position of the Fermi level. The occupation of the states depicted roughly corresponds to a mean electron density of $N_D = 1.2$ per atom. **a)** Occupation with positive and **b)** with negative correlation energy, **c)** position of the Fermi level with positive and **d)** with negative correlation energy. Occupied states are indicated in dark blue. The dotted line in (a) represents the density of states of the double occupied states, in (b) the density of states of the single occupied states.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Let us first consider amorphous semiconductors such as a-Si or a-Ge, in which the correlation energy has a positive sign, as shown in Figure 10.22a. The S^--states consequently lie above the neutral or positively charged defect states. The electron transfer that leads to paired electrons costs energy and therefore normally does not take place, so only $S^•$-defects are found. This makes the observed strong electron spin resonance signal of the tetrahedron-coordinated semiconductors understandable. If we try to dope these amorphous semiconductors, the numerous intrinsic defects prevent any noticeable shift of the Fermi level. Figure 10.22c shows the behavior of such semiconductors when doped as a function of the mean number of electrons $N_•$ per defect. The starting point is the intrinsic state with electron number $N_• = 1$, in which only S^0-defects are present. If acceptors are introduced, the average number of electrons per defect falls, but this does not change the position of the Fermi level, since the S^+- and $S^•$-defects have approximately the same energy. The S^0-defects act as donors, which results in a compensation effect.

If we now dope the amorphous semiconductor with donors and increase the number of electrons in the defect states above the value $N_• = 1$, then the doubly occupied states become populated, i.e. S^--defects are formed. The Fermi energy "jumps" from E_S to $E_S + U$, but owing to the small correlation energy, this means only a slight shift of the Fermi level. The high concentration of intrinsic impurities and their position near the center of the energy gap means that doping cannot have a significant effect on the Fermi level! The number of free charge carriers cannot be changed substantially by the concentration of the dopant atoms as happens in crystalline semiconductors.

It should be emphasized that we have mixed the one- and two-electron states here in the representation of the density of states in Figure 10.22. The upper states, occupied by two electrons, can only exist when the lower ones are already full. They are therefore only created by the occupation of defects with the second electron. Their density is shown in Figure 10.22 as dotted lines. With the appearance of doubly occupied states, the single occupied states disappear. With $N_• = 2$ the maximum at E_S would have completely disappeared in this representation.

A targeted and effective doping of amorphous silicon can only be achieved if the defect states are neutralized, i.e. made "harmless". This can be achieved by the chemical saturation of the dangling bonds. In the case of amorphous silicon, for example, the film is deposited in an atmosphere containing hydrogen, so that hydrogen atoms can bond to the unsaturated defects (see Figure 5.27). Since silicon and hydrogen form a strong bond, the silicon-hydrogen defect states that now appear are located deep in the valence band, so that these defects have no influence on the electrical properties of the sample. The number of remaining unsaturated bonds can be varied via the hydrogen concentration, which depends on the manufacturing conditions. At a concentration of about 10%, the signal from the electron spins disappears and the optical properties change. Amorphous silicon prepared in this way can be used for electronic components. In the manufacture of solar cells, for example, a-Si has the advantage that considerably less material is required than for cells made of crystalline

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

silicon, since the optical absorption (see Figure 10.19) in the spectral range of sunlight is higher than for crystalline silicon. Solar cells made of a-Si are relatively cheap, but their efficiency is much lower than that of crystalline cells. They are currently mainly used in pocket calculators, watches or toys. More recently, so-called a-Si/c-Si hetero solar cells with very high efficiency have been developed, based on a combination of amorphous and crystalline silicon.

Of course, not all defects have the same energies, because owing to the different defect environments, the structural relaxation is also differently effective. This leads to a broadening of the densities of states, which for the sake of clarity were drawn very narrowly in Figure 10.22. Frequently the two maxima overlap as in the schematic diagram in Figure 10.18 and fill a substantial part of the energy gap. It is also remarkable that the doping process is less effective in amorphous semiconductors than in crystalline materials. This is due to the fact that the dopant atoms, which have a different valence than the host material, do not necessarily assume the host coordination when incorporated into the amorphous network. Thus, phosphorus in amorphous silicon may well have three or five neighbors. In order to achieve a noticeable doping effect, much higher concentrations of the doping material are required in amorphous semiconductors.

For chalcogenide glasses, the correlation energy is negative. We have already given a vivid explanation for this in Section 5.3 when discussing defect configurations in amorphous solids. In chalcogenide glasses, a charged defect pair S^+ and S^- is formed from two neutral S^0-defects with a gain in energy but leaving no unpaired electrons. In Section 5.3 we called these defects $C_1^{\bullet}$, C_1^- and C_3^+. In the intrinsic state, i.e. at $N_{\blacksquare} = 1$, there are therefore no S^0-defects present. This makes clear why no localized spins are observed in these materials.

The negative correlation energy has an interesting consequence for the dopability of semiconductors, which is explained in Figures 10.22b and 10.22d. If we introduce two electrons into the semiconductor, the first one occupies the state at E_S, the second one causes a shift of the state to the lower energy $E_S - |U|$. So the states are always occupied in pairs, so that no surplus of occupied states occurs. In our drawing this density of states is therefore shown by dots. The density of states at $E_S - |U|$ was drawn larger, because there is room for two electrons per defect. The Fermi level therefore always lies between the energies of the two states and is thus fixed. There are practically no unsaturated spins that can be detected in electron spin resonance measurements. Without doping, i.e. in the intrinsic state with $N_{\blacksquare} = 1$, defects already exist with two electrons. Only when all S^--states are occupied, can the Fermi level be shifted.

As we have seen, in amorphous semiconductors isolated unsaturated bonds are of great importance. Using the simple example of amorphous Se, we have already discussed the formation of defect states in Section 5.3. This approach is useful for a basic understanding, even with negative correlation energy. In systems like chalcogenide glasses, the defects are more complicated. However, we will not pursue this aspect further here.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

10.4 Inhomogeneous Semiconductors

In this section, we discuss the basic physical principles necessary for understanding the technical applications of semiconductors. In this context we speak of **inhomogeneous semiconductors** whose doping and/or chemical composition varies spatially. The starting point of our discussion will be the p-n junction. This will be followed by a short introduction to the physics of heterostructures and superlattices. At the end of the chapter we will choose some simple, interesting devices from among the large number of electronic components and explain their mode of operation.

10.4.1 p-n Junction

For the technical application of semiconductors in solid-state electronics, a spatial variation of the impurity concentration is an indispensable prerequisite. The desired variation can be achieved by diffusing or implanting the dopant atoms into certain regions of the starting material. For the lateral confinement of the doping with dimensions in the micro- and nanometer ranges, various methods are applied. By far the most important is *photolithography*. We will not go into detail about this extremely demanding technique, since we are dealing with the basic understanding of solid state physics here .

The p-n Junction in Equilibrium. In the following discussion we will assume *abrupt transitions* in the changes in doping. The technical realization of an idealized step-like transition is, of course, only approximately possible. Figure 10.23a shows the initial

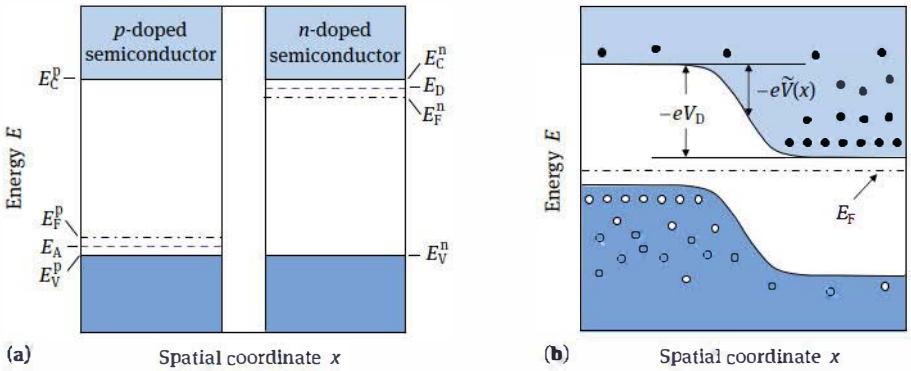


Fig. 10.23: Position of the impurity levels, the band edges and the Fermi level in the p-n transition. **a)** Position of the energy levels in separate p- and n-doped crystals. **b)** p-n transition in equilibrium. The band edges are shifted against each other by the diffusion voltage V_0. The points indicate electrons, the open circles indicate holes.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

materials. E_C^p and E_C^n indicate the position of the conduction band edges, E_V^p and E_V^n the edges of the valence bands in the p- and the n-conductor and E_F^p and E_F^n the energy of the Fermi levels in the two types of crystal. At room temperature, depending on the type of doping, the Fermi level lies just below the donor or above the acceptor level. If p-doped and n-type semiconductors are brought into contact, the Fermi levels on the two sides must be the same, this is achieved by the diffusion of charge carriers from the respective areas of high to low concentration, i.e. electrons diffuse from the n- to the p-type semiconductor and vice versa for the holes. As shown in Figure 10.23b, the so-called **diffusion voltage** is built up, which shifts the chemical potentials on the two sides to bring the Fermi levels into coincidence, i.e. that $E_F^p = E_F^n$. This forces a local bending of the bands , which in turn gives rise to currents that counteract the diffusion. The shape of E_C causes electrons to flow to the right and the shape of E_V causes a flow of holes to the left. The profile of the band can be described by means of a position-dependent, one-dimensional macroscopic potential $\tilde{V}(x)$, which we will now discuss in more detail.

We will first calculate the diffusion voltage V_D, i.e. the potential difference between the differently doped semiconductors, which is characteristic of p-n junctions. It is determined by the difference of the original Fermi levels and thus is essentially equivalent to the band gap of the semiconductor under consideration. From the fact that the Fermi levels equalize, it follows for the diffusion voltage V_D, which builds up between the two differently doped semiconductors in the saturation regime, that:

$$\begin{aligned} eV_D = E_F^n - E_F^p &= E_C - k_B T \ln\left(\frac{\mathcal{N}_C}{n_D}\right) - E_V - k_B T \ln\left(\frac{\mathcal{N}_V}{n_A}\right) \\ &= E_g - k_B T \ln\left(\frac{\mathcal{N}_C \mathcal{N}_V}{n_D n_A}\right) = k_B T \ln\left(\frac{n_D n_A}{n_i^2}\right) . \end{aligned} \tag{10.38}$$

A look at the result tells us that to first approximation $eV_D \approx E_g$.

Charge carriers located in regions where the doping corresponds to that type of carrier, i.e. electrons in an n-region and holes in a p-region, are called *majority charge carriers*. If they diffuse into the oppositely doped region, they are then known as *minority charge carriers*. In the following, we label the charge carrier concentration with an index indicating the region in which they are situated. For the majority charge carriers we therefore write n_n and p_p, and for the minority carriers n_p and p_n.

At greater distance from the transition region: $n_n \approx n_D^+ \simeq n_D$ and $p_p \approx n_A^- \simeq n_A$. The law of mass action $n_p p_p = n_n p_n = n_i p_i$ = constant ensures that at the usual doping concentrations the majority charge carrier density is much higher, and the minority charge carrier density is much lower, than the concentrations of the intrinsic charge carriers.

The charge carrier density changes very rapidly in the transition region. If we select the boundary between the n- and p-doped material as the coordinate origin and, as in Figure 10.23, denote the potential as $\tilde{V}(x)$, then we see directly that the profile of the conduction band edge is given by $E_C(x) = [E_C^p - e\tilde{V}(x)]$. We find a corresponding

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

expression for the valence band edge. Thus, taken with the two equations (10.10) and (10.11), we find the density of the free charge carriers to be:

$$n(x) = n_p e^{e\widetilde{V}(x)/k_B T} \qquad \text{and} \qquad p(x) = p_p e^{-e\widetilde{V}(x)/k_B T} . \tag{10.39}$$

The spatial dependence of $n(x)$ and $p(x)$ is shown schematically in Figure 10.24 on a logarithmic scale. Here it is assumed that the semiconductors are in the saturation regime, with the impurities completely ionized.

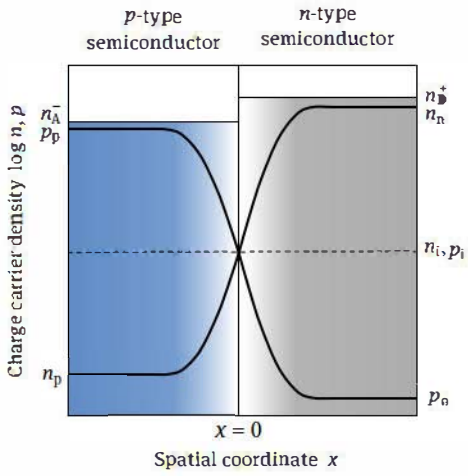


Fig. 10.24: Concentration of free charge carriers and ionized impurities at the p-n junction on a logarithmic scale spanning several orders of magnitude. It is assumed that the donor density shown in grey exceeds the blue density of the acceptors. The depletion zone is indicated schematically by lighter colors.

The figure shows that the number of free charge carriers, following the law of mass action $n(x)p(x) = \text{constant}$, is strongly reduced in the immediate vicinity of the transition. This area is therefore also called the **depletion zone**. Here, the charge of the ionized donors or acceptors is no longer compensated by free charge carriers, which results in the build-up of a space-charge. For the **space-charge density** we can write

$$\varrho(x) = e[n_D^+ - n_n(x) + p_n(x)] \qquad x > 0 \quad \text{n-region}, \tag{10.40}$$

$$\varrho(x) = -e[n_A^- + n_p(x) - p_p(x)] \qquad x < 0 \quad \text{p-region}. \tag{10.41}$$

The potential $\widetilde{V}(x)$ and space-charge density $\varrho(x)$ are connected to each other via the Poisson equation:

$$\frac{\partial^2 \widetilde{V}(x)}{\partial x^2} = -\frac{\varrho(x)}{\varepsilon_r \varepsilon_0} . \tag{10.42}$$

Difficulties arise in solving this equation, because the space-charge density depends on the potential curve itself, meaning that the solution would have to be determined self-consistently. We use the fact that the concentration of free charge carriers in the space-charge zone (cf. Figure 10.24) is very low and to first approximation can be neglected. The actual course of $\varrho(x)$ can therefore be replaced by a step function as

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

shown in Figure 10.25. This simplification is called the **Schottky model** of the space-charge zone. Since this zone must be neutral overall, the two rectangles have the same area. For the space-charge density we will therefore use the following expression:

$$\varrho(x) = \begin{cases} 0 & \text{for} \quad x < -d_\mathrm{p} \,, \\ -en_\mathrm{A} & \text{for} \quad -d_\mathrm{p} < x < 0 \,, \\ en_\mathrm{D} & \text{for} \quad 0 < x < d_\mathrm{n} \,, \\ 0 & \text{for} \quad x > d_\mathrm{n} \,. \end{cases} \tag{10.43}$$

Here, d_n and d_p are the thicknesses of the respective space-charge zones, which we will calculate in the following.

Figure 10.25 shows schematically the distribution $\varrho(x)$ of the space-charge and the resulting spatial dependence of the field strength $\mathcal{E}_x$ and the potential $\widetilde{V}(x)$.

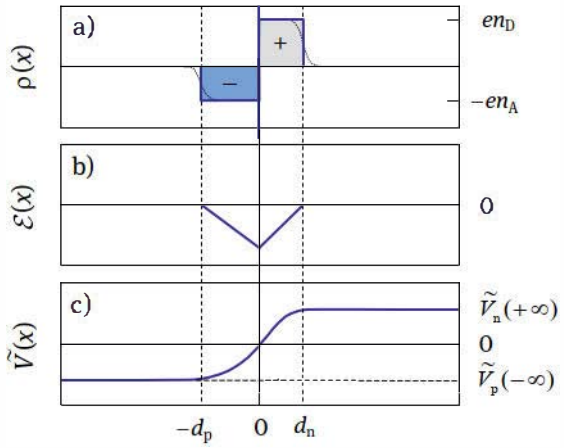


Fig. 10.25: The Schottky model of the space-charge zone. **a)** Rectangular approximation. The more realistic curve is shown as a dotted line. **b)** The spatial dependence of the electric field strength $\mathcal{E}(x)$ and **c)** of the potential $\widetilde{V}(x)$.

When assuming rectangular space-charge zones, the Poisson equation can be integrated separately for each. For the n-conducting part, i.e. over the interval $0 < x < d_\mathrm{n}$, this results in:

$$\frac{\partial^2 \widetilde{V}(x)}{\partial x^2} \approx -\frac{en_\mathrm{D}}{\varepsilon_\mathrm{r}\varepsilon_0} \,, \tag{10.44}$$

$$\mathcal{E}_x = -\frac{\partial \widetilde{V}(x)}{\partial x} = -\frac{en_\mathrm{D}}{\varepsilon_\mathrm{r}\varepsilon_0}(d_\mathrm{n} - x) \,, \tag{10.45}$$

$$\widetilde{V}(x) = \widetilde{V}_\mathrm{n}(\infty) - \frac{en_\mathrm{D}}{2\varepsilon_\mathrm{r}\varepsilon_0}(d_\mathrm{n} - x)^2 \,. \tag{10.46}$$

There are corresponding equations for the p-conducting part of the p-n junction.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The thicknesses d_n and d_p of the space-charges can be calculated from the neutrality condition $n_D d_n = n_A d_p$ and the continuity condition for $\bar{V}(x)$ at point $x = 0$,

$$\frac{e}{2\varepsilon_r\varepsilon_0}(n_D d_n^2 + n_A d_p^2) = \bar{V}_n(\infty) - \bar{V}_p(-\infty) = V_D \tag{10.47}$$

yielding:

$$d_n = \sqrt{\frac{2\varepsilon_r\varepsilon_0 V_D}{e}\,\frac{n_A/n_D}{n_A + n_D}} \quad \text{and} \quad d_p = \sqrt{\frac{2\varepsilon_r\varepsilon_0 V_D}{e}\,\frac{n_D/n_A}{n_A + n_D}}\,. \tag{10.48}$$

Putting in typical values for the band gap $eV_D \approx E_g \approx 1\,\text{eV}$ and an impurity concentration $n_A \approx n_D \approx 10^{20} \ldots 10^{24}\,\text{m}^{-3}$ we find for the thickness of the space-charge zone $d_n \approx d_p \approx 1\,\mu\text{m} \ldots 10\,\text{nm}$ and thus electric field strengths in the range $\mathcal{E} \approx 10^6\,\text{V/m} \ldots 10^8\,\text{V/m}$.

The mode of operation of devices based on p-n junctions becomes clearer when we examine the currents involved more closely. The so-called **diffusion current** is a consequence of the difference in the charge carrier concentration on the two sides of the p-n junction. In the literature this current is often referred to as the **recombination current**. This term indicates that after passing through the space-charge zone, the charge carriers recombine with the oppositely charged charge carriers present in large numbers on the other side of the junction. The diffusion current causes the diffusion voltage discussed above to build up. This voltage is associated with the electric field in the space-charge zone, as described by equation (10.45). This voltage leads to a **field current** which flows in the opposite direction to the diffusion current. This current also has various names in the literature, In this context it is called the drift or **generation current**, the latter name indicating that the minority carriers, which cause the field current, are being constantly generated anew.

In thermal equilibrium the two currents compensate each other. In other words, denoting the field current density as j^f and the diffusion current density as j^d, we have:

$$j^f + j^d = 0\,. \tag{10.49}$$

Both currents consist of both electrons and holes, so that we can write $j^f = (j_n^f + j_p^f)$ and $j^d = (j_n^d + j_p^d)$. The arguments for the disappearance of the total current also apply to the partial currents, since neither electrons nor holes accumulate in any region. From this follows:

$$j_n^f + j_n^d = 0 \quad \text{and} \quad j_p^f + j_p^d = 0\,. \tag{10.50}$$

Let us now take a closer look at the partial currents. We begin with the minority charge carriers which are drawn into the region of the majority charge carriers by the electric field in the p-n junction. Electrons from the p-region move into the n-region and holes move in the opposite direction. For simplicity, let us assume that the space-charge zone is thin and that the recombination rate there is small, so that we can neglect any recombination of electrons and holes in the space-charge zone. This means that

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

all charge carriers which reach the field of the space-charge region continue to move into the oppositely doped part of the semiconductor. Therefore, we assume that the diffusion length is large compared to the thickness of the space-charge zone, in which case the current flow does not depend on the potential.

We now consider the majority charge carriers, e.g. electrons from the n-region, which diffuse into the p-region, run up against the potential difference at the junction. The same applies in the opposite direction for the holes. The height of the barrier is given by eV_D, i.e. by the diffusion voltage. The Boltzmann factor determines the fraction of the "successful" charge carriers which can cross, so that the diffusion current is given by $j^d = a(T)\exp(-eV_D/k_BT)$. The pre-factor $a(T)$ is weakly dependent on temperature, but this dependence can be neglected in comparison with the exponential factor. Taking equation (10.49) into account, we find for the current densities:

$$|j^f| = |j^d| = a(T)\, e^{-eV_D/k_BT} . \tag{10.51}$$

The p-n Junction with an External Voltage. If an external voltage is applied to a p-n junction, the resulting current depends on its direction. The p-n junction can therefore be used as a current rectifier. An external electric voltage disturbs the equilibrium of field and diffusion current discussed above, so that equilibrium thermodynamics can no longer be applied to the system. If, however, the p-n junction is in a stationary state and not too far from thermal equilibrium, a relatively simple approach provides a good description: the applied voltage U primarily drops across the space-charge zone or depletion zone, which has a high resistance because there are only few free charge carriers present, the rest of the semiconductor being largely field-free. We can therefore write $[\tilde{V}_n(\infty) - \tilde{V}_p(-\infty)] = (V_D - U)$. We have chosen the sign of the applied voltage such that a positive voltage opposes the diffusion voltage and reduces the potential difference. With regard to the rectifier properties of p-n junctions, we speak of the **forward direction** when the p-type semiconductor has a positive and the n-type semiconductor a negative polarity. In the opposite case, the polarity is said to be in the **reverse direction**.

Figure 10.26 shows the effect of an external voltage on the bands and the Fermi level. Since the charge carriers in the space-charge zone are not in equilibrium, no common Fermi level can be defined. However, since the electrons are in equilibrium with each other, as are the holes, we can define two separate **quasi-Fermi levels** which can be treated independently. We will not go into this aspect in any further detail here, as it is irrelevant to the further discussion. However, an interesting consequence of different quasi-Fermi levels will be seen later in connection with the *semiconductor laser*.

What influence has the applied voltage on the two partial currents? To first approximation the field current is not influenced. Every charge carrier that comes into the region of the electric field at the junction, whether electron or hole, is pulled over, regardless of the strength of the field, and crosses the space-charge zone. In terms of

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

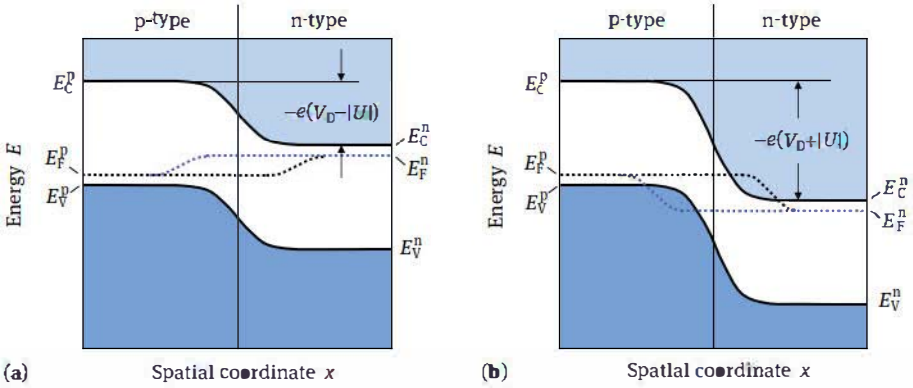


Fig. 10.26: A p-n junction with an applied external voltage. The quasi-Fermi level of the electrons is shown by the blue-dotted line, and that of the holes by the black-dotted line. **a)** Forward direction: the applied voltage reduces the potential difference. **b)** Reverse direction: the potential difference is increased by the applied voltage.

electrons we can therefore write:

$$j_n^f(U) \approx j_n^f(0) \,. \tag{10.52}$$

Since the barrier height changes with the applied voltage, the majority charge carriers no longer face the voltage V_D, but $(V_D - U)$. This changes the diffusion current. The two partial currents, diffusion and field current, no longer compensate each other. We can then write for the diffusion current:

$$j_n^d(U) = a(T)\, e^{-e(V_D-U)/k_B T} = j_n^d(0)\, e^{eU/k_B T} \,. \tag{10.53}$$

Noting that the field and diffusion currents flow in opposite directions, the resulting electron current density can be calculated by:

$$j_n(U) = j_n^d(U) - j_n^f = j_n^f\, e^{eU/k_B T} - j_n^f = j_n^f \left(e^{eU/k_B T} - 1\right) \,. \tag{10.54}$$

We have taken into account that with no voltage, the field and diffusion currents are equal in magnitude, i.e. $|j_n^d(0)| = |j_n^f(0)|$. Since both electrons and holes contribute to the charge transport, we obtain for the current density the final result:

$$j(U) = \left(j_n^f + j_p^f\right)\left(e^{eU/k_B T} - 1\right) = j_s \left(e^{eU/k_B T} - 1\right) \,, \tag{10.55}$$

where j_s represents the sum of the two field currents.

The current-voltage characteristics of a p-n junction is shown in Figure 10.27. The pronounced non-linear behavior gives the p-n junction the properties of a rectifier. If the voltage is applied in the forward direction, the height $e(V_D - U)$ of the potential barrier decreases and the current can increase "arbitrarily" with increasing voltage. However, if the polarity is reversed, at most the small field current flows. Any further

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

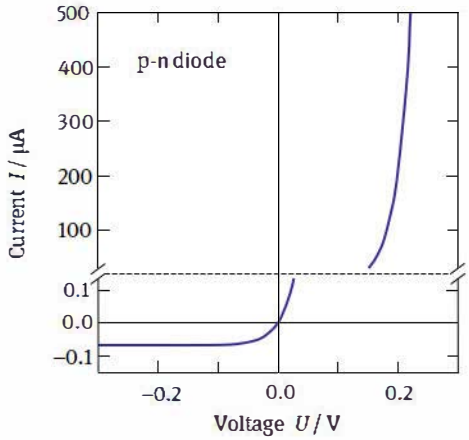


Fig. 10.27: The current-voltage characteristic of a p-n diode. Note the different scales in the reverse and forward directions.

growth in the current is not possible, because the generation of minority charge carriers, i.e. of electrons in the p-type semiconductor or of holes in the n-type semiconductor, cannot be changed.

Space-charge Capacity. The space-charge of the p-n junction is associated with a capacitance which can be influenced by an external voltage through the thickness of the space-charge zone. Using the Schottky model again, we can replace the diffusion voltage V_D by $(V_D - U)$ in equation (10.48). Taking the thicknesses without external stress as $d_n(0)$ and $d_p(0)$, we obtain:

$$d_n(U) = d_n(0)\sqrt{1 - \frac{U}{V_D}} \quad \text{and} \quad d_p(U) = d_p(0)\sqrt{1 - \frac{U}{V_D}}\,. \tag{10.56}$$

If the voltage dependence of the thickness of the space-charge zones is known, the capacitance C can be easily calculated. With the cross-sectional area A of the p-n junction, the stored-charge is $Q = en_D\, d_n(U)A$ and using (10.56), the capacity of the space-charge is thus given by:

$$C = \frac{dQ}{dU} = en_D A\left[\frac{d}{dU}d_n(U)\right] = \frac{A}{2}\sqrt{\frac{n_A n_D}{n_A + n_D}\,\frac{2e\varepsilon_r\varepsilon_0}{(V_D - U)}}\,. \tag{10.57}$$

It is a useful property that the capacitance of the space-charge zone depends on the number density of impurities, since the voltage dependence of the *junction capacitance* can be used for the experimental determination of the impurity concentrations. Furthermore, p-n junctions in electronic circuits often serve as adjustable capacitances. When used for this purpose they are known as *varactors* or *varicaps* and are used for tuning the the resonance in filter and oscillator circuits.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

10.4.2 Metal-semiconductor Junction

The contact between a metal and a semiconductor, i.e. a metal-semiconductor junction, is an important component in electronic circuits. In the ideal case, electrons can enter or leave the semiconductor unhindered. In practice, this simple case is rather the exception. Very often, instead of an *ohmic* contact, a *rectifying* junction occurs, which greatly impairs the flow of current. This is due to the different *work function* ϕ of the two materials, which is determined by the position of the Fermi level with respect to the vacuum level. In principle the following applies: if for an n-type semiconductor $\phi_{semi} > \phi_{metal}$, then the interface acts as an ohmic contact, whereas if $\phi_{semi} < \phi_{metal}$, then it is rectifying. In the case of a p-type semiconductor the reverse pertains.

We consider an n-type semiconductor with a work function greater than that of the metal. The initial situation without galvanic contact between the two materials is depicted in Figure 10.28a, which shows the position of the vacuum level, the band

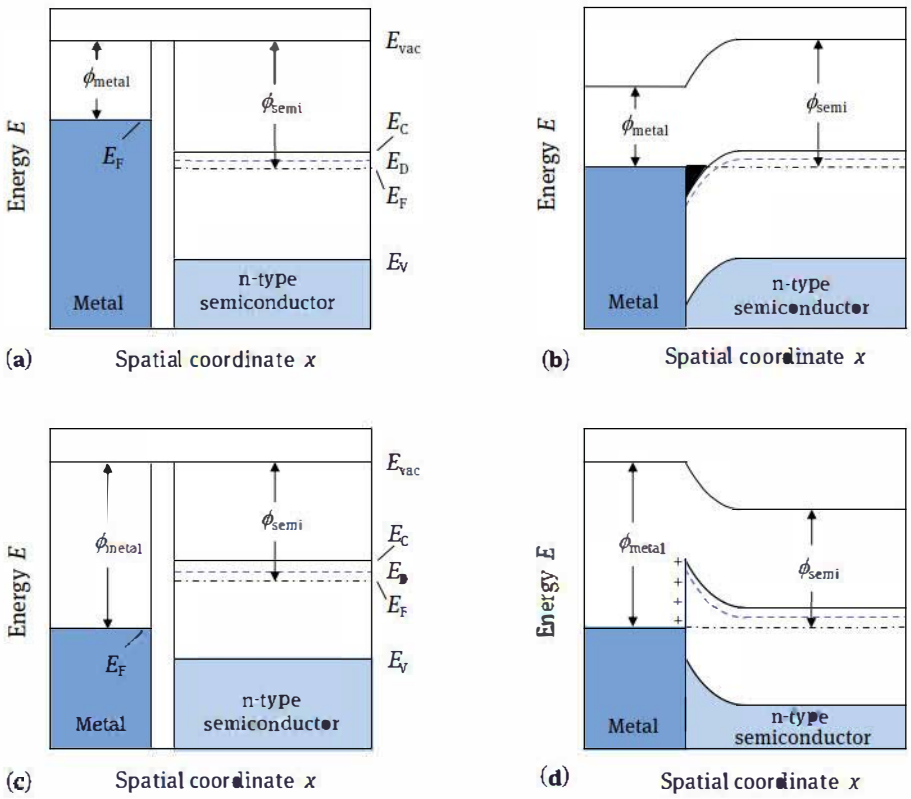


Fig. 10.28: Metal–n-type semiconductor junctions before (left) and after (right) the contact. Above: ohmic contact. The free electrons at the interface are drawn in black. Below: Schottky contact. The positive space-charge in the depletion zone is indicated by plus signs.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

edges, the donor states, the Fermi levels and the work function. As soon as the two materials come into contact, electrons flow from the metal into the semiconductor. This changes the potentials of the two materials until the Fermi levels are equalized. As in the p-n junction, the bands of the semiconductor bend near the contact (see Figure 10.28b). At the interface there is a strong accumulation of electrons in the semiconductor, shown in black in the figure, because the Fermi-level is located *within* the conduction band near the interface. If an external voltage is applied, the electrons flow through the contact without hindrance, i.e. the junction shows an ohmic characteristics.

If the work function of the metal is greater than that of the semiconductor, electrons flow from the semiconductor into the metal when contact is made. Now, as shown in Figure 10.28d, equalizing of the Fermi levels causes a high-resistance *depletion zone* in the semiconductor near the interface. The contact has a blocking effect and is usually referred to as a **Schottky barrier**. The height of the potential barrier ϕ_b between semiconductor and metal should be given by the difference in work function and the distance of the Fermi level from the conduction band, as shown in Figures 10.28c and 10.28d. Surprisingly, however, this is not confirmed by experiments. Although the barrier height observed experimentally depends on the type of semiconductor, it is almost independent of the metal selected or the level of doping. The reason for this is that interface states are formed at the metal-semiconductor junction whose energetic position is fixed in relation to the band edges. The appearance of these states is related to the fact that the wave functions of the delocalized electron states of the metal extend into the semiconductor with a smooth decay with distance from the interface. This leads to a superposition with the band states of the semiconductor. Since the density of these "mixed" states is relatively high, they determine the Fermi level. The resulting Schottky barrier ϕ_b is therefore only weakly dependent on the metal and the actual conditions at the interface. For example, we find typically $\phi_b = 0.95$eV for p-GaAs and $\phi_b = 0.47$ eV for n-GaAs.

Schottky barriers can be thought of as one half of a p-n junction, where the metal acts as the p-type semiconductor. In the description, the formulas for the p-n junction are used. If the Schottky contact consists of an n-type semiconductor and a metal, the expressions for $n_D \ll n_A$ apply, since the number of states in the metal exceeds that of the p-type semiconductor. If it is a transition to a p-type semiconductor, then the arguments just reverse. The rectifying properties of the Schottky diode were already exploited in early 20th century radio technology in the form of *crystal detectors*.

10.4.3 Semiconductor Heterostructures and Superlattices

By means of molecular beam epitaxy or MBE (**M**olecular **B**eam **E**pitaxy) and metal organic vapor phase epitaxy or MOVPE (**M**etal **O**rganic **C**hemical **V**apor **P**hase **E**pitaxy) layers of different semiconductors can be deposited on top of each other with almost perfect continuous crystalline structure, forming what are known as heterostruc-

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

tures. A prerequisite for this is that the lattice parameters of the two semiconductors differ as little as possible. For example, this is the case for the systems GaP/Si, GaAs/Ge or InAs/GaSb. Ternary or quaternary mixing systems such as $Al_xGa_{1-x}As$ or $Ga_xIn_{1-x}As_yP_{1-y}$ are also suitable. The band gap can be adjusted to the requirements by an appropriate mixing ratio. With the $Al_xGa_{1-x}As$ system, the energy gap can be engineered continuously between 1.4 eV (GaAs) and 2.2 eV (AlAs).

One of the many interesting questions that arise with these systems is that of the structure of the bands. Bringing two semiconductors with different band gaps together, as shown schematically in Figure 10.29, results in a **band discontinuity** and associated **band bending**. The transition from one band gap to the other occurs very abruptly, as shown in this figure, i.e. within the thickness of an atomic layer. Thus, electric fields of the order of atomic fields are connected thereby, i.e. around 10^{10} V/m. The band discontinuity occurs at the valence *and* conduction band edges. In the GaAs/Ge system, for example, we find values of $\Delta E_V = 0.49$ eV and $\Delta E_C = 0.28$ eV. The theoretical derivation of these numerical values is complex and will not be explained here. Naturally, in thermal equilibrium a uniform chemical potential must be established throughout the semiconductor. As a result, band bending occurs as in the p-n junction, which, depending on the charge carrier density, can extend over distances of a few hundred Ångström. The resulting fields are in the region of 10^7 V/m. Heterojunctions can be treated as classical p-n junctions. However, we have to note that the two semiconductors have different dielectric constants and the continuity of the electric displacement field has to be maintained.

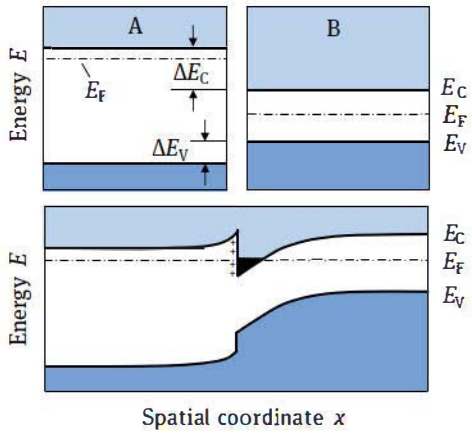


Fig. 10.29: Heterojunction consisting of two n-type semiconductors with different band gaps. Semiconductor A is strongly doped, semiconductor B is only weakly doped. At the top: bands and Fermi levels of the starting materials. Below: on contact, the Fermi levels become equalized. At the transition, band discontinuities ΔE_C and ΔE_V as well as band deflections occur. The degenerate electron gas in semiconductor B is indicated in black.

We now briefly look at a so-called *isotopic heterojunction* in which two different semiconductors with the same doping type are in contact, as shown in Figure 10.29. Since in thermal equilibrium the Fermi level must have the same value over the whole sample, in the semiconductor B with the smaller band gap, free electrons accumulate in the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

vicinity of the transition. Under suitable conditions, the Fermi level at the junction can even lie in the conduction band, so that the semiconductor is degenerate. On the other side of the junction, a depletion zone appears in semiconductor A. Electrons in the conduction band are transferred from the heavily doped semiconductor A to the semiconductor B, which has a potential well with a lower energy. This argument is also valid if the semiconductor B is not doped, i.e. is intrinsic with hardly any impurities. Then surprisingly a high concentration of electrons occurs in the *undoped* semiconductor, originating from the doped semiconductor. While in homogeneous semiconductors a high charge carrier density is always associated with high doping and thus with a high concentration of impurities, here the electrons are located in a largely defect-free material. At low temperatures the mobility of electrons in classical semiconductors is limited by strong impurity scattering. However, in the heterostructures discussed here, it is to be expected that there will be very high mobilities even at high electron density. In fact, enormous values for the electron mobility are observed at low temperatures in such systems. Figure 10.30 shows how the measured mobilities in AlGaAs/GaAs heterostructures have increased over time owing to improvements in molecular beam epitaxy. The sample with the highest mobility shown here had a complicated layer structure to reduce scattering processes at the interface. The doping was done with silicon atoms in the $Al_{0.35}Ga_{0.65}As$ layer, which has a larger band gap than GaAs. The electrons of the donor atoms are transferred to the energetically lower conduction band of the adjacent intrinsic GaAs material. Since epitaxially produced GaAs is largely free from defects, an increase in mobility by about four orders of magnitude has been observed compared to conventionally doped GaAs crystals.

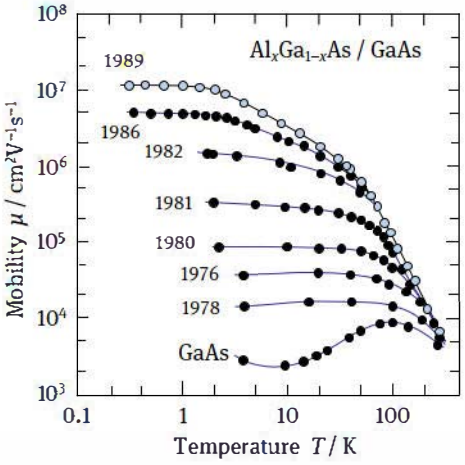


Fig. 10.30: Time line of the enhancement of the electron mobility in AlGaAs/GaAs heterostructures caused by improving the production conditions. The numbers labelling the curves indicate the year of production. The highest mobility values (until 1989) were achieved with an $Al_{0.35}Ga_{0.65}As$/GaAs system. (After L. Pfeiffer et al., Appl. Phys. Lett. **55**, 1888 (1989).)

By periodically lining up different materials a **superlattice** is created. A periodic repetition of the heterostructure just discussed leads to a doping-modulated *compositional*

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

superlattice. As shown in Figure 10.31, in this example layers of semiconductor A with high n-doping alternate with layers of the almost intrinsic semiconductor B. As a result of the new band structure, electrons migrate into the potential wells of the undoped semiconductor B. This results in a high-resistance depletion zone in each of the heavily doped layers of semiconductor A.

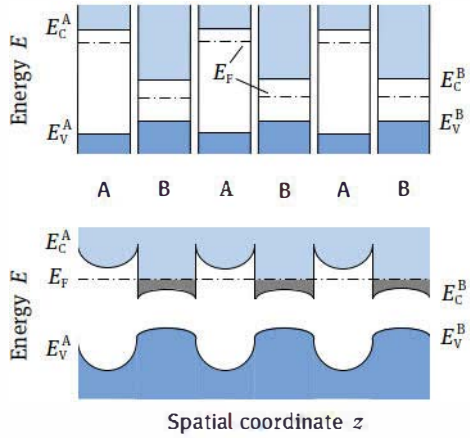


Fig. 10.31: Band structure of a compositional superlattice. The upper part shows the position of the band edges and the Fermi levels of the starting materials. In the superlattice (below) potential wells filled with a two-dimensional electrical gas (shown in grey) are created.

The behavior of doping-modulated compositional superlattices is also remarkable in other respects. The electrons trapped in the potential wells have completely different properties parallel and perpendicular to the interfaces. Parallel to the interfaces, the wave function of the electrons has the character of extended Bloch waves. In the perpendicular direction, which we refer to in the following as the z-direction, the motion is strongly restricted by the narrow potential well. To discuss the eigenvalues of the electrons, we start from a single *quantum well* and approximate it in the z-direction by a square potential. This is exactly the situation we discussed in Section 7.1 in connection with two-dimensional systems. The energy eigenvalues are given by (8.13) which we repeat here:

$$E_j(k_x, k_y) = \frac{\hbar^2(k_x^2 + k_y^2)}{2m_{xy}^*} + E_j \,, \tag{10.58}$$

where m_{xy}^* is the effective mass associated with the electron motion in the xy-plane and E_j is the transverse energy.

Since the electrons can only move within the xy-plane, these heterostructures represent the realization of a two-dimensional electron gas.[8] The density of states is constant and given by equation (8.15). For each sub-band j the simple expres-

8 The same observation can be made for holes in the maxima of the valence band.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

sion $D_j(E) = m^*_{xy}/\pi\hbar^2$ therefore holds. The existence of a step-like density of states can be verified by optical absorption or photoluminescence experiments.

At this point, we should mention another effect that can be observed at such superlattices. The transverse energy levels E_j are only sharp if the individual potential wells of the superlattice are sufficiently far apart. If the distance between the potential wells is less than 100 Å, the overlap between the wave functions becomes noticeable and causes a broadening of the individual levels into bands. This process is completely analogous to the broadening of the atomic levels in the solid, as described in Section 8.4 for the "strongly bound" electrons. The resulting bands are called **minibands**. In superlattices, it is possible to vary systematically the distance between the potential wells and thus also vary the overlap of the wave functions over wide ranges. The measurements of the Bloch oscillation discussed in Section 9.1 were performed on such structures.

Another interesting superlattice type is the **doping superlattice**, the structure of which is shown in Figure 10.32. This consists of a semiconductor with alternating n- and p-doping. The period can vary over a wide range, but is typically of a few hundred Ångströms. Since each n- and p-doped region sandwiches a thin intrinsic layer in between, these superlattices are also known as n-i-p-i or nipi structures. As Figure 10.32 indicates, these lattices have wave-like band edges. This shape is again forced by the Fermi level, which has a constant value throughout the semiconductor. Due to the alternating doping type, the conduction and valence band edges alternately lie closer to the Fermi level. This has the result that excited free electrons collect in the minima of the conduction band and holes in the valence band maxima. The charge carrier types are thus spatially separated from each other. This makes recombination more difficult, considerably increasing the lifetime of the electrons and holes. In addition, the modulation of the bands causes a reduction of the effective optical band gap - similar to

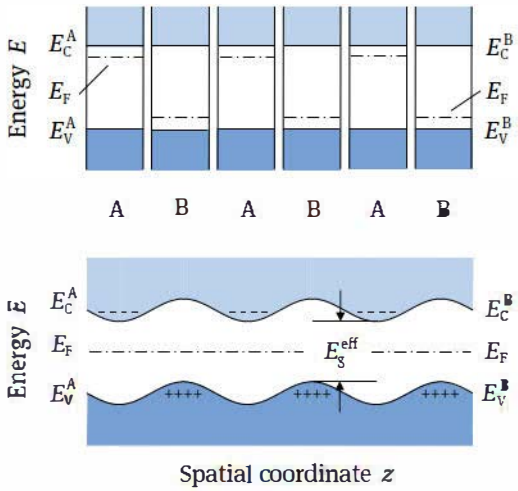


Fig. 10.32: The band structure in a doping superlattice. The upper part shows the position of the band edges and the Fermi levels of the starting materials. Below, showing the superlattice a wave-like potential variation is seen with the effective band gap indicated. The free charge carriers are indicated by plus and minus signs.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the behavior in amorphous semiconductors. A further effect arises from the fact that the band modulation is caused by space-charges. Thus, the "valleys" are formed in the n-type regions, since positively charged donors reside in these locations. In the region of the "hills" the space-charge has just the opposite sign. If many more free electrons and holes are created by intensive optical irradiation, they reduce the effective space-charge. This also reduces the modulation strength, in other words, the effective band gap can be increased by light irradiation. This effect can be demonstrated by means of photoluminescence, which shifts to higher photon energies as the intensity of the stimulating light increases.

10.5 Devices

Today's information technology is based on data processing, data storage and data transfer. Data processing is almost exclusively carried out by integrated circuits with silicon as the base material. In contrast, optoelectronic devices, which are frequently used for data transfer, are usually based on III-V semiconductors. These include optical detectors, light-emitting diodes and semiconductor lasers.

Two types of devices can be distinguished: *two-terminal devices* operate by the manipulation of the current flow between the two contacts, such as in the case of diodes. They include most optoelectronic components. On the other hand, *transistors* are used in data processing, storage and in power electronics. These are *three-terminal devices* in which currents or voltages between two contacts are controlled by applying an external voltage to a third contact. Basically, a distinction is made between *unipolar* and *bipolar* components, depending on whether one type of charge carrier or both are involved. In the following we will go into some examples of diodes and transistors and close the chapter with a description of the semiconductor laser.

10.5.1 Technical Application of the p-n Junction

There are several devices based on the application of a single p-n junction. As well as the rectifying properties of the p-n diode, which we have already covered, these include the Zener diode, the solar cell and the photodiode, which we briefly discuss below. There are further devices such as the *backward diode* or *tunnel diode*, also called the Esaki diode[9] after the name of its inventor. This element has a negative differential resistance over a limited voltage range and is often used in microwave technology as an amplifier, a switch, and also as an oscillator.

9 Leo Esaki, *1925 Osaka, Nobel Prize 1973

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Zener Diode.[10] We consider a p-n junction in a highly doped semiconductor biassed in the reverse direction. If the applied voltage is increased, a surprisingly strong current rise sets in. In the case of silicon diodes with a concentration of $10^{25}\,\mathrm{m}^{-3}$ dopant atoms or higher, the critical voltage is about 2 V. Since very high field strengths occur in the thin barrier layer with high doping, electrons from the valence band of the p-conductor can tunnel through the band gap into the largely empty conduction band of the n-conductor so long as the conduction band edge lies below the valence band edge of the p-conductor, as shown in Figure 10.33. With increasing field strength, the high field causes a sudden high current to flow, known as a *Zener breakdown*. Since the strong current flow starts very abruptly, this effect is often used in electrical circuits to provide a reference voltage.

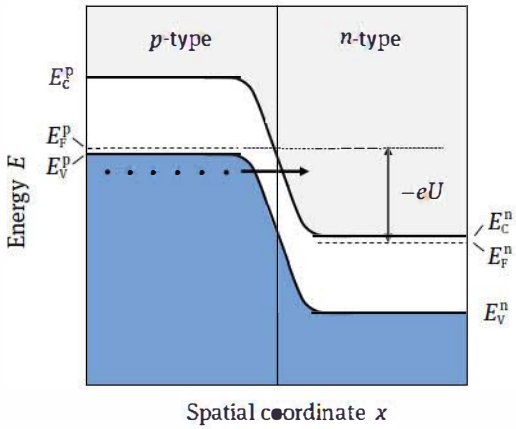


Fig. 10.33: The Zener diode. At a sufficiently high voltage in the reverse direction, electrons can tunnel through the depletion zone and cause a sharp increase in the current flow. The tunneling of the electrons (black dots) is indicated by the horizontal arrow.

Solar Cell. p-n junctions are the heart of a solar cell. When a photon is absorbed in a semiconductor, an electron-hole pair is created. If the absorption takes place in the space-charge zone of a p-n junction, the two charge carriers are separated by the electric field. Light irradiation therefore causes an additional current flow I_C in the p-n junction. If we insert this contribution into equation (10.55) and replace the current density by the current we obtain:

$$I = I_s\left(e^{eU/k_BT} - 1\right) - I_C\,. \tag{10.59}$$

In the classic silicon solar cell, sunlight falls on an about 1 µm thick n-conducting layer, which has been produced on top of the approximately 0.6 mm thick p-conducting substrate. The n-layer is so thin that the sunlight is preferably absorbed at the p-n junction. The p-layer is relatively thick, so that even light penetrating deeper can still contribute

10 Clarence Melvin Zener, *1905 Indianapolis, †1993 Pittsburgh

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

to charge separation. At the same time, this layer gives the solar cell the necessary mechanical stability. Typical doping concentrations are 2×10^{25} phosphorus atoms/m^3 and 5×10^{22} boron atoms/m^3. Of course, for a more detailed analysis of the current flow, we have to take into account the fact that the recombination of the charge carriers which takes place in the large size of the space-charge zones present here plays an important role, which we have neglected in our previous discussion of the p-n junction.

Figure 10.34 shows the current-voltage characteristic of a silicon solar cell with an area of 4 cm^2 when exposed to light. The parameter values $I_L = 100$ mA and $I_s = 1$ nA were used in calculating the characteristic curve. Also drawn is the equivalent circuit with load resistance R_L the optimum value of which is determined by the following considerations.

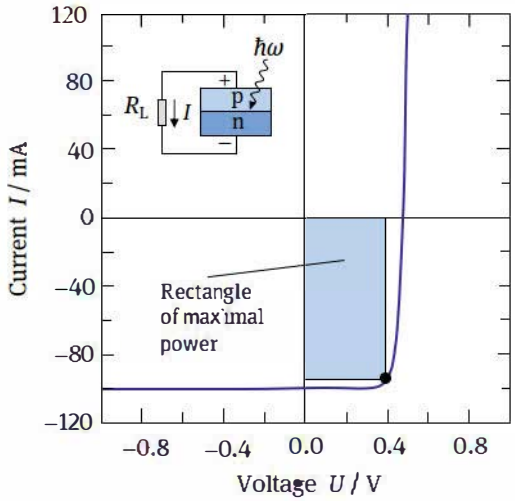


Fig. 10.34: Current-voltage characteristic of a silicon solar cell under illumination. The optimum operating point is found when the area of the blue rectangle, and thus the delivered electrical power, is at the maximum. The basic circuit with the load resistance R_L is also drawn.

The circuit is characterized by two parameters: the open circuit voltage and the short circuit current. If a solar cell is operated under the no-load condition, i.e. at $I = 0$, then according to equation (10.59) the voltage across the p-n junction is $U \approx (k_B T/e) \ln(I_L/I_s)$. Typical no-load voltages lie in the range of 0.5 V. In the case of a short circuit, i.e. when $U = 0$, the current is given by I_L which is determined exclusively by the light-induced contribution and is proportional to the illumination strength. The solar energy is optimally used when the power $P = UI$ delivered to a consumer is at its maximum value. The operating conditions must therefore be selected so that the area of the blue rectangle in Figure 10.34 is as large as possible. With the usual characteristic curves, this is the case when the working voltage is about 80% of the open-circuit voltage. In order to achieve high efficiency, the load resistance R_L must therefore be adapted to the parameters of the solar cell.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Why is the achievable efficiency of solar cells relatively low? On the one hand, photons with an energy smaller than the band gap make no contribution, and on the other hand, the energy surplus of the energy-rich photons is lost because they only produce *one* electron-hole pair. The efficiency therefore depends on the energy gap of the semiconductor used and the profile of the illumination spectrum. Figure 10.35 shows the ideal efficiency of solar cells, taking into account the relative position of the solar spectrum and the band gap of the device in question. The radiation spectrum in the illustration is the AM 1.5 Standard Spectrum taken when the sun is 48.2° above the horizon. The diagram shows that without further precautions such as light concentration or a more complicated structure consisting of a series of semiconductors with different energy gaps, the maximum efficiency is limited to 31%. This efficiency can theoretically be achieved if the energy gap corresponds to the photon energy at the peak of the solar spectrum. However, the efficiencies actually reached are much smaller. With solar cells consisting of amorphous silicon we can obtain up to 10%, with polycrystalline silicon 15% and with crystalline silicon up to 20%. A high efficiency is also achieved with GaAs, up to 25% with solar cells consisting of several layers.

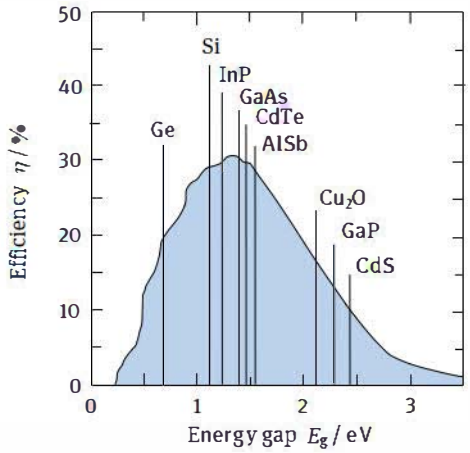


Fig. 10.35: The ideal efficiency η of solar cells as a function of the energy gap. The efficiency is given by the intersection of the line indicating the energy gap of the semiconductor and the black profile reflecting the solar spectrum penetrating the Earth's atmosphere. The weak oscillations are caused by atmospheric absorption. The radiation spectrum AM 1.5 was assumed (see text).

Photodiode. The p-n junction is also often the central element in photodiodes. As we have seen, irradiating a p-n junction with light generates a current flow I_L. Since the absorption of a photon produces an electron-hole pair, the current in the p-n junction is proportional to the intensity of the incident light at a given wavelength. As shown in Figure 10.36, in incident light the characteristic curve shifts downwards, changing the voltage drop across the working resistor R_L. To obtain a signal that is largely independent of the bias voltage, the photodiode is operated in the reverse direction, typical workling points being indicated by the circles in the figure.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

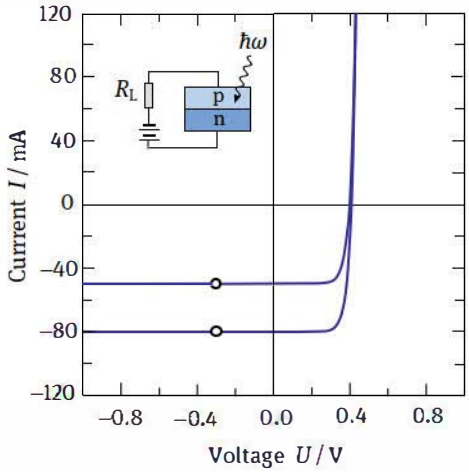


Fig. 10.36: Current-voltage characteristic of a photodiode at two light intensities. With increasing intensity the characteristic curve shifts downwards. The two circles mark typical working points. The equivalent circuit with load resistance R_L is also shown.

Light Emitting Diode (LED). Light Emitting Diodes (LEDs) are widely used in everyday life. To generate light, we exploit the effect that in a diode in the forward direction, the majority charge carriers after crossing the space-charge zone recombine within the diffusion length. In direct semiconductors, an electron can often jump from the conduction band into a hole in the valence band with the emission of a photon. This process is the reverse of the optical absorption process discussed in Section 10.1. In semiconductors with an indirect band gap, recombination usually takes place *without radiation*, i.e. the full recombination energy is transferred to the lattice in the form of phonons. In order to use such semiconductors as light emitting diodes, suitable *recombination centers* are built in. These are impurities with an energy level near the valence band to which a radiative transition can occur.

Light emitting diodes for all wavelengths, from infrared to ultraviolet, have been available on the market for several years. Since the wavelength of the emitted light is determined by the energy gap, semiconductors with different band gaps are used. For example, GaAs with $E_g = 1.42\,\text{eV}$ emits in the infrared, GaN with $E_g = 3.37\,\text{eV}$ in the blue and AlN with $E_g = 6.13\,\text{eV}$ in the ultraviolet. Noteworthy is the mixed system $In_xGa_{1-x}N$, which has a direct band gap, but with a value varying between 0.7 eV and 3.37 eV depending on the composition. LEDs based on heterostructures are also manufactured.

The availability of light-emitting diodes operating at very different wavelengths allows white light to be generated using this technology. For this purpose, various methods of additive colour mixing have been developed. One possibility is to place light-emitting diodes with red, green and blue light in one housing and to control them appropriately in each case. It is also possible to change the color of such lamps continuously by appropriately varying the control voltages. The maximum efficiency is currently around 85 %.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

10.5.2 Transistors

Transistors are used to switch and amplify currents and voltages. There are two types: the unipolar and the bipolar. We first take a brief look at the bipolar transistor, also known as the bipolar junction transistor, which is basically a combination of two p-n junctions. A more comprehensive theory of this transistor is relatively complex, but here we will deal only with the principle mode of operation. Then we will describe the mode of operation of a MOS field-effect transistor, which is one of the unipolar devices.

Bipolar Transistor. The classic bipolar transistor was invented in 1947 by *J. Bardeen*[11], *W. Brattain*[12] and *W.B. Shockley*[13]. A p-n-p transistor consists of *emitter*, *base* and *collector*, as shown in Figure 10.37. For the construction of the transistor it is important that the thickness of the base ($d < 1\mu m$) is so small that the recombination of the charge carriers does not play a role there. The same applies to the n-p-n transistor, where electrons carry the main current rather than holes as in the p-n-p transistor.

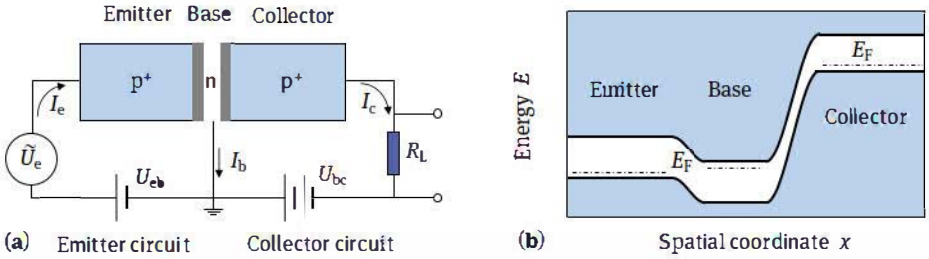


Fig. 10.37: Bipolar p-n-p transistor. **a)** Schematic structure and circuitry of a p-n-p transistor. The space-charge zones are marked in grey, the relevant voltages and partial currents are indicated. The + sign in the doping information indicates high doping. **b)** Band scheme of the transistor. The position of the Fermi level in the individual areas is indicated by dashed dotted lines.

The forward-biased emitter-base junction "emits" holes into the base where they diffuse until they reach the reverse-biased base-collector junction. There the holes, being minority charge carriers, can pass unhindered. They are "collected" at the collector and then flow through the load resistance R_L of the collector circuit. It is important, as already mentioned, that the holes, whose generation is controlled by the emitter-base voltage U_{eb}, do not recombine in the base with the large number of electrons present, but instead enter the collector as unhindered as possible. This is achieved by making the base zone small compared to the diffusion length. The base current I_b is then very

11 John Bardeen, *1908 Madison, †1991 Boston, Nobel Prize 1956 and 1972
12 Walter Houser Brattain, * 1902 Amoy (China), †1987 Seattle, Nobel Prize 1956
13 William Bradford Shockley, *1910 London, †1989 Stanford, Nobel Prize 1956

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

small and the emitter current flows largely into the collector. The emitter and collector currents are therefore approximately equal and thus independent of the base-collector voltage U_{bc} and thus also of the resistance R_L. The voltage $U_L = R_L I_k$ across the load resistor can therefore be much higher than the signal used at the input in the emitter circuit, in other words, the signal is amplified.

The gain can be roughly determined as follows: if $U_e = U_{eb} + \tilde{U}_e$ is the voltage and I_e is the current in the emitter circuit, as shown in Figure 10.37, then from our discussion of the p-n junction from equation (10.55) we obtain:

$$I_e = I_e^S \left(e^{eU_e/k_B T} - 1\right) \approx I_e^S \, e^{eU_e/k_B T} , \tag{10.60}$$

where I_e^S is the saturation current of the emitter in the reverse direction. Using α to denote the fraction of the holes that can diffuse through the base without recombining and arrive at the collector, then we can calculate the collector current:

$$I_c = I_c^S + \alpha I_e \approx \alpha I_e . \tag{10.61}$$

The final approximation applies because the base-collector diode is reverse biassed and the saturation current I_c^S is therefore relatively small. Thus we obtain for the voltage U_L across the load resistor:

$$U_L = R_L I_c = \alpha R_L I_e . \tag{10.62}$$

With these equations the amplification factor is obtained immediately:

$$\frac{dU_L}{d\tilde{U}_e} = \frac{e\alpha R_L I_e}{k_B T} . \tag{10.63}$$

Using typical values of $I_e = 10\,\text{mA}$, $k_B T/e \approx 0.025\,\text{V}$, $R_L = 1\,\text{k}\Omega$ and $\alpha \approx 1$, we find a value 400 for the amplification factor. Here we have looked at the *base circuit* where the base is at ground potential. The transistor then acts as a voltage or power amplifier. As we can easily see, in the case of the *emitter circuit*, where the emitter is grounded, the transistor acts as a current amplifier.

MOSFET. Before we deal with the operation of the *metal-oxide-semiconductor field effect transistor*, (MOSFET), we first consider a metal-oxide-semiconductor interface as shown schematically in Figure 10.38 for a weak p-type semiconductor. If a positive voltage U_g is applied to the metal electrode, the holes are repelled and the electrons are attracted. Due to the presence of negatively charged acceptors, a negative space-charge zone is formed, as in p-n junctions. This leads to band bending near the interface and to a *depletion* of holes. With increasing voltage, the bands at the interface are further lowered and the electron density increases. Since in this thin layer the conduction mechanism changes from the original p-conduction to n-conduction, it is called an **inversion layer.** The n-conducting layer is separated from the p-conducting substrate by the insulating depletion zone. When the voltage is further increased, the conduction band edge finally dips below the Fermi level and a degenerate Fermi gas with metallic

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

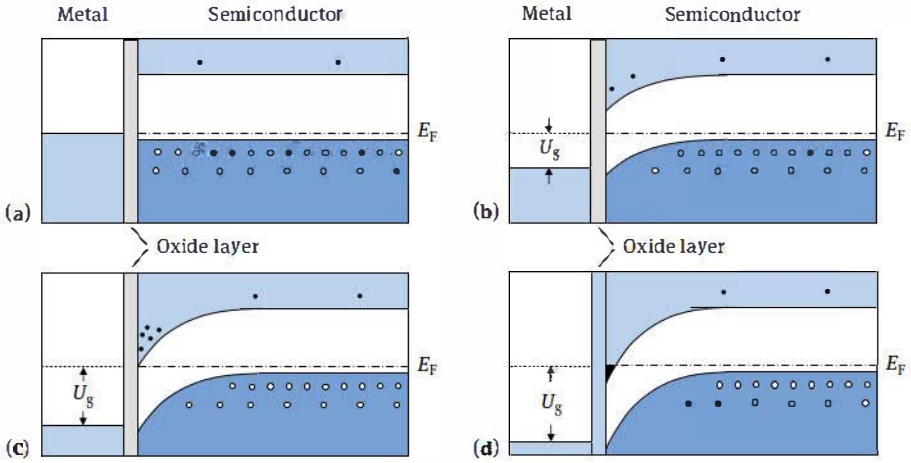


Fig. 10.38: The mode of operation of a metal-oxide-semiconductor interface. Electrons in the conduction band are represented by black circles and the holes in the valence band by open circles. A positive voltage U_g is being progressively applied (the counter-electrode is not shown). **a)** Position of the band edges and the Fermi level before application of the voltage, **b)** depletion layer caused by a small positive bias voltage, **c)** n-inversion layer with large positive bias voltage, **d)** inversion layer with degenerate electron gas in the conduction band, indicated in black.

character is formed. Since this well-conducting channel is relatively narrow, the inversion layer represents an experimental realization of a two-dimensional electron gas. The quantum Hall effect, discussed in Section 9.3, was observed for the first time in such a structure.

Now we are in a position to understand the switching characteristics of a MOSFET, the structure of which is shown schematically in Figure 10.39. Two highly doped n-regions, covered by a thin oxide layer (mostly SiO_2), are embedded in p-type material. The oxide layer is penetrated by two contacts which establish a connection to the

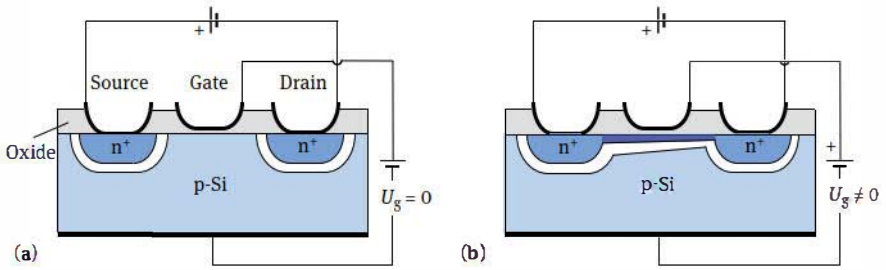


Fig. 10.39: MOSFET. The dark metal electrodes are in direct contact with the n^+ regions of source and drain. **a)** Without gate voltage U_g, these are separated from each other and from the p-conducting substrate by the depletion zones shown in white. **b)** A positive gate voltage gives rise to the dark blue conductive channel, which is also isolated from the substrate by a depletion zone.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

n-doped electron *source* and the similarly n-doped electron *drain*. The third contact to the *gate* remains separated from the p-doped substrate by an oxide layer about 100 nm thick. Together with the oxide layer, this contact forms the MOS interface as discussed immediately above.

With no bias voltage at the gate, one of the two p-n junctions is blocked and causing a high resistance. If a sufficiently large positive gate voltage is applied, the low-resistance, then the metallically conductive channel described above is formed. This channel, the source and the drain are separated from the p-conducting substrate by depletion zones. The width of the depletion layer varies with the local potentiaL The transistor can be easily set to the conductive state "on" or the non-conductive state "off" via the gate voltage.

The function of MOSFETs depends crucially on the quality of the SiO_2 layer and the interface. Any defects are also recharged when the gate voltage changes, so that only part of the voltage is effective for controlling the inversion layer. The required high quality is only achievable with silicon. With GaAs, so many defects occur that no functional MOSFETs can be produced with this material.

10.5.3 The Semiconductor Laser

An interesting, technically very important, application of heterostructures is the semiconductor laser, whose operating principle we briefly explain here. We will look at the p-n junction again. As already mentioned in the discussion of the light-emitting diode, light emission occurs when the diode is operated in the forward direction due to the recombination of electrons and holes after crossing the space-charge zone. A prerequisite for the onset of *laser activity* in a p-n junction is the occurrence of population inversion. Since the electrons and holes are located at the edges of the band, we must consider the occupation probability at the edges E_C and E_V. A condition for the necessary population inversion is therefore that $f(E = E_C) > f(E = E_V)$. The corresponding probabilities are given by:

$$f(E_C) = \frac{1}{e^{(E_C - E_F^n)/k_B T} + 1} \quad \text{and} \quad f(E_V) = \frac{1}{e^{(E_V - E_F^p)/k_B T} + 1} \tag{10.64}$$

so that this condition is fulfilled if

$$E_F^n - E_F^p > E_C - E_V = E_g \,. \tag{10.65}$$

This means that the quasi-Fermi levels, as sketched in Figure 10.40, must actually be within the bands. This is achieved by very high doping, indicated by p^{++} and n^{--} in the figure. Then in GaAs diodes, laser activity will actually occur if the injection current is sufficiently high.

The required high injection current can be significantly reduced by using a double heterostructure. Referring to Figure 10.41 we now explain the operation of an

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

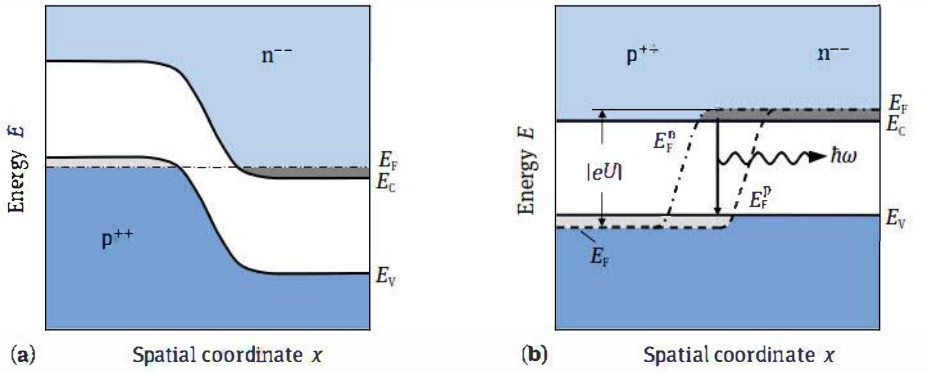


Fig. 10.40: Band diagram of a p^{++}-n^{--}-junction. **a)** Position of the Fermi level in thermal equilibrium with no applied external voltage. **b)** p^{++}-n^{--}-junction with applied voltage U in the forward direction. Occupation inversion occurs, leading to stimulated emission and laser activity.

AlGaAs/GaAs/AlGaAs laser. The active layer in the middle consists of lightly doped GaAs sandwiched between two layers of AlGaAs, which has a larger band gap. The AlGaAs layer on the left-hand side is p-doped and that on the right is n-doped. The band structure which forms with no voltage, is shown in Figure 10.41a. If a sufficiently high voltage is then applied in the forward direction, band minima and maxima are created in the GaAs layer. The quasi-Fermi levels of electrons and holes lie within the bands in the active GaAs layer so that a population inversion occurs. Holes from the p-doped region and electrons from the n-doped region flow into this zone. The band discontinuities prevent the charge carriers from flowing out of this zone and thus cause an increased radiative recombination. This effect is also called *electrical confinement*.

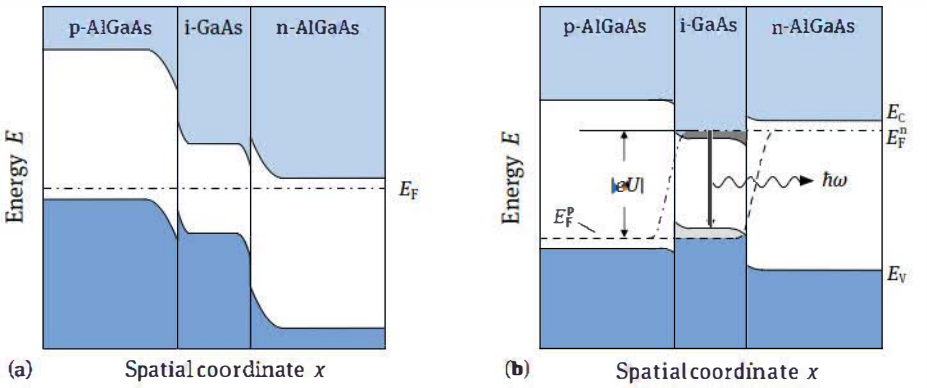


Fig. 10.41: Double heterostructure of p-AlGaAs/i-GaAs/n-AlGaAS basis for the semiconductor laser. **a)** Band diagram of the heterojunctions in thermal equilibrium. **b)** Band diagram when a voltage U is applied.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In addition, *optical confinement* also occurs. The reason for this is the difference in the refractive index of AlGaAs and GaAs, so that the laser light is kept preferentially in the active zone by reflection. The highly reflective interfaces between the semiconductor and air create an optical resonator requiring no extra optical mirrors. With such devices, a breakthrough was achieved leading to the widespread commercial use of semiconductor lasers.

10.6 Exercises and Problems

1. Fermi Level. We consider an intrinsic semiconductor with an energy gap $E_g = 0.6\,\text{eV}$ and effective masses $m_n^* = 0.04m$ and $m_p^* = 0.07\,m$ for the electrons and holes. Where is the Fermi level at room temperature?

2. Silicon. Calculate the following quantities for an undoped silicon crystal at 300 K:
(a) the position of the Fermi level with respect to the upper edge of the valence band,
(b) the occupation probability for a state in the conduction band,
(c) the charge carrier concentration,
(d) the resistivity.

If we now dope the silicon crystal with 10^{23} boron atoms / m^3. Calculate
(e) the new resistivity at 300 K,
(f) the temperature at which the Hall constant disappears.

Hint: the effective masses of electrons and holes would really have to be "suitably" averaged in this calculation, where the averaging would have to depend on the quantity concerned (i.e. carrier concentration, charge transport, ...). For simplicity, you may set $m_n^* = m$ and $m_p^* = 0.5m$.

3. Intrinsic Semiconductor. The following resistance values were measured in an undoped semiconductor sample at various temperatures (marked by the subscript): $R_{350\,\text{K}} = 98.2\,\Omega$, $R_{420\,\text{K}} = 2.53\,\Omega$, $R_{490\,\text{K}} = 0.17\,\Omega$, $R_{560\,\text{K}} = 0.023\,\Omega$, $R_{630\,\text{K}} = 0.0046\,\Omega$. How large is the energy gap? Which semiconductor was being measured? Would it be transparent at a wavelength of 1 μm?

4. Doped Semiconductor. If gallium arsenide ($\varepsilon = 13.13$) is doped with silicon, the silicon atom acts as a donor or acceptor depending on the particular lattice site occupied. In the simple hydrogen atom model, no distinction is made between the two cases. Use this model to calculate the ionization energy of the silicon donors and the orbital radius of the electrons in the ground state. At what concentration of silicon atoms is an impurity band formed?

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

5. Doped GaAs. We consider a GaAs sample doped with $5 \times 10^{22}\,\mathrm{m}^{-3}$ tellurium and $10^{20}\,\mathrm{m}^{-3}$ carbon atoms. Tellurium atoms act as donors, carbon atoms as acceptors. The impurity levels are respectively 0.03 eV and 0.02 eV away from the corresponding band edges.

(a) Determine the temperature range over which the transition from the ionization regime to the saturation regime occurs.
(b) Calculate the fraction of the donors ionized at 120 K.
(c) Determine the Fermi level at this temperature and at 300 K.
(d) Calculate the density of charged acceptors at room temperature.
(e) Calculate the electrical conductivity at room temperature.

6. p-n Junction. Calculate the capacitance of the space-charge zone of a silicon diode with cross-sectional area of $0.5\,\mathrm{mm}^2$ with no applied voltage. The doping is given by $n_\mathrm{D} = 1.5 \times 10^{22}\,\mathrm{m}^{-3}$ and $n_\mathrm{A} = 2 \times 10^{23}\,\mathrm{m}^{-3}$. How large is the capacitance when a voltage of $U = 0.3\,\mathrm{V}$ is applied in the reverse direction?

7. Solar Cell. In a silicon solar cell when irradiated with light of sufficiently short wavelength, 0.1 charge carriers per second and per silicon atom are generated in the space-charge zone. Calculate:

(a) the short-circuit current,
(b) the open-circuit voltage,
(c) the optimum working voltage at 22°C.

Assume an active area of $100\,\mathrm{cm}^2$ and a reverse current of 2×10^{-9} A. Use the numerical values in Section 10.5 for the space-charge zone.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

11 Superconductivity

As discussed in detail in Chapter 9, the electrical conductivity of metals is limited by the scattering of the conduction electrons with phonons, by scattering with other electrons, and by crystal defects. As these scattering processes occur at all finite temperatures, it might be expected that the ohmic resistance only disappears in perfect crystals and, if at all, only at absolute zero. Surprisingly, zero ohmic resistance is actually displayed by many metals at finite and not particularly low temperatures. This phenomenon is called **superconductivity**. It was discovered as long ago as 1911 by *H.K. Onnes*[1] when he was measuring the resistance of a mercury filament at low temperatures. As can be seen in Figure 11.1, the resistance initially decreased slowly when the sample was cooled down and dropped abruptly to an immeasurably small value at about 4.2 K. Similar observations, although at different temperatures, were later made for many other metals.

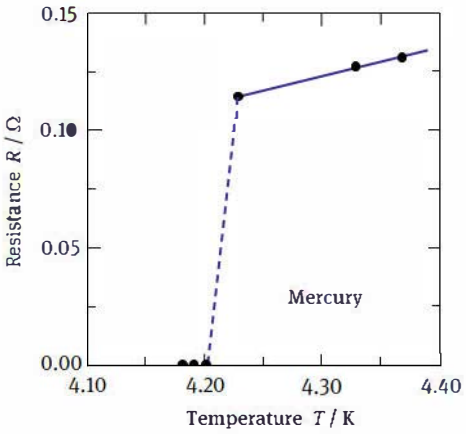


Fig. 11.1: The electrical resistance of a mercury filament as a function of temperature. (After H.K. Onnes, Leiden Commun. **124c** (1911).)

11.1 Phenomenological Description

For a theoretical understanding of superconductivity, the answer to the following question is of great importance: is the resistance really zero below the **transition temperature** T_c, or is it just very small? An elegant and very sensitive test of this question can be carried out by persistent current experiments: a magnetic field is generated in a ring of superconducting material above the transition temperature, the temperature

1 Heike Kamerlingh Onnes, *1853 Groningen, †1926 Leiden, Nobel Prize 1913

https://doi.org/10.1515/9783110666502-011

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

is lowered and the magnetic field is switched off below T_c. At finite resistance R, the induced current should decay according to the equation $I(t) = I_0 \exp(-Rt/L)$, where L is the inductance of the ring. However, the current in the superconducting ring does not change within the measuring accuracy. The results from the most sensitive measurement mean that the resistance drops by at least 14 orders of magnitude at the transition temperature T_c. It is therefore assumed that there is no electrical resistance actually in the superconducting state, in other words, a superconductor is an *"ideal conductor"*.

Some general empirical relationships are known from which we can guess whether various materials will become superconducting. For example, it is known that elements with small atomic volumes become superconducting, and that in alloys, the electron density plays an important role. However, quantitative predictions cannot be made from these observations. Surprisingly, most metals become superconducting, and their structural order appears to be of little importance, as evidenced by the fact that superconductors can occur in pure crystals, alloys or even amorphous materials. That the long-range order does not play a crucial role can be seen from the behavior of bismuth, in which the amorphous phase becomes superconducting, but not the crystalline phase.

The transition temperatures of elements all fall below 10 K, as summarized in Table 11.1. At 9.25 K. Niobium has the highest transition temperature of all elements. Other elements can also become superconducting under high pressure.

Tab. 11.1: The transition temperatures T_c, critical magnetic fields B_c and Debye temperatures Θ of superconducting elements at normal pressure. (After C. Fischer et al., *Springer Handbook of Condensed Matter and Material Data*, W. Martienssens, H. Warlimont, eds., Springer, 2005.)

Element	T_c (K)	B_c(mT)	Θ(K)	Element	T_c (K)	B_c(mT)	Θ(K)
Al	1.175	10.49	420	Pb	7.196	80.34	105
Be	0.026	0.11	1390	Re	1.697	20.1	415
Cd	0.517	2.81	209	Rh	0.0003	0.005	512
Ga	1.083	5.93	325	Ru	0.493	6.9	580
Hf	0.128	1.27	256	Sn	3.722	30.55	195
Hg	4.154	41.1	87	Ta	4.47	82.9	258
In	3.409	28.15	109	Tc	7.77	141	411
Ir	0.113	1.6	425	Th	1.374	16.0	165
La	4.87	9.8	151	Ti	0.40	5.6	415
Lu	0.1	35.0	210	Tl	2.38	17.65	87.5
Mo	0.916	9.69	460	V	5.46	140	383
Nb	9.25	206	276	W	0.015	0.115	383
Os	0.66	7.0	500	Zn	0.857	5.41	310
Pa	0.43	5.6	185	Zr	0.63	4.7	290

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

For alloys and metallic compounds, transition temperatures of over 20 K can be found, with MgB_2 having the highest value at $T_c = 39$ K. Table 11.2 lists the transition temperatures of some of these materials.

Tab. 11.2: The transition temperatures T_c of some superconducting compounds. (Various sources.)

Compound	T_c (K)	Compound	T_c (K)	Compound	T_c (K)	Compound	T_c (K)
Nb_3Sn	18.5	V_3Si	17.1	PbLi	7.2	Cs_3C_{60}	40
Nb_3Ge	23.2	V_3Ga	16.8	Pb_3Na	5.6	K_3C_{60}	19.5
Nb_3Al	17.5	MoC	14.3	MgB_2	39.0	$CeRu_2$	6.1
AuPb	7.0	$PbMo_6S_8$	14.7	MnU_6	2.3	La_3In	10.4

More recently, we have found the class of materials which are *high-temperature superconductors*, only discovered in 1986 with transition temperatures of up to 135 K. In addition to the oxidic high-temperature superconductors, other new and interesting classes of superconductors have been discovered in recent years, including *heavy-fermion systems*, *ruthenates* and *pnictides*. We will discuss some of these superconductors in Section 11.5, which also has a table listing the transition temperatures for some of these materials.

At first glance, it is astonishing that superconductivity has not yet been found in "good" metals with high conductivity such as Ag, Au, Cu or Na. For the understanding of superconductivity, it is also important that the ferromagnetic elements such as Fe, Ni or Ho remain normal conducting. In fact, magnetic properties play an important role. If, for example, the composition of high-temperature superconductors is varied, a transition from the superconducting to an antiferromagnetic phase is observed (see Section 12.4). In some heavy-fermion systems, to which we have already briefly referred in Section 8.2, the appearance of superconductivity is even supported by magnetism. Impurities, apart from those which are magnetic, have only a weak influence on the transition temperature. Although the theoretical principles of "classical" superconductors are well understood, it is not possible to calculate transition temperatures even if the structure and material parameters of the substances concerned are known.

11.1.1 Meissner-Ochsenfeld Effect, London Equations

As mentioned above, persistent current experiments show that superconductors behave like ideal conductors. The question arises whether superconductors are more than just ideal conductors, i.e. whether the electromagnetic properties of superconductors are described simply by Maxwell's equations, but with a vanishing resistance. The answer to that question is provided by experiments with magnetic fields. To give more insight,

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

we will conduct a thought experiment with a sample that becomes ideally conductive at the transition temperature T_c and compare the results with the properties that we find in real superconductors. We first consider an arbitrarily-shaped closed loop inside our sample. The law of induction states that a change with time in the magnetic flux penetrating the area enclosed by the loop results in an electric field around the contour of this area. However, since electric fields are perfectly short-circuited in materials with zero resistance, there is no possibility of any such electric field appearing, and therefore the enclosed magnetic flux cannot change. If we apply this consideration to the entire sample, consequences arise as shown in Figures 11.2a and 11.2b. The thought experiment starts on the left side of the figure and ends on the right side. Temperatures and magnetic fields are shown above and below the picture.

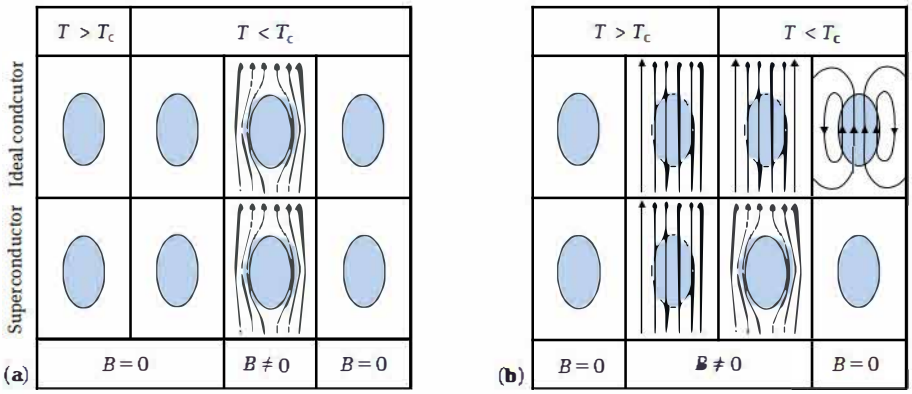


Fig. 11.2: The response of a n "ideal conductor" and of a superconductor to the application of a magnetic field. The two panels show two different procedures. In **a)** we first cool the samples through the transition temperature and *then* a magnetic field is applied, and then removed again. In **b)** the field is applied at the outset and the samples are then cooled through the transition. The behavior during these two procedures is quite different. In **a)** both samples remain field-free during the entire experiment. In **b)** the magnetic field when applied above the transition temperature clearly penetrates both samples. However, when the temperature falls below T_c, the magnetic field remains trapped in the "ideal conductor" while it is expelled from the superconductor. At this point if the field is switched off, two different final states are reached.

Figure 11.2a assumes that both samples are first brought to a temperature below T_c without a magnetic field and that a magnetic field is then applied. The ideal conductor and the superconductor react in the same way: the magnetic field cannot penetrate because as the field is increased currents are induced in the surface of the sample which shield the field. Because of the lack of resistance, these shielding currents do not dissipate, but persist, with the result that the interior of the sample remains permanently free of magnetic fields. If we switch off the field, the initial state is restored in both cases.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Obviously, ideal conductors and superconductors cannot be distinguished with this experimental procedure.

Now, as shown in Figure 11.2b, if we apply the magnetic field above the transition temperature T_c, the field penetrates both samples, since the shielding currents decay when there is a finite resistance. However, as we cool through T_c, the situation does not change in the ideal conductor; the magnetic flux remains constant and the sample is therefore penetrated by the magnetic field. If the field is switched off, currents are induced in accordance with Lenz's law, which prevent a change in magnetic flux in the sample. The ideal conductor now has a permanent magnetic moment, which only disappears again when the temperature rises to $T > T_c$. A superconductor behaves differently: when the temperature falls below the transition temperature, the field is expelled from the sample. Obviously, the field disappears from the interior of the sample and does not remain constant as required by the Maxwell equations for an ideally conducting sample. If we switch off the applied field, the initial state is restored. Depending on the sequence in which the experiment is conducted, two different final states occur in the ideal conductor, whereas the final state is uniquely defined in the superconductor. The expulsion of the magnetic field at the phase transition, the so-called **Meissner-Ochsenfeld effect**, was discovered in 1933 by *W. Meissner*[2] and *R. Ochsenfeld*[3].

When discussing the magnetic properties of superconductors, we have to take into account that the magnetic field inside and in the immediate vicinity of the sample depends on the sample shape, since a *demagnetization field* appears when the sample is exposed to a magnetic field. In order to avoid this complication, which will be discussed in Section 13.1 in the case of electric fields, we always consider long cylinders whose axes are parallel to the magnetic field when discussing superconductivity, because for this geometry the demagnetization field disappears.

Small magnetic fields are completely shielded by superconductors, because currents are induced at the sample surface, which cause a magnetization whose field completely counteracts the applied field $\mathbf{B}_a = \mu_0\mathbf{H}$. Since the magnetic flux is completely expelled, we can write $\mathbf{B}_i = \mu_0(\mathbf{H} + \mathbf{M}) = 0$ for the field or magnetic induction $\mathbf{B}_i$ inside the sample. This means that the magnetisation $\mathbf{M}$ is given by $\mathbf{M} = -\mathbf{B}_a/\mu_0$ and the susceptibility by

$$\chi = \frac{\mu_0\mathbf{M}}{\mathbf{B}_a} = -1 \,. \tag{11.1}$$

Superconductors are not only ideal conductors, they are also **ideal diamagnets**.

If the external field is increased, the shielding breaks down at a **critical magnetic field** $\mathbf{B}_c$ and a transition into the normal state takes place. The shielding behavior below the critical field is shown schematically in Figure 11.3.

2 Walter Meissner, *1882 Berlin, †1974 Munich

3 Robert Ochsenfeld, *1901 Helberhausen, †1993 Helberhausen

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

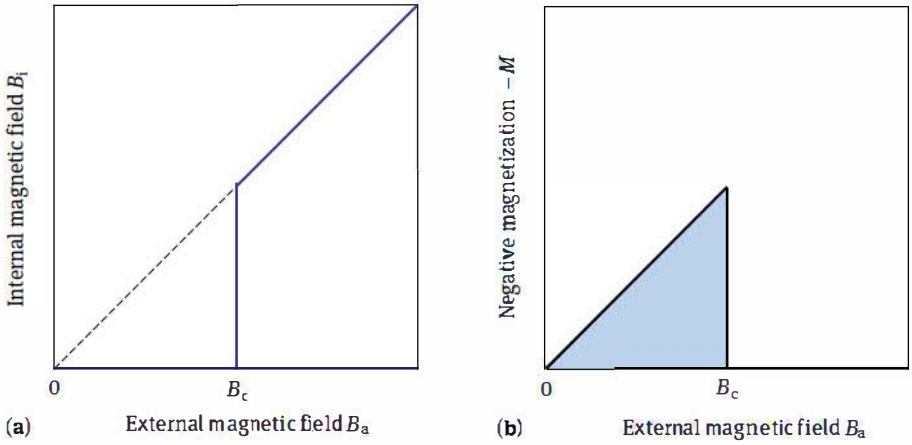


Fig. 11.3: **a)** The magnetic field $\mathbf{B}_i$ in a long superconducting cylinder. At $\mathbf{B}_c$ the superconductivity breaks down. **b)** The *negative* magnetization as a function of the applied field $\mathbf{B}_a$.

If the demagnetization field has to be taken into account, the situation is considerably less clear. An **intermediate state** can occur in which the sample is partly superconducting and partly normal. The geometrical arrangement of the different areas can be very complex. However, we will not pursue this interesting aspect of superconductivity further here.

The critical field depends on temperature, as depicted in Figure 11.4 for a number of superconductors. The temperature dependence can be approximated by the equation

$$B_c(T) = B_c(0)\left[1 - \left(\frac{T}{T_c}\right)^2\right] . \tag{11.2}$$

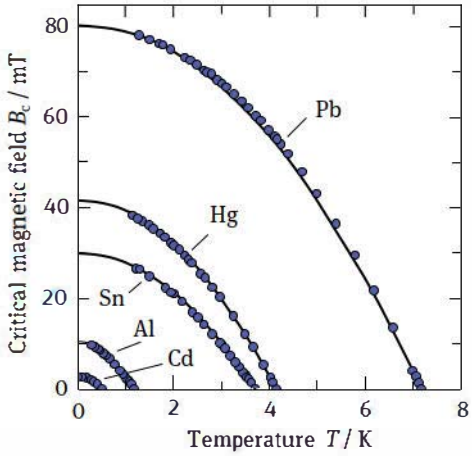


Fig. 11.4: The critical magnetic field as a function of temperature for a number of superconductors. The solid lines show the behavior expected from equation (11.2). (The data were taken from various publications).

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As can be seen in the figure, the critical field strength $B_c(0)$ is approximately proportional to the transition temperature T_c, which is often referred to as *Transition temperature*.

Now we will discuss the question, how the vanishing resistance and ideal diamagnetism can be dealt with in the framework of Maxwell's equations? The first requirement is easily met by omitting the scattering term in the equation of motion (9.19) resulting in $m\dot{\mathbf{v}} = -e\,\mathcal{E}$. We put $\mathbf{j} = -en\mathbf{v}$ into this relationship and for the time derivative of the supercurrent density $\mathbf{j}_s$ we get:

$$\frac{\mathrm{d}\mathbf{j}_s}{\mathrm{d}t} = \frac{n_s e_s^2}{m_s}\,\mathcal{E}\,. \tag{11.3}$$

This is the **first London equation**[4]. We have "temporarily" given the abbreviations for the charge carrier density n_s, the charge e_s and the mass m_s the subscript "s" because at this stage it is not clear what entity carries the supercurrent. We will get the answer to this question when we discuss the microscopic origin of superconductivity in Section 11.2. Here it is not the current density which is proportional to the electric field as in the case of Ohm's law, but rather its change with time. The equation gives the impression that the current density could increase arbitrarily. However, the infinitely high direct current conductivity of superconductors means that in the stationary case, the voltage drop along the sample and thus the change in current over time disappears. The current is then determined by the current source.

Now we put equation (11.3) into the Maxwell equation $\mathrm{curl}\,\mathcal{E} = -\partial\mathbf{B}/\partial t$ and get

$$\frac{\partial}{\partial t}\left(\mathrm{curl}\,\mathbf{j}_s + \frac{n_s e_s^2}{m_s}\mathbf{B}\right) = 0\,. \tag{11.4}$$

This equation applies to all materials with ideal conductivity and permits both constant and exponentially decaying solutions. Meissner-Ochsenfeld effect tells us that in a superconductor, not only the time derivative of the expression in the bracket must be zero, but the expression itself. Thus we obtain the **second London equation**[5], which connects the current flow with the magnetic field in the superconductor:

$$\mathrm{curl}\,\mathbf{j}_s = -\frac{n_s e_s^2}{m_s}\mathbf{B}\,. \tag{11.5}$$

Of course Maxwell's equations remain valid, but in the case of superconductors, they are supplemented by the second London equation.

In describing the Meissner-Ochsenfeld effect, we have argued that the magnetic field is kept completely out of the superconductor. However, this cannot be the case due to (11.5), since shielding currents also require the presence of magnetic fields.

4 Heinz London, *1907 Bonn, †1970 Oxford

5 In the literature the naming is not fully consistent. Occasionally, this equation alone is referred to as London equation.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

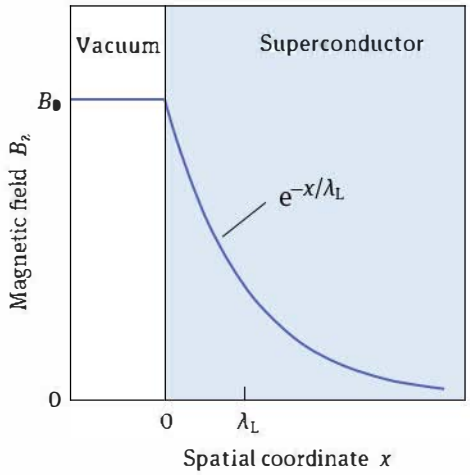


Fig. 11.5: Variation of the magnetic field inside a superconductor occupying the half space $x > 0$. The exponential decrease of the field strength is determined by the London penetration depth λ_L.

In fact, the magnetic field penetrates the sample slightly at the surface. To treat this phenomenon, we consider a superconductor as shown in Figure 11.5, occupying the half space $x > 0$ and subject to a magnetic field B_0 in the z-direction. If we insert (11.5) into the Maxwell equation $\operatorname{curl} \mathbf{B} = \mu_0 \mathbf{j}_s$, for the given geometry we obtain the differential equation:

$$\frac{d^2 B_z(x)}{dx^2} - \frac{\mu_0 n_s e_s^2}{m_s} B_z(x) = 0 \,. \tag{11.6}$$

This equation is readily solved yielding the following result for the magnetic field and the density of the shielding current:

$$B_z(x) = B_0 \, e^{-x/\lambda_L} \qquad \text{and} \qquad j_{s,y}(x) = j_{s,0} \, e^{-x/\lambda_L} \,, \tag{11.7}$$

where the **London penetration depth** λ_L is determined by

$$\lambda_L = \sqrt{\frac{m_s}{\mu_0 n_s e_s^2}} \,. \tag{11.8}$$

The current density and the magnetic field decrease exponentially into the superconductor, with λ_L playing the role of a characteristic length. The applied field and the maximum current density are linked by the relation $j_{s,0} = B_0/\mu_0 \lambda_L$.

To make a crude estimate of the penetration depth, we assume that at absolute zero, all the conduction electrons contribute to the superconductivity. We then obtain a value of the order of $\lambda_L \approx 15$ nm for the penetration depth. It can be determined by various experimental methods, for example by measuring the susceptibility of small superconducting samples with a well-defined shape. The result of such a measurement on thin lead cylinders is shown in Figure 11.6. From this experiment, we see that, in agreement with theoretical models not discussed here, the penetration depth diverges

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

at the transition temperature like

$$\lambda_L(T) = \frac{\lambda_L(0)}{\sqrt{1-(T/T_c)^4}} \tag{11.9}$$

and thus according to (11.8) the superconducting carrier density n_s must disappear at the transition temperature, as we would expect.

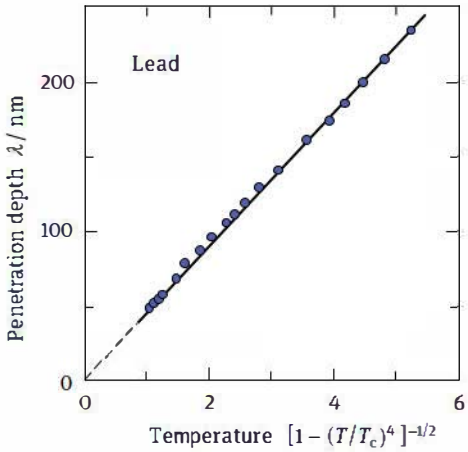


Fig. 11.6: Temperature dependence of the penetration depth λ in lead. The straight line represents the theoretical prediction. (After R.F. Gasparovic, W.L. McLean, Phys. Rev. B **2**, 2519 (1970).)

Surprisingly, however, in many experiments a greater penetration depth is found than predicted by London's theory. This is not due to a fundamental error in this theory, but arises from the fact that the current and the magnetic field can change drastically over very short distances. A similar problem occurs when applying Ohm's law at high frequencies. As soon as the mean free path of the electrons becomes comparable with the wavelength of the electromagnetic wave, the electric field must be averaged over the order of magnitude of the mean free path when calculating the resistance. A *non-local description* of superconductivity, in which the current density is not determined by the local value of the magnetic field but by an averaged value, was devised by *B. Pippard*[6] which leads to good agreement with experiment.

We discuss Pippard's arguments briefly here without deriving the corresponding theoretical relations. As we have seen, according to equation (11.5), in superconductors the current density $\mathbf{j}$ is linked to the magnetic field $\mathbf{B}$ (or the vector potential $\mathbf{A}$). If large changes in the vector potential occur over short distances, we must take into account that the current flow $\mathbf{j}$ at location $\mathbf{r}$ is also influenced by the field in the neighborhood. An important consequence of this phenomenon is that the penetration depth λ of the magnetic field can deviate considerably from the London penetration depth λ_L.

6 Alfred Brian Pippard, *1920 London, †2008 Cambridge

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

It follows from the non-local theory that when calculating the current at location **r**, the magnetic flux up to a distance r_0 must be taken into account. This relation is given by

$$\frac{1}{r_0} = \frac{1}{\ell} + \frac{1}{\xi_0} , \tag{11.10}$$

where two parameters define this distance, one, the mean free path ℓ of the conduction electrons, which is limited by the scattering at defects, two, the **coherence length** ξ_0, which is defined by

$$\xi_0 = 0.18 \frac{\hbar v_F}{k_B T_c} . \tag{11.11}$$

The coherence length reflects the extent of the so-called "Cooper pairs" which form the basis of superconductivity, as we will see in the next section.

The above corrections can be neglected if $\lambda \gg \xi_0$ and $\ell \gg \xi_0$. The superconductor under consideration is then in the *London limit* in which the penetration depth is described by (11.8). Interestingly, while ξ_0 and ℓ are temperature-independent variables, λ_L increases with increasing temperature according to (11.9) and reaches an infinite value at T_c. Therefore, there is always a temperature range in which the London limit applies to the superconductor under consideration.

If $\lambda \ll \xi_0$ and $\ell \gg \xi_0$, then the Pippard limit provides the correct description. In this case, the penetration depth can be approximated by $\lambda \approx 0.65\,(\lambda_L^2\,\xi_0)^{1/3}$. This means that the actual penetration depth is greater than that predicted by the London theory. Many "classical" superconductors, such as pure aluminum with $\lambda = 45$ nm and $\xi_0 = 1550$ nm, fall into this category (see Table 11.3). In the *dirty limit*, where the mean free path is reduced by defects as in alloys for example, $\ell \ll \lambda$. In this case, Pippard's theory yields the relation $\lambda \approx \lambda_L(1 + \xi_0/\ell)^{1/2}$. It follows that the reduction of the mean free path increases the penetration depth.

Tab. 11.3: London penetration $\lambda_L(0)$ and coherence length ξ_0 for several superconducting metals. (After C.P. Poole, *Handbook of Superconductivity* Academic Press, 2000.)

Material	T_c (K)	$\lambda_L(0)$ (nm)	ξ_0 (nm)	$\lambda_L(0)/\xi_0$
Al	1.18	45	1550	0.03
Sn	3.72	42	180	0.23
Pb	7.20	39	87	0.48
Nb	9.25	52	39	1.3
Nb_3Ge	23.20	90	3	30
K_3C_{60}	19.4	240	2.8	95

Figure 11.7 shows the temperature dependence of the penetration depth, according to the non-local theory, for the three special cases discussed in the text. Here the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

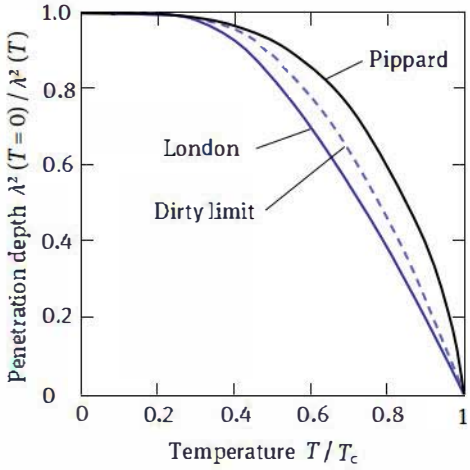


Fig. 11.7: Penetration depth as a function of temperature. The temperature variation of $1/\lambda^2$ is plotted for the three limiting cases described in the text. (After I.R. Waldram, Superconductivity of Metals and Cuprates, IOP Publishing, 1996).)

normalized quadratic penetration depth $\lambda^2(T=0)/\lambda^2(T)$ is plotted as a function of temperature.

Figure 11.8 shows the penetration depth of the magnetic field in tin as a function of the electron mean free path. The penetration depth clearly decreases with increasing mean free path, i.e. with increasing purity of the metal.

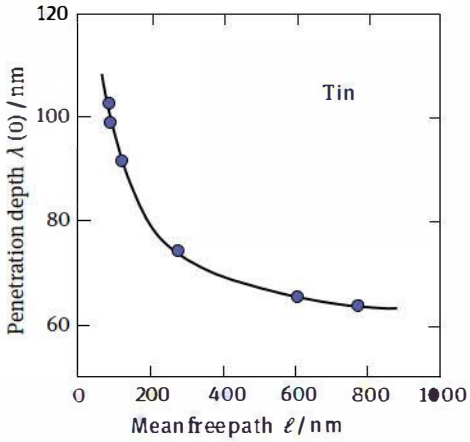


Fig. 11.8: The variation of the penetration depth of tin as a function of the electron mean free path. (After B.A. Pippard, Proc. R. Soc. A **216**, 547 (1953).)

11.1.2 Thermodynamic Properties

As we have seen, high magnetic fields destroy superconductivity. A simple thermodynamic argument shows that there must be a critical magnetic field B_c above which

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the superconductivity cannot exist. For the magnetic properties, the *Gibbs free energy*[7] $G(T, p, B) = U - TS + pV - V\mathbf{M}\cdot\mathbf{B}$ determines the thermodynamic state function. Given that in superconductors the magnetization **M** always runs parallel or antiparallel to the field **B**, we can omit the vector notation in this chapter and thus simplify our formulas. With no magnetic field, the superconducting state below T_c is more stable than the normal conducting state, and thus $G_n(B = 0, T) > G_s(B = 0, T)$. Here the indices "n" and "s" denote the normal and the superconducting states, respectively. For the change dG, from internal energy considerations $\mathrm{d}U = T\,\mathrm{d}S - p\,\mathrm{d}V + VB\,\mathrm{d}M$, we find the expression:

$$\mathrm{d}G = -S\mathrm{d}T + V\mathrm{d}p - VM\,\mathrm{d}B\,. \tag{11.12}$$

In the following, we can assume constant pressure so that the term $V\mathrm{d}p$ disappears, and again, we can avoid the complication of the demagnetization field by taking a long cylinder whose axis is parallel to the magnetic field. Using the susceptibility $\chi = -1$ for the superconductor, equation (11.12) at constant pressure thus reduces to:

$$\mathrm{d}G_s = -S\,\mathrm{d}T + \frac{V}{\mu_0}\,B\,\mathrm{d}B\,. \tag{11.13}$$

The equation is readily integrated to yield:

$$G_s(B,T) = G_s(0,T) + \int_0^B \frac{V}{\mu_0}\,B'\,\mathrm{d}B' = G_s(0,T) + \frac{VB^2}{2\mu_0}\,. \tag{11.14}$$

Clearly the expulsion of the magnetic field must cause an increase in the Gibbs energy, the magnetic field dependence of which is shown in Figure 11.9 for both normal metals and superconductors.

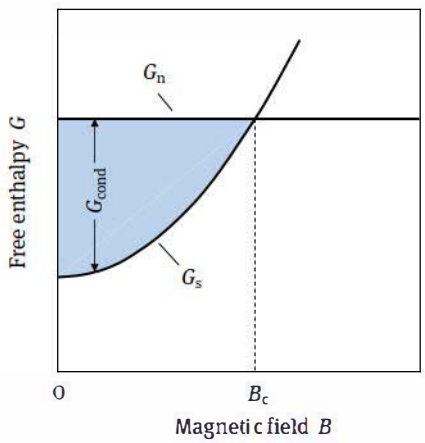


Fig. 11.9: The Gibbs free energy for normal and superconducting metals as a function of the magnetic field. The energy difference between the two states, G_{cond}, represents the condensation energy gained when the system becomes superconducting. The intersection of the two curves defines the critical field B_c.

7 Josiah Willard Gibbs, *1839 New Haven (Conn.), †1903 New Haven (Conn.)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

For non-magnetic normally conducting metals, the magnetic field-dependent contribution to the Gibbs free energy is very small, since free electrons only contribute via the Pauli spin susceptibility and the Landau diamagnetism (cf. Section 12.2). We neglect this contribution and write $G_n(B,T) \approx G_n(0,T)$. With increasing field, the Gibbs free energy G_s of the superconductor increases and finally reaches the value G_n for that of the normal metal. At this point, the energy gained by the creation of the superconducting state is fully compensated by the energy required for field expulsion. Above this field the superconducting phase is no longer stable and a transition to the normal state occurs. For the Gibbs free energy at the phase transition we can therefore write $G_s(B_c,T) = G_n(B_c,T) = G_n(0,T)$. Thus the important relationship follows from Equation (11.14)

$$G_n(0,T) - G_s(0,T) = G_{cond} = \frac{V B_c^2(T)}{2\mu_0} \,. \tag{11.15}$$

Since the temperature dependence of the critical magnetic field is known, this gives us the temperature dependence of G_s, which is shown in Figure 11.10a. The Gibbs free energy difference $G_{cond} = VB_c^2/2\mu_0$ plays an important role in the theory of superconductivity and is called the **condensation energy** for reasons we will discuss later. Here it should be noted that G_{cond} is relatively small, since a critical field of 50 mT results in $G_{cond} \approx 10^3\,\mathrm{Jm}^{-3}$.

From a knowledge of the condensation energy, we can draw a number of conclusions about the various temperature dependencies of the entropy in the normal and superconducting state through the relationship $S = -(\partial G/\partial T)$. Differentiating (11.15), we obtain:

$$\Delta S = S_n - S_s = -\frac{VB_c}{\mu_0}\frac{dB_c}{dT} \,. \tag{11.16}$$

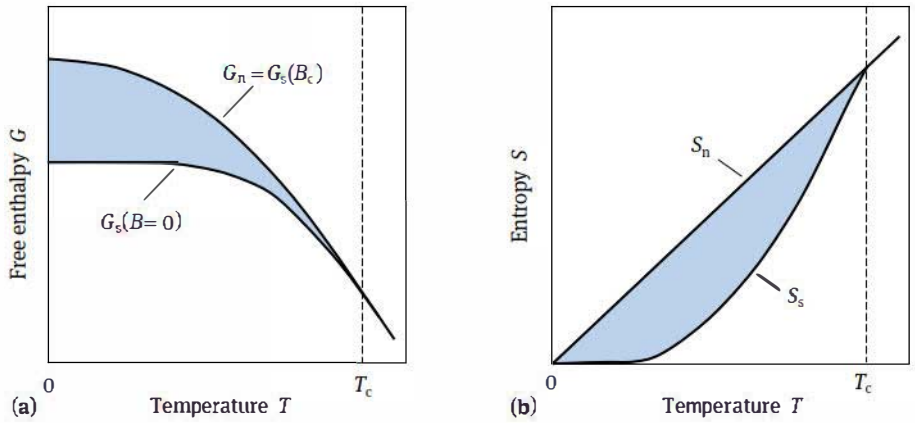


Fig. 11.10: Schematic representation of the Gibbs free energy **(a)** and the entropy **(b)** as a function of temperature for a metal in the normal and superconducting state.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

At low temperatures the entropy S_n of normal metals is determined by the free electrons. Since the same mathematical expression describes both the specific heat and the entropy for normal conductors, S_n also increases linearly with temperature. Figure 11.10b compares the temperature dependence of the entropy of a material in the normal and the superconducting state.

We can draw some interesting conclusions from Equation (11.16): the critical magnetic field B_c disappears at the transition temperature and with it the entropy difference. Consequently, the latent heat $\Delta Q = T_c(S_n - S_s) = 0$. This means that the transition from the normal to the superconducting state is not a first-order but a **second-order phase transition**. This is the starting point of a phenomenological theory of superconductivity by *V.L. Ginzburg* and *L.D. Landau*, which we will discuss later. For $T \to 0$, both dB_c/dT and ΔS go to zero in accordance with the third law of thermodynamics. Since in the temperature range between zero and T_c, $dB_c/dT < 0$ and thus $\Delta S > 0$, it follows directly that in the superconducting state the material is more ordered than in the normal state. Since $\Delta S \neq 0$, this implies that in magnetic fields there will be emission or absorption of latent heat during a transition between the two states, and the phase transition is therefore of first order in a magnetic field.

The specific heat is given by $C_p = T(\partial S/\partial T)_{p,B}$ and thus

$$\Delta C = C_s - C_n = \frac{VT}{\mu_0}\left[B_c \frac{d^2 B_c}{dT^2} + \left(\frac{dB_c}{dT}\right)^2\right] . \qquad (11.17)$$

From this it follows that the specific heats of the normal and the superconducting states differ at T_c, because at the transition temperature $dB_c/dT \neq 0$. We therefore expect that the specific heat of a superconductor will show a discontinuity at T_c. This is confirmed by the results of Figure 11.11, showing the specific heat of aluminum in the normal and superconducting states. We will come back to this figure in Section 11.2. For the moment, the main thing to note is that the specific heat at T_c changes abruptly, as expected.

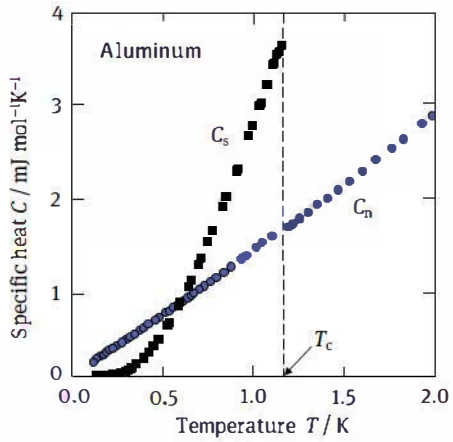


Fig. 11.11: The specific heat of aluminum in both the normal and the superconducting state. To allow the specific heat to be measured in the normal state, the superconductivity was suppressed by an external magnetic field of about 50 mT. The data in the normal state only deviate slightly from linear, showing that the contribution from the lattice plays only a minor role at these temperatures. (After N.E. Phillips, Phys. Rev. **114**, 676 (1959).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

11.2 Microscopic Description

The development of a microscopic theory proved to be very difficult. The description of a phenomenon that can be found in solids with completely different structures required entirely new concepts going beyond the one-electron approximation. During the long interval between the actual discovery of superconductivity and realizing a microscopic explanation, the theory has advanced in several stages. The first theories had a phenomenological character and were closely linked to the names of *F.* and *H. London*, *C.J. Gorter*[8] and *H.B.G. Casimir* in the 1930s, and *V.L. Ginzburg*[9] and *L.D. Landau* in the 1950s. A certain degree of completion was achieved in 1957 with the BCS theory by *J. Bardeen, L.N. Cooper*[10] and *J.R. Schrieffer*[11], the microscopic description of superconductivity.

11.2.1 Cooper Pairs

In normal conductors, the electrical conductivity is finite because the excitation spectrum of the free electrons allows for the transfer of arbitrarily small energies between the electrons and the lattice. It is therefore reasonable to assume that the electron spectrum of superconductors must be modified in some way compared to that of normal conductors. A possible mechanism was suggested by *L.N. Cooper* in 1956. He was able to show that in the case of an (arbitrarily small) *attractive* interaction between two electrons, the ground state of the Fermi gas is no longer stable and the energy of the two electrons is lowered. In other words: if two electrons which do not interact are brought to the surface of the "Fermi sea", i.e. at the energy E_F into a metal, they must remain there because the lower lying states are occupied. If there is an attractive interaction between the two electrons, then on approaching they can form a bound pair, reducing their energies to lie below the Fermi energy. Due to the Coulomb repulsion, however, an attractive interaction between two electrons seems difficult to imagine at first.

From the observation of the **isotope effect** it has been known since 1950 that the transition temperature of a superconducting metal depends on the mass of the atoms and thus on the lattice properties. This observation turned out to be of fundamental importance in the development of the microscopic theory of superconductivity. A particularly good example of this is the measurement of T_c for various mixtures of tin isotopes, as shown in Figure 11.12. The transition temperature is clearly proportional to the square root of the reciprocal atomic mass M and thus proportional to the Debye frequencies of the samples investigated (see equations (6.28) and (6.85)).

8 Cornelius Jacobus Gorter, *1907 Utrecht, †1980 Leiden
9 Vitaly Lazarevich Ginzburg, *1916 Moscow, †2009 Moscow
10 Leon Neil Cooper, *1930 New York, Nobel Prize 1972
11 John Robert Schrieffer, *1931 Oak Park, †2019 Tallahassee, Nobel Prize 1972

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

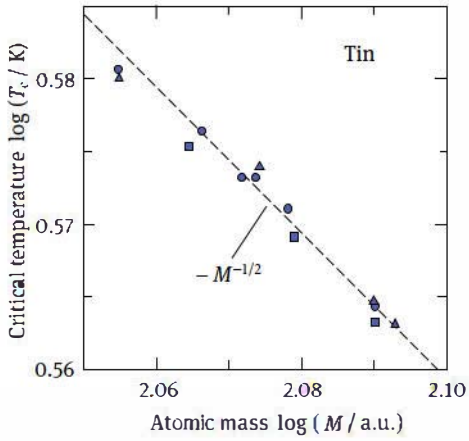


Fig. 11.12: The isotope effect in tin. The transition temperature is plotted as a function of the isotopic mass on a logarithmic scale. The data are taken from various publications and include measurements on both isotropically pure and mixed samples.

That the lattice can mediate an attraction between electrons can be illustrated by a rough picture as follows: as an electron moves through the lattice, it attracts the positively charged ions as it passes. The heavier ions respond more slowly with the result that the electron leaves a positively-charged cloud in its "wake" which in turn can attract other electrons. While the force exerted when an electron passes propels the ions in the direction of the electron path, the highest resulting positive charge density is only reached after a quarter of the period of the ion oscillation. In the meantime, the first electron has traveled about 100 nm. The delayed reaction of the ions results in the interacting electrons being relatively far apart. Thus the Coulomb repulsion between these electrons is relatively weak. It should also be noted that in metals, the range of the Coulomb repulsion between two electrons is greatly reduced by the shielding effects. As we have seen in Section 8.3, the Coulomb repulsion is replaced by the expression (8.40) derived by Thomas and Fermi.

In the theoretical description of the interaction potential, the concept of particle exchange is used. The two interacting electrons with wave vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ exchange *virtual phonons* with wave vector $\mathbf{q}$. After a phonon exchange the two electrons with the original wave vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ have the wave vectors $\mathbf{k}_1' = (\mathbf{k}_1 + \mathbf{q})$ and $\mathbf{k}_2' = (\mathbf{k}_2 - \mathbf{q})$ respectively. The total momentum $\mathbf{K}$ is maintained so that $(\mathbf{k}_1 + \mathbf{k}_2) = (\mathbf{k}_1' + \mathbf{k}_2') = \mathbf{K}$. At absolute zero, all states below the Fermi energy are occupied and for the two interacting electrons only states above E_F are accessible. The interaction thus takes place in the energy range from E_F to $(E_F + \hbar\omega_D)$, where $\hbar\omega_D$ stands for the energy of the phonons at the Debye frequency. In $\mathbf{k}$-space, this energy range corresponds to a spherical shell with the thickness $\delta k = (m\omega_D/\hbar k_F)$. This restriction is illustrated in Fig. 11.13. Since, at a given $\mathbf{K}$, only electron pairs whose wave vectors begin and end in the dark blue overlap areas fulfill the conservation of momentum, it is clearly evident that for $\mathbf{K} = 0$ the phonon exchange occurs with the greatest possible probability. Then the electrons involved have access not only to a small section hut to the entire spherical shell. This

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

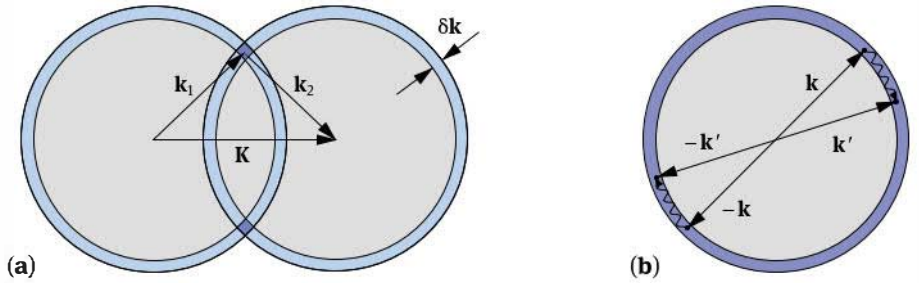


Fig. 11.13: **a)** Sketch of the conservation of momentum for pair interaction. The conservation of momentum is only fulfilled if the wave vectors of the electrons involved begin respectively end within the two dark blue regions. **b)** Typical scattering process for a Cooper pair. Here the wave vector **K** of the center of gravity motion is zero.

results in the important statement that for the wave vectors of the two electrons of the **Cooper pairs** the condition $\mathbf{k}_1 = -\mathbf{k}_2$ must be fulfilled. In the further course of the discussion we will use the notation $(\mathbf{k}, -\mathbf{k})$ for such a pair.

As in the case of the H_2-molecule, which also has a *pair of electrons* in its bond, a two-particle wave function $\Psi(\mathbf{r}_1,\mathbf{r}_2)$ is required to describe the Cooper pairs. As ansatz, we take a linear combination of single-particle functions, whereby we assume plane waves for the single particle functions. Since the electrons of the pair have opposite momenta, the center of gravity remains at rest and only relative coordinates appear in the description. The structure of the wave function becomes clear if we take an electron pair with the wave vectors $\mathbf{k}_1 = \mathbf{k}$ or $\mathbf{k}_2 = -\mathbf{k}$, represent it by plane waves and form the two-particle function or the *pair wave function*: $\Psi = A\,\exp(\mathrm{i}\mathbf{k}_1 \cdot \mathbf{r}_1) \cdot \exp(\mathrm{i}\mathbf{k}_2 \cdot \mathbf{r}_2) = A\,\exp[\mathrm{i}\mathbf{k}(\mathbf{r}_1 - \mathbf{r}_2)] = A\,\exp(\mathrm{i}\mathbf{k} \cdot \mathbf{r})$. Forsimplification, we have replaced the position vectors in the last expression with $\mathbf{r} = (\mathbf{r}_1 - \mathbf{r}_2)$.

Through interaction with the lattice, the electron pairs are constantly scattered into new states with different wave vectors. We therefore use a superposition of such pair states as a solution of the Schrödinger equation:

$$\Psi(\mathbf{r}) = \sum_{\mathbf{k}} A_{\mathbf{k}} \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \,. \tag{11.18}$$

$|A_{\mathbf{k}}|^2$ is a measure for the probability of finding an electron pair in the state $(\mathbf{k}, -\mathbf{k})$. The wave number of the interacting electrons is in the range $k_F < k < [2m(E_F + \hbar\omega_D)/\hbar^2]^{1/2}$ if we restrict ourselves to the limiting case $T = 0$. The coefficient $A_{\mathbf{k}}$ therefore is zero for all values of the wave vector $\mathbf{k}$ outside of this range.

To calculate the eigenvalue E we use the Schrödinger equation

$$\left[-\frac{\hbar^2}{2m}(\Delta_1 + \Delta_2) + \widetilde{V}(\mathbf{r}_1, \mathbf{r}_2)\right] \Psi(\mathbf{r}_1, \mathbf{r}_2) = E\,\Psi(\mathbf{r}_1, \mathbf{r}_2) \,. \tag{11.19}$$

The potential $\widetilde{V}(\mathbf{r}_1,\mathbf{r}_2)$ has two components: the attractive interaction between the pair of electrons via the lattice as just discussed, and the Coulomb repulsion between

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the two electrons. The latter is, however, considerably reduced by the large distance between the interacting electrons and by shielding effects as we have argued before. The actual potential curve is only roughly known, but is of no great importance for our further discussion.

We solve the Schrödinger equation (11.19) with the usual methods of quantum mechanics: we use the ansatz (11.18), multiply by $\exp(-i\mathbf{k}' \cdot \mathbf{r})$ and integrate over the sample volume. As in the derivation of the diffraction condition in Section 4.4, the integral over $\exp[i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}]$ disappears for $\mathbf{k} \neq \mathbf{k}'$ and is equal to the sample volume V_s for $\mathbf{k} = \mathbf{k}'$. Since the interaction depends only on the distance between the electrons, we may write $\tilde{V}(\mathbf{r}_1, \mathbf{r}_2) = \tilde{V}(\mathbf{r})$ and obtain the solution

$$\frac{\hbar^2\mathbf{k}^2}{m} A_{\mathbf{k}} + \frac{1}{V_s} \sum_{\mathbf{k}'} A_{\mathbf{k}'} \int \tilde{V}(\mathbf{r})\, e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} dV = E A_{\mathbf{k}} , \tag{11.20}$$

$$\left(\frac{\hbar^2\mathbf{k}^2}{m} - E \right) A_{\mathbf{k}} = -\frac{1}{V_s} \sum_{\mathbf{k}'} A_{\mathbf{k}'} \tilde{V}_{\mathbf{k}\mathbf{k}'} . \tag{11.21}$$

In the last equation, the abbreviation

$$\tilde{V}_{\mathbf{k}\mathbf{k}'} = \int \tilde{V}(\mathbf{r})\, e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} dV \tag{11.22}$$

was introduced to simplify the representation of the matrix element. The exact course of $\tilde{V}(\mathbf{r})$ is not important for the understanding of most phenomena, so that we make the simplifying assumption that in the energy range $E_F < \hbar^2k^2/2m, \hbar^2k'^2/2m < (E_F + \hbar\omega_D)$ the matrix element can be represented by a constant $\tilde{V}_{\mathbf{k}\mathbf{k}'} = -\tilde{V}_0$. Assuming an attractive interaction $\tilde{V}_0$ is a positive constant. With this strong simplification, the right side of equation (11.21) is independent of $\mathbf{k}$, and is thus constant. If we rearrange with respect to $A_{\mathbf{k}}$, we find

$$A_{\mathbf{k}} = \frac{\tilde{V}_0}{V_s} \frac{1}{(\hbar^2\mathbf{k}^2/m - E)} \sum_{\mathbf{k}'} A_{\mathbf{k}'} . \tag{11.23}$$

This equation is further simplified if we sum over all $\mathbf{k}$ vectors. Since the result of the summation does not depend on the naming of the wave vectors, $\sum_{\mathbf{k}} A_{\mathbf{k}} = \sum_{\mathbf{k}'} A_{\mathbf{k}'}$ applies. Thus we obtain

$$1 = \frac{\tilde{V}_0}{V_s} \sum_{\mathbf{k}} \frac{1}{(\hbar^2\mathbf{k}^2/m - E)} . \tag{11.24}$$

We replace the remaining summation over the wave vectors by an integration about the energy, as we have already done repeatedly in the previous chapters. Using the abbreviation $z = \hbar^2\mathbf{k}^2/2m$ for the kinetic energy of an electron, the resulting equation can be written as:

$$1 = \tilde{V}_0 \frac{D(E_F)}{2} \int_{E_F}^{E_F+\hbar\omega_D} \frac{dz}{2z - E} . \tag{11.25}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In this expression we have assumed the density of states in the vicinity of the Fermi edge to be constant which allows us to put it in front of the integral. The factor 1/2 occurs because equation (11.19) refers to pairs, but $D(E)$ describes the density of states in the one-electron approximation. If we perform the integration and calculate the energy change ΔE that an electron pair in the Fermi sea experiences due to the attractive interaction, we obtain

$$\Delta E = E - 2E_F = \frac{2\hbar\omega_D}{1 - e^{4/[\tilde{V}_0 D(E_F)]}} \approx -2\hbar\omega_D \, e^{-4/[\tilde{V}_0 D(E_F)]} . \tag{11.26}$$

The energy change ΔE is negative, i.e. the energy of the Cooper pairs is reduced by the indirect interaction between the two electrons. The last expression is only valid in the limiting case of weak coupling $\tilde{V}_0 D(E_F) \ll 1$, but this condition is usually fulfilled. At the surface of the Fermi sea, two-electron states whose energy is lowered by δE (compared to the energy of the free electrons at $T = 0$) are formed. This instability of the Fermi sea causes a transition from the ground state of the normal conductor to a new one, the **BCS ground state**, which we discuss briefly in the next section.

The wave function (11.18) tells us that no defined wave vectors can be assigned to the electrons of a Cooper pair. The wave function contains all wave vectors in the energy range from E_F to $(E_F + \hbar\omega_D)$. It follows from (11.23) that $A_{\mathbf{k}}$ is greatest for those states whose kinetic energy $z = \hbar^2 k^2/2m$ is comparable with E_F. Of course this result is also valid for electron pairs scattered from states below the Fermi energy to states above. Although their kinetic energy increases, the decrease in potential energy outweighs this, so the electrons remain in the bound state.

Equation (11.26) allows an explanation of the seemingly paradoxical observation that superconductivity has not yet been found in "good" metals such as silver or copper. These metals have a high electrical conductivity at room temperature because the electrons only weakly couple to phonons and are therefore hardly scattered. However, the weak electron-phonon coupling also means that the attraction between pairs of electrons via the exchange of virtual phonons is not very strong. In these metals, superconductivity will therefore only occur at extremely low temperatures, if at all.

Of course we still have to consider that the two electrons are indistinguishable fermions. The total wave function of the Cooper pairs must therefore be antisymmetrical. The wave function of the solution (11.18) is symmetrical with respect to the exchange of the two electrons. Therefore the spin part of the wave function (not represented in equation (11.18)) needs to be anti-symmetric, i.e. the spins are oppositely oriented. We indicate this symbolically by $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ and speak of a *singlet pair*, since the total angular momentum of the pair is zero. Outwardly, the Cooper pairs therefore behave like bosons[12]. In particular, the Cooper pairs can assume a *common*

12 It should be pointed out, however, that on closer inspection, there are certain differences compared to "real" bosons, but we will not go into these aspects in more detail here.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

quantum mechanical state, which is not possible for fermions due to the antisymmetric wave function. We will discuss the consequences of this in the following sections.

If the interaction between the electrons is not isotropic, as was previously assumed, a spin alignment in the same direction is also possible, resulting in *triplet pairs*. In this case, the spatial wave function of the pairs must be antisymmetric, like in the case of a p-wave function. In fact, such pair formation is found in the **suprafluid He^3**, whose exotic (naturally non-metallic) properties at temperatures around 1 mK are based on the existence of such Cooper pairs formed by He^3 atoms. The structure of Cooper pairs is also more complicated for a number of intermetallic compounds, namely for **heavy fermion systems** and for high-temperature superconductors. While the Cooper pairs in **high-temperature superconductors** have a d-wave function, there are indications that some heavy fermion systems have Cooper pairs with a p-wave function.

If one calculates with the help of (11.18) and (11.19) the average distance between the electrons of a pair, one finds a value which can be easily estimated: from the energy uncertainty δE of a Cooper pair follows for the uncertainty of the wave number $\delta k \approx (m\,\delta E/\hbar^2 k_F)$ and thus a minimum extension $\delta x \approx 1/\delta k$ of the Cooper pairs of the order $\delta x \approx (\hbar^2 k_F/m\,\delta E)$. Since the energy uncertainty cannot be greater than the binding energy of the Cooper pairs, and the binding energy is comparable with the thermal energy at the transition temperature, we set $\delta E \approx k_B T_c$ as a rough approximation and obtain $\delta x \approx (\hbar v_F/k_B T_c)$. Except for the numerical pre-factor, this result corresponds to the expression for the coherence length ξ_0 in equation (11.11), which is obviously determined by the size of the Cooper pairs. If we use numerical values, we find that the extension of the pairs is in the range of 100 to 1000 nm! Cooper pairs are therefore relatively large. Between the two electrons of a pair there are millions of other electrons.

11.2.2 BCS Theory

BCS Ground State. In the previous section it was shown that an attractive interaction between free electrons can lead to pair formation, which is associated with an energy reduction. The theoretical description of the whole system is mathematically more complex than the treatment of a single Cooper pair, because the behavior of *all* free electrons is involved. We do not want to reproduce the theory in its mathematical details here, but only present the basic features and try to make them plausible. First of all, we consider the BCS ground state, i.e. superconductivity at absolute zero.

Starting point of the theoretical description of the BCS ground state is a suitable ansatz for the wave function. Of course, the one-electron states cannot be used for this. Instead the pair states which were used for the construction of the wave function of the Cooper pairs can be used. If we denote the probability for the occupation of the state $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ with $v_\mathbf{k}^2$, $u_\mathbf{k}^2 = (1 - v_\mathbf{k}^2)$ is the probability, that this pair state is not occupied. The total energy W_0 of the BCS ground state, which involves *all* Cooper pairs is not determined solely by the kinetic energy of the electrons as in the case of free

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

electrons. The (negative) interaction energy resulting from the pair formation must also be taken into account. Therefore we write

$$W_0 = \sum_{\mathbf{k}} 2v_{\mathbf{k}}^2 \eta_{\mathbf{k}} - \frac{\widetilde{V}_0}{V_s} \sum_{\mathbf{k}',\mathbf{k}} v_{\mathbf{k}} u_{\mathbf{k}'} u_{\mathbf{k}} v_{\mathbf{k}'} \,. \tag{11.27}$$

The first term reflects the kinetic energy, where $\eta_{\mathbf{k}} = (\hbar^2\mathbf{k}^2/2m - E_F)$ stands for the difference between the kinetic energy of the electrons and the Fermi energy. The factor 2 takes into account that Cooper pairs consist of two electrons with opposite spin. The second term represents the interaction energy based on the scattering of the two-particle states $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ into the states $(\mathbf{k}'\uparrow, -\mathbf{k}'\downarrow)$. For the scattering to occur, the state $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ must be occupied, the state $(\mathbf{k}'\uparrow, -\mathbf{k}'\downarrow)$ must be empty. The probability amplitude for the initial state is therefore given by $v_{\mathbf{k}} u_{\mathbf{k}'}$, the final state by $v_{\mathbf{k}'} u_{\mathbf{k}}$.

Minimizing the energy W_0 with respect to $v_{\mathbf{k}}$ and $u_{\mathbf{k}}$ taking into account the relationship $(u_{\mathbf{k}}^2 + v_{\mathbf{k}}^2) = 1$ leads to

$$2u_{\mathbf{k}} v_{\mathbf{k}} \eta_{\mathbf{k}} - \Delta(u_{\mathbf{k}}^2 - v_{\mathbf{k}}^2) = 0 \,, \tag{11.28}$$

with the abbreviation

$$\Delta = \frac{\widetilde{V}_0}{V_s} \sum_{\mathbf{k}'} u_{\mathbf{k}'} v_{\mathbf{k}'} \,. \tag{11.29}$$

Now we will perform some simple transformations and introduce the new variable $E_{\mathbf{k}}$ which is defined by

$$E_{\mathbf{k}}^2 = \eta_{\mathbf{k}}^2 + \Delta^2 \,. \tag{11.30}$$

Using (11.28), the probability amplitudes $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ can be expressed as follows:

$$u_{\mathbf{k}}^2 = \frac{1}{2}\left(1 + \frac{\eta_{\mathbf{k}}}{E_{\mathbf{k}}}\right) \qquad \text{and} \qquad v_{\mathbf{k}}^2 = \frac{1}{2}\left(1 - \frac{\eta_{\mathbf{k}}}{E_{\mathbf{k}}}\right) \,. \tag{11.31}$$

With the new variables the energy W_0 of the ground state can be written in the following form:

$$W_0 = \sum_{\mathbf{k}} \eta_{\mathbf{k}} \left(1 - \frac{\eta_{\mathbf{k}}}{E_{\mathbf{k}}}\right) - \frac{\Delta^2 V_s}{\widetilde{V}_0} \,. \tag{11.32}$$

If we insert the relationship (11.30) in (11.31), we obtain for the probability $w_{\mathbf{k}} = v_{\mathbf{k}}^2$ that the pair state $(\mathbf{k}\uparrow, -\mathbf{k}\downarrow)$ is occupied the expression

$$w_{\mathbf{k}} = \frac{1}{2}\left(1 - \frac{\eta_{\mathbf{k}}}{\sqrt{\eta_{\mathbf{k}}^2 + \Delta^2}}\right) \,. \tag{11.33}$$

Figure 11.14 compares the occupation probability $w_{\mathbf{k}}$ of the pair states at $T = 0$ and the Fermi function $f(E,T)$ at T_c. The curves show amazing similarity. Even at absolute zero, the occupation of the states in the energy range $(E_F \pm \Delta)$ is "softened" in the superconductor. This is to be expected, since the interaction between electrons and

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

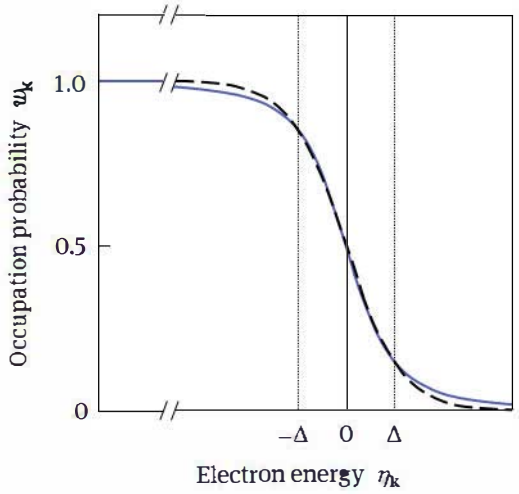


Fig. 11.14: Occupation probability of the pair states (blue curve) at absolute zero as a function of the single particle energy η_k. It is very similar to the Fermi function for the temperature $T = T_c$, which is shown as a dashed black line.

phonons can only occur in the vicinity of the Fermi edge. Therefore, the kinetic energy of the electrons in the superconductor is higher than in the normal conductor, when added up over all conduction electrons.

Even though the kinetic energy of the electrons of the superconductor is increased, there is still an overall energy gain at the transition from the normal to the superconducting phase, since the potential energy is lowered due to the interaction. If we denote the internal energy of a normal conductor by $W_0^n = 2\sum_{|k|<k_F} \eta_k$, the energy gain, the **condensation energy**, is given by $W_{cond} = (W_0 - W_0^n)$. After a series of algebraic transformations we find

$$\frac{W_{cond}}{V_s} = -\frac{1}{4} D(E_F)\Delta^2 . \tag{11.34}$$

This energy is released during the "condensation" of the conduction electrons into the superconducting state. This energy corresponds to the difference in Gibbs free energy between the normal and superconducting states, for which we have derived the expression $G_{cond}(0) = VB_c^2/2\mu_0$ in the previous section with the help of thermodynamic considerations. The different sign is due to the different choice of the reference system. Thus we have established a connection between the macroscopic and microscopic views.

In the next step we want to show that an *energy gap* appears in the energy spectrum of the superconductor. The expression (11.29) for the constant Δ can be rewritten by inserting (11.31) and we obtain

$$\Delta = \frac{1}{2}\frac{\widetilde{V}_0}{V_s}\sum_{\mathbf{k}} \frac{\Delta}{\sqrt{\eta_{\mathbf{k}}^2 + \Delta^2}} . \tag{11.35}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

If one replaces the sum by an integral and assumes again that the density of states is constant in the vicinity of the Fermi energy, this results in

$$1 = \frac{\widetilde{V}_0}{2} \int_{-\hbar\omega_D}^{\hbar\omega_D} \frac{D(E_F)}{2} \frac{d\eta}{\sqrt{\eta_{\mathbf{k}}^2 + \Delta^2}} = \frac{\widetilde{V}_0 D(E_F)}{2} \operatorname{arc\,sinh}\left(\frac{\hbar\omega_D}{\Delta}\right) . \tag{11.36}$$

Consequently, Δ is given by

$$\Delta = \frac{\hbar\omega_D}{\sinh\left[\frac{2}{\widetilde{V}_0 D(E_F)}\right]} \approx 2\hbar\omega_D \, e^{-2/[\widetilde{V}_0 D(E_F)]} . \tag{11.37}$$

The last expression is only valid for weak coupling, i.e. for $\widetilde{V}_0 D(E_F) \ll 1$. The great similarity with equation (11.26) is striking, in which the energy reduction during pair formation is described. As we will see, Δ represents an energy gap in the energy spectrum of the electrons, which is easily accessible to experiments and closely related to the binding energy of the Cooper pairs. We will discuss the exact meaning and consequences in detail. With the known Debye frequencies of superconductors and the experimental values of Δ, we typically find values around 0.6 for the parameter $\widetilde{V}_0 D(E_F)$.

Now we turn to the question of the significance of the quantity Δ and what kind of excited states occur in a superconductor. As mentioned above, Δ is the **energy gap** that occurs in the excitation spectrum of the electrons. In other words, the energies of the excited states are at least Δ above the ground state energy W_Φ. The simplest imaginable excited state is one in which only one state of the pair $(\mathbf{k}'\uparrow, -\mathbf{k}'\downarrow)$ is occupied. The energy of this excited state can be calculated with the help of equation (11.32), if we do not perform the summation over all values of $\mathbf{k}$, but only over $\mathbf{k} \neq \mathbf{k}'$. After some algebraic transformations, we get for the energy δE, which is needed to break up a pair

$$\delta E = 2E_{\mathbf{k}'} = 2\sqrt{\eta_{\mathbf{k}'}^2 + \Delta^2} . \tag{11.38}$$

Since $\eta_{\mathbf{k}'} = (\hbar^2 k'^2/2m - E_F)$ can be arbitrarily small, the minimum energy required to excite a superconductor is $\delta E_{min} = 2\Delta$. The factor 2 therefore means that when a Cooper pair is broken up, two unpaired electrons are always generated.

With this result the meaning of equation (11.30) becomes also clear. It is the **dispersion relation of unpaired electrons** and **holes** in the superconductor and replaces the relation $E = \hbar^2 k^2/2m = (\eta_{\mathbf{k}} + E_F)$ for free electrons in the normal conductor. This curve is shown in Figure 11.15. The nature of the excited electrons is not as obvious as the simple structure of the equation suggests. The excitations described are partly electrons, partly holes and are generally called **quasiparticles**. For large values of $\eta_{\mathbf{k}}$, i.e. far away from the Fermi surface, the dispersion relation (11.30) for $\eta_{\mathbf{k}} > 0$ turns into the one of free electrons. Quasiparticles with $\eta_{\mathbf{k}} < 0$ have a hole-like character. They are "pure" holes if $\eta_{\mathbf{k}} \ll \bullet$. For $\eta_{\mathbf{k}} \approx 0$, i.e. near the Fermi energy, quasi-particles are a mixture of an electron with the wave vector $\mathbf{k}$ and a hole with $-\mathbf{k}$. This becomes clearer when we consider the break up of a Cooper pair. Let us assume that an electron with

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

wave vector $\mathbf{k}$ from its pair state was scattered into the state $\mathbf{k}'$. Then an unoccupied state remains, a hole at $\mathbf{k}$ which interacts with the second electron of the pair with the wave vector $-\mathbf{k}$. Conversely, the electron now in state $\mathbf{k}'$ interacts with the hole at $-\mathbf{k}'$.

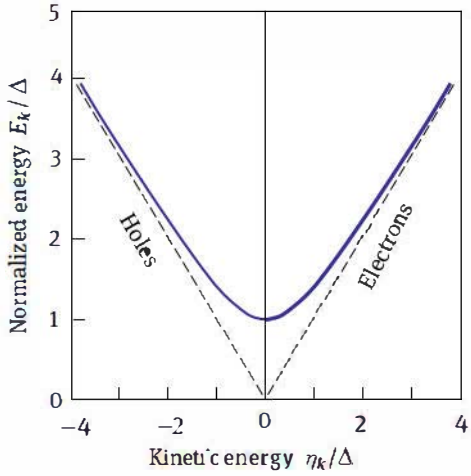


Fig. 11.15: Excitation energy of the quasiparticles near the Fermi energy. Hole-like states are located on the left side of the zero point, electron-like states on the right side. The dashed straight lines indicate the excitation energies of the electrons in the normal conductor.

In addition, it should be noted that, as with normal conductors, for every electron with the wave vector $\mathbf{k}$ deep in the Fermi sphere there is an electron with the wave vector $-\mathbf{k}$. In contrast to the Cooper pairs, they do not interact with each other, since they cannot exchange virtual phonons due to the lack of free one-electron states. Therefore, their potential energy is not lowered compared to the normal conducting state.

The *common ground state* of the Cooper pairs is separated from the states of the quasiparticles by the energy gap Δ. The density of states $D_s(E_\mathbf{k})$ of the quasiparticles follows directly from the density of states of the normal state, since the number of states is not changing in the transition to the superconductor, i.e. $D_s(E_\mathbf{k})\mathrm{d}E_\mathbf{k} = D_n(\eta_\mathbf{k})\mathrm{d}\eta_\mathbf{k}$, if $D_n(\eta_\mathbf{k})$ is the density of states of the normal conductor. In the vicinity of the Fermi energy we can set $D_n(\eta_\mathbf{k}) \approx D_n(E_F) = \text{const.}$ and get

$$D_s(E_\mathbf{k}) = D_n(\eta_\mathbf{k})\frac{\mathrm{d}\eta_\mathbf{k}}{\mathrm{d}E_\mathbf{k}} = \begin{cases} D_n(\eta_\mathbf{k})\dfrac{E_\mathbf{k}}{\sqrt{E_\mathbf{k}^2 - \Delta^2}} & \text{for} \quad E_\mathbf{k} > \Delta \\ 0 & \text{for} \quad E_\mathbf{k} < \Delta\,. \end{cases} \tag{11.39}$$

Figure 11.16a shows the predicted state density $D_s(E_\mathbf{k})$. It diverges at $E_\mathbf{k} = \Delta$ and changes for $E_\mathbf{k} \gg \Delta$ to that of the free electron gas. The density of the quasiparticles can be measured, for example, by tunnel junction spectroscopy, which we will discuss in the next section. Figure 11.16b shows the density of states of lead measured in this way. The agreement with the theoretical prediction is impressive.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

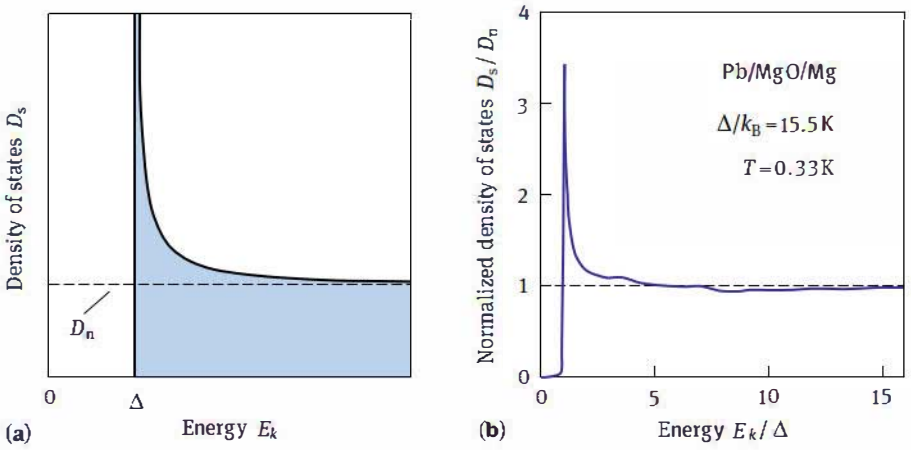


Fig. 11.16: **a)** Density of states of quasi-particles as a function of excitation energy. **b)** Experimentally measured density of states of lead as a function of single-particle energy in relation to the density of states at the Fermi energy. This measurement was performed with a Pb/MgO/Mg tunnel junction. (After I. Giaever et al., Phys. Rev. **126**, 941 (1962).)

BCS State at Finite Temperature. At finite temperature, not all electrons on the Fermi surface are paired, because thermal excitations break up Cooper pairs and produce quasi-particles. They occupy states that are no longer accessible to the interacting electrons that form Cooper pairs, thus hindering the exchange of virtual phonons. This reduces the interaction energy of the BCS state. The energy gap $\Delta(T)$ decreases with increasing temperature until it finally disappears completely at T_c. The energy gap $\Delta(T)$ at finite temperatures can be calculated very similar to the energy gap at $T = 0$ but now the contribution of the quasi-particles to the free energy must also be taken into account. The result is an expression that looks very similar to equation (11.36), but can only be evaluated numerically. The resulting temperature dependence of the gap energy can be seen in Figure 11.17 together with experimental results.

The data points result from experiments with different techniques and agree surprisingly well with the prediction of the BCS theory. This is remarkable because no free parameter was used for curve fitting. The occurrence of small deviations is not only due to measurement errors. They are also due to the fact that the assumption of a coupling parameter $\tilde{V}_0$ that is constant in the limited range $E_F < \hbar^2 k^2/2m, \hbar^2 k'^2/2m < (E_F + \hbar\omega_D)$ can only be a rough approximation.

If one takes into account that the energy gap disappears at T_c, one obtains the relationship

$$k_B T_c = 1.14\,\hbar\omega_D\, e^{-2/[\tilde{V}_0\, D(E_F)]}\,. \tag{11.40}$$

If one inserts this expression in (11.37), the important, material-independent result follows

$$\Delta(0) = 1.764\, k_B T_c\,. \tag{11.41}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

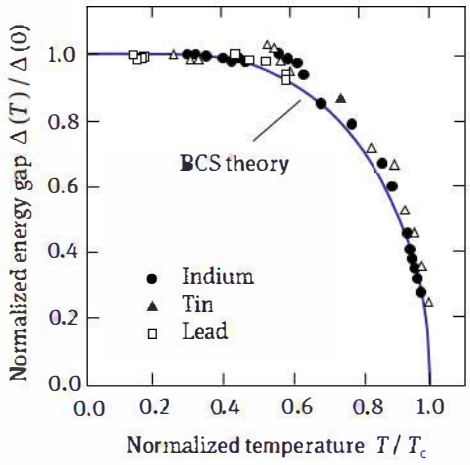


Fig. 11.17: Temperature dependence of the normalized energy gap $\Delta(T)/\Delta(0)$ as a function of the normalized temperature T/T_c. Beside the theoretical curve the experimental data for indium, tin and lead are shown. (After I. Giaever, K. Megerle, Phys. Rev. **122**, 1101 (1961).)

Table 11.4 shows the measured value of $2\Delta(0)/k_B T_c$ for some superconductors. In most cases the prediction agrees well with the experimental results. However, there are superconductors in which this ratio is significantly higher. In these cases we speak of *strongly coupling superconductors*. To describe these materials as well, *G.M. Eliashberg*[13] modified the original BCS theory. His expression for the transition temperature includes two material-specific, frequency-dependent parameters, namely the phonon density and the effective electron-phonon coupling. With this extension, good agreement between theory and experiment can be achieved even in case of strong coupling. However, we will not pursue this aspect further here.

Tab. 11.4: Experimental values of $2\Delta(0)/k_B T_c$ of some superconductors. (After R. Mersevey, B.B. Schwartz, *Superconductivity*, R.D. Parks, ed., Dekker, 1969.)

Superconductor	Al	Cd	Hg	In	Nb	Pb	Sn	Ta	Tl	Zn
T_c (K)	1.18	0.52	4.15	3.40	9.25	7.20	3.72	4.47	2.38	0.86
$2\Delta(0)/k_B T_c$	3.5	3.2	4.6	3.5	3.6	4.3	3.5	3.5	3.6	3.2

11.2.3 Experimental Evidence for an Energy Gap

A characteristic of superconductivity is the existence of an energy gap in the excitation spectrum of the electrons. This gap determines the number of excited quasiparticles at

13 Gerasim Matveevich Eliasberg, *1930 Saint Petersburg, †2021 Chernogolovka

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

a given temperature. The quasiparticles contribute to the specific heat and are clearly visible in their temperature dependence. If the temperature is increased, the thermal energy leads to breaking up more Cooper pairs and thus generating additional quasiparticles. Because of the energy gap $\Delta(T)$ the probability of the occupation of excited states and thus the specific heat is proportional to the Boltzmann factor $\exp[-\Delta(T)/k_B T]$. If the specific heat C_s is plotted logarithmically as a function of the inverse temperature, as in Figure 11.18a for vanadium and tin, the expected straight line is observed.

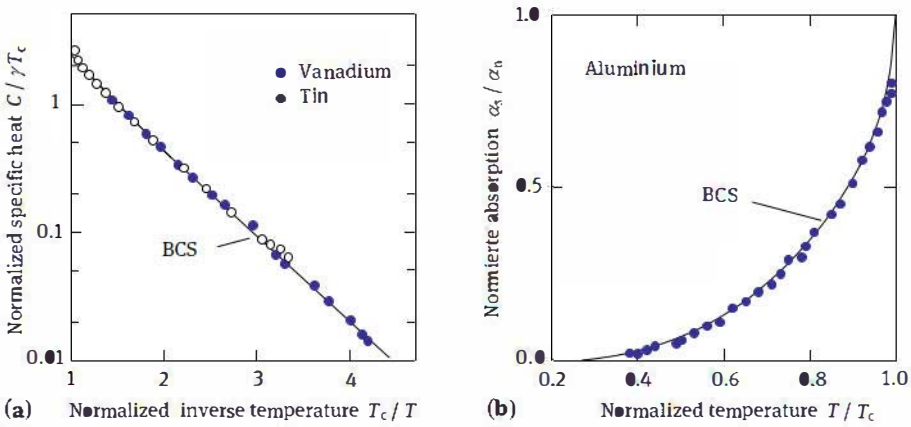


Fig. 11.18: **a)** Normalized specific heat of vanadium and tin as a function of the normalized inverse temperature T_c/T after subtraction of the phonon contribution. (According to M.A. Biondi et al., Rev. Mod. Phys. **30**, 1109 (1958).) **b)** Normalized ultrasonic absorption of aluminum as a function of the normalized temperature T/T_c. (After R. David, N.J. Ponlis, *Proc. 8th Int. Conf. Low Temp. Phys.*, R.O. Davies, ed., Butterworth, 1962.) The solid curves show the prediction of the BCS theory.

The data deviate from an exponential curve only in the vicinity of the transition temperature. This is not surprising, because the energy gap decreases rapidly as the temperature approaches the transition temperature. The BCS theory predicts the value $(C_s - C_n)/C_n = 1.43$ for the jump in specific heat at T_c shown for aluminum in Figure 11.10. In most cases, the experimental values agree with this figure, but there are also deviations in the case of strongly coupling superconductors (see Table 11.5).

Tab. 11.5: Experimental values of $[(C_s - C_n)/C_n]_{T_c}$ of some superconductors. (After R. Mersevey, B.B. Schwartz, *Superconductivity*, R.D. Parks, ed., Dekker, 1969.)

Superconductor	Al	Cd	Hg	In	Nb	Pb	Sn	Ta	Tl	Zn
T_c (K)	1.18	0.52	4.15	3.40	9.25	7.20	3.72	4.47	2.38	0.86
$[(C_s - C_n)/C_n]_{T_c}$	1.4	1.4	2.4	1.7	1.9	2.7	1.6	1.6	1.5	1.3

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The number of quasi-particles is also reflected in many transport properties. Cooper pairs are in the quantum-mechanical ground state and therefore carry no entropy and thus do not contribute to heat transport. With the number of quasi-particles, the thermal conductivity of superconductors thus decreases drastically with decreasing temperature below T_c. Another example is the ultrasound absorption in aluminum shown in Figure 11.18b. In pure metals, the mean free path of the ultrasonic phonons is limited by scattering processes with free electrons. Since Cooper pairs can only interact with phonons whose energy is sufficient to break the pair, e.g. at ultrasonic frequencies in the MHz range, scattering can only occur with thermally excited quasiparticles. Thus ultrasonic damping is proportional to the number of excited quasiparticles and rises with temperature.

How can the energy gap be measured "directly"? Microwave and infrared experiments are immediately obvious candidates. The gap can be determined via absorption using these techniques. If in such experiments the energy of the irradiated photons is not sufficient to break up the Cooper pairs, no absorption occurs. If, however, $\hbar\omega > 2\Delta$, the radiation is strongly absorbed and two quasi-particles are generated for each photon. Such measurements are relatively complex and are not easy to analyze, since other absorption mechanisms exist in metals (see Section 13.4). However, the results are in very good agreement with the predictions of the BCS theory.

We want to focus here on the conceptually extremely simple and elegant **tunnel junction spectroscopy**, which was mainly developed by *I. Giaever*[14]. The schematic structure and the principle circuit diagram of such a tunnel junction experiment are shown in Figure 11.19. The two metal strips are successively vapor-deposited onto the insulating substrate and are separated from each other in the overlap region by a thin insulating layer about 3 nm thick. If the intermediate layer is thin enough, electrons can tunnel from one metal to the other.

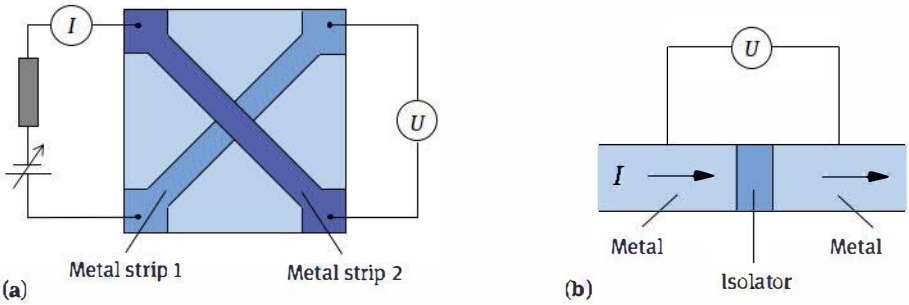


Fig. 11.19: Tunnel junction. **a)** The cross-shaped metal strips are separated from each other by a thin insulating layer. **b)** Schematic diagram. At a given current I the voltage drop is measured at the two metal strips.

14 Ivar Giaever, *1929 Bergen, Nobel Prize 1973

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

When a voltage U is applied, the Fermi levels of the two metals are shifted against each other by eU. As shown in Figure 11.20 for two normal conductors, electrons can tunnel from occupied states of one metal into the unoccupied states of the other. Since the shift eU of the Fermi energy is proportional to the voltage, the current-voltage characteristic is linear, i.e., following Ohm's law $I \propto U$.

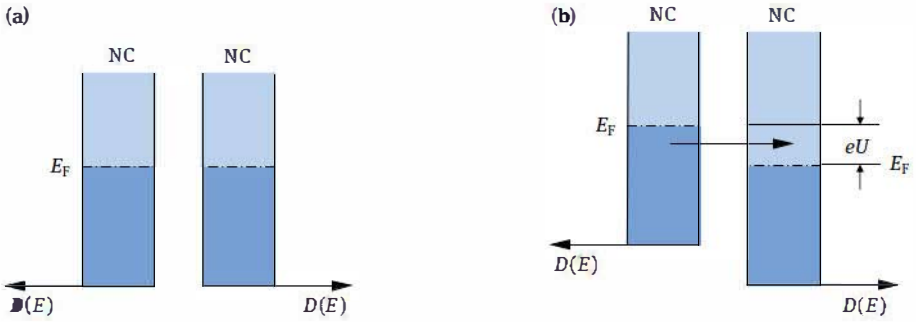


Fig. 11.20: Energy level diagram of a tunnel junction consisting of two normal conductors. The densities of states in the vicinity of the Fermi energy are shown for $T = 0$. **a)** $U = 0$, no current flows. **b)** $U \neq 0$, electrons tunnel from occupied states to unoccupied ones.

If the tunnel junction consists of a normal conductor and a superconductor, no current can flow for voltages $U < \Delta/e$ at $T = 0$, since there are no free states in the superconductor due to the energy gap. The corresponding densities of states are shown schematically in Figure 11.21. Current flow occurs when the critical voltage $U_c = \Delta/e$

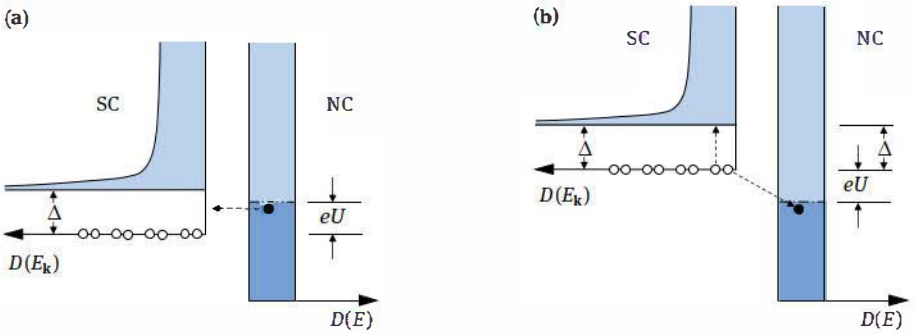


Fig. 11.21: Densities of states of a superconductor-normal conductor junction with an applied voltage $|U| < \Delta/e$. Cooper pairs are indicated by two open circles, quasiparticles or unpaired electrons by black dots. **a)** The Fermi energy of the normal conductor is raised, but the electrons do not find free states in the superconductor. **b)** The Fermi energy of the normal conductor is lowered, but the energy gain eU of the tunneling electron is not sufficient to lift the electron remaining in the superconductor into the band of the quasiparticles. The indicated process is therefore prohibited at this voltage.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

is exceeded. This conclusion is immediately obvious if the normal conductor is at a higher potential, as shown in Figure 11.21a. Electrons from the occupied states below E_F can only tunnel through the insulator layer into the empty quasi-particle states of the superconductor when $U > U_c$.

If the polarity is reversed (cf. Figure 11.21b), one may think that electrons could leave the BCS ground state at lower voltages and change over to the normal conductor. However, this is not correct, because a Cooper pair must be broken up and the remaining electron must be lifted into the band of quasiparticles. As long as $eU < \Delta$, the required energy is not available. If, on the other hand, the applied voltage is $U > U_c$, Cooper pairs can be broken up: the energy released is used to lift the second electron into the quasiparticle band.

The contribution of the quasi-particles with energy E to the current from the superconductor to the normal conductor is proportional to their number $D_s(E_\mathbf{k})f(E)$ in the superconductor and to the number of empty states $D_n(E + eU)[1 - f(E + eU)]$ in the normal conductor. The same applies to the current flow in the opposite direction. The occurring partial currents are therefore given by the two equations

$$I_{s\to n} = I_0 \int D_s(E_\mathbf{k}) f(E) D_n(E + eU)[1 - f(E + eU)]\,\mathrm{d}E\,,$$

$$I_{n\to s} = I_0 \int D_n(E + eU) f(E + eU) D_s(E_\mathbf{k})[1 - f(E)]\,\mathrm{d}E\,. \tag{11.42}$$

The following therefore applies to the total current

$$I = I_{s\to n} - I_{n\to s} = I_0 \int D_s(E_\mathbf{k}) D_n(E + eU)[f(E) - f(E + eU)]\,\mathrm{d}E\,. \tag{11.43}$$

The constant I_0 depends on the geometry and condition of the junction. Because the density of states of normal conductors is constant near the Fermi energy and the probability of occupation at absolute zero is a step function, equation (11.43) gives the simple relationship $\mathrm{d}I/\mathrm{d}U \propto D_s(E_\mathbf{k} = eU)$ for measurements at sufficiently low temperatures. Thus, the measurement of the current-voltage characteristic allows for a direct determination of the density of states of the quasi-particles. An example of this was already shown in Figure 11.16.

The expected curve of the current-voltage characteristic of a superconductor-normal conductor junction is sketched in Figure 11.22. The vertical rise at the onset of current flow reflects the singularity of the quasi-particle density of states at $(E_F \pm \Delta)$. At finite temperatures, excited electrons and quasi-particles are present in normal conductors and superconductors, which can already tunnel through the barrier at $U < U_c$ and thus cause a weak current flow. The resulting current-voltage characteristic is also shown in Figure 11.22.

Of course, the described experiment can also be performed with two superconducting metals. As one can easily guess, in this case the critical voltage is given by $eU_c = (\Delta_1 + \Delta_2)$. An interesting aspect is that a current maximum occurs at finite temperatures at $U = |\Delta_1 - \Delta_2|/e$, because at this voltage the maxima of the densities of states of the quasi-particles in the two superconductors are at the same level.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

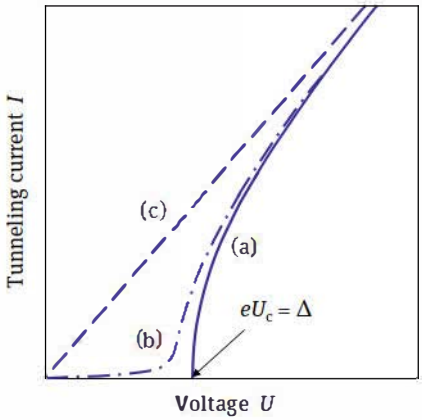


Fig. 11.22: Current-voltage characteristics of a superconductor-normal conductor junction at different temperatures. Curve (a): at the absolute zero, a steep current rise begins at $U_c = \Delta/e$. Curve (b): at $0 < T < T_c$ a weak current flow due to thermally excited quasi-particles already occurs at $U < \Delta/e$. Curve (c): at $T > T_c$, both metals are normal conducting. The tunnel junction behaves like an ohmic resistance.

It should be pointed out here that in tunnel junction spectroscopy a modified representation of the state densities is usually used, namely the so-called *semiconductor representation*. Here, the quasi-particles with $\eta_{\mathbf{k}} < 0$ are regarded as holes, despite their complicated structure. As we have seen in Section 9.1 when introducing the hole concept, hole states can be understood as electronic states with negative excitation energy. Applying this concept, negative energies must be allowed and the left branch of the dispersion curve of the quasi-particles in Figure 11.15 must be inverted at the coordinate origin. At $T = 0$, all states of the lower branch of the curve are occupied in this representation, while the upper ones are free. At finite temperature there are also occupied states in the upper branch and unoccupied states in the lower one, i.e. holes.

11.2.4 Current Transport Through Interfaces

Now we will briefly look at a current-carrying wire that is composed of two parts: the first part consist of a normal conductor, the second of a superconductor. While single electrons move in the normal conductor, Cooper pairs flow in the superconductor. This raises an interesting question: which mechanism is in play in generating the Cooper pairs at the interface? To simplify the problem, we consider the process at absolute zero.

As indicated in Figure 11.23a, there is a possibility that the incident electron will be reflected at the interface if the energy is smaller than the energy gap, since no free single electron states exist in the superconductor at the relevant energy. If this were the only possibility, the consequence would be that the interface would stop the current flow at $T = 0$. But there is a second possibility, the **Andreev reflection**[15] which allows the conversion of single electrons to Cooper pairs and thus makes the super

15 Alexander Fyodorovich Andreev, *1939 Saint Petersburg

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

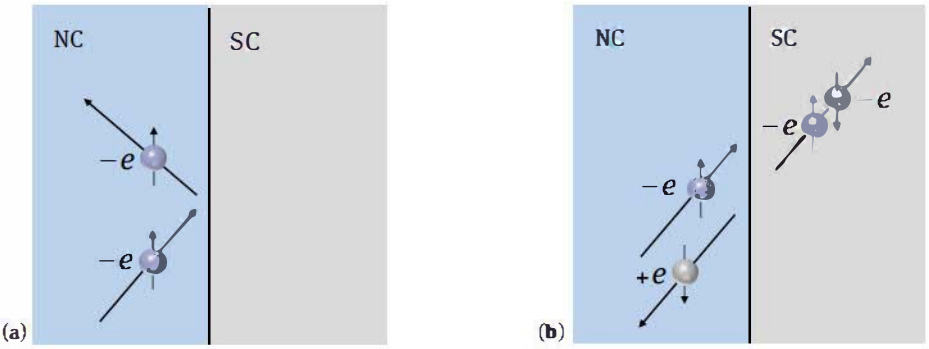


Fig. 11.23: Andreev reflection. a) An electron in the normal conductor is reflected at the interface when it impacts the superconductor. b) An incoming electron is reflected as a hole. At the same time, a Cooper pair is formed in the superconductor, which continues to move in the direction of the incident electron.

current possible. As shown in Figure 11.23b, an electron hits the interface, connects with another electron and enters the superconductor unhindered as a Cooper pair. Since the charge must of course be conserved in this process, a hole remains. From the conservation of momentum it follows that this hole carries the exact opposite momentum of the incident electron and thus has the wave vector $-\mathbf{k}$. For the same reason the spin of the hole is opposite to the spin of the incident electron.

11.2.5 Critical Current and Critical Magnetic Field

In the course of this chapter we found that on one hand, the shielding currents prevent magnetic fields from penetrating the superconductor, but that on the other hand these fields cannot be arbitrarily large since a critical magnetic field already exists for thermodynamic reasons. Now we want to show that a **critical current density** is also associated with the critical field, which in turn depends on the size of the energy gap.

Before we derive this relation, we will estimate the maximum current that can flow through a long superconducting wire with radius R. The magnetic field at the surface of a wire is given by

$$B = \mu_0 \frac{I}{2\pi R} . \tag{11.44}$$

This results in the critical current

$$I_c = \frac{2\pi R}{\mu_0} B_c . \tag{11.45}$$

Using this relationship, we find a relatively small critical current of $I_c = 150\,\text{A}$ for a tin wire with radius Radius $R = 1\,\text{mm}$, since the critical field of tin is only 30 mT. Since $I_c \propto B_c$, both quantities have the same temperature dependence. It is remarkable

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

that the critical current does *not* increase proportionally to the wire cross section, but linearly with its radius.

The current flow in superconductors results from the motion of center of mass of the Cooper pairs. Their velocity $\mathbf{v} = \hbar\,\delta\mathbf{K}/m$ is directly related to the change $\delta\mathbf{K}$ of the wave vector of the electrons involved. We can describe the Cooper pairs in this state by $(\mathbf{k} + \delta\mathbf{K}\uparrow, -\mathbf{k} + \delta\mathbf{K}\downarrow)$. It should be mentioned that the motion of center of mass of the pairs does not change $\widetilde{V}_{\mathbf{k}\mathbf{k}'}$ or Δ itself. The energy gap thus moves with the Fermi sphere in $\mathbf{k}$-space.

It is plausible that superconductivity is destroyed when the kinetic energy of the electrons exceeds the condensation energy (11.34), because scattering processes become possible which lead to the breaking of Cooper pairs. We can therefore assume for the critical center-of-mass velocity $\mathbf{v}_c$ the condition

$$\frac{1}{2} n_s m_s v_c^2 = \frac{1}{4} \mathcal{D}(E_F) \Delta^2 \,. \tag{11.46}$$

Inserting $\mathcal{D}(E_F)$, we find

$$v_c = \sqrt{\frac{3}{2}} \frac{\Delta}{m_s v_F} \,. \tag{11.47}$$

If we take into account that the density of the supercurrent is given by $\mathbf{j}_s = -n_s e_s \mathbf{v}$, then we find for the critical current density

$$j_c = \sqrt{\frac{3}{2}} \frac{e_s n_s \Delta}{\hbar k_F} \,. \tag{11.48}$$

With the values for tin we obtain for the critical current density $j_c \approx 1.5 \times 10^8 \mathrm{A/cm^2}$.

A magnetic field $\mathbf{B}$ is associated with the current, the strength of which we will calculate for the critical current in the wire. The field at the wire surface is $2\pi RB = \mu_0 \int \mathbf{j} \cdot \mathrm{d}\mathbf{f}$, where R is the wire radius and $\mathrm{d}\mathbf{f}$ is an area element of the wire cross-section. Under the condition $R \gg \lambda_L$ and with equation (11.7) we obtain for the current density $\int \mathbf{j} \cdot \mathrm{d}\mathbf{f} = 2\pi\, j_{s,0}\, \lambda_L\, R$. We use this relation to express the critical current density at the surface and find for the critical field

$$B_c = \mu_0 \lambda_L j_c = \sqrt{\frac{3}{2}} \frac{e_s n_s \mu_0 \lambda_L \Delta}{\hbar k_F} \,. \tag{11.49}$$

Critical magnetic field strength and critical current are therefore directly linked, regardless of whether the current density is caused by shielding or transport currents. A corresponding hypothesis was already put forward by *F.B. Silsbee*[16] in 1916.

The processes that take place in a superconducting wire when the critical current is exceeded are relatively complex. At first, one might think that the superconductivity simply retracts into the interior of the wire. Then, however, the current would also flow

16 Francis Briggs Silsbee, *1889 Lawrence, †1967 Washington DC

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

inside the conductor and create a field that is larger than the original field at the surface. The constriction would therefore continue until the whole wire is normal conducting. At this point, the current would be uniformly distributed over the cross section and the field would be smaller than B_c in most of the cross section. However, this area would then have to be superconducting.

As already mentioned, the phenomena that actually occur are difficult to describe. *F. London* and also *C.J. Gorter* have each developed a model to explain this process. Let us first look at the static model of *London*. He assumed that perpendicular to the wire axis specially shaped superconducting lamellae are formed which are not connected to each other. The current must then also flow through normally conducting areas, so that the wire has a finite resistance. In *Gorter's model* it is assumed that superconducting tubes are formed in the wire, which move towards the wire axis. This motion generates magnetic fields that vary in time, which cause electric fields in the normally conducting areas and thus produce a finite resistance. The processes that actually take place are probably a mixture of both. Here it should be emphasized that the experimental situation is not really clear either. The experiments are complicated by the fact that heat is dissipated during the transition to a finite resistance state. It is therefore extremely difficult to exclude temperature effects.

At this point, we would like to make some remarks about the disappearance of the electrical resistance in superconductors. In normal conductors the resistance is caused by the scattering of electrons at defects and phonons. In superconductors, charge transport is based on the common motion of all Cooper pairs, characterized by the additional wave vector $\delta\mathbf{K}$. Scattering of a Cooper pair is equivalent to leaving the common BCS state and thus breaking up the pair. The energy must be larger than the binding energy so that elastic scattering processes are excluded in the first place. Inelastic scattering is possible with phonons of high energy. The binding energy is provided by the absorbed phonon and the Cooper pair is destroyed. However, the reverse process also occurs, in which a Cooper pair is formed while emitting a phonon. In equilibrium, both processes cancel each other out, since the newly formed Cooper pairs condense into the states that were previously released by pair breaking. This raises the question of why the thermally excited quasi-particles do not cause any losses. The answer is simple: in the stationary state, electric fields are short-circuited. The quasiparticles are therefore not accelerated and do not contribute to the transport of current. This argumentation does not apply when an alternating voltage is applied, since in this case an electric field exists according to the first London equation. The quasi-particles are accelerated, and interact with the lattice causing losses. Loss-free current transport therefore only occurs with direct currents.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

11.3 Macroscopic Wave Function

In the discussion so far, we have spoken of a *common quantum state* in some places, but we have not explicitly made use of the **macroscopic wave function** of superconductors associated with it. Now we want to take a closer look at this wave function and get to know its meaning.

Cooper pairs do not carry a total spin. They therefore behave like bosons and can condense into a common many-particle state, the BCS state, which is characterized by the macroscopic wave function

$$\Psi = \Psi_0 \, e^{i\varphi(\mathbf{r})} = \sqrt{n_s} \, e^{i\varphi(\mathbf{r})} \tag{11.50}$$

The amplitude of the wave function is given by the density of the Cooper pairs $\Psi\Psi^* = |\Psi_0|^2 = n_s$. The real-valued function $\varphi(\mathbf{r})$ describes the phase of the wave function and has a well-defined value for superconductors over macroscopic distances. The existence of a macroscopic wave function has considerable consequences for the behavior of superconductors, which become particularly obvious when a magnetic field is applied. We will first discuss flux quantization and then the *Josephson effect*.

11.3.1 Flux Quantisation

If we bring a multiple-connected superconductor, for example a ring as shown in Figure 11.24, into a magnetic field and cool it below the transition temperature, the magnetic field is maintained in the inner region of the ring, while the superconducting ring itself is field-free except for a thin layer on the sample surface. The trapped magnetic flux remains trapped even after the external field is switched off.

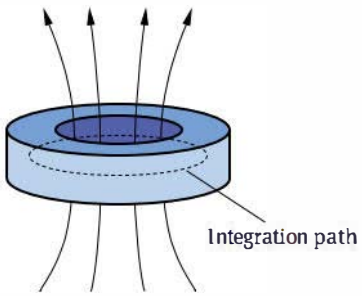


Fig. 11.24: Superconducting ring with enclosed magnetic flux. Shielding currents flow only at the surface of the superconductor. The dotted line indicates an integration path that runs in the current-free, inner area of the superconductor.

Let us first consider the phase of the wave function. The phase difference $\Delta\varphi = (\varphi_2 - \varphi_1)$ between the two locations 1 and 2 is given by the line integral $\Delta\varphi = \int_1^2 \operatorname{grad}\varphi(\mathbf{r}) \cdot \mathrm{d}\mathbf{s}$. Since the wave function has a defined value at each point, the phase difference after passing through a closed loop within the superconductor can only assume the values $2\pi p$, where p is an integer.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

This quantization condition has significant consequences for the current flow we will now consider. We use the quantum mechanical expression for the current density in a magnetic field

$$\mathbf{j}_s = \mathrm{i}\frac{\hbar q}{2M}\left(\Psi^*\nabla\Psi - \Psi\nabla\Psi^*\right) - \frac{q^2}{M}\mathbf{A}\,\Psi^*\Psi \tag{11.51}$$

and replace q by $-2e$ and M by $2m$ in this equation. Using the wave function (11.50) and the London penetration depth λ_L, this results in

$$\mu_0\,\lambda_L^2\,\mathbf{j}_s = \left(\frac{\hbar}{e}\nabla\varphi - 2\mathbf{A}\right) . \tag{11.52}$$

Now we build the loop integral over the current density:

$$\mu_0\,\lambda_L^2 \oint \mathbf{j}_s \cdot \mathrm{d}\mathbf{s} = \frac{\hbar}{e}\oint \nabla\varphi \cdot \mathrm{d}\mathbf{s} - 2\oint \mathbf{A}\cdot\mathrm{d}\mathbf{s} . \tag{11.53}$$

We have already discussed the integral over the phase gradient. The line integral over the vector potential can be transformed into a surface integral over the magnetic induction using Stokes' theorem. One finds $\oint \mathbf{A}\cdot\mathrm{d}\mathbf{s} = \int_\Sigma \mathbf{B}\cdot\mathrm{d}\mathbf{f} = \Phi$. Σ stands for the area enclosed by the integration path. Thus we obtain

$$\mu_0\,\lambda_L^2 \oint \mathbf{j}_s \cdot \mathrm{d}\mathbf{s} + 2\Phi = p\frac{h}{e} , \tag{11.54}$$

where p, as mentioned above, stands for an integer. If we choose, as indicated in Figure 11.24, a closed integration path in the center of the ring, then $\mathbf{j}_s = 0$ and the integral disappears. We then get

$$\Phi = p\,\frac{h}{2e} = p\,\Phi_0 . \tag{11.55}$$

The magnetic flux Φ through a closed superconducting loop is thus quantized[17], so that only multiples of the **flux quantum**

$$\Phi_0 = \frac{h}{2e} = 2.067833\,848\ldots \times 10^{-15}\ \mathrm{Vs} \tag{11.56}$$

appear. The flux quantum Φ_0, also called *fluxon*, is extremely small and has an exact value which is set by the new International System of Units. Thus a hollow cylinder in the earth field, that encloses *one* flux quantum, has a diameter of only about 5 µm.

It is not only the flux quantization that is important but also the fact that the charge $2e$ of the Cooper pair appears in (11.56). The measurement of the flux quantum in 1961 by *R. Doll*[18] and *M. Näbauer*[19] as well as *B.S. Deaver*[20] and *W.M. Fairbank*[21] was

17 The possibility of flux quantization was pointed out by Fritz London as early as 1950.
18 Robert Doll, *1923 Munich
19 Martin Näbauer, *1919 Karlsruhe, †1962 Munich
20 Bascom Sine Deaver, Jr. *1930 Macon
21 William Martin Fairbank, *1917 Minneapolis, †1989 Palo Alto

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

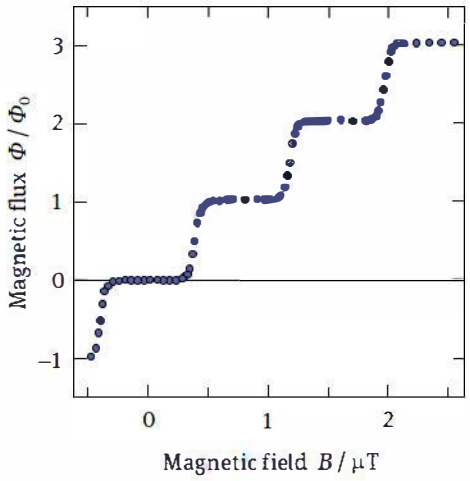


Fig. 11.25: Magnetic flux in a thin superconducting tin cylinder (length 24 mm, diameter 56 μm) as a function of the magnetic field in which the cylinder was cooled down. (After W.L. Goodman et al., Phys. Rev. B 4, 1530 (1971).)

therefore a strong indication of the existence of Cooper pairs. In the experiments, thin superconducting cylinders were cooled in a very weak magnetic field. The magnetic field was then switched off and the magnetic moment of the cylinders was determined for different external magnetic fields. The measurements showed that the flux trapped in the cylinder is indeed quantized and the flux is given by (11.55).

Figure 11.25 shows the result of a more recent measurement on a hollow tin cylinder.[22] The hollow cylinder had a diameter of 56 μm and was cooled in different magnetic fields. It is obvious that the frozen magnetic flux does not increase proportionally to the applied magnetic field, but is quantized. The rounding that occurs is due to the fact that under the experimental conditions of this measurement, a flux quantum does not always run along the entire length inside the cylinder.

The quantization of the magnetic flux results in the quantization of the current in a current loop. Since the phase of the wave function can only change by an integer multiple of 2π, only jumps in current are allowed. In principle, a superconductor can change into a state with a smaller number of flux quanta. However, this requires overcoming such a high energy barrier that the probability of this effect occurring is negligible. In fact, no decay of persistent currents has been observed in corresponding experiments for years.

22 The measurement was carried out with a SQUID magnetometer, a device that we will discuss at the end of this section.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

11.3.2 Josephson Effect

In Section 11.2 we have already considered experiments with tunnel junctions and discussed the tunneling of quasi-particles. If the thickness of the insulating layer between the two superconductors is reduced to about 1 nm, the wave function of one superconductor reaches noticeably into the region of the other. As a result, the wave functions of the two superconductors are coupled and a tunneling of *Cooper pairs* through the insulating layer is observed.

There are several possibilities to realize a weak coupling between two superconductors. Besides the oxide barrier between deposited films, *point contacts* and *micro-bridges* are used as so-called *weak links*. For example, a niobium point contact can be created by grinding a niobium wire to a point, allowing the tip to oxidize and then pressing it against a solid niobium sample. A micro-bridge is obtained by etching a thin superconducting film so that it consists of two parts connected by a very narrow bridge. The basic circuit diagram for studying the Josephson effect is very simple and is sketched in Figure 11.26.

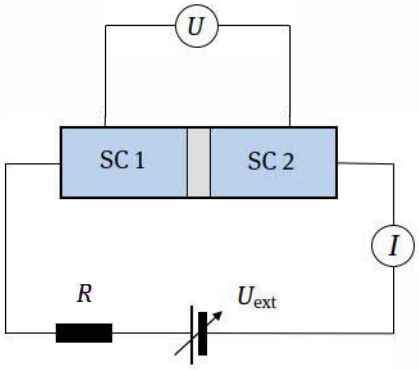


Fig. 11.26: Circuit diagram for the investigation of the Josephson effect.

The overlapping of the wave functions results in some surprising effects, which were already predicted in 1962 by *B.D. Josephson*[23]. If the two superconductors are separated from each other, the time dependence of their wave functions is described by the Schrödinger equations $i\hbar\dot{\Psi}_1 = H_1\Psi_1$ and $i\hbar\dot{\Psi}_2 = H_1\Psi_2$ with the eigenvalues E_1 and E_2. We can treat the coupled superconductors in terms of perturbation theory and write

$$i\hbar\dot{\Psi}_1 = E_1\Psi_1 + \mathcal{K}\Psi_2 \qquad \text{and} \qquad i\hbar\dot{\Psi}_2 = E_2\Psi_2 + \mathcal{K}\Psi_1 \,. \tag{11.57}$$

The coupling between the two superconductors is taken into account by the additional term with the coupling parameter $\mathcal{K}$. For simplicity, we assume that the two superconductors are made of the same material and therefore have the same Cooper pair

23 Brian David Josephson, *1940 Cardiff, Nobel Prize 1973

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

density $n_{s1} = n_{s2} = n_s$. In this case, E_1 and E_2 are also equal. However, if the voltage U drops across the insulating layer the eigenvalues shift and $(E_2 - E_1) = -2eU$ applies.

We use the wave function (11.50) for the respective superconductor and allow a time-dependent development of the Cooper pair density and the phase of the wave function. After the separation of the real and imaginary parts, the four equations are obtained:

$$\dot{n}_{s1} = \frac{2\mathcal{K}}{\hbar}\sqrt{n_{s1}n_{s2}}\sin(\varphi_2 - \varphi_1)\,, \qquad \dot{n}_{s2} = -\frac{2\mathcal{K}}{\hbar}\sqrt{n_{s1}n_{s2}}\sin(\varphi_2 - \varphi_1)\,, \tag{11.58}$$

$$\dot{\varphi}_1 = \frac{\mathcal{K}}{\hbar}\sqrt{\frac{n_{s2}}{n_{s1}}}\cos(\varphi_2 - \varphi_1) - \frac{E_1}{\hbar}\,, \quad \dot{\varphi}_2 = \frac{\mathcal{K}}{\hbar}\sqrt{\frac{n_{s1}}{n_{s2}}}\cos(\varphi_2 - \varphi_1) + \frac{E_2}{\hbar}\,. \tag{11.59}$$

Now we calculate the difference from the last two equations and find the **second Josephson equation** for the time evolution of the phase

$$\hbar(\dot{\varphi}_2 - \dot{\varphi}_1) = -(E_2 - E_1) = 2eU\,. \tag{11.60}$$

Let us first consider the case where there is no potential difference between the superconductors, i.e. $U = 0$. According to equation (11.60), in this case the phase difference $(\varphi_1 - \varphi_2)$ between the two wave functions is independent of time. Thus $\dot{n}_{s1} = -\dot{n}_{s2}$ follows. Accordingly, a current should flow between the two superconductors, which should immediately lead to an electrical charging of the superconductors. However, we must not forget that the two superconductors are a part of a circuit that ensures that n_{s1} and n_{s2} remain constant. Therefore, equation (11.60) is still valid and with equation (11.58) we obtain for the current through the junction the **first Josephson equation**

$$I_s = I_J \sin(\varphi_2 - \varphi_1)\,. \tag{11.61}$$

This means that although a direct current I_s flows through the tunnel junction, no voltage drop occurs at the insulating layer. This amazing phenomenon is called the **DC Josephson effect.** The critical current I_J depends on the density n_s of the Cooper pairs, the contact area A (typical value 0.1 mm^2) and above all the coupling strength $\mathcal{K}$ and thus the thickness of the insulator layer.

The current-voltage characteristic of a Pb/PbO$_x$/Pb-Josephson contact is shown in Figure 11.27. As long as $I_s < I_J$, the "supply" from the current source (see Figure 11.26) determines the current and thus the phase difference $(\varphi_2 - \varphi_1)$ between the two macroscopic wave functions.

If the output of the current source, is increased by raising the voltage U_{ext} and reaches the critical current I_J, the voltage across the junction jumps to a finite value, which is determined by the quasi-particle characteristic curve. Now $(E_2 - E_1) = -2eU$, and the phase difference increases linearly with time. The integration of equation (11.60) gives

$$(\varphi_2 - \varphi_1) = \frac{2eU}{\hbar}t + \varphi_0 = \omega_J t + \varphi_0\,. \tag{11.62}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

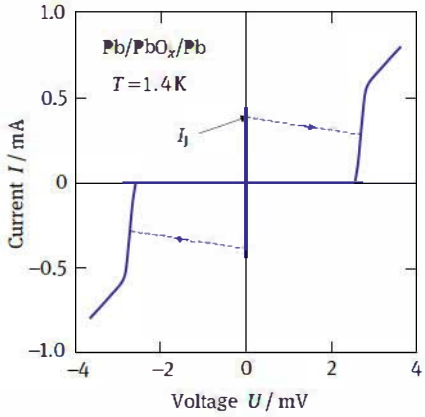


Fig. 11.27: Current-voltage characteristic of a Pb/PbO$_x$/Pb-tunnel junction. If the current through the tunnel junction is increased, at I_J the operating point jumps to the quasi-particle characteristic. (After K. Schwidtal, R.D. Finnegan, Phys. Rev. B**2**, 148 (1970).)

If we insert this result into equation (11.61), we obtain

$$I = I_J \sin(\omega_J t + \varphi_0) \qquad \text{with} \qquad \omega_J = \frac{2eU}{\hbar} . \tag{11.63}$$

Surprisingly, an alternating current with the *Josephson frequency* ω_J appears. This is the **AC Josephson effect**. In the DC characteristic shown in Figure 11.27, this current is of course not visible. The frequencies that result are relatively high. At a voltage of 10 μV, the frequency is already at 48 GHz. Since the voltage and the frequency are directly linked by e/h, the ratio of e/h can be determined very precisely with the help of the Josephson effect. On the other hand, if the constant e/h is known, voltages can be determined with high precision via a frequency measurement. Today's voltage standards are therefore based on the Josephson effect and have replaced the Weston element introduced in 1908.

The direct detection of the high-frequency Josephson alternating current is hard and unsuitable for practical applications, because the emitted microwave power is typically in the microwatt range. In addition, the impedance matching of the low-resistance junctions and thus the coupling out of the microwaves is extremely difficult. This problem can be avoided by radiating microwaves of frequency ω_{micro} into a junction with DC voltage U_0 and exploiting the existing non-linearities. If the resulting effective voltage $U = U_0 + U_{micro} \cos(\omega_{micro} t)$ is used in equation (11.60), then after the integration we find

$$(\varphi_2 - \varphi_1) = \omega_J t + \frac{2eU_0}{\hbar\omega_{micro}} \sin(\omega_{micro} t) + \varphi_0 . \tag{11.64}$$

The second term reflects the phase modulation that appears. Besides the alternating current with the Josephson frequency, sidebands with the frequency $(\omega_J \pm p\omega_{micro})$, where p stands for an integer, appear. Whenever the Josephson frequency ω_J is tuned to a multiple of the microwave frequency ω_{micro}, a sideband occurs at the frequency $\omega = 0$. Thus, a direct current which can be easily detected is generated. Most measurements of the Josephson alternating current effect are based on this or similar principles.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Josephson Junctions in Magnetic Fields. We now look at the influence of a magnetic field on the tunneling of the Cooper pairs. Since the derivation of the relevant equations is somewhat laborious in the case of a single tunnel junction, we discuss here two identical Josephson junctions connected in parallel, as shown schematically in Figure 11.28. The magnetic field is assumed to be perpendicular to the drawing plane. The enclosed area should be so large that the extension of the thin contact points can be neglected in the following considerations.

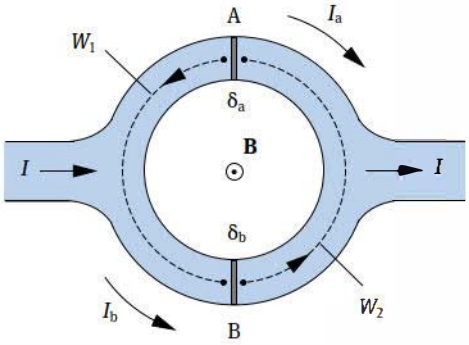


Fig. 11.28: Current flow through two identical Josephson junctions A and B connected in parallel. The magnetic field B is oriented perpendicular to the drawing plane. The phase differences δ_a and δ_b occur across the junctions. The phase differences are calculated along the dashed paths W_1 and W_2.

The total current I_s consists of the partial currents through the two individual Josephson junctions:

$$I_s = I_J(\sin\delta_a + \sin\delta_b) = 2\,I_J \cos\left(\frac{\delta_a - \delta_b}{2}\right) \sin\left(\frac{\delta_a + \delta_b}{2}\right). \tag{11.65}$$

Here, $\delta_a = (\varphi_{a1} - \varphi_{a2})$ and $\delta_b = (\varphi_{b1} - \varphi_{b2})$ denote the phase differences at the junctions A and B. As we will see, the difference $(\delta_a - \delta_b)$ is determined by the magnetic flux Φ through the ring. When calculating the phase, we proceed as we did when discussing flux quantization and place the integration path again inside the superconducting ring where no current flows. For the phase difference along the paths W_1 and W_2 then follows according to equation (11.53):

$$\varphi_{a1} - \varphi_{b1} = \frac{2e}{\hbar} \int_{W_1} \mathbf{A} \cdot \mathrm{d}\mathbf{s} \qquad \text{and} \qquad \varphi_{b2} - \varphi_{a2} = \frac{2e}{\hbar} \int_{W_2} \mathbf{A} \cdot \mathrm{d}\mathbf{s}\,. \tag{11.66}$$

If we add the two equations, we get

$$\delta_a - \delta_b = \frac{2e}{\hbar} \oint \mathbf{A} \cdot \mathrm{d}\mathbf{s} = \frac{2e\Phi}{\hbar}\,. \tag{11.67}$$

We have neglected the contributions of the field in the junctions, as mentioned above. If we insert equation (11.67) in (11.65), we find

$$I_s = 2I_J \sin\left(\frac{\delta_a + \delta_b}{2}\right) \cos\left(\frac{\pi\Phi}{\Phi_0}\right). \tag{11.68}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The phase angle $(\delta_a + \delta_b)$, is magnetic field independent and adapts to the experimental conditions. The cosine term describes the oscillation of the current with the magnetic field. The experimental Josephson current through a double junction as a function of the applied magnetic field is shown in Figure 11.29a. It is easy to see that the current flow reacts to extremely small changes in the magnetic field, corresponding even to fractions of Φ_0. The envelope of the tunnel current is caused, as we will see in a moment, by the finite extension of each Josephson junction.

If one investigates the magnetic field dependence of the tunnel current through one junction, one has to consider for the description not only the magnetic field in the *Josephson junction* but also the finite penetration depth of the magnetic field into the superconductor. After a somewhat longer calculation we find

$$I_S = I_J \left| \frac{\sin(\pi\Phi/\Phi_0)}{(\pi\Phi/\Phi_0)} \right| , \tag{11.69}$$

where ϕ stands for the total flux through the junction. As Figure 11.29b shows, there is very good qualitative agreement between this prediction and the experimental results.

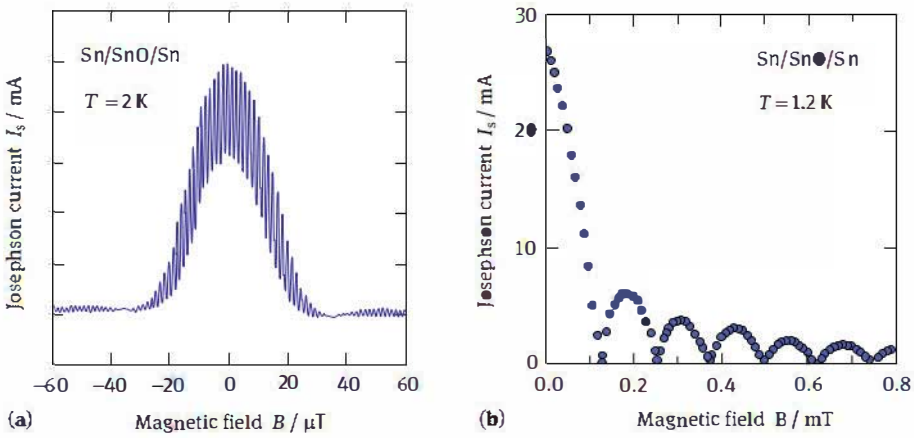


Fig. 11.29: Josephson junctions in a magnetic field. **a)** Current through two Josephson junctions connected in parallel as a function of the magnetic field. (After R.C. Jaklevic et al., Phys. Rev. **140**, A1628 (1965).) **b)** Magnetic field dependence of the current through a single Josephson junction. The applied magnetic field was oriented parallel to the insulator layer of the Sn/SnO/Sn tunnel junction. (After D.N. Langenberg et al., Proc. IEEE **54**, 560 (1966).)

The equations described and the experimental results demonstrate that there is great similarity with phenomena from optics. Obviously there is a formal relationship with light diffraction at single slit or double slit. This becomes particularly clear if we look at the current through the double junction again. As in the optical double-slit experiment, the envelope of light intensity is determined by the slit dimensions, so here the envelope of the tunnel current is determined by the dimensions of the individual Josephson

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

junctions. The common phenomenological basis for these observations in very different fields of physics is the interference of waves.

The high magnetic field sensitivity of the macroscopic wave function of superconductors is used in commercially available measuring devices. These so-called **SQUID**s (**S**uperconductive **Qu**antum **I**nterference **D**evice) are suitable for detection of smallest magnetic field changes down to 10^{-14} T and can serve beyond that as extremely sensitive ammeters and voltmeters. SQUIDs are essentially superconducting rings or cylinders with one or two "weak links", i.e. areas of weak coupling which act like Josephson junctions through which the magnetic field can enter the ring.

11.4 Ginzburg-Landau Theory and Type-II Superconductors

11.4.1 Ginzburg-Landau Theory

So far we have assumed that the Cooper pair density in the sample is constant, but there is a simple argument against the validity of this assumption near interfaces, for example at the boundary between a superconductor and a normal conductor or at the surface of a superconducting sample. It is well known that the curvature of the wave function reflects the kinetic energy. A sudden change in the wave function would therefore result in a drastic increase in energy. This means that the wave function of the superconductor must disappear at the boundary of the sample and increase steadily towards the interior of the superconductor. Since the density of the Cooper pairs according to (11.50) is given by $|\Psi|^2$, it follows that at the surface the Cooper pair density is smaller than in the rest of the sample.

In 1950, *V.L. Ginzburg* and *L.D. Landau* developed a phenomenological theory that takes this aspect into account. In 1959, *L.P Gorkov*[24] showed that the **Ginzburg-Landau theory** can be traced back to the BCS theory and that its validity is not, as originally assumed, limited only to temperatures in the immediate vicinity of the transition temperature.

The Ginzburg-Landau theory is a further development of the Landau theory of the second order phase transition, in which the Gibbs free energy is expanded with respect to an order parameter up to the fourth order. While in classical Landau theory the order parameter is real and does not change spatially, in Ginzburg-Landau theory the order parameter, the wave function $\Psi(\mathbf{r})$, is complex and can vary spatially. In addition, the theory must contain a term for the magnetic field energy and takes into account the coupling to the supercurrent. The expression for the Gibbs free energy per volume has the complicated form

$$g_s = g_n + \alpha|\Psi(\mathbf{r})|^2 + \frac{1}{2}\beta|\Psi(\mathbf{r})|^4 + \frac{1}{2\mu_0}|B_a - B_i|^2 + \frac{1}{2m}\left|(-i\hbar\nabla + 2e\mathbf{A})\,\Psi(\mathbf{r})\right|^2 . \qquad (11.70)$$

24 Lew Petrowitsch Gorkov, *1929 Moscow, †2016 Tallahassee

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The terms with uneven exponents are omitted in this expansion for reasons of symmetry. Since $\Psi(\mathbf{r})$ is position dependent, the expression contains a term proportional to $|-i\hbar\nabla\Psi(\mathbf{r})|^2$, which, as mentioned, prevents sudden changes of the wave function. While the constant α depends on the temperature and disappears at $T \to T_c$, the coefficient β is approximately temperature independent.

The further procedure consists of calculating the Gibbs free energy by integration over the sample volume and minimizing it with the help of the variation method regarding $\Psi(\mathbf{r})$ and $\mathbf{A}$. The result of this calculation are the **Ginzburg-Landau equations**

$$\alpha\Psi + \beta|\Psi|^2\Psi + \frac{1}{2m}(-i\hbar\nabla + 2e\mathbf{A})^2\Psi = 0\,, \tag{11.71}$$

$$\mathbf{j}_s = \frac{ie\hbar}{m}(\Psi^*\nabla\Psi - \Psi\nabla\Psi^*) - \frac{4e^2}{m}|\Psi|^2\mathbf{A}\,. \tag{11.72}$$

As we will see, the theory contains two characteristic lengths, namely the **penetration depth** λ and the Ginzburg-Landau **coherence length** ξ_{GL}. To obtain an expression for the penetration depth, we first consider an extended, magnetic field-free sample. In this case Ψ = const. and equation (11.71) is simplified to $|\Psi|^2 = -\alpha/\beta$. If we insert this result in (11.72), we obtain for the supercurrent

$$\mathbf{j}_s = -\frac{2}{m}\frac{|\alpha|}{\beta}\mathbf{A}\,. \tag{11.73}$$

A comparison with the second London equation (11.5) results in the following expression for the penetration depth

$$\lambda = \sqrt{\frac{m\beta}{4\mu_0 e^2|\alpha|}}\,. \tag{11.74}$$

The *coherence length* ξ_{GL} reflects the characteristic length over which the wave function can change. To connect this quantity with the parameters of the Ginzburg-Landau equations, we consider a superconductor that occupies the half space $x > 0$. Without a magnetic field, (11.71) is simplified in this case to

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} + \alpha\Psi + \beta\Psi^3 = 0\,. \tag{11.75}$$

Now we introduce the function $\tilde{\Psi}(x) = \Psi(x)/\Psi_{\infty}$. The index ∞ indicates that Ψ_{∞} is the solution for $x \to \infty$, i.e. the solution deep inside the sample. Further we use the already mentioned Ginzburg-Landau coherence length (compare to (11.11))

$$\xi_{GL} = \frac{\hbar}{\sqrt{2m|\alpha|}}\,. \tag{11.76}$$

Thus (11.75) takes the form

$$\xi_{GL}^2\frac{d^2\tilde{\Psi}(x)}{dx^2} + \tilde{\Psi}(x) - \tilde{\Psi}^3(x) = 0\,. \tag{11.77}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

With the boundary conditions

$$\tilde{\Psi}(0) = 0, \qquad \lim_{x\to\infty} \tilde{\Psi}(x) = 1 \qquad \text{and} \qquad \lim_{x\to\infty} \frac{d\tilde{\Psi}(x)}{dx} = 0 \tag{11.78}$$

we finally find

$$\tilde{\Psi}(x) = \tanh \frac{x}{\sqrt{2}\,\xi_{GL}} . \tag{11.79}$$

This result is illustrated in Figure 10.30. As indicated, the charge carrier density $n_s(x)$ increases steadily from zero at the interface to $n_s(\infty) = |\Psi_\infty|^2$. The range over which this increase occurs is determined by the coherence length ξ_{GL}.

In addition, it should be noted that the wave function of the superconductor also extends somewhat into the normal conductor or insulator. This behavior is called the *proximity effect*. The proximity effect gives rise to a number of interesting phenomena, one particular example is the Josephson effect, which is based on the penetration of the wave function through the insulating layer.

At the end of our very short introduction to the Ginzburg-Landau theory we add an important result without deriving it. As we will see in the next section the **Ginzburg-Landau-Parameter**

$$\kappa = \frac{\lambda}{\xi_{GL}} = \sqrt{\frac{m^2\beta}{2\mu_0\hbar^2 e^2}} \tag{11.80}$$

plays an important role in the theory (see (11.83)). It contains only the expansion coefficient β, which, in contrast to the parameter α, is almost independent of temperature. The value of the Ginzburg-Landau parameter κ is a characteristic quantity for the superconductor under consideration.

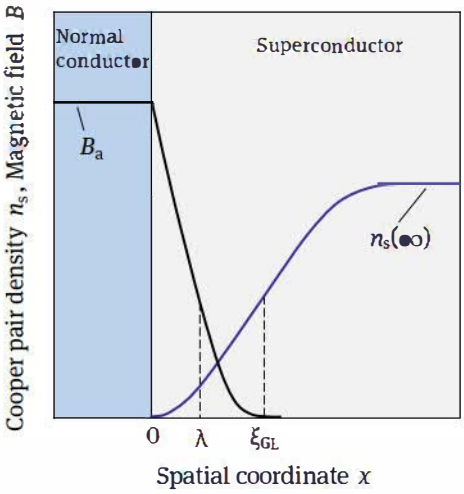


Fig. 11.30: Schematic spatial variation of the magnetic field B and the density n_s of the Cooper pairs at a normal conductor-superconductor interface. The course of the magnetic field (black solid line) is determined by the penetration depth λ, the course of the Cooper pair density (blue solid line) by the coherence length ξ_{GL}. The proximity effect was not taken into account here.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

11.4.2 Type-II Superconductors and Interface Energy

In our discussion of superconductivity, we assumed that superconductors behave like ideal diamagnets up to the critical field strength B_c and then lose their superconducting properties. In fact, this behavior is only found in **type-I superconductors**. But there are also superconductors, namely the **type-II superconductors**, which differ significantly in their reaction to external magnetic fields as shown in Figure 11.31. They react to small fields in the same way as the type-I superconductors discussed before, whose sample interior is field-free. This region, in which the two types of superconductors do not differ, is called the **Meissner phase**. From the *lower critical field* B_{c1} onwards, the magnetic field penetrates the superconductor gradually, so that the external field is only partially shielded. This state is called **Shubnikov phase**[25] or mixed state. When the *upper critical field* B_{c2} is exceeded the sample becomes fully normally conducting.

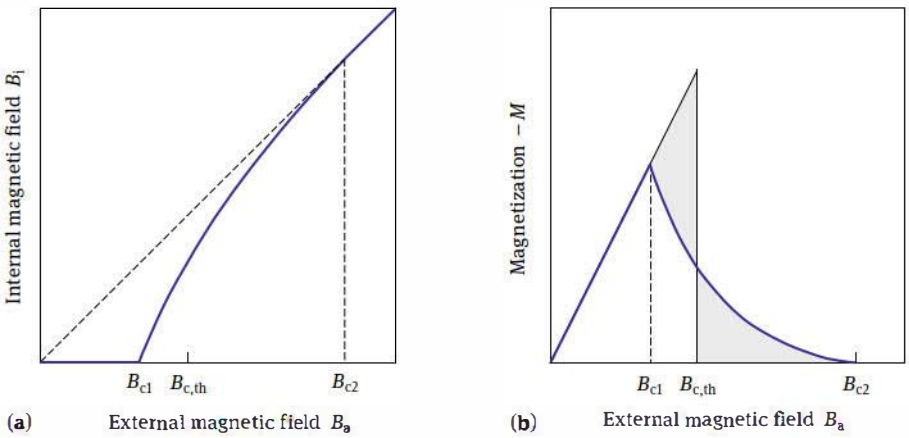


Fig. 11.31: a) Internal magnetic field B_i in a type-II superconductor and **b)** the associated (negative) magnetization M as a function of the applied field B_a. The Meissner phase occurs below B_{c1}, the mixed state or Shubnikov phase in the region $B_{c1} < B < B_{c2}$. At B_{c2} the superconductivity breaks down. Also shown is the thermodynamic critical field $B_{c,th}$. The two grey areas are of equal size and define the thermodynamic critical field.

The *thermodynamic critical field* $B_{c,th}$ is given by the relationship

$$\frac{B_{c,th}^2}{2} = \int_0^{B_{c2}} \mu_0 M \, dB_a \,. \tag{11.81}$$

It follows that the two grey areas in Figure 11.31h must be of equal size. For type-I superconductors $B_{c,th}$ is identical with the critical field strength B_c. As can be seen in

25 Lew Vasilyevich Shubnikov, *1901 Saint Petersburg, †1937 (Ukraine)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 11.32 for an indium-bismuth alloy, the critical field strengths show a very similar temperature dependence. In a first approximation, only their prefactors differ.

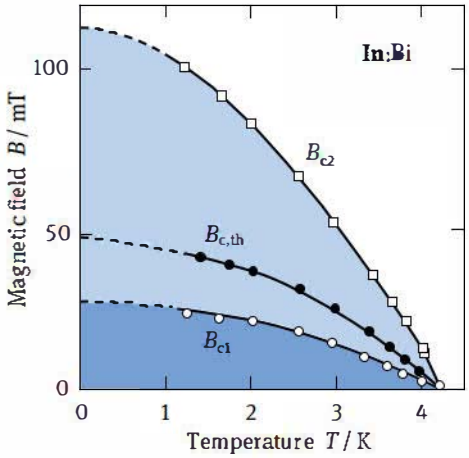


Fig. 11.32: Critical magnetic field strengths of an indium-bismuth alloy. Besides the two critical field B_{c1} and B_{c2} also the calculated thermodynamic critical field $B_{c,th}$ is indicated. (After T. Kinsel et al., Rev. Mod. Phys. **36**, 105 (1964).)

The question now arises what causes the different behavior of the two types of superconductors. The explanation for this is given by a simple consideration regarding the interface between normal and superconductors within the framework of the Ginzburg-Landau theory. Interfaces reduce the condensation energy, since the Cooper pair density and thus the gain in interaction energy is reduced there. It would therefore be expected that superconductors avoid the formation of interfaces as much as possible. However, there is an opposite effect, the energy cost for expelling a magnetic field in order to keep the interior of the superconductor field-free.

We want to estimate the relative contribution of the two effects. For this purpose we replace the steady increase of the Cooper pair density at the interface by a step function. We assume that a layer at the interface with the thickness of the coherence length is free of Cooper pairs. With equation (11.15) the reduction of condensation energy ΔE_{cond} can be given: for an interface with area A, we find $\Delta E_{cond} = A\xi_{GL}B_{c,th}^2/2\mu_0$. When a magnetic field is applied, the field-expulsion energy $E_B = VB^2/2\mu_0$ is needed according to equation (11.14) to keep the superconductor field-free. The formation of interfaces reduces this energy. Again, we roughly approximate the steady change of the magnetic field in the superconductor by a step function. We assume that the superconductor is penetrated by the field at the interface in a layer, which is given by the penetration depth λ, so we can calculate the energy to expel the magnetic field by $\Delta E_B = A\lambda B_{c,th}^2/2\mu_0$. Thus in this simplified model an interface causes a net energy change given by

$$\Delta E_{inter} = \Delta E_{cond} - \Delta E_B = (\xi_{GL} - \lambda)A\frac{B_{c,th}^2}{2\mu_0} . \qquad (11.82)$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

If $\xi_{GL} > \lambda$, ΔE_{inter} remains positive for all magnetic fields, i.e. the formation of interfaces is suppressed. This is the situation in type-I superconductors. If, on the other hand, $\xi_{GL} < \lambda$, as in type-II superconductors, the formation of interfaces is energetically more favorable.

A detailed calculation yields a slightly different numerical value of the Landau-Ginzburg parameter κ, which marks the boundary between type-I and type-II superconductors:

$$\kappa = \frac{\lambda}{\xi_{GL}} \quad \begin{cases} < \frac{1}{\sqrt{2}} & \text{for type-I superconductors ,} \\ > \frac{1}{\sqrt{2}} & \text{for type-II superconductors .} \end{cases} \tag{11.83}$$

The previous discussion could give the impression that the magnetic field only penetrates into type-II superconductors above the thermodynamic critical field $B_{c,th}$, but Figure 11.31 shows that the limit is B_{c1}. As mentioned above, the formation of interfaces is energetically favorable as long as $\xi_{GL} B_{c,th}^2/2\mu_0 - \lambda B^2 < 0$. If we also consider the factor $\sqrt{2}$ in equation (11.82), the magnetic field is expected to enter when

$$B^2 > \frac{B_{c,th}^2}{\kappa\sqrt{2}} . \tag{11.84}$$

The magnetic field therefore penetrates the sample already at field strengths $B < B_{c,th}$.

After discussing the energy of a normal conductor-superconductor interface, we now turn to the questions of how the magnetic flux is spatially distributed in type-II superconductors and how the interfaces look like. The magnetic flux distribution in the Shubnikov phase is illustrated schematically in Figure11.33a. The magnetic field penetrates the sample in small normal conducting channels called **flux tubes** or **flux vortices**. Each flux tube carries one flux quantum, since the arguments presented in Section 11.3 for superconducting rings also apply here. In perfect crystals, the flux tubes arrange themselves regularly in an **Abrikosov lattice**[26]. After decorating a type-II superconductor with fine iron particles, this flux tube lattice was made visible for the first time with the help of an electron microscope in 1967.

For some time now, such flux tube lattices can also be investigated with scanning tunneling microscopes. An example is shown in Figure 11.33b with the image of flux tubes in $NbSe_2$. The optical contrast in the image is due to the different work function of the electrons and thus to the different tunnel current characteristics in normal and superconducting regions. In fact, today in such measurements even the electronic density of states can be determined spatially-resolved in and around the flux tubes.

Typical representatives of type-II superconductors are alloys, transition metals, metallic glasses and also the new types of cuprate or oxide superconductors, which we will discuss in more detail in the next section. The value of B_{c2} can be much larger than

26 Alexei Alexejewitsch Abrikosov, *1928 Moskau, †2017 Palo Alto, Nobel Prize 2003

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

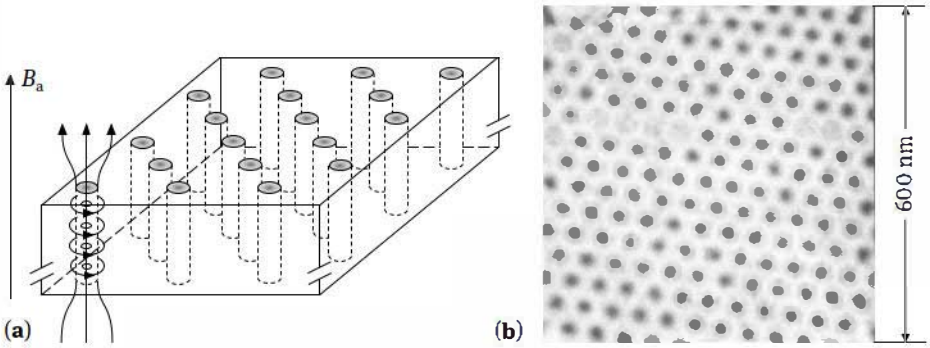


Fig. 11.33: **a)** Arrangement of the flux vortices in the Shubnikov phase. For one flux tube magnetic field lines and shielding currents are indicated. **b)** Image of an Abrikosov lattice measured with a scanning tunneling microscope. The image of the flux tube lattice in $NbSe_2$ was obtained at 1.8 K and a magnetic field $B = 1$ T. (From H.F. Hess et al., Phys. Rev. Lett. **62**, 214 (1989).)

the critical field of type-I superconductors. For example, for the high-field superconductor $PbMo_6S_8$, which belongs to the Chevrel compounds,[27] the value $B_{c2} = 60$ T has been observed at $T \approx 0$ (see Table 11.6). Nb_3Sn, which belongs to the A15 compounds[28] and has the high critical parameters $T_c = 18.7$ K and $B_{c2}(T = 0) = 25$ T, is of great technical importance for the production of magnets with very high field. Since it is difficult, to process this material, wires made of NiTi alloys are commonly used for magnets for smaller fields.

Tab. 11.6: Transition temperature T_c and critical field B_{c2} of some type-II superconductors

Superconductor	NbTi	Nb_3Sn	Nb_3Ge	Nb_3Ga	V_3Si	$PbMo_6S_8$	$PbMo_6Se_8$
T_c(K)	9.5	18.7	23.2	20.3	17.1	15.3	6.7
B_{c2}(T)	15	28	38	34	25	60	7

In type-I superconductors, electric direct currents flow without loss. This is of course also true for type-II superconductors in the Meissner phase, but the picture changes drastically with magnetic fields greater than B_{c1}. In a perfect type-II superconductor in the Shubnikov phase, electrical resistance occurs even at very low currents. This is due to the fact that the current exerts Lorentz forces on the flux tubes and moves them through the sample. The relatively complex loss mechanisms associated with this

27 Chevrel compounds have the composition MMo_6X_8, where M is a rare earth metal and X is sulfur or selenium.

28 A15 compounds have the composition A_3B and crystallize in the β-tungsten-structure. In Nb_3Sn the niobium atoms are arranged in a chain and have a distance which is smaller than in metallic niobium.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

motion will not be discussed here. In real samples, the flux tubes are usually not really free to move, as they are held in preferred places called pinning centers. If the Lorentz force is too small to rip off the flux tubes, the current flows without resistance even in the type-II superconductor. This phenomenon is illustrated in Figure 11.34, which shows the current-voltage characteristics of two $Nb_{0.5}Ta_{0.5}$ samples with different defect concentrations. The experiment was performed in the Shubnikov phase at 3 K and a field of 0.2 T. In the sample with the larger disorder, no voltage drop up to a critical current of 1.2 A could be detected, whereas in the sample with few defects a voltage drop occurred already at above 0.2 A. Lattice defects, such as those produced during cold working, or even small crystallites and precipitates in the sample act as pinning centers.

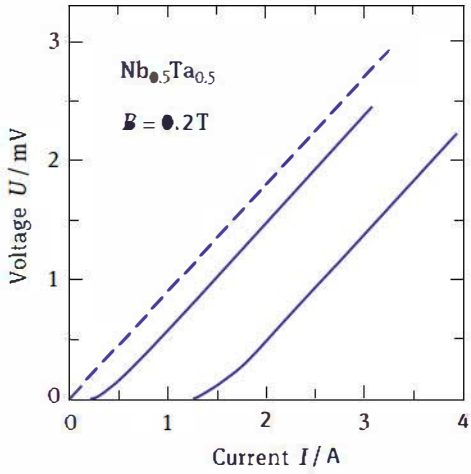


Fig. 11.34: Current-voltage characteristic in the mixed state. The two solid curves were measured on samples with different defect concentrations. The characteristic for a defect-free sample is shown by a dashed line. (After A.R. Strnad et al., Phys. Rev. Lett. **13**, 794 (1964).)

At the end of this section it should be noted that penetration depth, coherence length and critical fields are closely related. A more detailed discussion of the Ginzburg-Landau theory reveals the interesting relationships

$$B_{c1} \approx \frac{\Phi_0}{4\pi\lambda^2}(\ln\kappa + 0.08) \ , \tag{11.85}$$

$$B_{c2} = \frac{\Phi_0}{2\pi\xi_{GL}^2} \ , \tag{11.86}$$

$$B_{c,th} = \frac{\Phi_0}{\sqrt{8}\pi\lambda\xi_{GL}} \ . \tag{11.87}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

11.5 Unconventional Superconductors

In general, a distinction is made between conventional and unconventional superconductors. *Conventional* superconductors are those that can be described by the BCS theory. This means that the formation of the Cooper pairs is caused by phonons and their orbital momentum and spin are zero, i.e. $L = 0$ and $S = 0$. These are the superconductors we have considered so far. The *unconventional* superconductors are materials that exhibit superconductivity that is not consistent with BCS theory. Their quantum states are different from those of BCS superconductors, respectively the attraction between the electrons is not caused by phonons. The most prominent representatives of the unconventional superconductors are the high-temperature superconductors. Furthermore, organic superconductors with a quasi one-dimensional structure, alkali metal fullerenes, boron carbides, ruthenates, pnictide and heavy fermion systems are also included. It is astonishing that, as shown in 2006 by *H. Hosono*[29], superconductivity also occurs in the iron-containing pnictides.

Research on unconventional superconductors is one of the most interesting areas of low-temperature physics, and many surprising results are still to be expected. We limit ourselves here to a relatively brief discussion of high-temperature superconductors and the alloy UPt_3, which belongs to the heavy fermion systems.

11.5.1 High-temperature Superconductors

It has always been a goal of low-temperature research to find or develop superconductors with a particularly high transition temperature, but the "right" approach was not known. If not at room temperature, T_c should be at least above 77 K which is the boiling point of liquid nitrogen. A crucial step in this direction was taken in 1986 by *J.G. Bednorz*[30] and *K.A. Müller*[31] when they investigated the Ba-La-Cu-O compound system and found a transition temperature of about 30 K, which was clearly above the values known up to that time. Somewhat later, T_c values of up to 92 K were measured for Ba-Y-Cu-O compounds, and values of up to 125 K and 135 K respectively for Tl-Ca-Ba-Cu-O and Hg-Ba-Ca-Cu-O compounds. Due to their composition, these materials are called **cuprate superconductors**. Occasionally the term *ceramic high-temperature superconductors* is also used. Table 11.7 lists the transition temperatures of some unconventional superconductors including several high-temperature superconductors. Their relatively complicated composition makes it easy to understand why it is difficult to produce samples with suitable stoichiometry and to grow single crystals.

29 Hideo Hosono, *1953 Kawagoe
30 Johannes Georg Bednorz, *1950 Neuenkirchen, Nobel Prize 1987
31 Karl Alexander Müller, *1927 Basel, Nobel Prize 1987

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Tab. 11.7: Transition temperature T_c of some unconventional superconductors.

Superconductor	T_c (K)	Superconductor	T_c (K)
Sr_2RuO_4	1.5	$YBa_2Cu_3O_7$	92
$RuSr_2(Gd,Eu,Sm)Cu_2O_8$	58	YPd_2B_2C	23
$SmFeAsO_{0.85}$	55	$Bi_2Sr_3Ca_2Cu_3O_6$	110
$GdFeAsO_{0.85}$	53.5	$Tl_2Ba_2Ca_2Cu_3O_{11}$	125
$La_{0.9}F_{0.2}FeAs$	28.5	$HgBa_2Ca_2Cu_3O_8$	134
$ET_2Cu[N(CS)_2]Br$[32]	10.4	$Hg_{0.8}Tl_{0.2}Ba_2Ca_2O_8$	138

The most thoroughly investigated system is $YBa_2Cu_3O_{6+x}$, which is usually referred to as **YBCO** or also as **Y123** because of the relative number of metal atoms. The composition of this compound can be varied widely. Figure 11.35 shows its structure that is typical for cuprates for the special cases $x = 1$ and $x = 0$. The crystals are orthorhombic or tetragonal and have a perovskite structure. The unit cell has the stacking sequence Y-CuO_2-BaO-CuO_x-BaO-CuO_2- ... . The CuO_2-planes, which are perpendicular to the c-axis and are responsible for the formation of superconductivity, are the most important feature of the cuprates. These planes are separated by yttrium atoms or BaO layers.

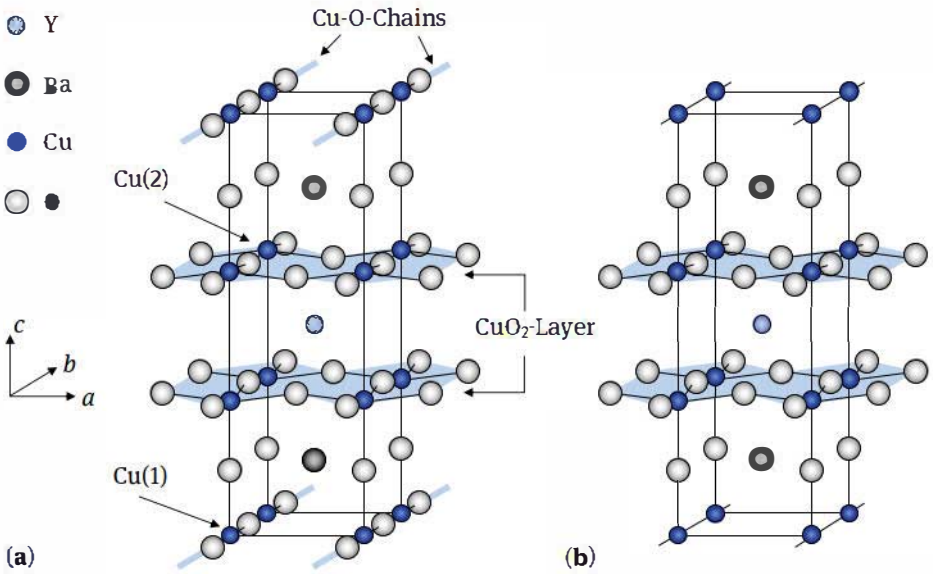


Fig. 11.35: Crystal structure of cuprates. **a)** $YBa_2Cu_3O_7$ (orthorhombic), **b)** $YBa_2Cu_3O_6$ (tetragonal).

32 ET is used as abbreviation for the acronym BEDT-TTF, which itself stands for bis(ethylenedithiolo)tetrathiofulvalene.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As shown in Figure 11.35a, $YBa_2Cu_3O_7$ has Cu-O chains along the b-axis. As the oxygen concentration decreases, vacancies appear there. With $x = 0$ the entire chains are free of oxygen. All cuprate superconductors have CuO_2 layers like YBCO, but differ in the number of these layers and in the structure of the intermediate layers.

The oxygen content determines the electrical properties. Thus, YBCO at $x = 0$ is an insulator in which the copper atoms are *antiferromagnetically* ordered, i.e. the spins of adjacent Cu_2^+ ions are oppositely oriented (cf. Chapter 12). At about $x \approx 0.4$ a metal-insulator transition occurs. In the metallic phase, the conductivity is based on the motion of *holes*. By doping with oxygen, electrons are removed from the layers with copper atoms, so that with increasing oxygen concentration the hole concentration also increases. Above $x \approx 0.4$, superconductivity sets in. The transition temperature increases with x from $T_c = 40\,K$ to 92 K. The "best" superconducting properties are observed at the oxygen concentration $x = 0.92$.

Due to their layer structure, cuprates have highly anisotropic properties, which are reflected, for example, in the electrical resistance of single crystals in the normal conducting state. Figure 11.36 shows measurement results on a $YBa_2Cu_3O_{6.9}$-sample sample for the three main crystal directions. The resistance in the a- and b-direction is slightly different. As could be expected due to the crystal structure, a much higher resistance occurs in the direction of the c-axis, i.e. perpendicular to the layers. In the case of $Bi_2Sr_2CaCu_2O_{8+x}$ the ratio ρ_c/ρ_a is even much higher than 100. Despite the pronounced anisotropic layer structure, the sample becomes superconducting in all directions at the same time, because superconductivity is a three-dimensional phenomenon.

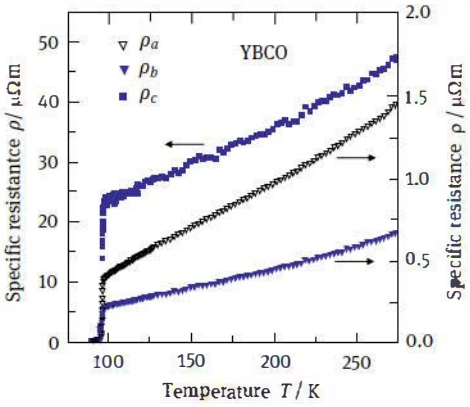


Fig. 11.36: Specific resistance of a YBCO single crystal in the direction of the crystal axes a,b and c. Note the different scales for ρ_a, ρ_b and ρ_c. (After T.A. Friedmann et al., Phys. Rev. B **42**, 6217 (1990).)

Not only the normal conducting, but also the superconducting properties show strong anisotropic effects. An example of this are the critical field B_{c1} and B_{c2}, which are difficult to access in the experiment. In particular, the high fields required for the measurement of B_{c2} cannot be generated experimentally. Therefore, indirect methods

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

are used, where these quantities are derived from measurements of susceptibility, specific heat and penetration depth using theoretical considerations. Figures 11.37a and 11.37b show the result of the analysis of such measurements on $YBa_2Cu_3O_7$ for fields in the direction of the c-axis and perpendicular to this direction. The critical field B_{c1} in the direction of the c-axis is significantly larger than perpendicular to it, while for B_{c2} it is just the other way round. However, this is not surprising, since this relationship follows from equations (11.85) – (11.87). The rather large deviation of one data point in Figure 11.37b close to T_c is due to experimental difficulties caused by fluctuation effects near the phase transition.

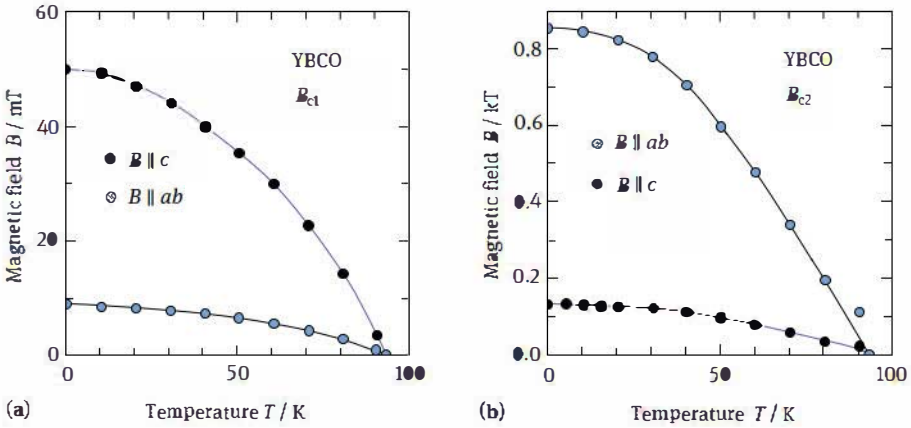


Fig. 11.37: Critical magnetic fields in YBCO parallel and perpendicular to the c-axis. The data were calculated from specific heat and penetration depth measurements. **a)** Lower critical field B_{c1}, **b)** upper critical field B_{c2}. (After D.N. Zheng et al., Phys. Rev. B **49**, 1417 (1994).)

As a result of the relatively small values of the critical magnetic field B_{c1}, fields penetrate deep into cuprate samples. For example, with $\lambda_c \approx 890$ nm in the direction of the c-axis and $\lambda_{ab} \approx 135$ nm perpendicular to it, values are observed which are considerably larger than those of conventional superconductors. Because of $B_{c1} \propto \lambda_L^{-2} \propto n_s$ it follows that in high-temperature superconductors, the density of Cooper pairs is very low. In order to describe the directional dependencies, one introduces directional effective masses in the Ginzburg-Landau theory, which is also applicable to high-temperature superconductors. However, we will not go into these details of the description here.

As mentioned, typical high-temperature superconductors have very high critical fields B_{c2}, but relatively small B_{c1}-values. This fact considerably limits the practical application of these fascinating materials. As we have seen, above B_{c1} the magnetic field penetrates into the superconductor and the Lorentz forces due to current transport cause the flux vortices to move leading to losses. In conventional superconductors, the motion of the flux tubes can be largely prevented by the introduction of pinning

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

centers. At higher temperatures, however, where high-temperature superconductors are particularly attractive for technical applications, the thermal motion of the flux lines is very strong and leads to a particularly poor pinning in the mixed phase.

According to equation (11.86) the large values of B_{c2} are a consequence of the small coherence length ξ_{GL}. For example, the coherence length in YBCO in the *ab*-plane is $\xi_{GL,ab} \approx 1.6$ nm, in the direction of the *c*-axiseven only $\xi_{GL} \approx 0.24$ nm. This means that in cuprates the Cooper pairs extend over only a few atoms. The very small coherence length gives rise to a number of peculiarities in the behavior of high-temperature superconductors. One important consequence is the occurrence of exceptionally strong fluctuation effects. An example is the specific heat of $YBa_2Cu_3O_7$ in the vicinity of the transition temperature, which is shown in Figure 11.38. Note that the lattice provides the main contribution, which is not visible due to the zero-point suppression of the *y*-axis in the plot. In a comparison with the aluminum data in Figure 11.11, it is noticeable that the transition from normal to superconductor does not take place in the form of an abrupt change. The rounded measurement curve is due to the mentioned fluctuation of the Cooper pair density in the vicinity of T_c. The significance of fluctuation effects depends on the number of Cooper pairs in the *coherence volume* ξ_{GL}^3. Whereas in conventional superconductors there are about 10^6 to 10^7 Cooper pairs in this volume, there are only about ten pairs in cuprates. Since the temperature range in which fluctuations play an important role is inversely proportional to the coherence volume, the influence of fluctuations in high-temperature superconductors is still noticeable far from the phase transition.

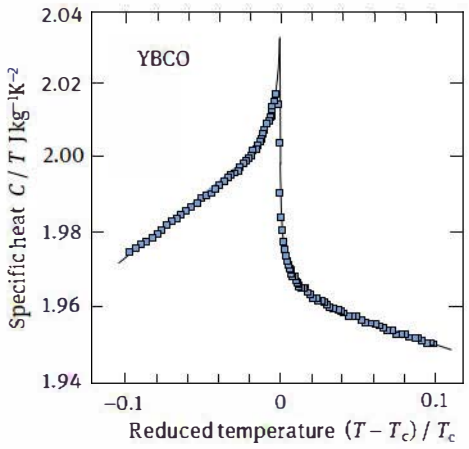


Fig. 11.38: Specific heat C/T of $YBa_2Cu_3O_7$ near the transition temperatures. Note that the zero point of the x-axis is suppressed. (After N. Overend et al., Phys. Rev. Lett. **72**, 3238 (1994).)

There is no doubt that the phenomenological considerations on superconductivity can also be applied to high-temperature superconductors. This is especially true of the London and Ginzburg-Landau theories, because these theories have not been derived from a microscopic point of view. Although an important point is that, unlike

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

classical superconductors, electric charge transport, as mentioned above, occurs by holes and not by electrons, this has no influence on the phenomenological description. Furthermore it was found in experiments that also in high-temperature superconductors the current is carried by Cooper pairs and the magnetic flux is quantized in units of $\Phi_0 = h/2e$. The energy gap was measured with different methods, whereby values for $\Delta(0)/k_B T_c$ between 3 and 4 were found. Although these values are noticeably larger than the BCS prediction, such large values are also found in strongly coupling conventional superconductors.

A very important difference between conventional superconductors and cuprates is the symmetry of the wave function. The Cooper pairs in conventional superconductors are singlet pairs. They have the orbital momentum $L = 0$ and the spin $S = 0$. In the cuprates, the Cooper pairs have the orbital momentum $L = 2$, so they have the symmetry of a d-state. Of the five possible different values of the z-component of the orbital angular momentum only one is actually realized due to the quasi-two-dimensional character of the high-temperature superconductors and the associated crystal field.

An important peculiarity of the d-state is that the energy gap is not isotropic (as in conventional superconductors). It is described by the *energy gap function*, which indicates the size of the energy gap as a function of the angle Φ. In the present case it has the form

$$\Delta_{\mathbf{k}} = \Delta_m \cos 2\Phi \,, \tag{11.88}$$

where Δ_m stands for the maximum energy gap and Φ for the angle between the a-axis and the wave vector $\mathbf{k}$ in the ab-plane.

The energy gap function for different types of superconductors is shown schematically in Figure 11.39. As can be seen in the figure, in s-wave superconductors the energy gap has a finite angle-independent. In the d-wave superconductors discussed here, however, it disappears in certain directions, namely in the ⟨110⟩-directions. In anticipation of the discussion in the next section, the energy gap function of a p-type superconductor is also shown.

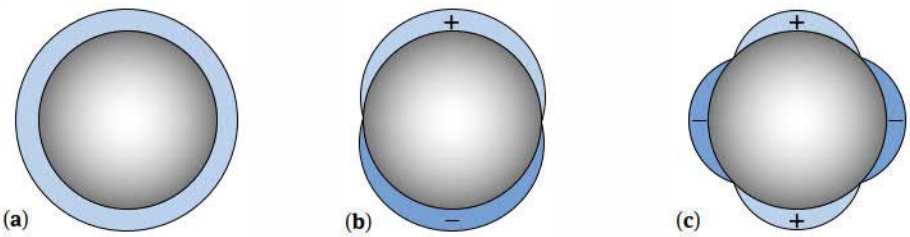


Fig. 11.39: Schematic representation of the energy gap function of **a)** s-wave, **b)** p-wave and **c)** d-wave superconductors. The Fermi surface is shown in grey.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Angle-resolving photoemission spectroscopy allows for the verification of such predictions. Figure 11.40 shows the result of such a measurement on Bi-2212 ($Bi_2Sr_2CaCu_2O_8$). The agreement with (11.88) is convincing.

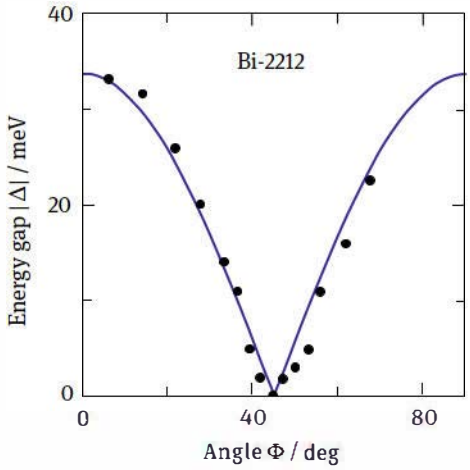


Fig. 11.40: Angular dependence of the energy gap of Bi-2212 measured by photoemission spectroscopy. The solid curve represents equation (11.88). (After H. Ding et al, Phys, Rev. B 54, R9678 (1996).)

At present, no clear statement can be made about the microscopic mechanism of high-temperature superconductivity, because the mechanism of coupling between the charge carriers is still unclear. Since an isotope effect has been observed in some systems, it is reasonable to conclude that phonons are involved in the coupling of the electrons. However, it is unlikely that coupling via the lattice is sufficient to explain the high values of T_c. Common to all cuprate superconductors is a perovskite layer structure with CuO_2-planes, in which the copper atoms are in a mixed Cu^{2+}-Cu^{3+}-state and carry a magnetic moment. There are indications that the spin exchange between the electrons plays an essential role in the coupling process. An important aspect in the theoretical treatment of Cooper pairs in cuprates is their small size, because this makes the Coulomb repulsion between the electrons (holes) more important, so that correlation effects must be taken into account.

11.5.2 Heavy Fermion Systems

As already mentioned, in addition to the cuprates, a number of other unconventional superconductors have been discovered and investigated in detail in recent years in addition to the cuprates. As an instructive example we choose the heavy-fermion system UPt_3. We will see that this alloy has three different superconducting phases, so that the familiar *s*-wave superconductivity can be excluded as the cause in at least two cases. When interpreting the measurement results we assume that the Cooper pairs have a *p*-wave character similar to the superfluid ^3He. Unfortunately, it has not yet been

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

clarified which mechanism is responsible for pair formation in this material. However, we can assume that coupling via the exchange of virtual phonons does not make the crucial contribution. It is assumed that the interaction via the electron spins contributes a substantial part of the pair bonding energy.

The occurrence of different superconducting phases is clearly evident in the data depicted in Figure 11.41. There, measurements of the specific heat and the ultrasonic absorption in the vicinity of the transition temperatures are shown. Let us first look at the specific heat data obtained at different magnetic fields. In the measurements at zero field, two steps unexpectedly appear which indicate phase transitions and thus the existence of different superconducting phases. As can be seen in Figure 11.41a, the step-like increase of the specific heat shifts to lower temperatures with increasing magnetic field. At the same time, one of the two observed steps gradually disappears.

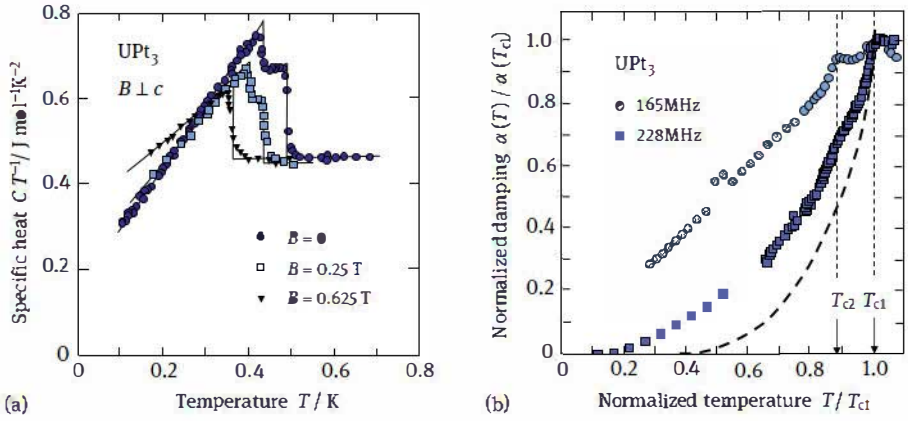


Fig. 11.41: Temperature dependence of the specific heat and the ultrasonic attenuation of UPt_3 in the vicinity of the transition temperatures. **a)** Normalized specific heat C/T as a function of temperature measured at different magnetic fields. (After K. Hasselbach et al., Phys. Rev. Lett. **63**, 93 (1989).) **b)** Normalized ultrasonic attenuation as a function of normalized temperature measured at 165 MHz and 228 MHz. (After M.J. Graf et al., Phys. Rev. B **62**, 14 393 (2000).) The black-dashed line shows the temperature dependence of the attenuation that is expected for a *s*-wave superconductor with $T_c = T_{c1}$.

The temperature dependence of the ultrasound absorption shown in Figure 11.41b supports this argument. The measurements shown were performed at about 200 MHz with transversal waves. The sound waves propagated along the *a*-axis and were polarized parallel to the *b*- or *c*-direction. As mentioned in Section 11.2, the attenuation in superconductors is proportional to the number of thermally-exited quasi-particles, i.e. the number of unpaired electrons. For comparison, the figure shows the absorption curve in a dashed line as predicted by the BCS theory for s-wave superconductors and measured, for example, on aluminum (see Figure 11.18b). Obviously, the attenuation is

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

much greater than expected for singlet superconductors and rises less steeply. This means that the number of quasi-particles that cause ultrasonic attenuation is relatively large in UPt_3. As mentioned in the discussion of specific heat, two transitions which can be clearly seen in ultrasonic data occur at normal pressure. The two transition temperatures are denoted T_{c1} and T_{c2} in Figure 11.41, in contrast to the designation in the specialist literature.

The measurements of thermodynamic quantities and transport properties can be explained by the phase diagram shown in Figure 11.42. Here the course of the phase boundaries was determined with the help of sound velocity measurements. Three superconducting phases can be distinguished, which are marked by 1, 2 and 3. The corresponding transition temperatures in the magnetic field-free case are $T_{c1} \approx 540$ mK and $T_{c2} \approx 490$ mK. The value of T_{c3} depends on the direction of the magnetic field. If the field is parallel to the c-axis, $T_{c3} \approx 380$ mK; if it is perpendicular to this axis, $T_{c3} \approx 430$ mK.

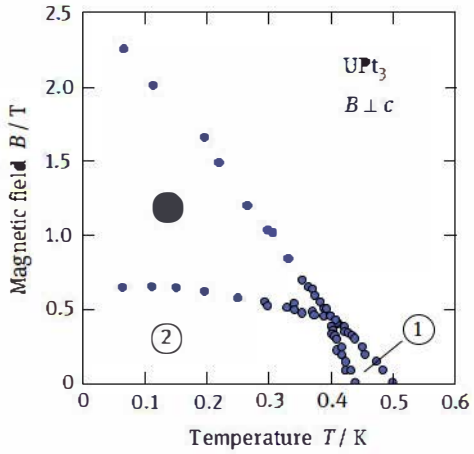


Fig. 11.42: Phase diagram of UPt_3 derived from sound velocity measurements. In these experiments the magnetic field was oriented perpendicular to the c-axis. (After S. Adenwalla et al., Phys. Rev. Lett. 65, 2298 (1990).)

The measurements of specific heat and ultrasonic damping mentioned here can only be understood if the energy gap has nodal lines along which the gap disappears. In this case the excitation of quasi-particles can be performed much more easily than in conventional superconductors. One of the energy gap functions discussed is shown in Figure 11.37b. However, it must be said that despite considerable theoretical efforts, it has not yet been possible to finally clarify the symmetry of the energy gap in the three superconducting regions.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

11.5.3 Technical Application of Superconductivity

At the end of this chapter, some technical applications of the superconductors are briefly mentioned. In principle, superconductors have a great potential for application and in fact they are used in many areas. After the discovery of high-temperature superconductors, great euphoria followed. It was believed that the expensive cooling of superconductors with liquid helium could be replaced by simpler cooling with liquid nitrogen, thus simplifying and expanding the application. However, it soon became apparent that not only the theory of high-temperature superconductors was very complex, but also that the manufacturing aspects were very difficult to master. A major obstacle proved to be that, with some exceptions, larger samples could only be produced in granular form. The further development of high-temperature superconductors therefore progressed much more slowly than originally thought.

Nevertheless, there are areas of application in which superconductors are used with great success. Nb, Nb_3Sn, NbTi and high-temperature superconductors such as YBCO or $Bi_2Sr_2Ca_2Cu_3O_{10}$ are mainly used in low-loss power transmission systems, electric motors, transformers, magnetic levitation, magnetic bearings and low-noise electronic circuits. We would like to emphasize the SQUID magnetometers mentioned before, can be used to detect very weak magnetic fields and smallest magnetic field changes.

A very important field of application is the generation of high magnetic fields. While conventional electromagnets have high losses in the coils, superconducting coils work without losses. Usually wires made of niobium compounds are used, since wires made of high-temperature superconductors still have major technological difficulties. A typical application of such magnetic coils is found in medicine in magnetic resonance imaging. Very large superconducting coils are used in nuclear fusion reactors. The required magnetic fields are generated, for example, in the recently commissioned "Wendelstein 7-X" stellarator in Greifswald Germany using 3.5 m high, complicated-shaped 3 T magnetic coils made of NbTi. It should also be mentioned that superconducting microwave cavity resonators are often built into particle accelerators.

Here we would like to briefly address a technical "problem" when using large coils. If a problem occurs in which the superconductivity in the coil collapses locally, the normally conductive area warms up further due to ohmic heating, and it expands. The coil heats up within a short time, becomes normally conducting, and the high magnetic field energy stored in the system can destroy the coil.

As we have seen in Section 11.2, the current flows in superconducting wires at the surface. It therefore makes sense to avoid large cross-sections and to use many, very thin superconducting wires embedded in copper. This has the advantage that if the superconductivity collapses, during the so-called "quench", the current can be absorbed by the copper and the coil is not destroyed.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

11.6 Exercises and Problems

1. Specific Heat. Depending on the temperature range, the specific heat of a superconductor is larger or smaller than that of the corresponding normal conductor.
(a) How does the difference in specific heat of normal and superconductors depend on the temperature?
(b) In which temperature range is the specific heat of the superconductor lead larger than the normal state specific heat of lead?

2. Condensation Energy. The condensation energy of superconductors follows from the BCS theory, but it can also be derived from thermodynamic considerations. Use both possibilities to calculate its magnitude for lead and compare the results. Does the effective mass play a role? The necessary data are included in this book.

3. Critical Power. The following values apply to lead at 4.2 K: critical magnetic field $B_c = 52.9$ mT, London penetration depth $\lambda_L = 41.5$ nm.
(a) Calculate the transition temperature $T_c(B = 0)$ of lead.
(b) Calculate the critical current I_c at 4.2 K for a lead wire of 3 mm diameter.
(c) Calculate the critical current density j_c at the surface of the wire.

4. Magnetic Field in a Superconducting Plate. A homogeneous magnetic field $\mathbf{B}_0 = B_0\hat{\mathbf{z}}$ is applied parallel to the surface of a thin superconducting plate filling the space $-d/2 \le x \le d/2$.
(a) Calculate the course of the magnetic field inside the plate and
(b) the shielding currents.
(c) How large is the critical field in a thin superconducting film ($d < \lambda_L$) compared to the critical field of a massive sample?

5. Thermodynamic Properties of Aluminum. Thermodynamics can be used to make interesting predictions about the properties of superconductors.
(a) Calculate the difference between the Gibbs free energy of an aluminum cube with an edge length of 1 cm in normal and superconducting state at 0.8 K.
(b) How large is the latent heat that occurs during the transition between the two phases at this temperature?
(c) How large is the specific heat at the temperature at which the specific heat of the normal and superconductors are equal? Aluminum is face-centered cubic with the lattice constant $a = 4.04$ Å.

6. Flux Tubes. A long cylinder of Nb_3Sn with a diameter of 4 mm is exposed to a magnetic field of 1 T at a temperature of 5 K. Estimate the number of flux tubes in the cylinder.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

7. Josephson Junctions in a Magnetic Field. Figure 11.43 shows the measurement curve depicted in Figure 11.29 and another curve taken with a second superconducting double junction. How large were the areas enclosed by the superconductor in each case?

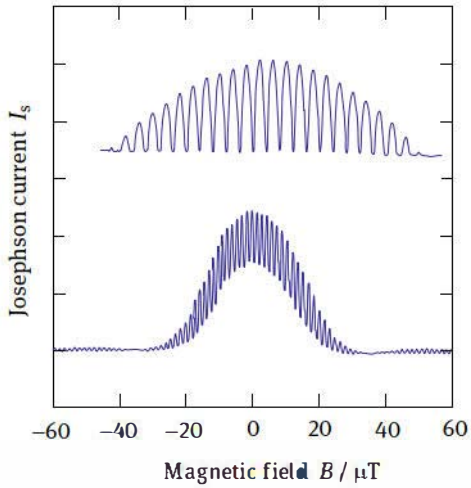


Fig. 11.43: Current through two parallel Josephson junctions as a function of the magnetic field. Two different double junctions were investigated. (After R.C. Jaklevic et al., Phys. Rev. 140, A1628 (1965).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

12 Magnetism

The magnetic properties of solids are closely related to the spin and the orbital motion of electrons. The contribution of the nuclei is comparatively small and only plays a role in solid state physics in special cases. These special cases include areas of low-temperature physics and nuclear spin spectroscopy. We will not consider these here – apart from one specific example.

Depending on whether the sign of the magnetic susceptibility is positive or negative, a sample is called *para-* or *diamagnetic*. Their magnetic responses differ greatly in strength. Diamagnetism is a property of all materials and opposes external magnetic fields, but is very weak. Paramagnetism, when present, is usually stronger than diamagnetism and produces magnetization in the direction of the external field, and proportional to the external field. We will discuss para- and diamagnetism in Section 12.3. *Ferromagnetic* and *ferrimagnetic* solids which exhibit a *spontaneous magnetization* because their magnetic moments are already aligned even without an external field, have a special role. Ferromagnetic effects are very large, producing magnetizations sometimes orders of magnitude greater than the external field and as such are much larger than either diamagnetic or paramagnetic effects. We will focus on these materials and the closely related *antiferromagnets* in Sections 12.3 and 12.4. In the last section we turn to *spin glasses*, in which the magnetic moments are statistically frozen during cooling.

12.1 General Remarks on Magnetic Quantities

Before we start the discussion of magnetic properties, let us recall some important general aspects regarding the denotation and definition of magnetic quantities. We write for the magnetic field, also called magnetic flux density or magnetic induction,

$$\mathbf{B} = \mathbf{B}_{\text{ext}} + \mu_0 \mathbf{M} = \mu_0 (\mathbf{H} + \mathbf{M}) \,, \tag{12.1}$$

where $\mathbf{B}_{\text{ext}}$ stands for the externally applied magnetic field and $\mathbf{H} = \mathbf{B}_{\text{ext}}/\mu_0$ for the magnetizing field, also referred to as magnetic field strength. The magnetization $\mathbf{M}$ is defined as the magnetic moment $\mathbf{m}$ per volume which can be expressed with the help of the mean magnetic dipole moment $\overline{\boldsymbol{\mu}}$

$$\mathbf{M} = \frac{\mathbf{m}}{V} = \frac{N\overline{\boldsymbol{\mu}}}{V} = n\,\overline{\boldsymbol{\mu}} \,, \tag{12.2}$$

where N is the number of magnetic dipoles and n is their density. The magnetic properties of a sample are determined by the *magnetic susceptibility*

$$[\chi] = \frac{\mathbf{M}}{\mathbf{H}} \,. \tag{12.3}$$

https://doi.org/10.1515/9783110666502-012

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

It follows from the definition that the susceptibility is a tensorial quantity. To simplify the equations, we will treat it as a scalar and consider the directional dependency only when necessary.

We will not go into the question of how large the *local magnetic field* is that actually is present at the location of a magnetic dipole moment in a solid. It differs from the applied field for two reasons. First, the *demagnetization field* which depends on the shape of the sample and changes the field inside the solid must be taken into account. Secondly, magnetic moments in the neighborhood influence the local field. Both effects also occur in the treatment of dielectric properties and are much more important there. Therefore we will only deal with this topic in Chapter 13.

12.2 Dia- and Paramagnetism

In paramagnetic materials, the applied field and the induced magnetic moment are aligned so that the external field is amplified. In diamagnetic substances, on the other hand, the induced field is opposite to the applied field and thus causes a weakening of the external field. In the following we will deal with these two phenomena in more detail.

12.2.1 Diamagnetism

We first consider diamagnetism in insulators, which is often referred to as Larmor diamagnetism[1]. A simple explanation for this behavior can be found with the help of classical electrodynamics, if we interpret the orbital motion of the electrons as circular currents. When the magnetic field is switched on, an additional current is induced according to Lenz's rule[2], which counteracts the change in magnetic flux. Since this effect occurs with all atoms or ions, diamagnetism is not a typical solid state property.

The classical and quantum mechanical treatment of diamagnetism lead to the same result. For substances composed of isotropic atoms or ions, one finds after a short calculation the expression

$$\chi_d = -\mu_0 \frac{ne^2}{6m_e} Z\overline{r^2}, \tag{12.4}$$

where Z is the number of electrons, $\overline{r^2}$ is the mean square distance of the electrons from the nucleus and m_e stands for the electron mass.[3] In this approximation, the

1 Joseph Larmor, *1857 Magheragall (Ireland), †1942 Holywood (Northern Ireland)

2 Heinrich Friedrich Emil Lenz, *1804 Tartu (Dorpat), †1865 Rome

3 As an exception, we denote the electron mass with an index to avoid confusion with the magnetic moment.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

diamagnetic susceptibility of solids is independent of direction and temperature. In the case of ionic crystals, the contributions of the different ions can simply be added and then, as with noble gas crystals, good agreement is found with the experimental data. Equation (12.4) is not applicable to covalent or mixed covalent ionic solids because for these the bonding electrons are preferentially located between adjacent atomic cores, leading to an anisotropy on an atomic level.

The *free electrons of the metals* also have diamagnetic properties, which we discussed in Section 9.3. Their circular motion in the magnetic field, which leads to the Landau levels, is disturbed by scattering at room temperature, but still causes a magnetic moment which is directed against the external field. Without going into the calculation of the **Landau diamagnetism**, we give the result here:

$$\chi_d = -\mu_0 \frac{n\mu_B^2}{2E_F}\left(\frac{m_e}{m^*}\right)^2 = -\frac{1}{3}\mu_0\mu_B^2 D(E_F)\left(\frac{m_e}{m^*}\right)^2 , \tag{12.5}$$

with m^* again denoting the effective mass of the electrons. If $m^* = m_e$, then the Landau-diamagnetism compensates for one third of the *Pauli paramagnetism* of the free electrons, which we will discuss in Section 12.2.2. When comparing the calculated values with the experimental ones for metals, the correction for the effective mass of the free electrons and the contribution of the atomic cores must also be taken into account. Since the latter contribution can be quite substantial, the predictions of the free electron model are only of limited use.

Table 12.1 shows experimental values of the molar susceptibility for a number of noble gas atoms and ions.[4] The wide range over which the values of susceptibility are distributed is remarkable.

Tab. 12.1: Molar susceptibility χ_d (in 10^{-6} cm^3/mol) of noble gas atoms and some selected ions. (After W.R. Myrers, Rev. Mod. Phys. **24**, 15 (1952).)

Element	χ_d (10^{-6} cm^3/mol)	Element	χ_d (10^{-6}cm^3/mol)	Element	χ_d (10^{-6} cm^3/mol)
He	−24	Li^+	−88	F^-	−1181
Ne	−85	Na^+	−767	Cl^-	−3042
Ar	−246	K^+	−1835	Br^-	−4337
Kr	−362	Rb^+	−2765	J^-	−6360
Xe	−552	Cs^+	−4412	Mg^{2+}	−541

4 When comparing with literature values, it should be noted that the susceptibility in SI and CGS units has the same dimension, but the SI values are larger by the factor 4π.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

12.2.2 Paramagnetism

The magnetic behavior fundamentally changes when the samples under investigation have permanent magnetic dipole moments. These occur with atoms and molecules having an uneven number of electrons or exhibit partially filled inner shells. Also lattice defects often carry a magnetic moment and must therefore be considered when investigating magnetic properties. As known from atomic physics, the magnetic moment $\boldsymbol{\mu}$ is composed of spin and orbital contributions:

$$\boldsymbol{\mu} = -g\mu_B \mathbf{J} \qquad \text{with} \qquad g = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \,. \tag{12.6}$$

Here g is the *Landé factor*[5], $\hbar\mathbf{J}$ the total angular momentum, L and S are the quantum numbers for orbital angular momentum and spin.

While diamagnetism is largely temperature-independent, the susceptibility of paramagnetic materials usually decreases with increasing temperature, following the **Curie law**[6], with $1/T$. An important condition for the validity of this law is that the interaction between the magnetic moments can be neglected. An example of this typical temperature dependence is shown in Figure 12.1, which depicts the inverse susceptibility of dysprosium sulfate as a function of temperature.

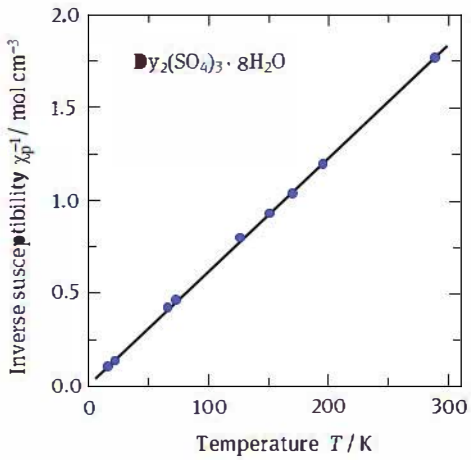


Fig. 12.1: Inverse molar magnetic susceptibility χ_p^{-1} of dysprosium sulfate octahydrate $Dy_2(SO_4)_3 \cdot 8\,H_2O$ as a function of temperature. (After L.C. Jackson, Proc. Phys. Soc. (London) **48**, 741 (1936).)

First we calculate the magnetization in classical approximation, as it was originally derived by *P. Langevin*[7] for an ideal magnetic gas. If a field is applied to a sample with permanent dipoles, it causes a partial alignment of the dipoles, since the potential

5 Alfred Landé, *1888 Wuppertal, †1976 Columbus, Ohio
6 Pierre Curie, *1859 Paris, †1906 Paris, Nobel Prize 1903
7 Paul Langevin, *1872 Paris, †1946 Paris

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

energy $U = -\boldsymbol{\mu}\cdot\mathbf{B} = -\mu B\cos\theta$ depends on the angle θ between field and dipole moment. If we disregard very low temperatures and very high fields, only a partial alignment is found, because under the usual experimental conditions, $\mu B \ll k_B T$ applies.

The mean value $\langle\cos\theta\rangle$ can be calculated using thermodynamics,

$$\langle\cos\theta\rangle = \frac{\int \cos\theta\, e^{-U/k_B T}\, d\Omega}{\int e^{-U/k_B T}\, d\Omega}\,, \tag{12.7}$$

where $d\Omega$ stands for the solid angle element. If one uses the new variable $x = \mu B/k_B T$, the *Langevin function* $L(x)$ is obtained:

$$\langle\cos\theta\rangle = \coth x - \frac{1}{x} \equiv L(x)\,. \tag{12.8}$$

For $\mu B \ll k_B T$ the Langevin function is $L(x) \approx x/3$, so that for a dipole density n one finds the following expression for the magnetization

$$M = n\mu\,\langle\cos\theta\rangle = n\mu\,\frac{\mu B}{3k_B T}\,. \tag{12.9}$$

For the susceptibility this results in the relationship

$$\chi_p = \mu_0 \frac{n\mu^2}{3k_B T}\,. \tag{12.10}$$

The quantum mechanical calculation takes into account that in magnetic fields the energy levels of the atoms split into $(2J+1)$ sublevels, with equal splitting $g\mu_B B$. The temperature dependence of the susceptibility is determined by the thermal occupation of the energy levels. The particularly simple case $J = 1/2$, i.e. the thermal population of two-level systems, has already been discussed in Section 6.2. In the general case, the contributions of the $(2J+1)$ equidistant levels with energy $E = g\mu_B J_z B$ must be added up:

$$M = n\frac{\sum_{J_z=-J}^{J} g\mu_B J_z e^{-g\mu_B J_z B/k_B T}}{\sum_{J_z=-J}^{J} e^{-g\mu_B J_z B/k_B T}}\,. \tag{12.11}$$

The result is:

$$M = ng\mu_B J\mathcal{B}(x)\,. \tag{12.12}$$

Here the dependence of the magnetization on temperature and magnetic field is contained in the argument $x = g\mu_B J B/k_B T$ of the **Brillouin function**

$$\mathcal{B}(x) = \frac{2J+1}{2J}\coth\left[\frac{(2J+1)\,x}{2J}\right] - \frac{1}{2J}\coth\left(\frac{x}{2J}\right)\,. \tag{12.13}$$

For very low temperatures, $x \to \infty$ and $\mathcal{B}(x) \to 1$, because only the ground state is occupied and all moments are aligned in field direction. In most cases, however, the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

splitting is small compared to the thermal energy $k_B T$, i.e. $x \ll 1$. Then the Brillouin function can be expanded and one gets

$$\chi_p \approx \frac{n\mu_0 g^2 J(J+1)\mu_B^2}{3k_B T} = \frac{n\mu_0 p^2 \mu_B^2}{3k_B T} = \frac{C}{T} . \tag{12.14}$$

This is the already mentioned **Curie law**, where C stands for the Curie constant and the quantity p for the *effective number of Bohr magnetons* $p = g\sqrt{J(J+1)}$. Obviously the quantum mechanical and the classical expression, i.e. equation (12.14) and (12.10), converge in the limiting case of high temperatures, if we set $\mu = p\mu_B$. The magnetic moment μ is identified as the maximum magnetic moment in the z-direction. The validity of the Curie law has been confirmed for paramagnetic materials in many experiments. If the inverse susceptibility is plotted as in Figure 12.1, the slope of the straight line directly indicates the value of the Curie constant C and thus the effective number of Bohr magnetons p.

With the help of the *Hund's rules*[8] the ground state of the atoms or ions involved can be determined at given values of the quantum numbers L, S and J. However, we will not go into this in detail but refer to the corresponding literature on atomic physics. We will take a closer look at some special aspects that are of particular importance in solid state physics.

Figure 12.2 shows the temperature and field dependence of the magnetization of paramagnetic salts with pure spin magnetism. The measurements were performed at low temperatures and high magnetic fields. At the highest fields, the *saturation magnetization* was reached. Obviously, the measured magnetization values follow the predictions of the Brillouin function almost exactly.

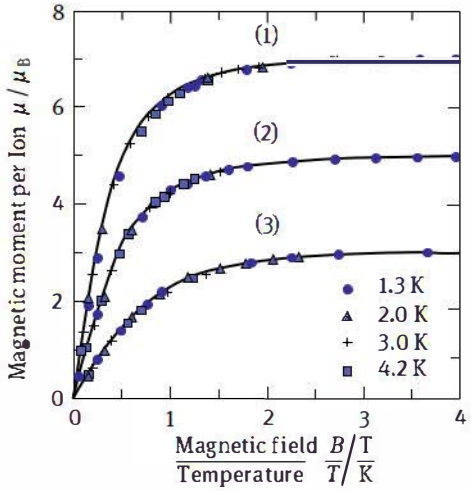


Fig. 12.2: Mean magnetic moment per ion as a function of B/T. Gadolinium sulfate octahydrate (1), iron ammonium alum (2) and potassium-chromium-alum (3) have the spin quantum number $S = 7/2$, $5/2$ and $3/2$ respectively, and the solid lines represent the Brillouin function for each of these values. (After W.E. Henry, Phys. Rev. **88**, 559 (1952).)

8 Friedrich Hund, *1896 Karlsruhe, †1997 Göttingen

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The magnetic properties of rare earth ions and transition metals of the fourth period, the iron series, have been studied in detail in the past. For rare earth ions, the magnetic moments are caused by the $4f$-electrons, which are shielded by the $5s$- and $5p$-electrons located further out. For these materials, the values known from atomic physics and the measured values of the effective number of Bohr magnetons correspond well. Such agreement is not found with the ions of the iron series. There, the contributions of the orbital motion are *quenched* and only the spin contributions remain. This can be understood as follows: in solids there is a spatially varying, strong, inhomogeneous electric field known as the **crystal field**, which is caused by the neighboring ions. In the case of the $4f$-electrons of the rare earths, this field has no great influence, since the $4f$-electrons are located deep in the atomic cores and the crystal field is shielded by the outer electrons. The $3d$-electrons of the ions of the iron series, on the other hand, form the surface of the atomic cores. Their movement is so strongly changed by the crystal field that even the LS-coupling is destroyed. In the inhomogeneous crystal field, the magnitude of the orbital momentum is maintained, but the z-component is no longer a constant of motion. With its temporal average value, the contribution of the orbital motion to the magnetic moment also disappears. As can be seen in Figure 12.2, only the spin contributions of the d-electrons are to be included in the consideration, i.e. $p = 2\sqrt{S(S+1)}$. For example, the measurements on the paramagnetic salt chrome alum are best represented by the quantum numbers $L = 0, S = 3/2$, although the Cr^{3+}-ions are in the ${}^4F_{3/2}$-state.

The situation is more complicated when the magnetic ions with d- or f-electrons are not in insulating materials but in metals with free electrons. A number of interesting effects occur, of which we will pick out one here: the *diluted alloy* **CuMn**, i.e. metallic copper with a low concentration of manganese ions, has paramagnetic properties. The magnetic susceptibility obeys the Curie law, from which the high effective Bohr magneton number $p = 4.9$ for the manganese ions can be derived. This is not unexpected, since free manganese atoms with the electron configuration $3d^5 4s^2$ and ground state ${}^6S_{5/2}$ are strongly magnetic. Surprisingly, however, the magnetic properties of the manganese ions disappear completely when aluminum serves as the host metal. The reason for this is that the localized d-electrons interact strongly with the delocalized s-electrons of the host metal. This leads to a hybridization of the states and thus to a broadening of the $3d$-levels. If the interaction is very strong due to the high density of free electrons, as in the case of aluminum the large broadening leads to a suppression of the local magnetic moments. The influence of the electron density is illustrated by the following observation: The ions of the iron series from vanadium to cobalt show the known magnetic properties in gold. If these ions are introduced in low concentrations in zinc with the higher density of free electrons, magnetic moments only occur for chromium and manganese. When aluminum is host metal, the magnetism is completely suppressed in all cases.

The magnetization of paramagnetic salts is often used in low-temperature physics to measure temperature, because the sensitivity to temperature changes increases

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

with decreasing temperature. A frequently used substance is cerium-magnesium-nitrate (CMN), where the $1/T$ dependence of the susceptibility is observed down to a few millikelvin. In experiments, however, the problem occurs that despite the best possible thermal contact, the thermal equilibrium within the thermometer and the equilibrium between sample and thermometer is only established very slowly. As the thermal conductivity (see Section 9.2) of metals is much higher than that of non-metals at low temperatures, metals are better suited as sensors. In recent years, therefore, metals doped with small amounts of spin-carrying impurities have increasingly been used as low-temperature thermometers. Examples are the systems **Pd**Fe and **Au**Er, where the doping with iron or erbium is in the concentration range of 10 – 100 ppm to avoid the influence of the interaction between the magnetic moments and thus deviations from the Curie behavior.

Figure 12.3a shows the result of a measurement of the susceptibility of **Au**Er at temperatures down to 250 μK (!). In this sample, deviations from the Curie behavior are observed only below 500 μK, which are hardly to see in the representation selected here. These deviations are result from the interaction between the relatively distant spins of the erbium ions. The socalled *RKKY interaction*, which is causing this effect will be discuss briefly in Section 12.3.2.

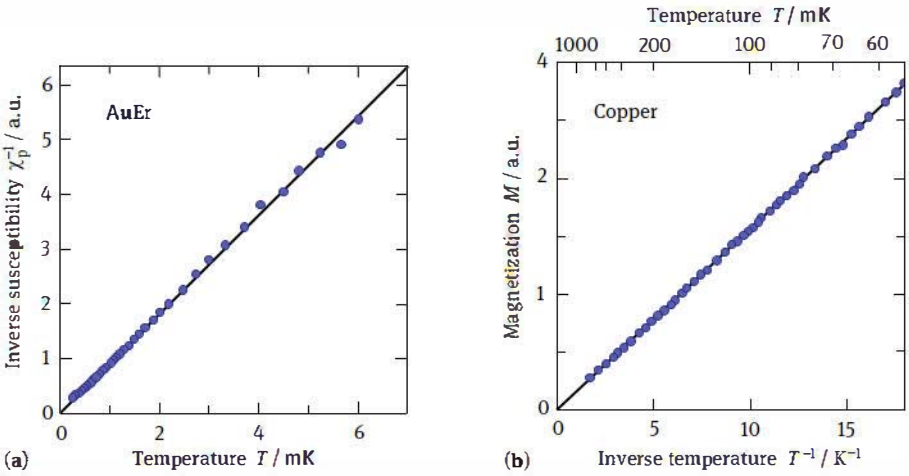


Fig. 12.3: a) Temperature dependence of the inverse magnetic susceptibility χ_p^{-1} of 60 ppm erbium in gold measured down to 250 μK. (After T. Herrmannsdörfer et al., Physica B **284 - 288**, 1698 (2000).) **b)** Magnetization of pure copper as a function of the inverse temperature $1/T$. The measurement was performed in a magnetic field of 0.25 mT. (After R.A. Buhrman et al., *Proc. 12th Int. Conf. Low Temp. Phys.* R.O. Kanda, ed., Academic Press, 1971.)

As mentioned at the beginning of the chapter, nuclei usually play no role in solid state physics, because their magnetic moment is about 1000-times smaller than that of

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the electrons. Nevertheless, nuclei can also be used for magnetization thermometry. Especially at temperatures below 1 mK, nuclear magnetism is suitable for temperature determination. However, longer thermal relaxation times have to be accepted, since electrons and nuclear spins are only weakly coupled. Despite of this, the measurement of the susceptibility of platinum using nuclear magnetic resonance has become a standard method of low-temperature thermometry.

With modern high-resolution SQUID magnetometers,[9] susceptibility measurements can be performed with high sensitivity and accuracy. The result of such a measurement on a highly pure copper sample in a magnetic field of 0.25 mT is shown in Figure 12.3b. The temperature dependence perfectly follows the expected Curie behavior. The absolute value also agrees well with the expected value. Nevertheless, care must be taken with such measurements, since even the smallest amounts of magnetic impurities can make a noticeable temperature-dependent contribution to the magnetization. For example, in small magnetic fields, 1 ppm iron would cause the same magnetization as the copper nuclei in the temperature range shown. Measurements are therefore often carried out at the highest possible fields, because then the contribution of the electron spins to the magnetization is saturated and independent of temperature.

Of course, the assumptions that led to the Curie law (12.14) can only be maintained as long as the interaction *between* the magnetic moments can be neglected. If the interaction energy is not small compared to the thermal energy, deviations from Curie's law occur. This leads to the *Curie-Weiss law*, which we will discuss in the next section.

Pauli Spin Susceptibility. An unsolvable problem of the classical theory of metals at the beginning of the 20th century was the paramagnetism of simple metals. Only the Fermi statistics, together with the knowledge of the magnetic moment of the electrons, allowed an explanation of the temperature-independent paramagnetic susceptibility of the free electron gas.

If an electron is brought into a magnetic field B, the two degenerate spin states split up. This results in a two-level system with an energy splitting $\delta E = g\mu_B B$, whereby we use the value $g \approx 2$ for the electronic g-factor. For electrons with spin 1/2, using equation (12.14) for the susceptibility we find

$$\chi_p \approx \mu_0 \frac{n\mu_B^2}{k_B T} . \tag{12.15}$$

If we use the numerical values for sodium, we get $\chi_{Na} = 6.9 \times 10^{-4}$ for the room temperature susceptibility. But the experimental value $\chi_{Na} = 8.6 \times 10^{-6}$ is much smaller and also temperature independent! A comparison with measurements on other metals shows that the paramagnetic susceptibility calculated with (12.15) is always too large by orders of magnitude.

9 The principle of SQUID magnetometers was discussed in Section 11.3.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As in case of the specific heat of metals, the Fermi statistics prevents the classical occupation of the two states. To illustrate the effect, we split the density of states (8.9) into two parts, as shown in Figure 12.4. The first part comprises of the electrons whose magnetic moments point in the direction of the field, while in the second part the moments oriented opposite to the field. Without a magnetic field, the contributions of the two subsystems to the magnetization just cancel each other out. The field causes the energy zero point of the two partial densities of states to be shifted against each other by $\delta E = g_0 \mu_B B \approx 2\mu_B B$. In equilibrium, the Fermi energy has the same value in both subsystems. This leads to a redistribution so that the number of spins in magnetic field direction increases.

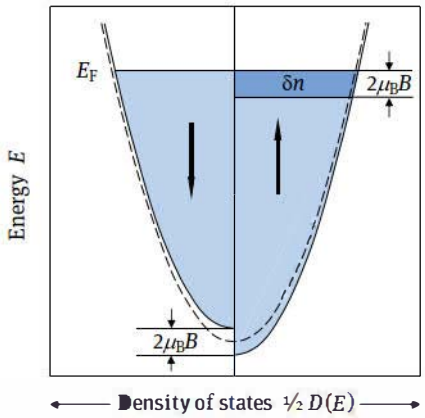


Fig. 12.4: Relative position of the density of states for different spin orientations in magnetic fields. On the left, the magnetic moment of the electrons, indicated by an arrow, is antiparallel; on the right, it is parallel to the magnetic field. The electrons in the dark blue area are responsible for the observed magnetization. The density of states without applied field is shown as a dashed line.

Because of $\mu_B B \ll E_F$ the fraction of spins redistributed is very small. The density of states can therefore be regarded as constant in the range $E_F \pm \mu_B B$. The electron concentration δn, whose magnetic moment is not compensated, can then be approximated by

$$\delta n = \frac{1}{2} D(E_F)\, 2\mu_B B\,. \tag{12.16}$$

The magnetization and the resulting susceptibility are therefore temperature-independent and relatively small:

$$M = \delta n \mu_B = D(E_F)\, \mu_B^2 B = \frac{3 n \mu_B^2 B}{2 k_B T_F}\,, \tag{12.17}$$

$$\chi_{\text{Pauli}} = \mu_0 D(E_F)\, \mu_B^2\,. \tag{12.18}$$

If we compare the two results (12.15) and (12.18), we find

$$\frac{\chi_{\text{Pauli}}}{\chi_p} = \frac{3T}{2T_F}\,. \tag{12.19}$$

As with the specific heat, the Fermi statistics causes a reduction of the susceptibility by a factor of T/T_F compared to the classical calculation, since only the fraction of spins

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

at the Fermi edge contribute to the magnetization that is not compensated by spins oriented in opposite directions.

The magnetic behavior caused by the spins of the free electrons is called **Pauli spin susceptibility** or **Pauli paramagnetism**. Judging the agreement with the experimental data, it must be taken into account that the conduction electrons also have diamagnetic properties based on their orbital motion described by equation (12.5). In the case that the effective mass does not differ from the free electrons, one finds

$$\chi = \chi_{\text{Pauli}} + \chi_{\text{d}} = \frac{2}{3}\mu_0 D(E_{\text{F}})\,\mu_{\text{B}}^2 . \tag{12.20}$$

As already mentioned, when comparing with experimental data, the contribution of the atomic cores must also be considered. In the case of transition metals, it is also important to note that it is not only the *s*-electrons that contribute to Pauli paramagnetism. The agreement between the experimental values and the prediction of equation (12.20) is therefore not convincing in many cases.

Furthermore, it should be noted that the admixture of excited magnetic states to the non-magnetic ground state of atoms or molecules can lead to a temperature-independent contribution, the so-called *Van Vleck paramagnetism*[10], which competes with Langevin's diamagnetism.

12.3 Ferromagnetism

Ferromagnetic materials are characterized by their spontaneous magnetization. This means that magnetic moments can already appear without an external field. Depending on their arrangement, a distinction is made between **ferro**-, **antiferro**- and **ferrimagnetic systems**. The meaning of these terms is illustrated in Figure 12.5, which shows the respective arrangement of the magnetic moments. However, in many cases the arrangement of the magnetic moments is much more complicated. The moments can be tilted against each other or arranged in a spiral. However, we will limit ourselves here to the simple basic forms mentioned.

Ferromagnetism is one of the phenomena of solid state physics for which there is no general microscopic theory. Whereas in dia- and paramagnetism the individual magnetic moments are regarded as independent of one another, ferromagnetism is based on collective phenomena. The formulation of a microscopic theory proves to be a very difficult task, since single and multi-electron aspects play an important role. Although different ferromagnets show the same phenomenological behavior, the microscopic causes can be different, although in all cases the **exchange interaction** is the driving force in the alignment of the magnetic moments. We will limit ourselves

10 John Hasbrouck Van Vleck, *1899 Middletown (Conneticut), †1980 Cambridge (Massachusetts), Nobel Prize 1977

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

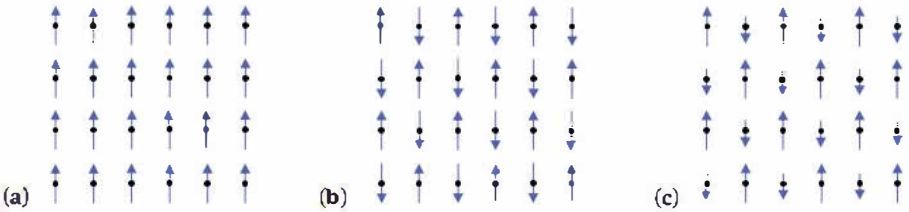


Fig. 12.5: Schematic representation of the three basic forms of magnetic order. **a)** Ferromagnetism: all magnetic moments are directed in the same direction. **b)** Antiferromagnetism: the magnetization of the sublattices cancel each other out. **c)** Ferrimagnetism: the magnetic moments of the two sublattices differ in magnitude.

here to fundamental aspects, with the emphasis on answering the question: why do magnetic molecules align without an external field?

12.3.1 Meanfield Approximation

In the first step, we describe the magnetization of ferromagnetic materials using the **mean field approximation**, which goes back to *P.-E. Weiss*[11] (1907). Since it does not rely on microscopic phenomena, it is generally applicable. The starting point of the mean field approximation is the assumption that the strong mean field $\mathbf{B}_M$, which is proportional to the magnetization $\mathbf{M}$, also acts on each dipole in addition to the external field $\mathbf{B}_{ext}$. However, the mean field is *not* a real magnetic field, but represents in this description the (non-magnetic) interaction of the atom with all other atoms. It is therefore not included in the Maxwell equations. The mean field approximation reduces the complex many-particle problem to a (modified) single-spin problem, i.e. to the behavior of a reference moment in the average "field" caused by its neighbors. For the effective field $\mathbf{B}_{eff}$ we use the approach

$$\mathbf{B}_{eff} = \mathbf{B}_{ext} + \mathbf{B}_M = \mathbf{B}_{ext} + \lambda\mu_0\mathbf{M} \,. \tag{12.21}$$

Here we have introduced the **mean field constant** λ, which is assumed to be temperature independent. Local fluctuations of the magnetization are ignored in this description, although they are extremely important near the transition from the ferromagnetic to the paramagnetic phase.

The aim of our theoretical considerations is the description of the spontaneous magnetization $\mathbf{M}_s(T)$, i.e. the magnetization which already occurs without an external magnetic field. The experimental data for nickel and iron in Figure 12.6 show that the magnetization is almost temperature-independent at low temperatures and then drops

11 Pierre-Ernest Weiss, *1865 Mulhouse, †1940 Lyon

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

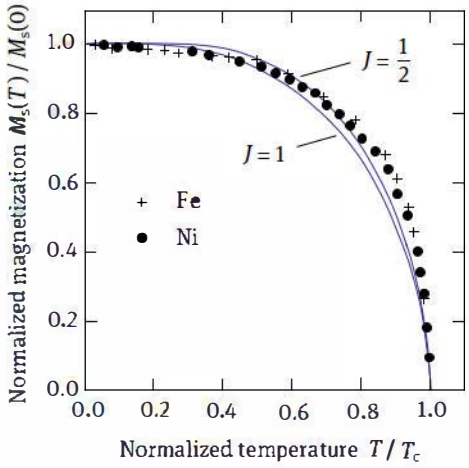


Fig. 12.6: Spontaneous magnetization of nickel and iron as a function of temperature, plotted in normalized units. The solid curves are obtained in the molecular field approximation for the values $J = 1/2$ and $J = 1$. (After P. Weiss, R. Forrer, Ann. Physique **5**,153 (1926); H. Potter, Proc. Roy. Soc. (London) **A 146**, 362 (1934).)

steeply towards the transition temperature or **ferromagnetic Curie temperature** T_c. Above T_c, the spontaneous magnetization vanishes and the ferromagnetic material enters the magnetically disordered, paramagnetic phase. Furthermore, we expect from the theory statements about the relationship between the experimental parameters that determine the temperature dependence of the magnetization and the Curie temperature. In particular, we will deal with the effect of the molecular field $\mathbf{B}_M$ and the significance of the molecular field constants λ.

First we discuss the temperature dependence of the spontaneous magnetization. We start from equation (12.12), which describes the magnetization in the paramagnetic state. In the ferromagnetic phase, which we now consider, in addition to the external magnetic field $\mathbf{B}_{ext}$, the molecular field $\mathbf{B}_M$ is also effective, which, as we will see, far exceeds the external field. We therefore neglect $\mathbf{B}_{ext}$ and replace the external field in the argument of the Brillouin function (12.13) by the molecular field $B_M = \lambda\mu_0 M$. Thus, for the magnetization we get the relation

$$M_s = ng\mu_B J\mathcal{B}(x) \qquad \text{with} \qquad x = \frac{g\mu_B J\lambda\mu_0 M_s}{k_B T} . \tag{12.22}$$

Due to the occurrence of the magnetization in the argument of the Brillouin function, the feedback effects are taken into account. The numerical solution of the equation leads to curves as shown in Figure 12.6 for $J = 1/2$ and $J = 1$. Considering how rough the approximations of molecular field theory are, the agreement with the experimental data for iron and nickel is quite satisfactory. A closer comparison of theory and experiment, however, reveals significant deviations not only near the Curie temperature, but also at low temperatures. There, a somewhat stronger variation of the magnetization with temperature is observed than the simple molecular field approximation predicts. An explanation for this will be given towards the end of this section.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As can easily be shown, equation (12.22) only has solutions for temperatures below a certain critical temperature. For this *ferromagnetic Curie temperature* T_c, (12.22) gives the expression

$$T_c = \frac{ng^2 J(J+1)\mu_B^2 \mu_0 \lambda}{3k_B} = C\lambda \,, \tag{12.23}$$

with the Curie constant C that was introduced with equation (12.14).

It has already been mentioned that for the metals of the iron series the contribution of the orbital angular momentum to the magnetic moment is suppressed. Therefore only the spin contribution needs to be included in the calculation of the effective Bohr magneton number, i.e. $p^2 = g^2 S(S+1)$. Using the data in Table 12.2 and the volume number density $n = 8.5 \times 10^{28}$ m^{-3} of iron, the value $\lambda \approx 1000$ is calculated for the mean field constant. The spontaneous magnetization of 1.75×10^6 A/m results in roughly 2000 T for the molecular field $B_M = \lambda \mu_0 M_s$. This field is much higher than any magnetic fields that can be produced today in continuous operation. If one calculates the dipole field, which is caused at the location of a reference atom by the magnetic moments of the neighboring spins, one finds $\mu_0 \mu_B / a^3 \approx 0.1$ T. The molecular field is thus far stronger than the magnetic field produced by the neighboring magnetic moments. As already emphasized at the beginning, the molecular field serves as a general description of the interaction between the atoms. The magnetic moments are aligned by forces that obviously do *not* originate in the magnetic interaction between the spins.

Tab. 12.2: Experimental values of the ferromagnetic Curie temperature T_c, the paramagnetic Curie temperature Θ, the saturation magnetization M_s at $T = 0$ and the effective Bohr magneton number p of some ferromagnetic materials. (After *American Institute of Physics Handbook*, D.W. Gray, ed., McGraw-Hill, 1963.)

Material	T_c (K)	Θ (T)	M_s(kA/m)	p
Fe	1043	1100	1750	2.22
Co	1395	1415	1450	1.72
Ni	629	649	510	0.60
Gd	289	302	2060	7.12
Dy	85	157	2920	6.84
EuO	69	78	1930	7.0
EuS	17	19	1240	7.0

Now we take another look at the paramagnetic phase of ferromagnets. We take advantage of the fact that the interaction that leads to spin alignment in the ferromagnetic phase does not suddenly disappear at T_c, i.e. the molecular field also causes an amplification of the external field in the paramagnetic phase. We modify the Curie law (12.14) to incorporate the feedback effect between the effective field B_{eff} and the magnetization M.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Instead of $\mu_0 M = \chi B_{\text{ext}} = C B_{\text{ext}}/T$ we write

$$\mu_0 M = \frac{C}{T}\,(B_{\text{ext}} + B_{\text{M}}) = \frac{C}{T}\,(B_{\text{ext}} + \lambda \mu_0 M)\ . \tag{12.24}$$

From this follows directly the **Curie-Weiss law** valid for ferromagnetic materials in the paramagnetic phase

$$\chi_{\text{p}} = \frac{\mu_0 M}{B_{\text{ext}}} = \frac{C}{T - \lambda C} = \frac{C}{T - \Theta}\ . \tag{12.25}$$

Similar to normal paramagnetic substances, we expect a steep increase in susceptibility with decreasing temperature, but which finally ends in a singularity at the **paramagnetic Curie temperature** $\Theta = C\lambda$. A comparison with (12.23) shows that in the context of molecular field theory the two characteristic temperatures Θ and T_{c} should be identical. In real systems, however, Θ is always slightly larger than T_{c}.

Figure 12.7 shows the inverse susceptibility of nickel in the paramagnetic phase. At high temperatures, i.e. far away from the phase transition, the agreement with (12.25) is very good, but deviations occur near the Curie temperature. Coming from high temperatures, the susceptibility increases less rapidly than predicted by the Curie-Weiss law. If one extrapolates the high temperature data to low temperatures, one finds the value $\Theta = 649$ K. The actual phase transition, however, takes place at $T_{\text{c}} = 630$ K. These deviations are mainly due to fluctuation effects which are typical for second order phase transitions and are not considered in the molecular field approximation. Theories for this phase transition predict, in agreement with the experiment, for the susceptibility at the ferromagnetic phase transition the proportionality $\chi \propto |T - T_{\text{c}}|^{-\eta}$. The parameter η assumes the values 4/3 and 1/3, depending on whether the critical point is approached from high or low temperatures. In the mean field approximation η has the values 1 or 1/2.

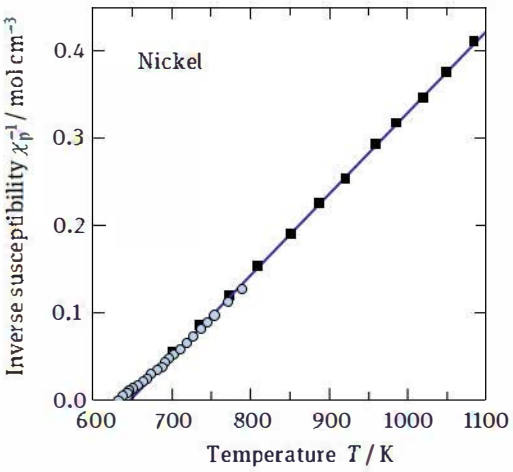


Fig. 12.7: Inverse susceptibility χ_{p}^{-1} of nickel in the paramagnetic phase. The extrapolated paramagnetic Curie temperature of nickel is $\Theta = 649$ K. (After P-E. Weiss, R. Forrer, Ann. Phys. 5, 153 (1926) (light blue circles); W. Sucksmith, R.R. Pearce, Proc. Roy. Soc. (London) A 167, 189 (1938) (black squares).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

12.3.2 Exchange Interaction

We postpone the discussion of the well-known ferromagnets like iron or nickel and first consider the exchange interaction between localized electrons which is mainly found in ferro- or antiferromagnetic insulators.

Exchange Interaction Between Localized Electrons. In this context, the interesting question arises: why is it that in some systems the spins are aligned parallel, and in others antiparallel? We want to discuss this question by means of a very simple system consisting of the two ions a and b and two electrons with the position vectors $\mathbf{r}_i$ and the spins $\mathbf{s}_i$. We have already discussed the essential aspects that arise in context of the covalent bond in Section 2.4. As explained there, the wave function of the electrons $\Psi(\mathbf{r}_1, \mathbf{s}_1; \mathbf{r}_2, \mathbf{s}_2)$ can be represented as a product of a orbital wave function and a spin function. Depending on the symmetry of the orbital wave function, the two spins are oriented either parallel or antiparallel. To describe the orbital wave function, we again use the approach of *W.H. Heitler* and *F. London*. The orbital wave function $\Psi_+ = \mathcal{N}_+[\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) + \psi_b(\mathbf{r}_1)\psi_a(\mathbf{r}_2)]$ of the singlet state is symmetric, the spins are antiparallel, the total spin is zero. The wave function $\Psi_- = \mathcal{N}_-[\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) - \psi_b(\mathbf{r}_1)\psi_b(\mathbf{r}_2)]$ of the triplet state is antisymmetric, the spins are aligned parallel, the total spin is one. The prefactors $\mathcal{N}_+$ and $\mathcal{N}_-$ are determined via the normalization $\int \Psi^*\Psi \, dV_1 \, dV_2 = 1$. We will use the approximation $\mathcal{N}_+ \approx \mathcal{N}_- = \mathcal{N}$.

Now we calculate the potential energy of the ground state. We assume that the potential $\tilde{V}(\mathbf{r}_1, \mathbf{r}_2)$, which describes the interaction of the electrons with the ions and between the electrons, is symmetrical regarding the exchange of the electrons; thus $\tilde{V}(\mathbf{r}_1, \mathbf{r}_2) = \tilde{V}(\mathbf{r}_2, \mathbf{r}_1)$ applies. Using the ansatz for the wave functions, we obtain the following expressions for the potential energy U_+ and U_-

$$U_\pm = 2\mathcal{N}^2 \int \psi_a^*(\mathbf{r}_1)\psi_b^*(\mathbf{r}_2)\tilde{V}(\mathbf{r}_1, \mathbf{r}_2)\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) \, dV_1 dV_2$$
$$\pm 2\mathcal{N}^2 \int \psi_a^*(\mathbf{r}_1)\psi_b^*(\mathbf{r}_2)\tilde{V}(\mathbf{r}_1, \mathbf{r}_2)\psi_a(\mathbf{r}_1)\psi_b(\mathbf{r}_2) \, dV_1 dV_2 \, . \qquad (12.26)$$

The positive sign in the second line applies to the singlet state, the negative sign to the triplet state. Since the contribution of kinetic energy in the two states is almost equal, the difference $(E_+ - E_-)$ of the energy eigenvalues is essentially given by $(U_+ - U_-)$. In our further discussion we therefore use for the **exchange constant** $\mathcal{J}$ the expression

$$\mathcal{J} = E_+ - E_- \approx 4\mathcal{N}^2 \int \psi_a^*(\mathbf{r}_1)\psi_b^*(\mathbf{r}_2)\tilde{V}(\mathbf{r}_1, \mathbf{r}_2)\psi_b(\mathbf{r}_1)\psi_a(\mathbf{r}_2) \, dV_1 dV_2 \, . \qquad (12.27)$$

The sign of $\mathcal{J}$ depends on the shape of the wave functions and the potential. For $\mathcal{J} > 0$, the spins are aligned parallel, if $\mathcal{J} < 0$ they are antiparallel.

We breakdown the potential as follows: $\tilde{V}(\mathbf{r}_1, \mathbf{r}_2) = \tilde{V}_i(\mathbf{r}_1) + \tilde{V}_i(\mathbf{r}_2) + \tilde{V}_{ee}(\mathbf{r}_1, \mathbf{r}_2)$. The first two terms describe the interaction of the electrons with the two ions, the last term represents the interaction between the electrons. Let us first consider the repulsion

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

between the two electrons, for which we can write $\widetilde{V}_{ee} = e^2/(4\pi\varepsilon_0 |\mathbf{r}_1 - \mathbf{r}_2|)$. It always causes a positive contribution to the exchange constant, i.e. the Coulomb interaction between the electrons tries to parallel align the spins. The interaction between electrons and ions is attractive and makes a negative contribution to $\mathcal{J}$. It therefore leads to a reduction of energy in antiparallel oriented spins. The denser the atoms are packed, the stronger the overlap of the wave functions and the greater the energy reduction. Which sign the exchange coefficient $\mathcal{J}$ actually has depends on the relative size of these oppositely acting contributions.

Considering that due to the Pauli principle, spin wave functions are unambiguously linked to the orbital wave functions, the energies can also be expressed by means of the spin states. If we use $\mathbf{s}_1$ and $\mathbf{s}_2$ to denote the spin operators of the two electrons, the relationship $\mathbf{S}^2 = |\mathbf{s}_1 + \mathbf{s}_2|^2 = 3\hbar^2/2 + 2(\mathbf{s}_1 \cdot \mathbf{s}_2)$ applies to the total spin $\mathbf{S}$. The operator $(\mathbf{s}_1 \cdot \mathbf{s}_2)$ assumes the eigenvalue $-3\hbar^2/4$ for the singlet state ($S = 0$) and $+\hbar^2/4$ for the triplet state ($S = 1$). We use this result to formulate a modified Hamiltonian operator H_{spin} to describe the two states, which only affects the spin function of the electrons:

$$H_{spin} = \frac{1}{4}(E_+ + 3E_-) - \frac{(E_+ - E_-)}{\hbar^2}(\mathbf{s}_1 \cdot \mathbf{s}_2) = \frac{1}{4}(E_+ + 3E_-) - \frac{\mathcal{J}}{\hbar^2}(\mathbf{s}_1 \cdot \mathbf{s}_2) \,. \qquad (12.28)$$

Using the eigenvalues of the operator $(\mathbf{s}_1 \cdot \mathbf{s}_2)$, it is easy to see that this Hamilton operator leads to the energy eigenvalues E_+ for the singlet and E_- for the triplet state. Formally, the eigenvalues only depend on the spin orientation. The term $(E_+ + 3E_-)/4$ is unimportant for further considerations, because it disappears when the energy zero point is chosen appropriately, while the second term expresses the energy difference between the different spin orientations.[12] If $\mathcal{J} > 0$, the spins are aligned parallel, if $\mathcal{J} < 0$, the alignment is antiparallel. In the first case the system is ferromagnetic, in the second antiferromagnetic.

The introduction of this Hamiltonian operator may seem somewhat artificial here, and also seems more involved than necessary in the context of the problem at hand, but this procedure can be generalized in a useful way. The equation (12.28) can be extended to any spin operators $\mathbf{S}_i$ and $\mathbf{S}_j$ and different exchange coefficients $\mathcal{J}_{ij}$. This is the starting point of the **Heisenberg model**[13] of ferromagnetism, in which the spin-dependent Hamilton operator has the following appearance

$$H_{spin} = -\sum_i \sum_{j \neq i} \mathcal{J}_{ij} (\mathbf{S}_i \cdot \mathbf{S}_j) \,. \qquad (12.29)$$

12 Many authors also write $-2\mathcal{J}\mathbf{s}_1 \cdot \mathbf{s}_2$ for the last part of the Hamilton operator. In this case, $\mathcal{J}$ represents half the energy difference.

13 Werner Heisenberg, *1901 Würzburg, †1976 Munich, Nobel Prize 1932

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The summation is done over all existing atoms i and all neighbors j. The Heisenberg operator is nonlinear; therefore, despite its simple appearance, the performance of calculations in concrete cases is limited even with rough approximations.[14]

Not in all cases do the magnetic ions interact *directly* with each other, i.e. there is not always an overlap of the electron shells of the spin-carrying ions. There are certainly ferromagnets in which the magnetic ions are separated from each other by a non-magnetic ion. In this case the exchange interaction is mediated by the diamagnetic ions in between. This *super exchange* is found in many ferromagnetic oxides and compound ions of the transition elements. A well-known example is MnO, where the interaction takes place via the diamagnetic O^{2-}-ions.

In the case of rare earths, *indirect exchange* is observed. In these metals, the $4f$-electrons carry the magnetic moments, but the overlap of their wave functions is small. The coupling takes place via the conduction electrons. The magnetic moment of the atom aligns the spins of the conduction electrons in the environment and these in turn align the moments of the neighboring ions. This **RKKY interaction**, named after *M.A. Ruderman, C. Kittel, T. Kasuya*[15] and *K. Yosida*[16], is proportional to $(1/r^3)\cos 2k_F r$, so it has a long-range and an oscillating characteristic. Depending on the distance between the ions, this interaction causes a parallel or antiparallel alignment of the adjacent magnetic moments.

We now want to establish the connection between the exchange constant J, the mean field constant λ and the Curie temperature Θ. For this purpose, we determine the potential energy that a spin has in the molecular field approximation and compare it with the exchange energy. For simplicity, we assume that the sample consists of atoms of the same kind and only the interaction with the z nearest neighbors has to be considered.

In the molecular field approximation a spin $\mathbf{S}_i$ with the magnetic moment $\boldsymbol{\mu}$ has the potential energy

$$U = -\boldsymbol{\mu} \cdot \mathbf{B}_{\text{eff}} = g\mu_B \mathbf{S}_i \cdot \mathbf{B}_{\text{eff}} \approx \mu_0 g \mu_B \lambda \, \mathbf{S}_i \cdot \mathbf{M} \,. \tag{12.30}$$

The last term is obtained by expressing $\mathbf{B}_{\text{eff}}$ through the molecular field $\mu_0 \lambda \mathbf{M}$ and neglecting the external field. When calculating the exchange energy, we replace the operator $\mathbf{S}_j$ in equation (12.29) with its expectation value $\langle \mathbf{S}_j \rangle$. This procedure is equivalent to that in mean field theory, in which $\mathbf{B}_{\text{eff}}$ reflects the *average* strength of the interaction. If we restrict ourselves to the interacting nearest neighbors and consider

14 A very well-known and often used simplification of the Heisenberg model is the *Ising model*, where one only considers the z-components of particles with spin 1/2, i.e. the spins have only two possible orientations.

15 Tadao Kasuya, *1927 Yokohama

16 Kei Yosida, *1922

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the relation between $\langle \mathbf{S}_j \rangle$ and the magnetization $\mathbf{M} = -ng\mu_B \langle \mathbf{S}_j \rangle$, we can write

$$U = -z\mathcal{J}\mathbf{S}_i \cdot \langle \mathbf{S}_j \rangle = \frac{z\mathcal{J}}{ng\mu_B}\mathbf{S}_i \cdot \mathbf{M} \,. \tag{12.31}$$

By comparing (12.30) and (12.31) we obtain for the molecular field constant the expression

$$\lambda = \frac{z\mathcal{J}}{n\mu_0 g^2 \mu_B^2} \,. \tag{12.32}$$

If we now use the relationship (12.25) between the molecular constant and the Curie temperature and insert the Curie constant (12.14), where J is replaced by S since in ferromagnets the orbital contribution to the magnetic moment is suppressed, we obtain

$$\mathcal{J} = \frac{3k_B\Theta}{zS(S+1)} \,. \tag{12.33}$$

The result is as expected: exchange energy and thermal energy at the phase transition are comparable.

Exchange Interaction in the Free Electron Gas. The briefly discussed concept of exchange interaction between localized magnetic moments is suitable for describing the ferromagnetism of many systems, but not of the most well-known ferromagnetic substances such as iron, cobalt or nickel. There, collective properties of the non-localized 3*d*-electrons in the bands play a crucial role. For their description a combination of band model and exchange interaction is required.

Surprisingly, there is already a tendency to align the spins in the free electron gas, if correlation effects are included. Although the following considerations cannot be applied directly to real ferromagnets, we will take a brief look at this effect to discuss the basic ideas. We pick out *two electrons with parallel spin* and, to satisfy the Pauli principle, we represent their common wave function by an antisymmetric superposition of plane waves:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = \mathcal{N}\left[e^{i\mathbf{k}_1\cdot\mathbf{r}_1} e^{i\mathbf{k}_2\cdot\mathbf{r}_2} - e^{i\mathbf{k}_1\cdot\mathbf{r}_2} e^{i\mathbf{k}_2\cdot\mathbf{r}_1}\right] = \mathcal{N} e^{i(\mathbf{k}_1\cdot\mathbf{r}_1 + \mathbf{k}_2\cdot\mathbf{r}_2)}\left[1 - e^{-i(\mathbf{k}_1-\mathbf{k}_2)\cdot(\mathbf{r}_1-\mathbf{r}_2)}\right] \,. \tag{12.34}$$

The factor $\mathcal{N}$ serves here again to normalize the wave function. The probability of finding electron 1 in the volume element dV_1 and electron 2 in the volume element dV_2 is given by the following expression:

$$|\Psi(\mathbf{r}_1, \mathbf{r}_2)|^2 \, dV_1 dV_2 = |\mathcal{N}|^2 \left\{1 - \cos\left[(\mathbf{k}_1 - \mathbf{k}_2)\cdot(\mathbf{r}_1 - \mathbf{r}_2)\right]\right\} dV_1 dV_2 \,. \tag{12.35}$$

This result is remarkable: the probability of finding two electrons with equal spin at the same location disappears for *arbitrary* wave vectors, without having to take the repulsive Coulomb forces into account! If we pick out an electron with a given spin, the probability of finding electrons with the same spin in the neighborhood is strongly

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

reduced. Since the total charge density a free electron sees is composed of electrons with the same and opposite spin, a so-called *exchange hole* is formed, as shown in Figure 12.8. At the position of the reference electron, the electron density is only half as high as at larger distances. A corresponding calculation shows that the radius of this hole is about $(2k_F)^{-1}$, i.e. 1 – 2 Å. This means that the free electrons do not screen the potential of the atomic cores as well as they would without taking the correlation into account. We have not considered this effect in Section 7.3 when discussing the shielding of the Coulomb field. The weaker shielding of the atomic core potential in turn causes a reduction of the energy of the reference electron, i.e. a higher binding energy. The parallelization of as many spins as possible therefore leads to an energy gain. This correlation effect acts like a collective exchange interaction with positive exchange constants.

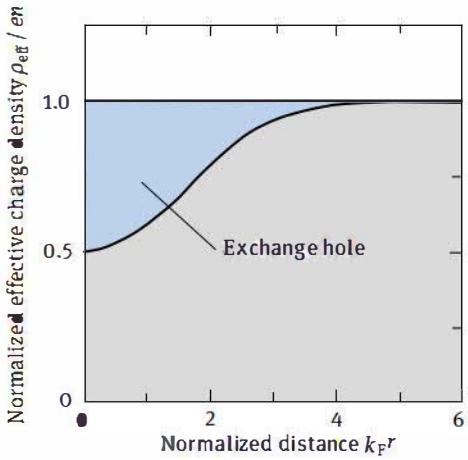


Fig. 12.8: Normalized effective charge density in a free electron gas. The exchange effect reduces the density of the electrons with the same spin orientation in the vicinity of the reference electron.

12.3.3 Band Ferromagnetism

In the model of *E.C. Stoner*[17] and the later advanced version by *E.P. Wohlfarth*[18], the tendency towards spin alignment is taken into account by attributing a positive contribution I to the exchange energy for each electron pair with opposite spin. For the energy of the electrons in the subbands with different spin alignment, the following ansatz is suitable

$$E_\uparrow(\mathbf{k}) = E(\mathbf{k}) + I\frac{n_\downarrow}{n} - \mu_B B \qquad \text{and} \qquad E_\downarrow(\mathbf{k}) = E(\mathbf{k}) + I\frac{n_\uparrow}{n} + \mu_B B\,. \tag{12.36}$$

17 Edmund Clifton Stoner, *1899 Esher, †1968 Leeds
18 Erich Peter Wohlfarth, *1924 Gleiwitz, †1988 London

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Here, $E(\mathbf{k})$ stands for the energy eigenvalue in the one-electron approximation and $E_\uparrow(\mathbf{k})$ or $E_\downarrow(\mathbf{k})$ for the energy after consideration of the exchange interaction. Additionally, the shift of the energy levels due to the Pauli susceptibility is considered. The variables $n_\downarrow$ and $n_\uparrow$ denote the densities of the electrons with the specified spin. The above approach is based on an occupation-dependent energy, which in turn has an effect on the occupation itself. Basically, the assumptions are similar to the molecular field approximation already discussed, because each spin pair acts on the entire subband.

We shift the energy zero point by $I/2$ and write $\tilde{E}(\mathbf{k}) = E(\mathbf{k}) + I/2$. We express the relative surplus of one spin type by $r = \left(n_\uparrow - n_\downarrow\right)/n$ and thus instead of (12.36) get

$$E_\uparrow(\mathbf{k}) = \tilde{E}(\mathbf{k}) - \frac{Ir}{2} - \mu_B B \qquad \text{and} \qquad E_\downarrow(\mathbf{k}) = \tilde{E}(\mathbf{k}) + \frac{Ir}{2} + \mu_B B\,. \tag{12.37}$$

In our simple consideration, the energy splitting between the two subbands is independent of the wave vector.

In thermal equilibrium, there must be no difference in chemical potential between the two subbands, i.e. the two subbands have the same Fermi energy, determined by the particle number density n. Following equation (8.24) we write

$$r = \frac{1}{n}\int \frac{D(E)}{2}\Big\{f[E_\uparrow(\mathbf{k})] - f[E_\downarrow(\mathbf{k})]\Big\}\,\mathrm{d}E\,. \tag{12.38}$$

The factor 1/2 occurs because the density of states $D(E)$ is related to the total number of electrons. We expand the two Fermi functions around E_F according to $(Ir/2 + \mu_B B)$ and stop the expansion in each case after the linear term. After summarizing both contributions we get

$$r = -\int \frac{D(E)}{2n}\frac{\partial f(\mathbf{k})}{\partial \tilde{E}(\mathbf{k})}(Ir + 2\mu_B B)\,\mathrm{d}E\,. \tag{12.39}$$

As in Section 9.2, we approximate the derivation of the Fermi function by a delta function, i.e. we set $\partial f/\partial\tilde{E} \approx -\delta(\tilde{E} - E_F)$ and find

$$r = \int \frac{D(E)}{2n}(Ir + 2\mu_B B)\,\delta[\tilde{E}(\mathbf{k}) - E_F]\,\mathrm{d}E = \frac{D(E_F)}{2n}(Ir + 2\mu_B B)\,. \tag{12.40}$$

This allows to express the magnetization $M = rn\mu_B$ and thus also the susceptibility

$$\chi = \frac{\mu_0 M}{B} = \frac{\mu_0\mu_B^2 D(E_F)}{1 - [ID(E_F)/2n]} = \frac{\chi_{\text{Pauli}}}{1 - [ID(E_F)/2n]}\,. \tag{12.41}$$

Here χ_{Pauli} stands for the Pauli spin susceptibility given by equation (12.18). A look at the denominator shows that the feedback mechanism amplifies the already existing paramagnetic behavior of the electron gas.

The *Stoner parameter* I can be calculated approximately from the electron configuration of the atoms involved, taking into account the exchange interaction. The result of this calculation is shown in Figure 12.9 together with the product $ID(E_F)/2n$ for the metallic elements up to atomic number 49. Particularly high amplification factors

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

are found for the metals calcium, scandium and palladium. The values of χ/χ_{Pauli} are 4.5 (Ca), 6.1 (Sc) and 4.5 (Pd).

With increasing value of $ID(E_F)$ the Pauli paramagnetism becomes more and more pronounced. If the **Stoner criterion** $ID(E_F)/2n > 1$ is fulfilled, the formal solution is obviously unstable. Due to the feedback, a spontaneous magnetization is formed. A look at Figure 12.9 shows that the described calculation actually correctly predicts the ferromagnetism of iron, cobalt and nickel. Figure 12.9h also makes it clear that the metals Ca, Sc and Pd mentioned above are at the limit of ferromagnetism. This is impressively demonstrated by the diluted alloy **Pd**Fe: even the addition of 0.15 % Fe to palladium causes the alloy to become ferromagnetic.

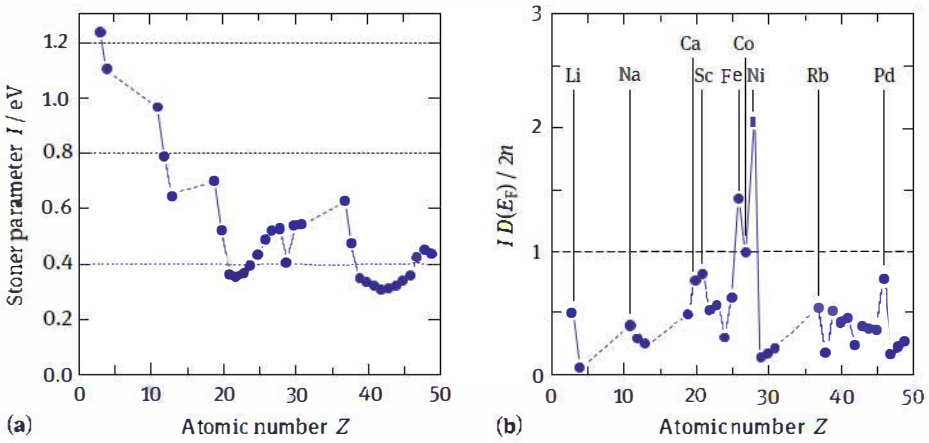


Fig. 12.9: a) Stoner parameter I of the metallic elements up to the atomic number $Z = 49$, **b)** Product $ID(E_F)/2n$ of the different elements. The Stoner criterion is shown as black-dashed line. (After J.F. Janak, Phys. Rev. B **16**, 255 (1977).)

Let us turn briefly to the question of what conditions the band structures must fulfill to make ferromagnetism possible. The question can be discussed particularly well on the basis of the properties of nickel and copper: the two metals are neighbors in the periodic table, but only nickel is ferromagnetic. Figure 12.10 shows the calculated electronic density of states. The general course is very similar: 4*s*-band and 3*d*-bands overlap. This has no consequence for the magnetic properties of copper with the atomic electron configuration $3d^{10}4s^1$, where the 3*d*-band is completely filled with ten electrons per atom and the 4*s*-band is half filled by the remaining electron. The Fermi level is therefore in the *s*-band and the exchange interaction does not change this state.

Much more subtle conditions are present in nickel, which as a free atom has the configuration $3d^8 4s^2$, but only 0.54 of the two 4*s*-electrons originally present in the metal remain in the *s*-band. In the paramagnetic phase, the Fermi level is near the upper edge of the 3*d*-band (cf. Figure 12.10b). In order to occupy it fully, all existing

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

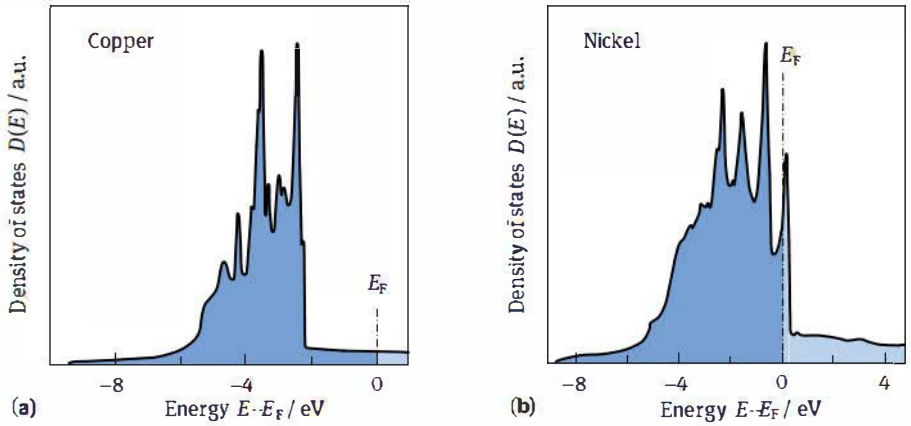


Fig. 12.10: **a)** Density of states $D(E)$ of copper. The d-band is completely filled, the s-band up to the Fermi energy. (After H. Eckhardt et al., J. Phys. F14, 97 (1984).) **b)** Density of states of nickel at a temperature above the Curie temperature. The Fermi energy lies in the d-band. (After J. Callaway, C.S. Wang, Phys. Rev. B 7, 1096 (1973).)

s-electrons would have to change over into this band. But since the lower end of the s-band is lower than the d-band, it is occupied by some of the electrons. These are now missing for complete occupation of the d-band. Below the Curie temperature, the two subbands with opposite spin orientation shift relative to one another due to the exchange interaction and the circumstances are changing fundamentally. To work out this difference, we consider the two subbands separately, as shown in Figure 12.11. The exchange interaction, or in other words the molecular field, raises or lowers the two subbands by more than 0.5 eV. This does not change the number of electrons in the s-band, but one subband of the $3d$-band is now completely filled up, while the

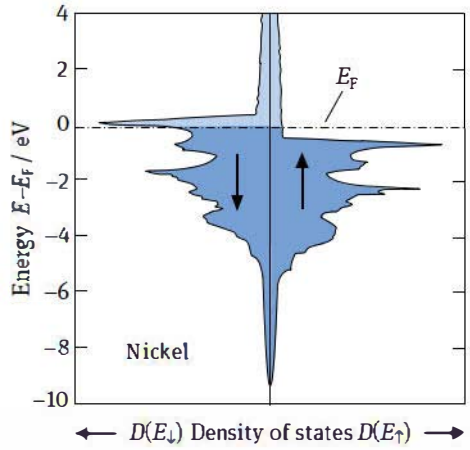


Fig. 12.11: Densities of states $D(E_\uparrow)$ and $D(E_\downarrow)$ of the subbands of nickel in the ferromagnetic phase. The right $3d$-subband is completely, the left is only partially filled. (After J. Callaway, C.S. Wang, Phys. Rev. B 7, 1096 (1973).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

other is even less filled compared to the situation without exchange interaction. Thus a spontaneous magnetization, which can be attributed to the 0.54 holes per atom in the $3d$-band is formed. Due to the high density of states at the Fermi energy, the exchange interaction leads to a large change in occupation and thus to a large magnetization.

On the one hand, this example shows that for ferromagnetism to occur, a very special constellation must exist with regard to the number of electrons present and the band structure. On the other hand, we can see that when band ferromagnetism occurs, the effective number of Bohr magnetons cannot be directly correlated with the spin quantum number of the individual atoms, even if no resulting orbital angular momentum has to be taken into account.

12.3.4 Spin Waves – Magnons

In addition to the electronic excitations and lattice vibrations discussed so far, in ferromagnetic materials there also exist collective magnetic excitations so-called **spin waves** or **magnons**. Initially, one would expect that the energetically lowest magnetic excitation would have to be the flipping of individual spins. An estimation of the required energy allows equation (12.31), according to which the necessary energy should be $\delta E = 2z\mathcal{J}S^2$. Considering $\mathcal{J} \approx k_B T_C$, it is clear that far below the Curie temperature only few single-spin excitations are thermally excited. It turns out that the excitation of a collective precession motion of spins, as schematically illustrated in Figure 12.12 is energetically more favorable.

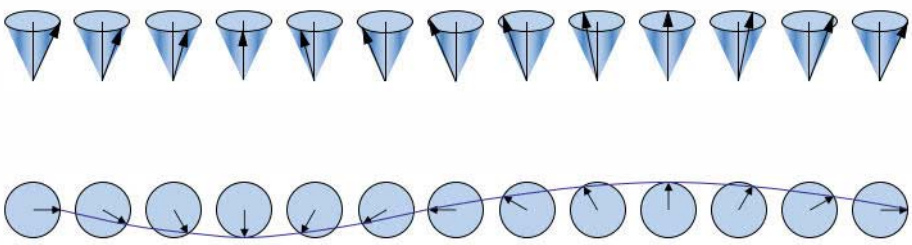


Fig. 12.12: Schematic representation of a spin wave (magnon) along an array of spins in perspective side view (above) and in top view (below).

We derive the equation of motion of the spins in a semi-classical consideration, where spin vectors are replaced by classical vectors. It can be shown that the quantum mechanical treatment of the problem leads to the same result. As with phonons, the amplitude of the deflection is quantized, so that the energy $E = \hbar\omega$ can be attributed to each magnon. From an angular momentum point of view, the excitation of a **magnon** corresponds to the flipping of a single spin 1/2, i.e. to a change of the angular momentum of the spin ensemble by $\hbar$.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

To derive the dispersion relation of spin waves, we consider for simplicity a linear chain of identical atoms spaced a apart, with spin $\mathbf{S}$ oriented in the z-direction, instead of a three-dimensional solid. We further assume that only the molecular field is effective, so that the external field $\mathbf{B}_{\text{ext}}$ does not have to be taken into account. If a spin with the magnetic moment $\boldsymbol{\mu}$ is deflected from the z-direction, a torque $(\boldsymbol{\mu} \times \mathbf{B}_{\text{M}})$ acts on it due to the molecular field $\mathbf{B}_{\text{M}}$, which is connected to the angular momentum via the relation $\boldsymbol{\mu} = -g\mu_{\text{B}}\mathbf{S}$. Since the time derivative of the angular momentum $\hbar\mathbf{S}$ is equal to the applied torque, the equation motion results in of form of $\mathrm{d}(\hbar\mathbf{S})/\mathrm{d}t = -g\mu_{\text{B}}\mathbf{S} \times \mathbf{B}_{\text{M}}$.

Nowwe have to find a suitable expression for the molecular field. If we consider only the interaction with the nearest neighbors, we get (12.29) for the potential energy U_m of spin $\mathbf{S}_m$ due to the exchange interaction the relationship $U_m = -\mathcal{J}\mathbf{S}_m \cdot (\mathbf{S}_{m-1} + \mathbf{S}_{m+1})$. On the other hand without external field equation (12.30) can be written in the form $U_m = -\boldsymbol{\mu} \cdot \mathbf{B}_{\text{M}} = g\mu_{\text{B}}\mathbf{S}_m \cdot \mathbf{B}_{\text{M}}$. By equating the two expressions, the molecular field $\mathbf{B}_{\text{M}} = (-\mathcal{J}/g\mu_B)(\mathbf{S}_{m-1} + \mathbf{S}_{m+1})$ can be determined. If we apply this result to the equation of motion given above, we obtain

$$\frac{\mathrm{d}\mathbf{S}_m}{\mathrm{d}t} = \frac{\mathcal{J}}{\hbar}\mathbf{S}_m \times (\mathbf{S}_{m-1} + \mathbf{S}_{m+1}) \,. \tag{12.42}$$

This differential equation is non-linear since products of the spin vectors occur. For small deflections in x- and y-direction, an ansatz of the form

$$\begin{aligned} S_{m,x} &= A\cos(mqa - \omega t) \,, \\ S_{m,y} &= A\sin(mqa - \omega t) \,, \\ S_{m,z} &= \sqrt{S^2 - A^2} \,, \end{aligned} \tag{12.43}$$

can be used, which assumes a propagating wave with amplitude A and wave vector q. If we insert the solution (12.43) into the equation of motion (12.42), we obtain the dispersion relation

$$\omega = \frac{2\mathcal{J}S}{\hbar}\Big[1 - \cos(qa)\Big] = \frac{4\mathcal{J}S}{\hbar}\sin^2\left(\frac{qa}{2}\right) \,. \tag{12.44}$$

For small wave vectors we can approximate this equation by

$$\omega \approx \frac{\mathcal{J}S}{\hbar}a^2q^2 \,. \tag{12.45}$$

At long wavelengths, the frequency increases proportional to the square of the wave vector; for phonons, this relationship is linear. If we perform the calculation not for a linear chain but for a cubic crystal, we will find a very similar dispersion relation:

$$\omega = \frac{\mathcal{J}S}{\hbar}\sum_{i=1}^{z}\Big[z - \cos(\mathbf{q} \cdot \mathbf{r}_i)\Big] \,. \tag{12.46}$$

The summation is performed over all z nearest neighbors which are connected to the atom via the position vectors $\mathbf{r}_i$. In the limiting case $qa \ll 1$ we come back to the expression (12.45), where a now stands for the lattice constant of the crystal.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Fig. 12.13a shows the dispersion curve of spin waves in the first Brillouin zone, according to equation (12.44). The experimental dispersion curve of cobalt alloyed with 8 % iron is shown in Figure 12.13h. The measurement was carried out using inelastic neutron scattering. Since the measurement was only carried out at relatively small wave numbers, a parabolic curve was observed independent of the direction of observation as expected. At $q = 0$ there is a small gap in the magnon spectrum which is caused by the anisotropy of the exchange interaction. However, we will not go into this additional effect here.

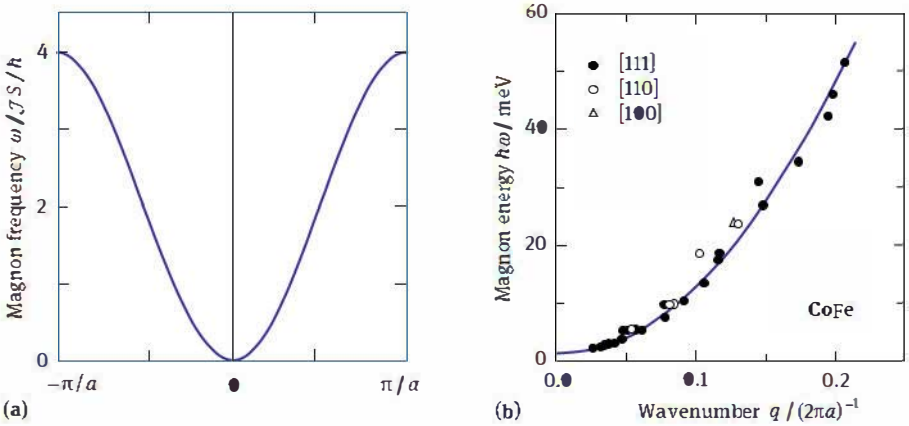


Fig. 12.13: Magnon dispersion curve. **a)** Schematic course of the dispersion curve of spin waves. **b)** Magnon dispersion curve of cobalt alloyed with 8% Fe at small wave numbers. The gap at $q = 0$ is due to the anisotropy of the exchange interaction. The blue solid line represents a parabolic fit of the data with the small gap accounted for. (After R.N. Sinclair, B.N. Brockhouse, Phys. Rev. 120 1638 (1960).)

We come back again briefly to the flipping of single spins mentioned at the beginning. This process, which occurs at higher energies and higher temperatures, is called *Stoner excitation*. It shortens the lifetime of the magnons and causes the dispersion curve to bend at larger wave vectors. The result of a measurement at 295 K of nickel using inelastic neutron scattering is shown in Figure 12.14. It can be clearly seen that for large wave vectors the experimental data deviate from the prediction of equation (12.44) represented by the dashed line.

12.3.5 Temperature Dependence of Magnetization

At finite temperatures, due to the precessional motion of the spins, magnons cause deviations from the perfect alignment of the magnetic moments and thus give rise to a reduction of the spontaneous magnetization. The number of excited magnons therefore

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

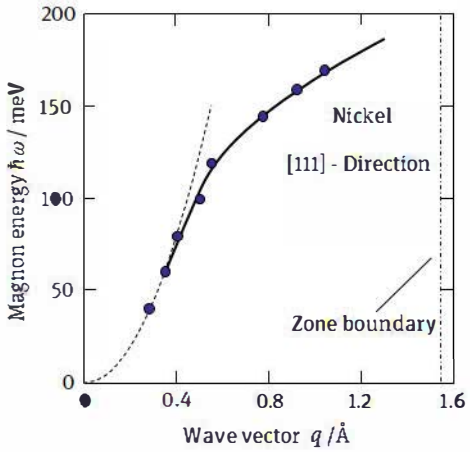


Fig. 12.14: Magnon dispersion in nickel at 295 K. The deviation from the expected dependence (12.46) due to Stoner excitations is clearly visible. (After H.A. Mook, D. McK. Paul, Phys. Rev. Lett. **54**, 227 (1985).)

determines the temperature dependence of the magnetization at low temperatures. At $T = 0$ the magnetization can be expressed by (12.22). If we set $J = S$, then $M_S(0) = ng\mu_B S$ applies. As already mentioned, the excitation of each magnon corresponds to the flipping of a spin 1/2 and thus a reduction of the total spin by $\hbar$ and the magnetization by $g\mu_B$ independent of the magnon energy $\hbar\omega$. We can therefore express the temperature dependence of the spontaneous magnetization if we use the number density n_{mag} of excited magnons per volume, because the following applies: $M_S(T) = M_S(0) - g\mu_B n_{mag}$.

When calculating n_{mag}, we proceed in the same way as with the calculation of the number of phonons in Section 6.4, because the wave vectors of magnons and phonons are subject to the same boundary conditions. Analogous to equation (6.99), the magnon density is given by $n_{mag} = \int D(\omega)\langle n(\omega, T)\rangle \mathrm{d}\omega$. Here, the Bose Einstein factor $\langle n(\omega, T)\rangle$ represents the average occupation of the magnon states and $D(\omega)$ their density of states, which we still have to calculate.

For the sake of simplicity, we assume that the density of states is isotropic. Then we can use equation (6.80), which we originally derived for phonons in isotropic solids. At moderately high temperatures only small wave vectors are important, so that we can use the dispersion relation (12.45). Thus we find with (6.80) for the density of states the following expression

$$D(\omega) = \frac{V}{2\pi^2}\frac{q^2}{v_g} = \frac{V}{4\pi^2}\left(\frac{\hbar}{\mathcal{J}Sa^2}\right)^{3/2}\sqrt{\omega}\,, \tag{12.47}$$

where v_g stands for the group velocity obtained by deriving (12.45).

Since at low temperatures only spin waves with low energies are excited, we can use for the upper limit of the integral infinity and find with $x = \hbar\omega/k_B T$ for the number of magnons

$$n_{mag} = \int D(\omega)\langle n(\omega, T)\rangle \mathrm{d}\omega = \frac{1}{4\pi^2}\left(\frac{k_B T}{\mathcal{J}Sa^2}\right)^{3/2}\int_0^\infty \frac{\sqrt{x}}{e^x - 1}\,\mathrm{d}x\,. \tag{12.48}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The integral has the numerical value $4\pi^2 \times 0.0587$. For the temperature dependence of the spontaneous magnetization this leads to the result

$$\frac{M_s(0) - M_s(T)}{M_s(0)} = \frac{0.0587}{nSa^3}\left(\frac{k_B T}{\mathcal{J}S}\right)^{3/2} . \tag{12.49}$$

This is **Bloch's $T^{3/2}$-law** for the magnetization of ferromagnets, which is in good agreement with the experiment at low temperature. In Figure 12.15 the experimental data of nickel are compared with the theoretical prediction. The deviation at higher temperatures is mainly due to the use of the approximation (12.45) for magnons with small wave vectors.

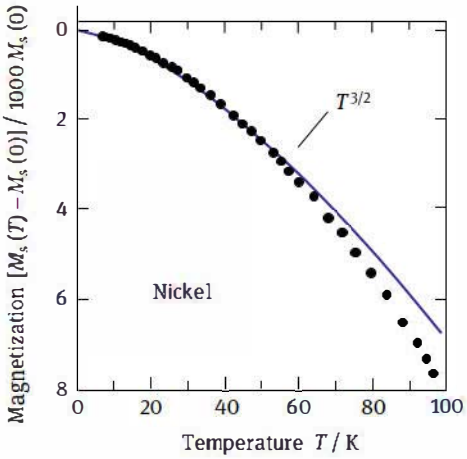


Fig. 12.15: Spontaneous magnetization of nickel as a function of temperature. At low temperatures the magnetization follows the $T^{3/2}$-dependence as expected. (After B.E. Argyle et al., Phys. Rev. **132**, 2051 (1963).)

12.3.6 Ferromagnetic Domains

Although ferromagnetic materials have spontaneous magnetization, the dipole moment of a macroscopic sample can be neglected in most cases if no external field is applied. The samples contain areas, so-called **domains**, whose magnetization points in different directions. In Figure 12.16 the domain boundaries of a 50 μm wide iron single crystal are made visible with the help of a fine magnetic powder which is deposited in the areas of high magnetic fields, i.e. on the domain walls.

Certain **directions of easy magnetization** occur, because the overlap integral of the wave functions and thus the exchange energy in crystals is not isotropic but directional. For body-centered cubic crystals such as iron, the ⟨100⟩-directions are preferred. The domain boundaries themselves are not sharp, because the transition of magnetization from one domain to the other occurs continuously within a relatively small but finite distance. In this region, the **Bloch wall**, the direction of one spin is slightly rotated

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

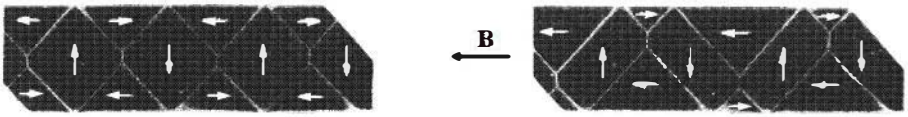


Fig. 12.16: Direction of the spontaneous magnetization in the domains of a 50 µm wide iron single crystal. The domain walls are made visible by means of a fine magnetic powder. The left picture was taken without, the right picture with magnetic field in the given direction. (After R.W. DeBois, C.D. Graham J. Appl. Phys. **29**, 931 (1958).)

relative to the neighboring one. Thus the magnetization changes direction within a few hundred spins. When an external field is applied, a macroscopic magnetization is created by shifting the Bloch walls. This process can be clearly seen in Figure 12.16. The original domains that pointed in the direction of the field have grown at the expense of the others.

Magnetic domains are created because they reduce the stray fields outside the sample and thus the magnetic field energy. This can be seen by looking at two bar magnets lying next to each other. The field energy is much lower when the opposite poles are adjacent and not the same. An alignment of the domains, as shown in Figure 12.16, therefore minimizes the field energy. On the other hand, forming Bloch walls costs energy even if the angle between adjacent spins is very small, i.e. the Bloch walls are very thick. The reason for this is that the spins of the wall do not point in the direction of the easy magnetization. The free energy of the sample is not only minimized by the formation of domains, but also an optimal angle between the directions of the adjacent spins and thus an optimal wall thickness is achieved.

12.4 Ferri- and Antiferromagnetism

In ferri- and antiferromagnetic materials, the magnetic moments are arranged in (at least) two sublattices whose spin orientation are opposite. A solid is antiferromagnetic when the magnetic moments of the sublattices just compensate one another, and ferrimagnetic when this is only partially the case. Below a critical temperature, magnetized domains spontaneously appear; above this temperature, the two substance classes are paramagnetic.

12.4.1 Ferrimagnetism

Phenomenologically, ferrimagnetic substances behave very similar to ferromagnets. The name *ferrimagnetism* comes from the term ferrite for materials with the composition $MO \cdot Fe_2O_3$. Here M stands for a bivalent metal, such as Cd, Co, Cu, Mg, Ni, Zn or Fe. In the latter case, it is magnetite Fe_3O_4, in which the iron atoms occur in di- *and* trivalent

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

form. Ferrites have the structure of the mineral spinel ($MgAl_2O_4$) with a relatively complicated cubic elementary cell, which has an extension of 6 to 9 Å depending on the chemical composition. The ions of the two metals sit on non-equivalent lattice sites in the unit cell. At low temperatures, the spins of one sublattice are aligned parallel to one of the edges of the cube, while the spins of the other sublattice are aligned in the opposite direction.

Ferrimagnetism occurs when the exchange constant $\mathcal{J}_{AB}$ between two nearest neighbors A and B with magnetic moments is negative, because then the spins of one type of ion are opposite to those of the other. In many cases the exchange energy between the ions of *one* sublattice is also negative, i.e. $\mathcal{J}_{AA}$ or $\mathcal{J}_{BB}$ is also negative. However, if the AB interaction is strongest, the spins within the sublattices are aligned in parallel. Table 12.3 shows the Curie temperature and the saturation magnetization of some ferrimagnetic materials.

Tab. 12.3: Curie temperature T_c and saturation magnetization M_s of ferrimagnetic materials. (After F. Keffer, *Handbook of Physics*, Volume 18, Springer, 1966.)

	Fe_3O_4	$CoFe_2O_4$	$NiFe_2O_4$	$CuFe_2O_4$	$MnFe_2O_4$	$Y_3Fe_5O_{12}$
T_c (K)	860	790	860	730	570	560
M_s (kA/m)	510	470	300	160	560	150

We will not continue to deal with the magnetic properties of ferrimagnets here, but will turn now to antiferromagnets, since the situation is somewhat clearer there.

12.4.2 Antiferromagnetism

When discussing the properties of antiferromagnets, we refer back to the results we obtained when treating ferromagnets. In the following considerations we assume that the two sublattices A and B consist of the same atoms and that without an external magnetic field the magnetic moments of the A atoms are antiparallel to the moments of the B atoms. An example of a simple antiferromagnetic substance is the tetragonal manganese fluoride, whose structure is shown in Figure 12.17. The spins of the two equivalent ferromagnetic manganese sublattices A and B are oriented antiparallel. We initially assume that there is no external field. To describe the exchange fields we use equation (12.21) and find under the assumption $\mathbf{B}_{ext} = 0$:

$$\mathbf{B}_{eff}^{A} = -\mu_0\lambda_{AA}\mathbf{M}_A - \mu_0\lambda_{AB}\mathbf{M}_B \quad \text{and} \quad \mathbf{B}_{eff}^{B} = -\mu_0\lambda_{BA}\mathbf{M}_A - \mu_0\lambda_{BB}\mathbf{M}_B\,. \tag{12.50}$$

The molecular field constants λ are positive. The minus signs reflect the fact that the occurring forces try to align the spins of the interacting neighbors antiparallel. In the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

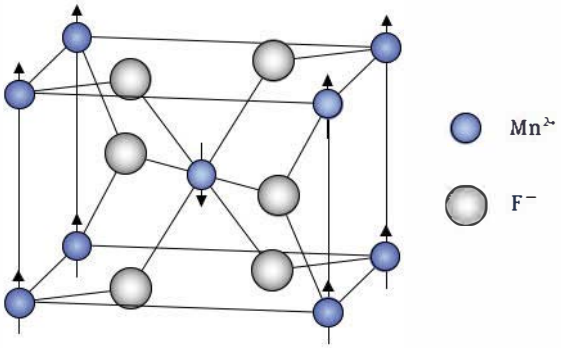


Fig. 12.17: Unit cell of the antiferromagnetic MnF_2. The spin directions are indicated by arrows.

simple case we are considering here, the following applies for reasons of symmetry: $\lambda_{AB} = \lambda_{BA}$ and $\lambda_{AA} = \lambda_{BB}$ and therefore $\mathbf{B}_{\text{eff}}^{A} = -\mathbf{B}_{\text{eff}}^{B}$.

The total magnetization $\mathbf{M} = (\mathbf{M}_A + \mathbf{M}_B)$ is composed of the contributions of sublattices A and B. Since the magnetization is zero in antiferromagnets, $\mathbf{M}_A = -\mathbf{M}_B$ applies. Thus the two equations for the molecular fields are simplified as follows:

$$\mathbf{B}_{\text{eff}}^{A} = \mu_0(\lambda_{AB} - \lambda_{AA})\mathbf{M}_A = -\mathbf{B}_{\text{eff}}^{B} . \tag{12.51}$$

If we replace in equation (12.21) the molecular field constant λ by $(\lambda_{AB} - \lambda_{AA})$, we get the equation just derived. Thus, the same laws apply to the magnetization of the sublattices as to the magnetization of ferromagnets. In particular, equation (12.25) for the critical temperature, here called **Néel temperature**,[19] at which the transition from the antiferromagnetic phase to the paramagnetic phase takes place, follows the relationship

$$T_N = C\,\frac{\lambda_{AB} - \lambda_{AA}}{2} . \tag{12.52}$$

Here C is the Curie constant given by (12.14). The factor 1/2 occurs because we relate the constant to *all* atoms with magnetic moments and not only to the atoms of one sublattice.

Above the Néel temperature the system is in the paramagnetic phase in which both sublattices contribute equally to the total magnetization, because then $\mathbf{M}_A = \mathbf{M}_B$ and $\mathbf{M} = 2\mathbf{M}_A$. If we use this relationship in (12.50) and (12.24) and take into account the external field B_{ext}, we obtain

$$\mathbf{B}_{\text{eff}}^{A} = \mathbf{B}_{\text{ext}} - \mu_0\frac{\lambda_{AB} + \lambda_{AA}}{2}\mathbf{M} \tag{12.53}$$

and

$$\mu_0\mathbf{M} = \frac{C}{T}\left(\mathbf{B}_{\text{ext}} - \mu_0\frac{\lambda_{AB} + \lambda_{AA}}{2}\mathbf{M}\right) . \tag{12.54}$$

19 Louis Eugéne Felix Néel, *1904 Lyon, †2000 Brive-la-Gaillarde, Nobel Prize 1970

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

For the paramagnetic susceptibility this results in:

$$\chi_p = \frac{C}{T + \Theta}, \tag{12.55}$$

where Θ stands for the *paramagnetic Néel temperature*, which is defined by

$$\Theta = C \frac{\lambda_{AB} + \lambda_{AA}}{2}. \tag{12.56}$$

Starting at high temperatures, the susceptibility of antiferromagnets in the paramagnetic phase is similar to that of ferromagnets. However, it does not diverge at the Néel temperature, the temperature at which the transition to the antiferromagnetic phase occurs. Since in the denominator of equation (12.55) a positive sign exists, χ_p diverges only at $T = -\Theta$, i.e. not at real temperatures.

Now we will address another interesting aspect here. The arguments presented so far assume that the applied magnetic field is infinitesimally small. In experiments this condition is usually not fulfilled and the behavior of antiferromagnets below the Néel temperature is somewhat more complicated.

The reaction to a magnetic field depends on the orientation of the sample under investigation. Let us first consider the case where the magnetic field is perpendicular to the spontaneous magnetization of the sublattices. As one can easily imagine (cf. exercise 7 at the end of this chapter), the magnetization of the sublattices is turned slightly in field direction. For the magnetization and the associated anti-ferromagnetic susceptibility, the following expression holds:

$$\chi_\perp = \frac{1}{\lambda_{AB}}. \tag{12.57}$$

Obviously the temperature is not involved here, so that we expect a temperature-independent susceptibility $\chi_\perp$. This is confirmed by the measurement on MnF_2, which is shown in Figure 12.18. Coming from high temperatures, the susceptibility initially increases following the Curie-Weiss law. If the temperature falls below the Néel temperature, the susceptibility $\chi_\perp$ remains constant during further cooling when the field is perpendicular.

If the field points in the direction of the aligned spins, we have $\chi_\parallel = 0$ at absolute zero, because the two sublattices are rigidly connected to each other. With increasing temperature, the susceptibility $\chi_\parallel$ increases exponentially, because the spins can now flip over thermally activated and then preferably point in the field direction.

Magnons also exist in the antiferromagnetic phase, but they are clearly different from the magnons of ferromagnets. A derivation of the equation of motion of the spins as in the case of ferromagnets, but with the additional complication that two coupled sublattices exist, leads to the dispersion relation

$$\omega = \frac{2S|\mathcal{J}|}{\hbar} |\sin qa|, \tag{12.58}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

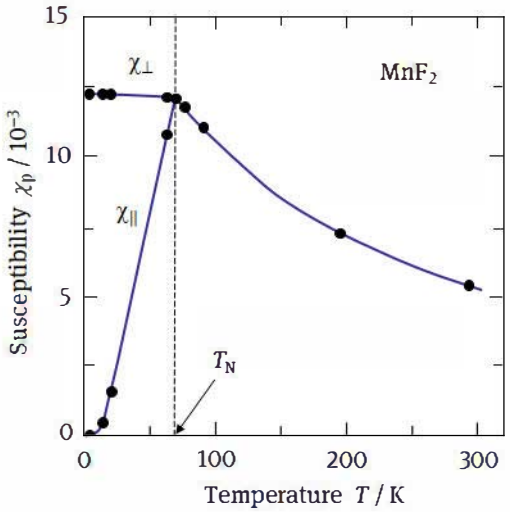


Fig. 12.18: Susceptibility of MnF_2. Below the Néeltemperature T_N the susceptibility depends on the field direction. (After H. Bizette, B. Tsai, Comptes rendus **238**, 1575 (1954).)

which looks very similar to the dispersion relationship of acoustic phonons. Again, for small wave vectors there is a linear relationship between frequency and wave vector, as neutron scattering measurements confirm. Figure 12.19 shows the measurement results on cubic $RbMnF_3$ crystals. In this crystal, the dispersion curves hardly differ for magnons propagating in [100]-, [110]- or [111]-direction, so that the data points were combined in one graph.

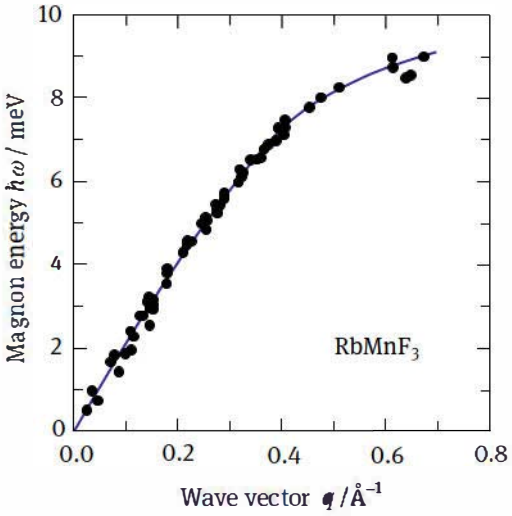


Fig. 12.19: Dispersion curve of the magnons in the antiferromagnet $RbMn_3$, measured by inelastic neutron scattering. (After C.G. Windsor, R.W.H. Stevenson, Proc. Phys. Soc. **87**, 501 (1960).)

Table 12.4 shows the Néel temperature T_N and paramagnetic Néel temperature Θ of various antiferromagnets.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Tab. 12.4: Néel temperature T_N and paramagnetic Néel temperature Θ of some antiferromagnets (After K. Kopitzki, P. Herzog, *Introduction to Solid State Physics*, Teubner, 2002; F. Keffer, *Handbook of Physics*, Volume 18, Springer, 1966).

	MnO	MnF_2	MnS	FeO	$FeCl_2$	CoO	$CoCl_2$	$NiCl_2$
T_N (K)	122	67	160	195	24	291	25	50
Θ (K)	610	82	530	570	48	330	38	68

12.4.3 Giant Magnetoresistance

In today's world, the digital processing, transmission and storage of information plays an outstanding role. In addition to texts and data, this also includes image, sound and video information. In many storage systems, the information is recorded as a direction of magnetization in magnetic domains. In the past, the commonly used inductive readout mechanism limited the storage density of hard disks. Since the end of the nineties, the *giant magnetoresistance* has been used in the majority of all hard disk heads to read the stored data.

Magnetoresistance is the change in electrical resistance that occurs when an external magnetic field is applied. In ferromagnetic materials, this effect originates from the spin-orbit interaction, which causes the resistance to depend on the direction of the current, i.e. on whether the current flows parallel or perpendicular to the magnetization. As we have seen, when a magnetic field is applied, the "incorrectly" oriented domains try to align themselves. The change of resistance is of the order of 1 %.

In the following, we consider structures which, as shown schematically in Fig. 12.20, are composed of three layers in the simplest case, whereby the two outer ones, with a typical thickness of about 10 nm, consist of a ferromagnetic material such as Fe, Co or Ni. The middle layer, about 1 nm thick, is not ferromagnetic and can be metallic, semiconducting or insulating. If there is no external field, the magnetization of the layers is antiparallel due to the coupling via the intermediate layer.

The *tunnel magnetoresistance* is the "precursor" of the giant magnetoresistance effect. It occurs when the interlayer consists of a semiconductor or insulator. This effect is based on the fact that electrons can tunnel from one conductor to the other when a voltage is applied between the two outer layers. We have already mentioned this phenomenon in connection with tunnel contact spectroscopy (see Figure 11.20).
The interest in such layer systems increased dramatically after the discovery of the *giant magnetoresistance* or GMR effect by *A. Fert*[20] and *P. Grünberg*[21]. They were able to demonstrate that in multiple layers an external magnetic field can cause changes of up to 50% in the electrical resistance of the layer system.

20 Albert Fert, *1938 Carcassonne, Nobel Prize 2007
21 Peter Grünberg, *1939 Pilsen, †2018 Jülich, Nobel Prize 2007

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

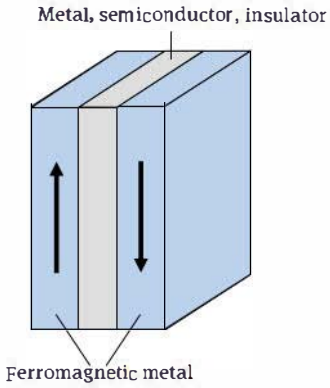


Fig. 12.20: Trilayer system. The two outer layers consist of Fe, Co or Ni. The intermediate layer is, depending on the experiment, metallic, semiconducting or insulating.

The result of a measurement of the electrical resistance at the three layer system Fe/Cr/Fe as a function of the applied magnetic field is shown in Figure 12.21. In the experiment, the two iron films were 12 nm, the intermediate layer of chromium was only 1 nm thick. It can be clearly seen that the resistance has a pronounced maximum when the magnetic field disappears and drops steeply with increasing field. For comparison, a measurement on a 25 nm thick iron film is also shown. It is obvious that with the same thickness of the iron layers, the resistance changes are much greater in the multilayer system. This shows that the electron spin can be of considerable importance for the electrical transport properties, especially for the current flow.

From the figure we can see that the resistance of the three-layer system is particularly high when the magnetization of the two ferromagnetic layers is antiparallel. The spin-dependent scattering process that causes the resistance will be discussed below. If an external magnetic field is applied in direction parallel to the layers of the sample, the magnetic moments which are in the opposite direction to the field are reversed

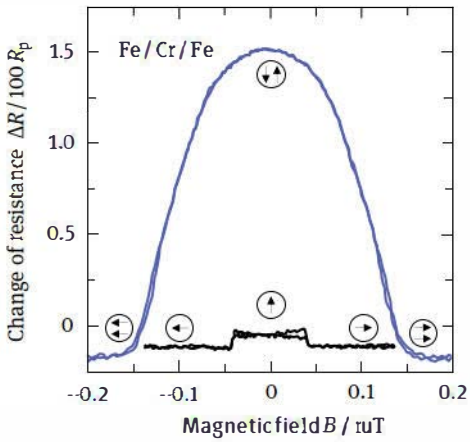


Fig. 12.21: Giant magnetoresistance of the three-layer system Fe/Cr/Fe (blue curve). The iron layers were 12 nm, the chromium layer 1 nm thick. The magnetoresistance of a 25 nm thick iron layer is shown in black. (After G. Binasch et al., Phys. Rev. B 39, 4828 (1989).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

with increasing field until all spins point in the same direction. During the increase of the field and the resulting spin orientation the resistance decreases steadily.

The giant magnetoresistance is calculated using the equation

$$\frac{\Delta R}{R_{\uparrow\uparrow}} = \frac{R_{\uparrow\downarrow} - R_{\uparrow\uparrow}}{R_{\uparrow\uparrow}}\,, \tag{12.59}$$

where $R_{\uparrow\uparrow}$ is the resistance for parallel and $R_{\uparrow\downarrow}$ for antiparallel magnetization. Table 12.5 shows the giant magnetoresistance values of three-layer Fe/Cr/Fe and Co/Cu/Co systems and the respective layer thicknesses. Even greater values up to $\Delta R/R_{\uparrow\uparrow} = 0.65$ were achieved with Co/Cu/Co multilayers.

Tab. 12.5: Giant magnetoresistance of three-layer sample at room temperature. In addition to the magnetoresistance $\Delta R/R_{\uparrow\uparrow}$ the layer thickness d_{mag} is given. (From J. Grünberg, *Kopplung macht den Widerstand*, Physik Journal 6, No. 8/9, 33 (2007).)

Multilayer	$\Delta R/R_{\uparrow\uparrow}$	d_{mag}
Fe/Cr/Fe	0.015	12
Fe/Cr/Fe	0.02	5
Co/Cu/Co	0.02	10
Co/Cu/Co	0.19	3
Co/Cu/Co	0.16	28

As already discussed in this chapter, current is carried mostly by s-electrons in metals with band-ferromagnetism. The resistance predominantly originates from the scattering of the *s*-electrons at the *d*-electrons, because their density of states at Fermi energy is much higher than that of the *s*-electrons. If we take the density of states of nickel shown in Fig. 12.11 as an example, we see that mainly the *d*-electrons contribute to the scattering, whose magnetic moment is directed downwards.

An explanation of giant magnetoresistance is possible on the basis of *Mott's two-current model* and spin-dependent electron scattering. For this purpose, the current is divided into two parallel-flowing partial currents with spin ↑ and spin ↓, i.e. parallel and anti-parallel with respect to the magnetization. The respective contribution to the total current is determined by the associated scattering rates. Electrons are scattered much more strongly in layers in which the magnetization direction is antiparallel to their spin than in layers in which magnetization and spin are oriented parallel. In a simple description, therefore, the spin inversion processes can be neglected.

Figure 12.22 shows the spin-dependent electron scattering schematically. Electrons with spin ↑ are not scattered in the left-hand part of the picture and cause a short circuit in this simple representation. In contrast, electrons with antiparallel orientation are subject to strong scattering. If, on the other hand, the spins of the layers are oppositely

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

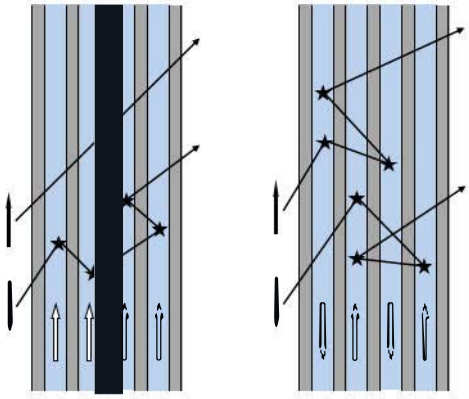


Fig. 12.22: Electrons are scattered more strongly in layers in which the magnetization (light arrows) is oriented antiparallel to the electron spin (black arrow) than in layers in which magnetization and spin are parallel to each other. (After P. Bruno, Physik in unserer Zeit **38**, 272 (2007).)

oriented, all electrons are scattered, i.e. the resistance is particularly high in this configuration.

In a more general description, the local scattering asymmetry parameter $\alpha = \varrho_\downarrow/\varrho_\uparrow$ is defined, where $\varrho_\downarrow$ and $\varrho_\uparrow$ each stands for the local resistance. Depending on whether electrons with spin upwards or downwards are scattered more strongly one has $\alpha < 1$ or $\alpha > 1$. If α is larger or smaller in both layers, this is called the "normal" giant magnetoresistance effect. If the sample consists of layers with $\alpha < 1$ and $\alpha > 1$, the resistance is greater when the magnetizations are aligned in parallel. This is called the "inverse" giant magnetoresistance effect.

12.5 Spin Glasses

In dilute alloys, an unusual state can be observed at low temperatures, which can be classified between the paramagnetic and ferromagnetic states. In these materials, the spins freeze in a randomly oriented manner during cooling and thus form a **spin glass**. The most prominent representatives are **Cu**Mn, **Ag**Mn, **Au**Fe and **Eu**Sr. We will discuss the behavior of spin glasses on the example of EuS, in which the exchange interaction between neighboring Eu^{2+} ions is *positive*, but *negative* for next but one neighbors. Depending on the distance from the reference spin, there is therefore a tendency for the neighboring spins to be parallel or antiparallel. If we use $\mathcal{J}_1$ and $\mathcal{J}_2$ to denote the exchange constants associated with this, then for EuS approximately $\mathcal{J}_1/\mathcal{J}_2 \approx -2$ applies, i.e. the interaction with the nearest neighbors is about twice as strong as that with the next but one neighbors. Since the interaction with the nearest neighbors is predominant, a ferromagnetic phase is established in EuS below 16.5 K. If Eu^{2+} ions are replaced by non-magnetic Sr^{2+} ions, the phase transition from the paramagnetic to the ferromagnetic state shifts to lower temperatures. The phase diagram of $Eu_xSr_{1-x}S$ is shown in Figure 12.23. As can be seen in this figure, below 2 K, the above-mentioned spin glass phase occurs when the Eu^{2+} ions are diluted by Sr^{2+} ions.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

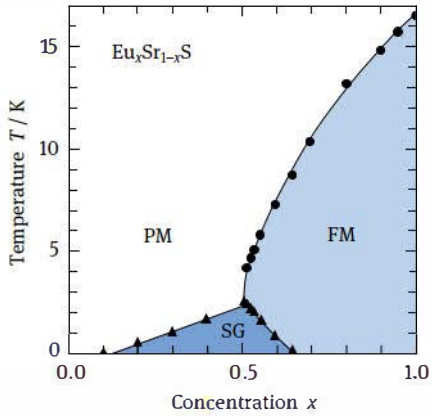


Fig. 12.23: Phase diagram of $Eu_xSr_{1-x}S$. The abbreviations PM and FM stand for the paramagnetic and ferromagnetic phase and SG for the spin glass phase. (After H. Maletta, J. Appl. Phys. **53**, 2185 (1982).)

The cause for the formation of this state below the **spin glass temperature** T_g can already be illustrated on the basis of a two-dimensional model. To simplify the conditions, we assume, as in the case of the **Ising model**, that the spins have only two possible orientations. If we further restrict the range of the interaction to the nearest and the second nearest neighbors, the sign of the exchange energy that a single spin has at given exchange constants and given configuration of the environment can be easily calculated. Again we assumed that $\mathcal{J}_1/\mathcal{J}_2 \approx -2$ applies as in EuS.

Let us look at the spin configuration in Figure 12.24a. Although the two spins in the grey background are opposite to the ferromagnetic order, this arrangement has the lowest energy for the given distribution of the non-magnetic ions. However, if the non-magnetic ions are arranged as shown in Figures 12.24b and 12.20c, the sum of the exchange energies of the two spins with grey backgrounds is independent of their orientation. The two configurations are energetically degenerate, there is no unique state with the lowest energy. Due to the competing interactions, the optimal alignment of one of the two spins is prevented by the presence of the second spin in both configurations. This is called the **frustration** of the interactions.

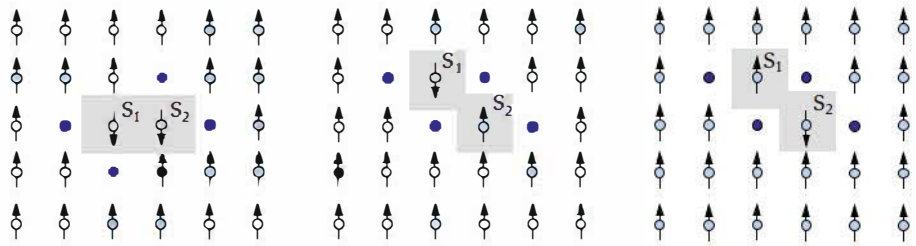


Fig. 12.24: Two-dimensional model of the frustration effect. The dark blue circles represent the non-magnetic ions. **a)** The two spins S1 and S2, shaded in grey, occupy the energetically lowest position. **b)** The presence of spin S2 prevents the ferromagnetic alignment of spin S1. **c)** This arrangement is energetically equivalent to arrangement b) despite the rotation of the two spins.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In the EuS, the number of configurations disturbing a uniform ferromagnetic order increases with increasing concentration of non-magnetic ions. At low temperatures and sufficiently high concentrations of Sr^{2+}-ions, therefore, a random distribution of spin orientations is frozen in. At a glance, this state appears very similar to the paramagnetic state, but there is a fundamental difference: in the paramagnetic phase, the energy difference between the two spin states is determined by the magnetic field and therefore has a well-defined value. The thermal motion reduces the degree of alignment of the spins. Ideally, the neighbors play no role, and in real systems, only a marginal role. In contrast, in spin glasses, the energy of a spin and thus its orientation depends on the spin configuration of its environment in a complicated way and is hardly influenced by the thermal energy. The argument does not change for a three-dimensional model where the spin orientations are not limited to two values.

Finally, we will briefly describe the behavior of the magnetic susceptibility of spin glasses. The reaction to alternating fields can be seen in Figure 12.25. The measurements were performed on the insulating spin glass $Fe_{0.5}Mn_{0.5}TiO_3$. The solid line represents a measurement in the constant field of the magnetometer. The sample was cooled down in steps of 0.1 K and held at this temperature for 1000 s each. All other data were recorded at frequencies between 5 mHz and 51 kHz. The interesting result is that as the frequency decreases, the maximum shifts to lower temperatures and becomes increasingly sharp.

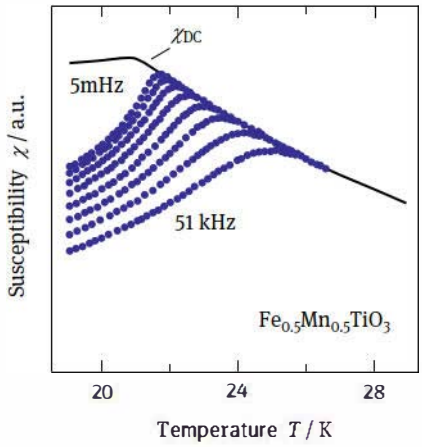


Fig. 12.25: Temperature dependence of the susceptibility of $Fe_{0.5}Mn_{0.5}TiO_3$. The points belong to measurements whose frequencies differ by a factor of 10. The black-solid curve shows the DC susceptibility obtained during cooling in the field. (After P. Nordblad, P. Svendlindh, *Spin Glasses and Random Fields*, A.P. Young, ed., World Scientific, 1998).)

The described behavior of the susceptibility can be understood qualitatively if one assumes that the free energy of the spin glasses is only at a minimum when the sample has been cooled in the DC field. If, on the other hand, the sample is cooled without a field, one of the many metastable configurations that exist due to the frustration effect is frozen in. Only by changing the local structure in many places a spin glass can reach lower free energy minima over time. Since potential barriers have to be overcome at the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

transitions from one spin configuration to another, new configurations, even if they are energetically more favorable, can only form slowly.

As explained, when the temperature decreases, spins or clusters of spins in certain configurations freeze and cannot reorient themselves as they continue to cool. Roughly speaking, the spins can be divided into two groups. The first group can follow the external AC field. For them, $\omega\tau < 1$ applies, where τ indicates their relaxation time. The spins of the second group with $\omega\tau > 1$ can no longer change their orientation within the time given by the period of the perturbation. They do not contribute to susceptibility. Thus it becomes understandable why a maximum is observed in χ during cooling, which shifts to lower temperatures with decreasing frequency.

The result of a DC field measurement at 0.59 T on **Cu**Mn is shown in Figure 12.26. As expected at high temperature, i.e. above the spin glass temperature T_g, the susceptibility follows the Curie-Weiss law. The spin system is in thermal equilibrium during this part of the measurement. Below the spin glass temperature, the susceptibility depends on the measurement process. If the field was applied above T_g and remained on during the entire experiment, the curve FC ("field cooling") was recorded. In a second experiment, the sample was cooled without field ($B < 5\,\mu T$) to the lowest temperature and then the field was applied. The corresponding curve is marked in the figure with "ZFC" for "zero field cooling". The observed cusp at the spin glass temperature is characteristic for the susceptibility of a spin glass. If the temperature is kept constant below T_g for a longer period of time, the DC susceptibility slowly increases and finally reaches the temperature-independent value FC. In such experiments, a remanent magnetization is observed below the spin glass temperature which disappears with time.

There are a number of other interesting effects in spin glasses, such as *aging*, *rejuvenation* or *memory effects*, to name just the keywords which we will not go into here. These are non-equilibrium phenomena that are only rudimentarily understood and represent a playground of statistical physics.

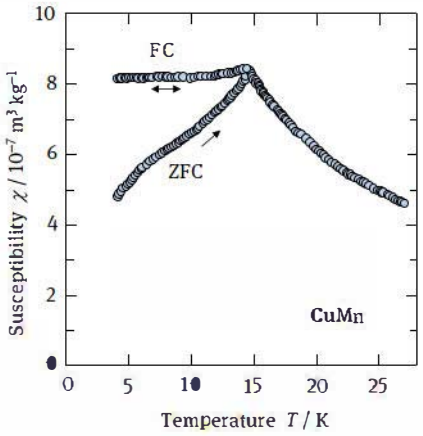


Fig. 12.26: Temperature curve of the susceptibility of copper alloyed with 2% manganese. The upper curve (FC) was measured during cooling in the field, the lower curve (ZFC) after cooling without field. (After S. Nagata et al., Phys. Rev. B **19**, 1633 (1979).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

12.6 Exercises and Problems

1. Dipole Interaction. Show that the magnetic dipole interaction between two magnetic moments of magnitude μ_B, which are 3 Å apart, only becomes important at temperatures below 100 mK.

2. Magnetization of $NiFe_2O_4$. Especially the Ni^{2+}-ions with their eight 3*d*-electrons contribute to the spontaneous magnetization of the ferrite $NiFe_2O_4$. The spins of the two iron ions compensate each other. How large is the magnetization in the ferrimagnetic phase when the densities are given by $\varrho_{Ferrit} = 9350 kg/m^3$ and $\varrho_{Ni} = 8908 kg/m^3$.

3. Exchange Coefficient. Calculate the molecular field constant, the molecular field and the exchange coefficient of the two ferromagnetic metals iron and nickel, which have body-centered cubic and face-centered cubic lattices with the lattice constants 2.87 Å and 3.52 Å respectively.

4. Ferromagnetism. As explained in Section 12.3, ferromagnetism is not based on the interaction between the magnetic moments of the electrons involved. This can be read directly from the magnitude of the interaction energy between neighboring spins. Compare this interaction energy with the thermal energy and the actual Curie temperature. Consider iron, cobalt and nickel, whose densities are given by $\varrho_{Fe} = 7874\ kg/m^3$, $\varrho_{Co} = 8900 kg/m^3$ and $\varrho_{Ni} = 8908\ kg/m^3$.

5. Ferromagnetic Properties. Assume that there are two ferromagnets that have the same exchange coefficient $\mathcal{J}$ but differ in their magnetic moments S_1 and S_2, with $S_1 = 1$ and $S_2 = 2$. Compare the molecular field and Curie constants, the Curie temperatures and the saturation magnetizations.

6. Spin Waves in Nickel. Nickel is a body-centered cubic ferromagnet with the lattice constant $a = 3.52$ Å and the Debye temperature $\theta = 450$ K. The dispersion relation of the magnons at long wavelengths is given by $\hbar\omega = Dq^2$ with $D = 6.4 \times 10^{-40}\ Jm^2$. Calculate the exchange constant $\mathcal{J}$ and the contribution of the spin waves to the specific heat at 5 K under the assumption that the spin quantum number is $S = 1/2$. Calculate the temperature where magnons and phonons contribute equally to specific heat.
(Note: $\int_0^\infty x^{3/2}/(e^x - 1)dx = 1.783$.)

7. Susceptibility of Antiferromagnets. If the magnetic field of antiferromagnets is perpendicular to the spontaneous magnetization of the sublattices, the external magnetic field causes a torque which is compensated by the exchange field. Show that in this case the antiferromagnetic susceptibility is given by (12.57).

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

13 Dielectric and Optical Properties

The interaction between electromagnetic fields and solids can be described *microscopically* and *macroscopically*. For example, we have used a microscopic picture to discuss the optical properties of semiconductors. Here a photon is absorbed and an electron-hole pair is generated in absorption processes. Depending on the band structure of the semiconductor and the momentum distribution, in addition, a phonon can also come into play. In the macroscopic description, the Maxwell equations are used and the sample is characterized by its material parameters. The combination and interconnection of these two approaches is the main goal of this chapter.

Under the influence of an electric field, positive and negative charges are driven in opposite directions and in turn generate an electric field. How solids react to electric fields depends on whether free or bound charges are present and on how the fields vary in space and time. Free electrons in metals shield static fields by forming surface charges, but this shielding mechanism loses its effect at high frequencies. The same applies in real space, because fields are not shielded at arbitrarily small distances, since characteristic shielding lengths exist. In insulators, electrons can only be moved over very small distances. An electrical polarization which does not completely shield static fields even over longer distances is built up. In general, the reaction of insulators and metals to electric fields depends in a complicated way on the frequency and the wavelength of the fields, i.e. on time and length scale. A variety of dielectric and optical processes are responsible for this. We discuss, the most important among them in this chapter. In the first sections, we deal with insulators and place special emphasis on the properties of ionic crystals. In the last section, we discuss the behavior of the free electrons in metals.

13.1 Dielectric Function and Optical Quantities

If an electric field $\mathcal{E}$ acts on an insulator, this leads to a shift of charges and thus to a polarization $\mathbf{P}$ of the sample. Both quantities are connected with each other via the *dielectric susceptibility* $[\chi]$:

$$\mathbf{P} = \varepsilon_0 [\chi] \, \mathcal{E} \, . \tag{13.1}$$

Directly linked to this is the *dielectric tensor*

$$[\varepsilon] = [\mathbf{1}] + [\chi] \, . \tag{13.2}$$

Both $[\chi]$ and $[\varepsilon]$ are symmetrical tensors of second order. For cubic crystals and amorphous solids, these quantities are scalars. These materials therefore behave "electronically isotropic". Since in most cases the tensor properties are not relevant, we replace the tensors in the following also for non-cubic crystals with scalar quantities in order to keep the equations simple. The *electric induction* $\mathbf{D}$, often also called *electric*

https://doi.org/10.1515/9783110666502-013

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

displacement field, is defined by

$$\mathbf{D} = \varepsilon_0 \mathcal{E} + \mathbf{P} \equiv \varepsilon_0 \varepsilon \mathcal{E} . \tag{13.3}$$

Here, ε stands for the **dielectric constant**, that is often also called **dielectric function**.

The quantities $\mathbf{D}$ and $\mathcal{E}$ in general are time-dependent and can each be represented by a Fourier series:

$$\mathcal{E}(t) = \int_{-\infty}^{\infty} \mathcal{E}(\omega) e^{-i\omega t} d\omega \qquad \text{and} \qquad \mathbf{D}(t) = \int_{-\infty}^{\infty} \mathbf{D}(\omega) e^{-i\omega t} d\omega . \tag{13.4}$$

The Fourier coefficients $\mathcal{E}(\omega)$ and $\mathbf{D}(\omega)$ are then linked via the frequency-dependent dielectric constant $\varepsilon(\omega)$: $\mathbf{D}(\omega) = \varepsilon_0 \varepsilon(\omega) \mathcal{E}(\omega)$.

In time-varying fields, the Maxwell equations include not only the current density $\mathbf{j}$, which is caused by the free charge carriers, but also the displacement current $\partial \mathbf{D}/\partial t$:

$$\operatorname{curl} \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t} . \tag{13.5}$$

If we use Ohm's law $\mathbf{j} = \sigma \mathcal{E}$, this equation can be expressed in the form

$$\operatorname{curl} \mathbf{H}(\omega) = \sigma \mathcal{E}(\omega) - i\omega \varepsilon_0 \varepsilon(\omega) \mathcal{E}(\omega) = \tilde{\sigma}(\omega) \mathcal{E}(\omega) . \tag{13.6}$$

With $\tilde{\sigma} = (\sigma - i\omega\varepsilon_0\varepsilon)$ we have introduced here the frequency-dependent, generalized conductivity. Instead of (13.6), one can also use

$$\operatorname{curl} \mathbf{H}(\omega) = -i\varepsilon_0 \tilde{\varepsilon}(\omega) \omega \mathcal{E}(\omega) = -i\omega \mathbf{D}(\omega) \tag{13.7}$$

and thus define the generalized dielectric constant $\tilde{\varepsilon}(\omega) = \varepsilon(\omega) + i\sigma/\varepsilon_0\omega$.

The interchangeability of the two descriptions is based on the fact that the distinction between the free and the bound charges becomes vague in the case of alternating fields, and is only clear in the case of DC fields. Of course, the quantities $\mathcal{E}$ and $\mathbf{D}$ can also vary spatially. The spatial changes can be accounted for by a superposition of plane waves. The expansion coefficients then depend on the wave vector $\mathbf{k}$. We do not want to pursue this here, because for the effects discussed in this chapter, only the wave vectors are of importance in the expansion that are small compared to a reciprocal lattice vector. In this case we may set $k \approx 0$.

Without proof, we would like to mention that the real and imaginary parts of the susceptibility and the dielectric function are linked, since they can be understood as the *response functions* of a linear, passive system for which the **Kramers[1]-Kronig[2] relations** are valid. For the dielectric function $\varepsilon = (\varepsilon' + i\varepsilon'')$ they have the form

$$\varepsilon'(\omega) - 1 = \frac{2}{\pi} \mathcal{P} \int_0^{\infty} \frac{\omega' \varepsilon''(\omega')}{\omega'^2 - \omega^2} d\omega' , \tag{13.8}$$

1 Hendrik Anthony Kramers, *1894 Rotterdam, †1952 Oegstgeest

2 Ralph Kronig, *1904 Dresden, †1995 Zeist

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

$$\varepsilon''(\omega) = -\frac{2\omega}{\pi}\,\mathcal{P}\int_0^\infty \frac{\varepsilon'(\omega') - 1}{\omega'^2 - \omega^2}\,\mathrm{d}\omega' \,. \tag{13.9}$$

Here, $\mathcal{P}$ stands for the principle value of the integral. If the real or imaginary part is measured over a wide frequency range with sufficient accuracy, the other part of the dielectric function can be calculated.

At the end of this section it should be pointed out that optical experiments play an important role in the investigation of the dielectric properties of solids. We therefore briefly explain the connection between the dielectric function and the optical quantities. Let us denote the *refractive index* with n' and the *extinction coefficient* with κ, then the following applies

$$\varepsilon' + \mathrm{i}\varepsilon'' = (n' + \mathrm{i}\kappa)^2, \tag{13.10}$$

with

$$\varepsilon' = n'^2 - \kappa^2 \qquad \text{and} \qquad \varepsilon'' = 2n'\kappa \,. \tag{13.11}$$

Measurements of n' and κ prove to be particularly difficult in the case of strong absorption, e.g. in the interesting range of resonance. For this reason, often the reflectivity R, also called reflectance, is determined for which the following relationship holds for waves incident perpendicular to the surface

$$R = \left|\frac{\sqrt{\varepsilon} - 1}{\sqrt{\varepsilon} + 1}\right|^2 = \frac{(n' - 1)^2 + \kappa^2}{(n' + 1)^2 + \kappa^2} \,. \tag{13.12}$$

However, the two important quantities n' and κ cannot be determined from the intensity ratio of reflected to incident beam. A measurement of the reflection would only be sufficient if in addition to the ratio of intensities, the phase shift between incoming and reflected waves could also be determined. As mentioned before, the *Kramers-Kronig relations* are a way out of this difficulty, because they connect ε' with ε'' or n' with κ. In order to be able to actually execute the numerical integration required for this, the measured quantity R of a thick sample must be determined at "all" frequencies, i.e. over a sufficiently wide frequency range.

13.2 Local Electrical Field

First we turn to the question of what field is present at a particular atom of an insulator when we apply an external field to a sample. We have omitted the equivalent question in the treatment of magnetic effects in the previous chapter, because this question is much more important for the electrical properties than for the magnetic ones.

If we put atoms in an electric field, they become polarized. In descriptive terms, the center of the electron cloud shifts relative to the nucleus of the atom. Thus the atoms themselves act as dipoles with the moment

$$\mathbf{p} = \varepsilon_0[\alpha]\mathcal{E} \,. \tag{13.13}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

$\mathcal{E}$ is the electric field at the location of the chosen atom and $[\boldsymbol{\alpha}]$ is the atomic polarizability,[3] which is a tensor quantity for non-spherical molecules. Due to the anisotropic structure and the directional bonds, the polarizability of many solids is dependent on the direction. We will treat polarizability as a scalar to keep the equations simple. However, a treatment of the directional dependence is possible without much effort. If equation (13.13) is applied to gases, the field strength $\mathcal{E}$ is identical with the applied field $\mathcal{E}_a$. In a solid, however, not only the applied field hut also the fields of the neighboring atoms contribute because of the small interatomic distances. The sum of both contributions, i.e. the field to which an atom is actually exposed, is called the **local field** $\mathcal{E}_{loc}$. On an atomic scale, the contribution of the atoms causes the electric field strength to vary. The macroscopic field $\mathcal{E}$, on the other hand, which is included in the Maxwell equations, is the average value over the locally varying fields.

We now want to establish the link between the local and the applied fields. For this purpose, we calculate the polarization $\mathbf{P}$, which an applied field causes in the microscopic and macroscopic description and compare the two results. Except for substances with permanent dipoles, which we will discuss separately, all induced dipole moments point in direction of the applied field and contribute equally to the polarization, i.e. $\mathbf{P} = n\mathbf{p} = n\varepsilon_0\alpha\,\mathcal{E}_{loc}$. On the other hand, in electrodynamics the polarization is linked to the macroscopically effective field via dielectric susceptibility $[\boldsymbol{\chi}]$, so that $\mathbf{P} = \varepsilon_0[\boldsymbol{\chi}]\mathcal{E}$. If we ignore the tensor properties of the susceptibility, then macroscopic and local field are connected via the relationship $\chi\mathcal{E} = n\alpha\,\mathcal{E}_{loc}$.

We derive the connection between the two fields within the framework of the **Lorentz approximation**. For this purpose, we split the local field into four parts, the meaning of which we discuss in turn:

$$\mathcal{E}_{loc} = \mathcal{E}_a + \mathcal{E}_D + \mathcal{E}_L + \mathcal{E}_s\,. \tag{13.14}$$

In the following discussion, we assume a homogeneously polarized dielectric sample, e.g. an insulating plate between a parallel-plate capacitor. We virtually divide the dielectric sample into two areas, one very small, spherical volume around the atom, where the local dipole-dipole interaction is taken into account, and the remaining part, which is described by the mean values of the fields. This situation is schematically sketched in Figure 13.1.

The induced charges on the surface of the sample cause a field $\mathcal{E}_D = -f\mathbf{P}/\varepsilon_0$ opposing the external field, and whose strength depends on the geometry of the sample. For the thin plate shown, the geometry-dependent depolarization factor is $f = 1$, for a sphere $f = 1/3$. If the field is applied along a thin cylinder or along a thin plate, the opposing field disappears, i.e. $f = 0$.

3 Occasionally the quantity $\varepsilon_0\alpha$ is defined as polarizability. As a consequence, the factor ε_0 appears in equations such as (13.20) or (13.21).

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

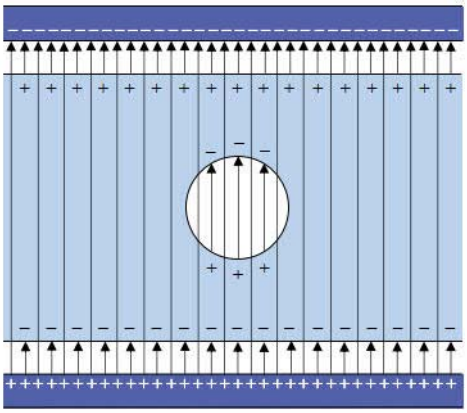


Fig. 13.1: Cross-section through a thin dielectric plate located in a parallel plate capacitor to derive the local electric field. The dark blue electrodes carry charges. The external field $\mathcal{E}_a$ in the "air gap" is reduced by polarization charges on the sample surface. The field inside the hole is larger than that in the surrounding medium.

In order to derive at geometry-independent expressions, the applied external field $\mathcal{E}_a$ and the **depolarisation field** $\mathcal{E}_D$ can be combined to form the macroscopically effective field $\mathcal{E} = (\mathcal{E}_a + \mathcal{E}_D)$.

Now we come back to the hole sketched in Figure 13.1. We imagine that a small sphere with the radius $\mathcal{R}$ was cut out of the sample. The sphere should contain a sufficiently large number of induced dipoles so that averaging over the atomic fields leads to the desired macroscopic field. On the surface of the spherical hole, there are charges due to the polarization of the plate. In the case of homogeneously polarized samples the surface charge density ϱ_p is given by $\varrho_p = -P\cos\theta$ for geometrical reasons. The dark blue ring surface in Figure 13.2 therefore carries the charge $dq = -P\cos\theta\ 2\pi\mathcal{R}\sin\theta\ \mathcal{R}\,d\theta$. For reasons of symmetry, the surface charge only generates a field in the direction of polarization. The ring-type surface element therefore contributes to the field inside the hole the part

$$d\mathcal{E}_L = -\frac{1}{4\pi\varepsilon_0}\cdot\frac{dq}{\mathcal{R}^2}\cos\theta\,. \tag{13.15}$$

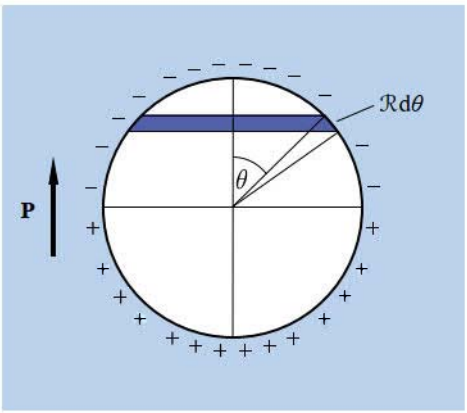


Fig. 13.2: Section through the sample with a hole in the center. The charge on the hole surface is due to the homogeneous polarization of the sample. The charge dq is located on the dark blue ring-type surface element.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The **Lorentz field** $\mathcal{E}_L$, which is caused by the polarization on the hole surface, is obtained by integration:

$$\mathcal{E}_L = \frac{P}{2\varepsilon_0} \int_0^\pi \cos^2\theta \sin\theta \mathrm{d}\theta = \frac{P}{3\varepsilon_0} . \tag{13.16}$$

We still have to discuss the influence of the neighboring atoms in the cut-out sphere on the field at the reference atom. Since the contributions of the neighboring atoms largely compensate each other, this field $\mathcal{E}_s$ is relatively small. $\mathcal{E}_s$ depends on the crystal structure and even disappears if the neighboring atoms have a cubic arrangement. To show this, we sum up the contributions from all atomic dipole fields in the sphere. Assuming that the electric field is applied in z-direction, the following expression is obtained

$$\mathcal{E}_s = \frac{1}{4\pi\varepsilon_0} \sum_m p_m \frac{3z_m^2 - r_m^2}{r_m^5} . \tag{13.17}$$

From the symmetry of the cubic lattice and the cut out sphere follows:

$$\sum_m \frac{x_m^2}{r_m^5} = \sum_m \frac{y_m^2}{r_m^5} = \sum_m \frac{z_m^2}{r_m^5} . \tag{13.18}$$

If we insert this expression in (13.17), we obtain the simple result $\mathcal{E}_s = 0$, which is plausible for geometrical reasons. With a cubic arrangement of the neighboring atoms, the dipole fields at the location of the atom just cancel each other out. Thus we get for the local field in a cubic lattice the **Lorentz relation**

$$\boldsymbol{\mathcal{E}}_{loc} = \boldsymbol{\mathcal{E}} + \frac{\mathbf{P}}{3\varepsilon_0} . \tag{13.19}$$

Although strictly speaking the field of the cut out sphere only disappears for cubic lattices, the Lorentz relation is a good approximation in most other cases as well.

Depending on the shape of the sample and the associated depolarization field, the polarization has an amplifying or weakening effect on the external electric field. If the electric field is perpendicular to a thin plate as shown in Figure 13.1, the local field is $\boldsymbol{\mathcal{E}}_{loc} = (\boldsymbol{\mathcal{E}}_a - 2\mathbf{P}/3\varepsilon_0)$. If, on the other hand, it runs parallel, we find $\boldsymbol{\mathcal{E}}_{loc} = (\boldsymbol{\mathcal{E}}_a + \mathbf{P}/3\varepsilon_0)$. In spherical samples, $\boldsymbol{\mathcal{E}}_D$ and $\boldsymbol{\mathcal{E}}_L$ cancel each other out.

Using the relationship $\chi\boldsymbol{\mathcal{E}} = n\alpha\,\boldsymbol{\mathcal{E}}_{loc}$ and the local field in the Lorentz approximation, we can now calculate the susceptibility assumed to be isotropic, and find with the help of (13.19)

$$\chi = \frac{n\alpha}{1 - n\alpha/3} . \tag{13.20}$$

Without consideration of the local field $\chi = n\alpha$. For the actually occurring dielectric constant $\varepsilon = (1 + \chi)$ the following expression results:

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{n\alpha}{3} . \tag{13.21}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

This connection between the experimentally measurable dielectric constant and the polarizability is the well-known **Clausius-Mossotti relation**[4,5]. With its help and equation (13.19), the local field that actually exists at the location of an reference atom can be calculated if the external field is known. The local field plays a very important role in the interpretation of measured data.

13.3 Dielectric Polarization

In this section we want to discuss the dielectric behavior of insulators in more detail. As an illustration, Figure 13.3 shows a schematic overview of the frequency response of the real part of the dielectric function of a polar crystal, i.e. a crystal with ionic bonding. Three contributions can be distinguished, each of which loses its significance beyond a characteristic frequency range as the frequency increases. Molecules with permanent dipole moment cause a *dipolar polarization* at low frequencies. As we will see, these contributions to the polarization typically vanish in the microwave range.

The *ionic polarization* is based on the shift of charged ions in the electrical field and is most pronounced in ionic crystals. The ionic contribution to the dielectric constant disappears in the infrared range. At even higher frequencies, only the *electronic polarization* is contributing. The electronic polarization can be understood as a shift of the electron cloud with respect to the nuclei. In the X-ray region, the contribution of the atomic shells finally also disappears and the dielectric constant approaches the

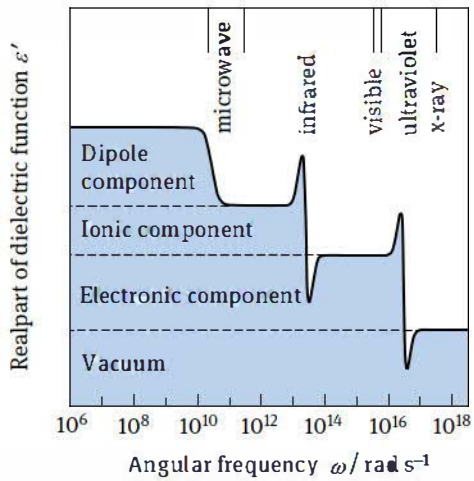


Fig. 13.3: Schematic representation of the frequency dependence of the dielectric constant of polar crystals. The size and exact position of the different contributions depends on the specific properties of the solid under consideration.

4 Rudolf Clausius, *1822 Köslin, †1888 Bonn
5 Ottaviano Fabrizio Mossotti, *1791 Novara, †1863 Pisa

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

vacuum limit $\varepsilon = 1$ from "below". The total polarization is the sum of the respective individual contributions, each of which we now discuss separately.

13.3.1 Electronic Polarizability

The optical properties of solids in the visible and ultraviolet are determined by *interband* and *intraband* transitions. In the case of isolators, which are the focus here, only interband transitions are possible because the bands are either completely occupied or completely empty. In contrast, intraband transitions between occupied and unoccupied states within bands also occur in metals. In the *tight-binding model* (cf. Section 8.4), optical interband transitions correspond to the transitions between the occupied and unoccupied energy levels of individual atoms, whereby the discrete levels are broadened into bands as a result of the interaction with the neighboring atoms. The dielectric function can be calculated by summing over all single excitations. We will not reproduce the quantum mechanical calculation here, but only give the result for the polarizability or the dielectric function.

As early as 1907, *H.A. Lorentz* succeeded in describing the electronic polarizability in a simple and proper way with the help of the **oscillator model**. The electrons are treated as negative charge clouds which surround the nucleus and are excited to harmonic oscillations by the incident electromagnetic wave. We will briefly derive the predictions of this model and compare them with the quantum mechanical result. With a slight modification, this model also allows for the description of ionic polarization, which we will discuss below.

We assume that an electron of an atom of the crystal lattice experiences a restoring force which is proportional to the displacement x. If an alternating electric field $\mathcal{E}_{\mathrm{loc}}(t) = \mathcal{E}^0_{\mathrm{loc}} \exp(-\mathrm{i}\omega t)$ interacts with an electron with mass m, an oscillation, is excited, which is damped by radiation effects. We therefore start with the equation of motion of a driven harmonic oscillator:

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + m\gamma\frac{\mathrm{d}x}{\mathrm{d}t} + m\omega_0^2 x = -e\mathcal{E}_{\mathrm{loc}} \,. \tag{13.22}$$

Here γ stands for the damping constant and ω_0 for the resonant frequency of the undamped oscillator. The stationary solution of this differential equation is

$$x(t) = -\frac{e}{m}\,\frac{1}{\omega_0^2 - \omega^2 - \mathrm{i}\gamma\omega}\,\mathcal{E}_{\mathrm{loc}}(t) \,. \tag{13.23}$$

The dipole moment $p = -ex$ is linked to the motion. If we insert this result in (13.13), we obtain for the polarizability

$$\alpha = \frac{e^2}{\varepsilon_0 m}\,\frac{1}{\omega_0^2 - \omega^2 - \mathrm{i}\gamma\omega} \,. \tag{13.24}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

For some atoms and ions, Table 13.1 shows the values of α. If we continue to use equation (13.19), the dielectric function $\varepsilon(\omega) = 1 + \chi(\omega) = 1 + P(\omega)/\varepsilon_0\mathcal{E}$ becomes

$$\varepsilon(\omega) = 1 + \frac{ne^2}{\varepsilon_0 m}\,\frac{1}{\omega_0^2 - [ne^2/3\varepsilon_0 m] - \omega^2 - i\gamma\omega} = 1 + \frac{ne^2}{\varepsilon_0 m}\,\frac{1}{\omega_1^2 - \omega^2 - i\gamma\omega}\,. \tag{13.25}$$

In the second expression we have introduced the new resonance frequency ω_1, which is given by $\omega_1^2 = (\omega_0^2 - ne^2/3\varepsilon_0 m)$. The local field, which is defined by the neighbors causes a shift in the resonance frequency. We will not pursue this effect here, since we have already dealt with the shift of energy levels in detail in Section 8.4 in the context of the tight-binding model.

Tab. 13.1: Polarizability of atoms and ions. (After A. Dalgarno, Adv. Phys. **11**, 281 (1962).)

	He	Ar	Kr	Xe	Na^+	K^+	F^-	Cl^-
Polarizability $\alpha/10^{24}\text{cm}^3$	0.2	1.6	2.5	4.0	0.2	0.9	1.2	3.0

Now we split the complex dielectric function $\varepsilon = (\varepsilon' + i\varepsilon'')$ into its real and imaginary part and obtain:

$$\varepsilon'(\omega) = 1 + \frac{ne^2}{\varepsilon_0 m}\,\frac{\omega_1^2 - \omega^2}{(\omega_1^2 - \omega^2)^2 + \gamma^2\omega^2}\,, \tag{13.26}$$

$$\varepsilon''(\omega) = \frac{ne^2}{\varepsilon_0 m}\,\frac{\gamma\,\omega}{(\omega_1^2 - \omega^2)^2 + \gamma^2\omega^2}\,. \tag{13.27}$$

Figure 13.4 shows the schematic frequency dependence of these two functions. Characteristic for resonant excitations is that the imaginary part ε'' only noticeably differs from zero in the vicinity of the resonance frequency. In this frequency range the real part ε' also strongly changes. With increasing frequency ε' reaches its maximum value

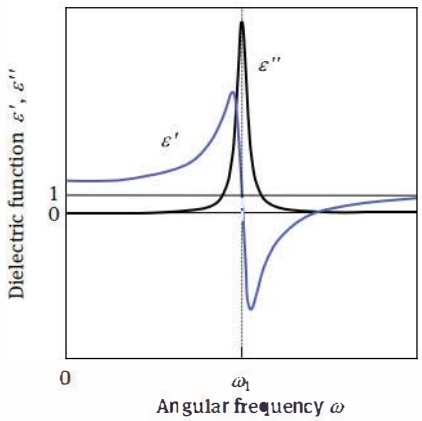


Fig. 13.4: Frequency dependence of the real (blue) and imaginary part (black) of the dielectric function. At weak attenuation the first zero-crossing of ε' is found approximately at the resonance frequency ω_1, i.e. near the maximum of ε''.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

just below ω_1 and then drops rapidly. At $\omega = \omega_1$ the real part ε' assumes the value of one. At even higher frequencies ε' becomes negative and passes through a minimum. After a further zero crossing, ε' increases to approach its vacuum value of one well above the resonance frequency. Because ε'' disappears outside the resonance range, the refractive index according to (13.11) is given there by $n' \simeq \sqrt{\varepsilon'(\omega)}$.

Atoms and thus also solids have several resonance frequencies, which have differently strong effects on the polarizability. To take this fact into account, the *oscillator strength* f_k is introduced for the individual transitions, the value of which is adapted to the experimental conditions. The dielectric function is then obtained by summing all resonances that occur. Instead of (13.25) we have

$$\varepsilon(\omega) = 1 + \frac{ne^2}{\varepsilon_0 m} \sum_k \frac{f_k}{\omega_k^2 - \omega^2 - i\gamma_k \omega} \,. \tag{13.28}$$

The same frequency dependence is found in the more complex quantum mechanical calculation. A comparison shows that the oscillator strength introduced here is essentially determined by the matrix element of the respective transition.

The calculation of the absorption spectrum of solids is a relatively involved problem. The probability for the occurrence of a certain transition depends not only on the corresponding matrix element and thus on the oscillator strength, but also on the electronic density of states of the initial and final states. This means that the $\varepsilon''(\omega)$ curve reflects the *combined density of states*. Furthermore, it must be taken into account that not only direct processes, but also, as discussed in Section 10.1, phonons-mediated processes can be involved in the absorption and enable additional transitions. Of particular interest are the absorption processes which are based on the transition of electrons from the valence band to the conduction band and thus require the least amount of energy among the electronic excitations in the insulator. These are called **fundamental absorption processes**.

In corresponding experiments, the reflectivity of the sample is first measured, and then ε' and ε'' are calculated using the Kramers-Kronig relations. From the course of the dielectric function, conclusions can then be drawn about critical points, the densities of states and thus about the band structure. Figure 13.5 shows the course of the real and imaginary parts of the dielectric function of germanium over a wide range of photon energies. Since germanium has neither ionic nor dipole polarization, the dielectric function over the entire frequency range is only determined by the contribution of the electronic transitions. Obviously, several absorption maxima occur, which can be assigned to critical points of the combined density of states.

With increasing photon energy, the oscillator strengths decrease rapidly, since the atomic core electrons hardly contribute to the polarizability. The absorption in the deep ultraviolet and X-ray range is therefore weak and hardly affects the optical properties at lower frequencies. If one calculates the refractive index n' from the data in Figure 13.5 using equation (13.11), we find that n' becomes smaller than one at a photon

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

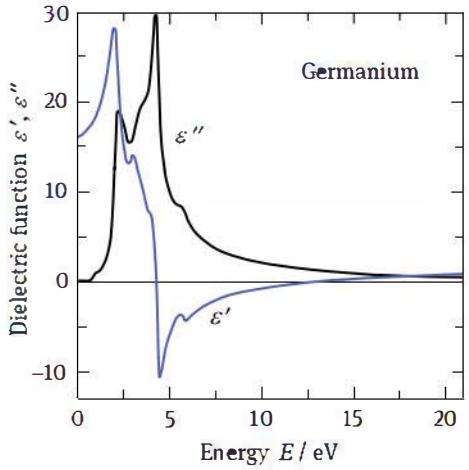


Fig. 13.5: Real (blue) and imaginary part (black) of the dielectric function of germanium over a wide energy range. (After H.R. Philipp, H. Ehrenreich, Phys. Rev. **129**, 1550 (1963).)

energy above 7 eV, although other resonances occur in the X-ray region. Above 14 eV, ε' becomes positive again and the refractive index is approaching one.

13.3.2 Ionic Polarization

The schematic representation in Figure 13.3 suggests that ionic polarization is also caused by resonance processes. As with the electronic polarization, we can also use the simple *oscillator model* to describe it.

Within the microscopic description, the interaction of infrared radiation with the ions of a lattice can be described as a scattering process of photons with optical phonons. Since the quasi-momentum is conserved for these processes, only phonons with very small wave vectors can participate. As we have seen in Section 6.2, the dispersion curve of the optical phonons near the Γ-point is almost horizontal. All phonons of one branch involved in the interaction therefore have almost the same frequency. At long wavelengths in the case of crystals with a diatomic basis, to which we limit ourselves here, the ions of one sublattice move rigidly in opposite phase to the ions of the oppositely charged one. For a macroscopic description it is therefore sufficient to pick out one ion pair and study its reaction to electromagnetic waves, since all ion pairs behave in the same way.

The ions of the two sublattices with the charges q and $-q$, respectively, are affected not only by elastic forces but also by electrical forces caused by the local electric field $\mathcal{E}_{\text{loc}}$ which is composed of the applied field and the field of the ions themselves. If we use $\mathbf{u}_1$ and $\mathbf{u}_2$ to denote the displacements of the ions with the masses M_1 and M_2,

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

then instead of equation (6.36) we obtain the expressions

$$\begin{aligned} M_1\ddot{\mathbf{u}}_1 + 2C\mathbf{u}_1 - 2C\mathbf{u}_2 &= q\mathcal{E}_{\text{loc}}\,, \\ M_2\ddot{\mathbf{u}}_2 + 2C\mathbf{u}_2 - 2C\mathbf{u}_1 &= -q\mathcal{E}_{\text{loc}}\,. \end{aligned} \tag{13.29}$$

Since we are only interested in the relative motion of the two sublattices, we introduce the relativ displacement $\mathbf{u} = (\mathbf{u}_1 - \mathbf{u}_2)$ of the sublattices and the reduced mass μ of the ion pairs and obtain

$$\mu\ddot{\mathbf{u}} + \mu\omega_0^2\mathbf{u} = q\mathcal{E}_{\text{loc}}\,. \tag{13.30}$$

Here we have used the abbreviation $\omega_0^2 = 2C/\mu$ for the resonance frequency of the system if only elastic forces are acting. The equation of motion is identical to that of a driven linear harmonic oscillator.

In this equation, we add an attenuation term that takes into account the finite lifetime of the optical phonons. The physical cause of damping is the interaction of the phonons with each other and the coupling to electromagnetic waves. If we denote the damping constant as γ, then instead of (13.30) we have the equation

$$\mu\ddot{\mathbf{u}} + \mu\gamma\dot{\mathbf{u}} + \mu\omega_0^2\mathbf{u} = q\mathcal{E}_{\text{loc}}\,. \tag{13.31}$$

If an alternating electric field $\mathcal{E}_{\text{loc}}(t) = \mathcal{E}^0_{\text{loc}}\exp(-\mathrm{i}\omega t)$ acts on the charge q, then the stationary solution to this differential equation is:

$$\mathbf{u}(t) = \frac{q}{\mu}\,\frac{1}{\omega_0^2 - \omega^2 - \mathrm{i}\gamma\omega}\,\mathcal{E}^0_{\text{loc}}\mathrm{e}^{-\mathrm{i}\omega t}\,. \tag{13.32}$$

Formally, the solution is identical with (13.23), but the quantities that appear have a different meaning.

13.3.3 Optical Phonons in Ionic Crystals

Before we go into the solution (13.32) in more detail, we first consider the case of free lattice vibrations to study the influence of local electric fields on the vibration spectrum of ionic crystals. As we have seen from the dispersion curves of silicon and lithium fluoride in Section 6.3, non-polar crystals and ionic crystals behave quite differently with respect to their optical phonons.[6] While in silicon (Figure 6.21), all optical phonons have the same frequency at the Γ-point, in lithium fluoride (Figure 6.20) there is a large gap between the transverse and longitudinal branches.

The motion of the ions causes the oscillating dipole moment $\mathbf{p}(t) = q\mathbf{u}(t)$ and thus the polarization $nq\mathbf{u}(t)$. In addition, there is the electronic contribution to the

6 In the following we will use the abbreviations LO and TO for *longitudinal optical* and *transverse optical* respectively.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

polarization. Since the resonant frequency of the oscillating ions is much lower than that of the electrons due to their larger mass, we can assume that the electronic polarizability (13.24) is constant in the frequency range under consideration. In further discussion we summarize the contributions of the positive and negative ions to the electronic polarizability $\alpha = (\alpha_+ + \alpha_-)$. Thus the total polarization of the sample can be expressed by

$$\mathbf{P}(t) = nq\mathbf{u}(t) + n\varepsilon_0\alpha\mathcal{E}_{\mathrm{loc}}(t) \,. \tag{13.33}$$

Due to the different motion of the ions in longitudinal and transverse waves, the local field differs considerably. As indicated in the schematic illustration in Figure 13.6, we can divide the sample into fictitious slices bounded by nodal planes defined by $\mathbf{u} = 0$. In case of long-wavelength lattice vibrations, there are many atomic layers between the nodal planes.

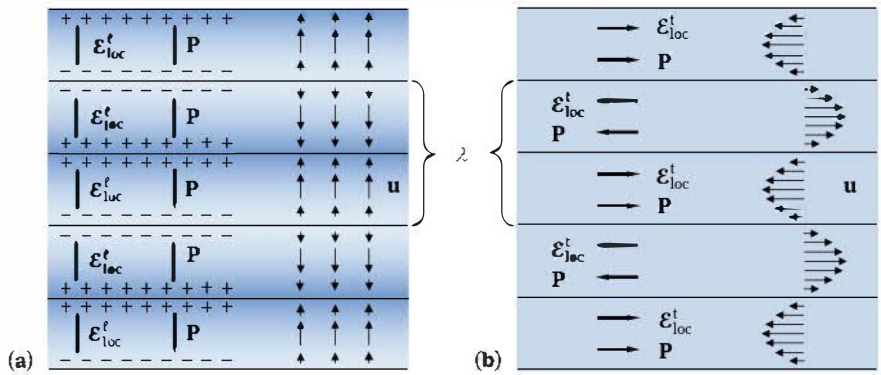


Fig. 13.6: For the derivation of the local field at optical lattice vibrations in ionic crystals. The nodal planes (solid lines) of long-wavelength phonons are many atomic layers apart. **a)** For the case of LO phonons, the local field is antiparallel to the polarization. **b)** For the case of TO phonons, the two fields are parallel to each other.

For LO phonons, the atomic displacement is parallel to the wave vector. Thus the local field and the polarization are perpendicular to the nodal planes. In this case, the depolarization factor is $f = 1$. If there is no external field, the electric field is given by:

$$\mathcal{E}^{\ell}_{\mathrm{loc}} = \mathcal{E}_{\mathrm{D}} + \mathcal{E}_{\mathrm{L}} = -\frac{\mathbf{P}_{\ell}}{\varepsilon_0} + \frac{\mathbf{P}_{\ell}}{3\varepsilon_0} = -\frac{2\mathbf{P}_{\ell}}{3\varepsilon_0} \,. \tag{13.34}$$

The local field and the polarization therefore change in opposite phase. Thus, the local field counteracts the relative displacement $\mathbf{u} = (\mathbf{u}_1 - \mathbf{u}_2)$ of the sublattices, tries to reduce the displacement and thus causes an additional restoring force.

With TO phonons, no depolarization field occurs because $\mathcal{E}_{\mathrm{loc}}$ and $\mathbf{P}$ run parallel to the slices. This results in

$$\mathcal{E}^{\mathrm{t}}_{\mathrm{loc}} = \mathcal{E}_{\mathrm{D}} + \mathcal{E}_{\mathrm{L}} = +\frac{\mathbf{P}_{\mathrm{t}}}{3\varepsilon_0} \,. \tag{13.35}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The local field, which also determines the displacement of the ions, therefore has opposite signs for LO and TO phonons. While in case of LO phonons it increases the repulsive forces, in case of TO phonons it counteracts the elastic forces and makes the material softer. If the local field is known, the frequency of the longitudinal and transverse phonons can easily be calculated with the help of (13.32) and (13.33). If we neglect the damping term in (13.32), since it exerts only an insignificant influence here, we obtain for the frequency of the transverse and longitudinal phonons at the Γ-point the two equations:

$$\omega_t^2 = \omega_0^2 - \frac{nq^2}{3\varepsilon_0\mu}\,\frac{1}{1-n\alpha/3}\,, \tag{13.36}$$

$$\omega_\ell^2 = \omega_0^2 + \frac{2nq^2}{3\mu\varepsilon_0}\,\frac{1}{1+2n\alpha/3}\,. \tag{13.37}$$

The electric field in the sample increases the frequency of the longitudinal vibrations and lowers the frequency of the transverse vibrations, i.e. $\omega_\ell > \omega_t$. Lattice distortion and polarization are closely related in ionic crystals.

It is worthwhile taking a brief look at the interaction between electromagnetic waves and optical phonons. The coupling between the two types of waves is only effective if they both travel in the same direction. Under this condition, the electric fields of the LO phonons are perpendicular to the fields of the electromagnetic wave. Therefore no coupling occurs between the two waves (inside crystals), so that LO phonons cannot be excited and detected by infrared radiation. Their detection is possible with the help of Raman scattering (see Figure 13.9) or inelastic electron scattering. TO phonons, on the other hand, can be excited by infrared radiation, since the electric fields of the two waves run parallel in this case. The interaction with TO phonons is the cause of the strong absorption of ionic crystals in the infrared.

13.3.4 Dielectric Function of Ionic Crystals

We now come back to solution (13.32) of the differential equation (13.31) and calculate with its help the dielectric function $\varepsilon(\omega) = 1 + \chi = 1 + |\mathbf{P}|/\varepsilon_0|\mathfrak{E}|$. For this we eliminate the quantities $\mathfrak{E}_{\text{loc}}(t)$ and $\mathbf{u}(t)$ in (13.33) with the help of (13.19) and (13.32) and obtain the lengthy expression

$$\varepsilon(\omega) = 1 + \frac{n\alpha}{1-n\alpha/3} + \frac{nq^2}{\varepsilon_0\mu}\left(\frac{1}{1-n\alpha/3}\right)^2\left[\omega_0^2 - \frac{nq^2}{3\varepsilon_0\mu(1-n\alpha/3)} - \omega^2 - \mathrm{i}\gamma\omega\right]^{-1}. \tag{13.38}$$

A brief inspection of the equation shows that the resonance of the forced oscillation does not occur at the frequency ω_0 of the purely elastic oscillation, but at the frequency ω_t of the transverse-optical lattice vibrations.

The somewhat cumbersome equation can be reduced to the simple form

$$\varepsilon(\omega) = \varepsilon_\infty + \frac{\omega_t^2(\varepsilon_{\text{st}} - \varepsilon_\infty)}{\omega_t^2 - \omega^2 - \mathrm{i}\gamma\omega}\,, \tag{13.39}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

if, in addition to the new resonance frequency (13.36), the abbreviations $\varepsilon_{\bullet\bullet}$ and ε_{st} which stand for the limiting values of the ionic contribution to the dielectric function at high and low frequencies are introduced. The index $\bullet\bullet$ does not actually mean infinitely high, but at frequencies that are large compared to the frequency of the ionic resonance, i.e. $\varepsilon_{\bullet\bullet}$ reflects the electronic polarization in the case that the frequency is small compared to that of the electronic resonance. On the other hand ε_{st} stands for the *static value* of the dielectric constant, i.e. for its value at frequencies significantly below ionic resonance. This value includes the contributions of ions *and* electrons. For substances without permanent dipole moments, where no orientation polarization occurs, ε_{st} actually corresponds to the value of the dielectric function at DC fields.

Now we split the complex dielectric function $\varepsilon = (\varepsilon' + i\varepsilon'')$ into its real and imaginary part and find:

$$\varepsilon'(\omega) = \varepsilon_{\infty} + \frac{(\varepsilon_{st} - \varepsilon_{\infty})\,\omega_t^2\,(\omega_t^2 - \omega^2)}{(\omega_t^2 - \omega^2)^2 + \gamma^2\omega^2}\,, \tag{13.40}$$

$$\varepsilon''(\omega) = \frac{(\varepsilon_{st} - \varepsilon_{\infty})\,\omega_t^2\,\gamma\,\omega}{(\omega_t^2 - \omega^2)^2 + \gamma^2\omega^2}\,. \tag{13.41}$$

Figure 13.7 shows the frequency response of these two functions schematically. Qualitatively the picture is identical with Figure 13.4, however now the quantities have a different meaning. Again, the imaginary part ε'' is only noticeably different from zero in the vicinity of the resonance frequency. At the same time there are strong changes in the real part ε', which twice becomes zero: at ω_ℓ and near ω_t.

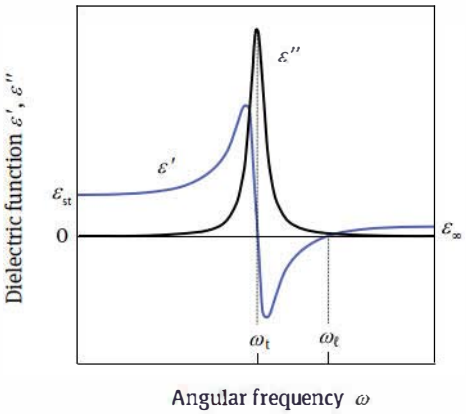


Fig. 13.7: Frequency response of the real (blue) and imaginary part (black) of the dielectric function. The zero crossings of ε' are located at ω_ℓ, and near ω_t.

To understand the meaning of the second zero crossing of ε', we consider a longitudinal-optical phonon propagating in x-direction with polarization $\mathbf{P}_\ell = \hat{\mathbf{x}} P_0 \exp[-i(\omega t - q_\ell x)]$. If we calculate the divergence, we see that $\mathrm{div}\,\mathbf{P}_\ell \neq 0$. According to the Maxwell equations, the divergence of the dielectric displacement of a neutral, space-charge-free

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

sample disappears and thus

$$\mathrm{div}\,\mathbf{D} = \mathrm{div}\,[\varepsilon_0\varepsilon(\omega)\,\mathbf{E}] = \varepsilon_\infty\,\mathrm{div}\,\mathbf{E} + \mathrm{div}\mathbf{P} = 0\,. \tag{13.42}$$

Since for phonons the divergence of $\mathbf{P}$ has a finite value, also $\mathrm{div}\,\mathbf{E} \neq 0$. The equation can only be fulfilled if the dielectric function $\varepsilon(\omega)$ disappears at the frequency of vibration and thus $\varepsilon(\omega_\ell) = 0$.

If we divide (13.36) by (13.37) and use the definition of ε_{st} and ε_∞, a short calculation results in

$$\frac{\omega_\ell^2}{\omega_t^2} = \frac{\varepsilon_{st}}{\varepsilon_\infty}\,. \tag{13.43}$$

This is the **Lyddane-Sachs-Teller relation**[7,8,9] which is in excellent agreement with experimental results. It shows that dielectric and elastic properties are closely related. An interesting aspect is that ε_{st} becomes very large if the eigenfrequency of the transverse-optical phonons is strongly reduced, i.e. when they become "soft". As we will see, this process plays an important role in ferroelectrics, which we will discuss later.

Table 13.2 shows the dielectric constants and the frequency of the longitudinal- and transverse-optical phonons of various crystals.

Tab. 13.2: Dielectric constants ε_{st} and ε_∞ and the frequency of the optical phonons of a number of dielectric crystals. (Most data were taken from E. Kartheuser, *Polarons in Ionic Crystals and Polar Semiconductors*, J.T. Devreese, ed., North Holland, 1972).

	LiCl	NaCl	NaJ	KCl	KBr	CsCl	GaAs	CdS	ZnSe	PbS
ε_{st}	11.95	5.9	7.28	4.85	4.52	6.68	12.83	8.42	8.33	190
ε_∞	2.79	2.40	3.15	2.22	2.43	2.69	10.90	5.27	5.90	18.50
$\omega_t/10^{13}$ Hz	4.16	3.35	2.34	2.67	2.15	2.02	5.14	4.60	3.90	1.26
$\omega_\ell/10^{13}$ Hz	8.19	5.10	3.43	4.07	3.18	3.16	5.58	5.80	4.63	4.03

13.3.5 Phonon Polaritons

As mentioned above, transverse-optical phonons can directly couple to electromagnetic waves. This interaction causes a mixture of the two wave types and thus leads to a drastic change in the phonon spectrum near the Γ-point. Vividly speaking, an

7 Russell Hancock Lyddane, *1913 Washington D.C., †2001 Chester County
8 Robert Green Sachs, *1916 Hagerstown, Maryland, †1999 Chicago
9 Edward Teller, *1908 Budapest, †2003 Stanford

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

electromagnetic wave causes polarization and thus a lattice distortion that moves with the wave. Conversely, a TO lattice wave is accompanied by an electromagnetic wave.

To make our discussion as clear as possible, we assume that the electromagnetic wave and the TO lattice wave propagate both in x-direction and are polarized in the y-direction. Since we are looking for a common solution for both wave types, the same frequency and the same wave number must occur in the ansatz for both functions. For the polarization of the lattice wave we write $\mathbf{P}_t = \hat{\mathbf{y}} P_0 \exp[-i(\omega t - q_t x)]$. The propagation of the electromagnetic waves can be calculated, if magnetic effects are neglected, with the wave equation

$$c^2 \Delta \mathcal{E} = \varepsilon(\omega) \ddot{\mathcal{E}} \,. \tag{13.44}$$

where c stands for the speed of light in vacuum. If we use a plane wave with the electric field strength $\mathcal{E} = \hat{\mathbf{y}} \mathcal{E}_0 \exp[-i(\omega t - q_t x)]$ as the solution, we obtain the known dispersion relation

$$\omega^2 = \frac{1}{\varepsilon(\omega)} c^2 q_t^2 \,, \tag{13.45}$$

whose frequency response is essentially determined by the frequency dependence of the dielectric function (13.39). We neglect the damping and insert the expression for the dielectric function in (13.45). The result is

$$\omega^2 \left[\varepsilon_\infty + \frac{\omega_t^2 (\varepsilon_{st} - \varepsilon_\infty)}{\omega_t^2 - \omega^2} \right] = c^2 q_t^2 \,. \tag{13.46}$$

The course of the dispersion relation is shown in Figure 13.8. The following limiting cases are of interest: If $\omega \ll \omega_t$, the equation is simplified to $\omega = cq_t/\sqrt{\varepsilon_{st}}$. The electromagnetic wave propagates at low frequencies with a velocity that is given by the static dielectric constant. The same applies to optical frequencies $\omega \gg \omega_t$, where $\omega = cq_t/\sqrt{\varepsilon_\infty}$ because there the electronic polarizability determines the wave propagation. Particularly interesting are the cases $\omega \to \omega_t$ and $\omega \to \omega_\ell$. In the first case we find that $q_t \to \infty$, in the second, however, taking equation (13.43) into account, we get $q_t \to 0$.

The strong coupling between photons and TO phonons leads to mixed states, which are called **polaritons**. If one increases the frequency coming from zero, as shown in Figure 13.8, there is a continuous transition from purely electromagnetic waves to pure lattice vibrations. This is followed by a frequency gap from ω_t to ω_ℓ. Above the gap, at the frequency ω_ℓ one finds again optical phonons with purely transverse (!) displacements. At very high frequencies the excitations then gradually change into purely electromagnetic waves. If waves with a frequency in the forbidden range fall onto a sample, they are totally reflected, because waves with this frequency cannot propagate in ionic crystals. The forbidden frequency range is not caused by the periodicity of the lattice, as in the case of the frequency gap of the pure lattice vibrations, but is a consequence of the resonant coupling where the degeneracy of electromagnetic and elastic waves is lifted. In non-polar solids, $\varepsilon(\omega)$ is approximately constant in this frequency range, and the dispersion curves of light and lattice vibrations cross each other without special features.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

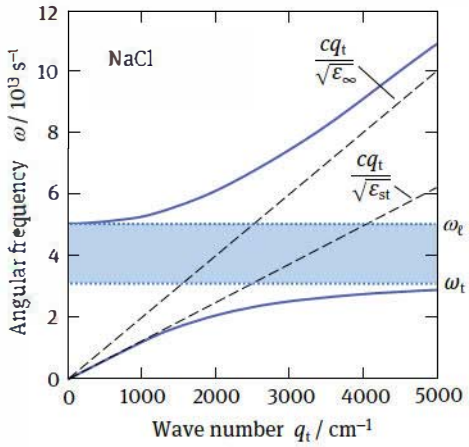


Fig. 13.8: Dispersion relation of polaritons. The dispersion curves of light in the two limiting cases mentioned in the text are shown as dashed lines. The dispersion curves of optical phonons without coupling are horizontal (dotted lines). The blue curves were calculated with the numerical values for NaCl: $\omega_t = 3.1 \times 10^{13}\ s^{-1}$, $\omega_\ell = 5.0 \times 10^{13}\ s^{-1}$, $\varepsilon_{st} = 5.9$ and $\varepsilon_\infty = 2.25$. The forbidden zone is highlighted in light blue.

Figure 13.9 shows data from Raman scattering at GaP, which can be used to determine the dispersion curve of phonon polaritons. The measured values, which belong to the polaritons, are represented by dark blue dots. In addition, measurement data for LO phonons are shown in light blue, which were also determined using Raman scattering. In the range of small wave numbers, the phonon frequency does not noticeably depend on the wave number.

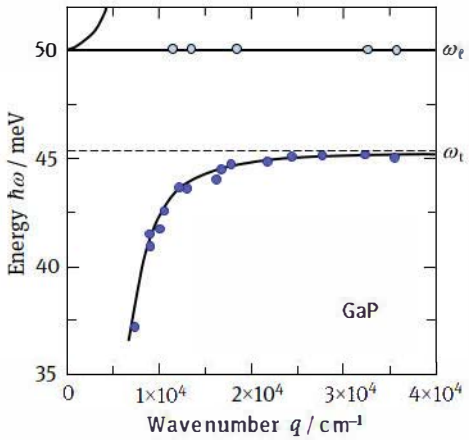


Fig. 13.9: Polaritons (dark blue points) in GaP measured by Raman scattering. In addition, measured values for LO phonones are shown in light blue. The dispersion curve of the uncoupled TO-phonons is indicated by a dashed line. (After C.H. Henry, J.J. Hopfield, Phys. Rev. Lett. 15, 964 (1965).)

In addition, it should also be noted that we started from large samples and ignored surface effects. The situation changes considerably if we go to thin layers or small grains, because there electronic boundary conditions are different. It is therefore not surprising that the oscillation frequencies and coupling strengths change when the wavelength of the exciting electromagnetic field becomes comparable with the sample dimension.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The existence of polaritons is reflected in the course of the dielectric function (13.39) and thus in the infrared properties of the ionic crystals, which we will discuss based on experimental data for *cadmium sulfide* (CdS). Although in a real experiment the starting point is the measured reflectivity in the infrared, in our example we want to discuss the infrared properties of cadmium sulfide in "reverse" order. Figures 13.10a and 13.10b show the frequency response of the imaginary and real part of the dielectric function of cadmium sulfide, calculated from the reflection data using the Kramers-Kronig relations.

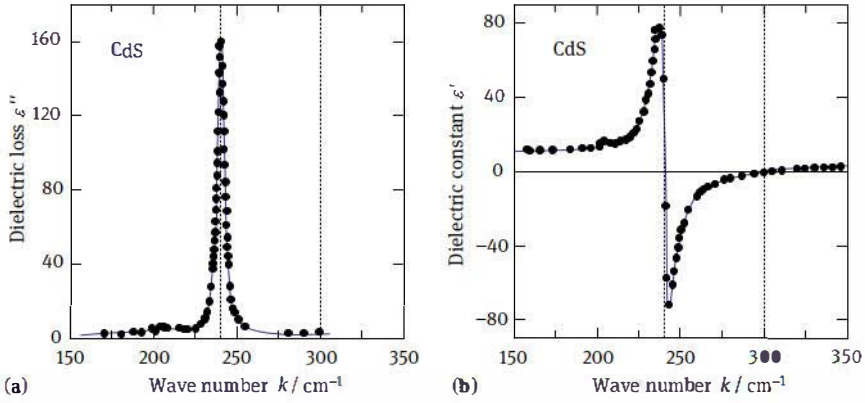


Fig. 13.10: Dielectric function of cadmium sulfide. **a)** Imaginary part ε'', **b)** Real part ε'. The dashed lines mark the values of ω_t and ω_ℓ. (After M. Balkanski, Optical Properties of Solids, F. Abelès, ed., North-Holland, 1972.)

In accordance with equation (13.41), a pronounced maximum of ε'' is observed at the wave number 240 cm^{-1}, i.e. at the frequency $\omega_t = 4.5 \times 10^{13}\ s^{-1}$ of the TO phonons. The small maximum at lower frequencies cannot be understood with our simple theory. It is based on anharmonic effects and is meaningless for our discussion of linear relations. Starting from the static value $\varepsilon_{st} = 8.4$, the real part ε' of the dielectric constant, as predicted by equation (13.40), first increases, passes a maximum and intersects the frequency axis near ω_t. Then ε' is negative and crosses the axis a second time at 301 cm^{-1} or at $\omega_\ell = 5.7 \times 10^{13}\ s^{-1}$. In the further course ε' approaches the value $\varepsilon_\infty = 5.3$.

From the curves, the course of the refractive index n' and the extinction coefficient κ can be calculated using (13.11). Figure 13.11a shows the result for κ, which has a similar frequency response as ε''. As can be seen in Figure 13.11b, the refractive index n', coming from small frequencies, first increases like ε'. Between ω_t and ω_ℓ, i.e. in the prohibited frequency range, ε' is negative. If ε'' were zero in this frequency, n' would also disappear.

The measured reflectivity from which the data displayed in Figures 13.10 and 13.11 were derived is shown in Figure 13.12. It first increases with increasing frequency as the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

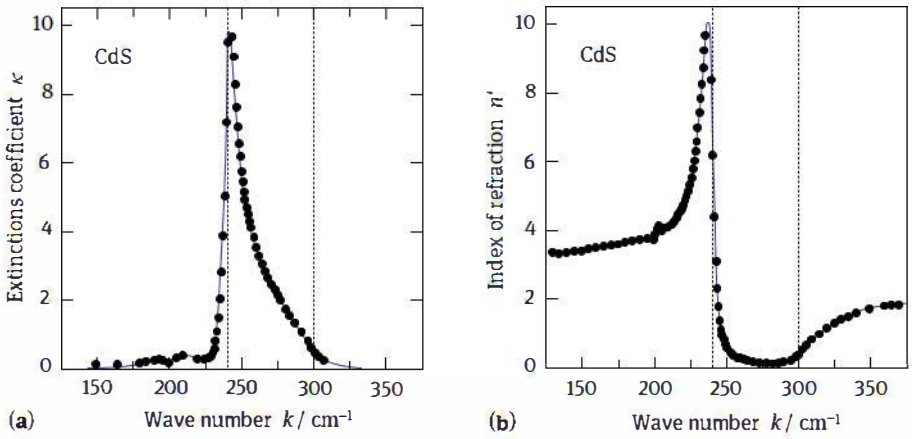


Fig. 13.11: Optical properties of cadmium sulfide. a) Extinction coefficient κ, b) Refractive index n'. The dashed lines mark the values of ω_t and ω_ℓ. (After M. Balkanski, *Optical Properties of Solids*, F. Abelès, ed., North-Holland, 1972.)

refractive index increases. In the forbidden frequency range the reflectivity is close to one. The finite damping causes a finite value of n' and thus a reflectivity less than one. At the frequency of $6.0\times10^{13}\ s^{-1}$, i.e. at 320 cm^{-1}, ε' and n' pass through the value one, so that after (13.12) the reflectivity almost disappears. Above this frequency the refractive index approaches the optical limit $\sqrt{\varepsilon_\infty}$ and the reflectivity increases accordingly. It should be emphasized once again that the experimental procedure is the reverse of the above. First, the reflectivity is measured very precisely. Then, with the help of the Kramers-Kronig relations (13.8) and (13.9) the other quantities are calculated.

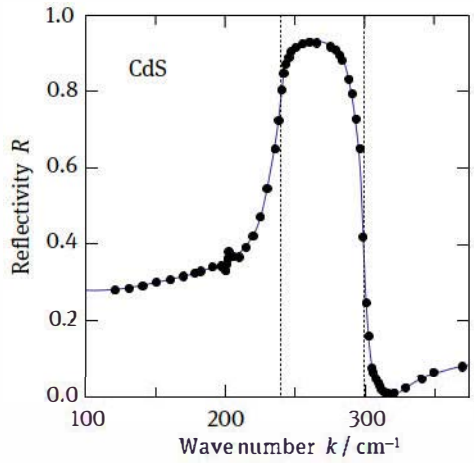


Fig. 13.12: Experimentally determined course of the reflectivity of cadmium sulfide in the infrared. In this measurement, the electric field of the incident radiation was perpendicular to the c-axis. The dashd lines mark the values of ω_t and ω_ℓ. (After M. Balkanski, *Optical Properties of Solids*, F. Abelès, ed., North-Holland, 1972.)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In the past, the dielectric properties of ionic crystals discussed above were used to generate infrared radiation with a relatively narrow frequency distribution. If an infrared beam with a broad frequency spectrum is reflected at an ionic crystal, after some reflections only radiation of the frequency interval with a high reflectivity remains. This procedure for generating a narrow frequency spectrum in the infrared is therefore called the **Reststrahlen method**.

13.3.6 Orientation Polarization

Static Polarization. Until now we have assumed that the electric field *induces* dipole moments. However, if a **permanent dipole moment** already exists, a static field causes a preferential orientation of the moments, since the potential energy $U = -\mathbf{p} \cdot \mathcal{E} = -p\,\mathcal{E}\cos\theta$ depends on the angle θ between the field and the dipole moment. The same kind of arguments we have already made in Section 12.2 in the discussion of paramagnetism. Apart from very low temperatures, only a partial alignment is achieved, because $p\,\mathcal{E} \ll k_B T$. If the molecules are free to rotate as it is the case in liquids or gases, the mean value $\langle\cos\theta\rangle = p\,\mathcal{E}/3k_B T$ can be calculated as for paramagnetism and expressions analogous to (12.8) and (12.9) can be found. For $p\,\mathcal{E} \ll k_B T$ follows therefore for the orientation polarization P_0 in the case of a DC field the **Langevin-Debye equation**

$$P_0 = np\,\langle\cos\theta\rangle \approx n\frac{p^2\mathcal{E}}{3k_B T}\,. \tag{13.47}$$

The temperature dependence of the associated dielectric susceptibility corresponds to the Curie law (12.14) for the magnetic susceptibility. As with paramagnetic substances, it was assumed that the interaction between the dipoles can be neglected, so that the susceptibility should only diverge at absolute zero. An example of the static dielectric

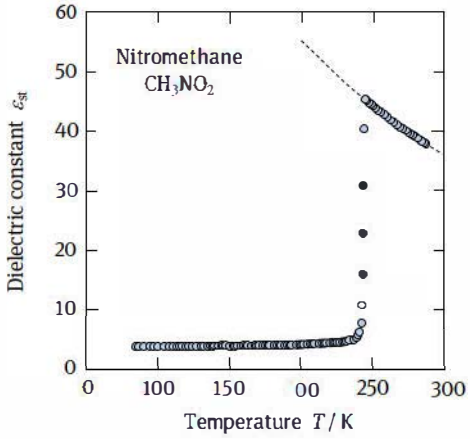


Fig. 13.13: Dielectric constant of the polar liquid CH_3NO_2 with the melting point at 244 K, measured at 115 kHz. The $1/T$ dependence is indicated by a dotted line. (After G. Kasper, A. Reiser, private communication.)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

constant of freely rotating molecules is shown in Figure 13.13. Coming from high temperature, the dielectric constant of the polar liquid nitromethane CH_3NO_2 increases up to the melting point. At the transition to solid state, the dielectric constant jumps to a much lower value which is almost independent of temperature. This means that the CH_3NO_2-molecules in the solid state can no longer rotate freely, assume fixed equilibrium positions and thus no longer contribute to orientation polarization. However, there are many solids in which the molecules still reorient themselves even in the solid phase. This is mostly the case when there are no covalent bonds and the molecules are approximately spherical.

Within solids, however, molecules in general cannot truly rotate freely, but reorient between preferred directions. In the following discussion we assume for the sake of simplicity that due to their shape and the crystal structure the dipoles can only occupy two equilibrium positions corresponding to a rotation of 180°. Furthermore, we assume that the field lies in the direction of the connecting line of the two equilibrium positions. The potential energy of the dipoles as a function of rotation angle then runs like sketched in Figure 13.14, where the field-free course is drawn as dashed line. Between the potential wells a potential barrier V_a exists.

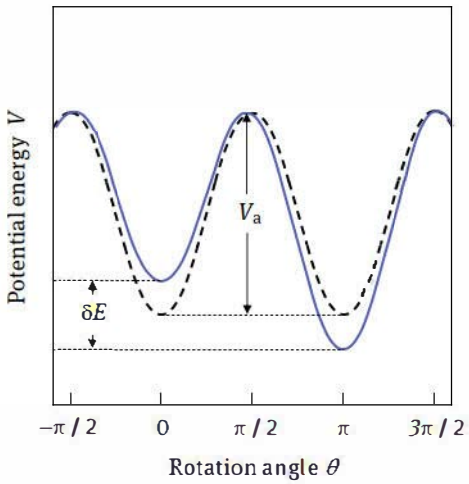


Fig. 13.14: Course of the potential energy of an electric dipole as a function of the angle of rotation between dipole moment and applied field. The curves show the potential energy with applied electric field (blue curve) and without (black-dashed curve).

If the electric field $\mathcal{E}$ is applied in the direction of the two wells, the energy difference $\delta E = 2\mathbf{p}\cdot\mathcal{E}$ occurs between the two equilibrium positions, without taking into account the local field correction. This is obviously a two-level system, the thermal occupation of which we have already derived in Section 7.2. We use the result (7.18) and obtain for the polarization $\mathbf{P}_0 = \mathbf{p}\delta n$ and the static dielectric constant ε_{st} the equations

$$P_0 = np \tanh \frac{p\mathcal{E}}{k_B T} \approx \frac{np^2\mathcal{E}}{k_B T} \tag{13.48}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

and

$$\varepsilon_{st} - 1 \approx \frac{np^2}{\varepsilon_0 k_B T} . \tag{13.49}$$

With the exception of a factor of 1/3, this result corresponds to the result for freely rotating molecules. There are no qualitative changes if the considerations are transferred to more complicated systems. Figure 13.15 shows the static dielectric constant of liquid and solid HCl. It can be clearly seen that the increase of ε_{st} continues with decreasing temperature, albeit with a slightly different slope, below the solidification temperature. The jump at the solidification is essentially a consequence of the density change at the phase transition.

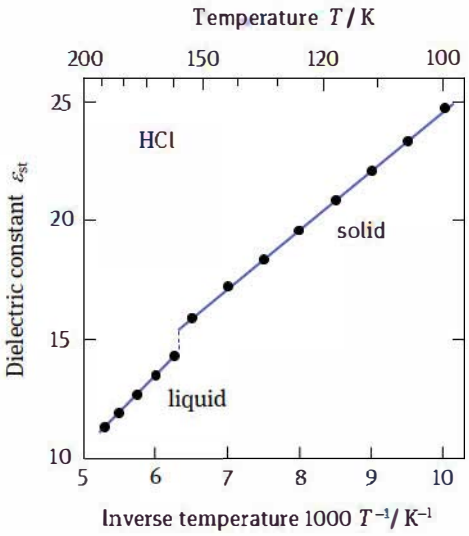


Fig. 13.15: Static dielectric constant ε_{st} of HCl as a function of the inverse temperature. An abrupt change occurs at the solidification. (After R.W. Swenson, R.H. Cole, J. Chem. Phys. 22, 284 (1954).)

Cooling crystals with permanent dipoles down to very low temperatures the quantum mechanical nature of the dipoles come into play. As we have already seen in Section 6.5, amorphous solids have tunneling systems which largely determine the low-temperature properties of this class of substances. Tunneling systems are also frequently found in crystals with point defects, whose dielectric behavior we will examine in more detail. Lithium atoms, which are incorporated as substitutional impurity atoms in potassium chloride crystals, serve as an example. It turns out that the energetically most favorable position for the Li^+-ions, whose diameter is less than half of the potassium ions, is not at the center of the K^+-lattice site, but somewhat shifted to so-called "off-center positions" in one of the ⟨111⟩-directions. The lithium ions can tunnel between these eight energetically equivalent positions. This leads to a system with 8 eigenstates, the term diagram of which is sketched in Figure 13.16. The separation of the levels is given by the *tunnel splitting* Δ_0. In total, only four different energy eigenvalues occur, since the two middle levels are both threefold degenerated. With off-center positions

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

of the Li^+-ions, a comparatively large dipole moment of about $p = 8.7 \times 10^{-30}$ Asm is associated. The resulting polarization in an electric field can be calculated with a little bit of computing effort using thermodynamics from the (negative) derivative of the free energy F with respect to the electric field. For $p\mathcal{E} \ll \Delta_0$ an expression is obtained which is very similar to the orientation polarization (13.48) of classical dipoles, namely

$$\chi = \frac{P_0}{\varepsilon_0 \mathcal{E}} = -\frac{1}{\varepsilon_0 \mathcal{E}} \frac{\partial F}{\partial \mathcal{E}} = \frac{2}{3} \frac{np^2}{\varepsilon_0 \Delta_0} \tanh\left(\frac{\Delta_0}{2k_B T}\right) . \tag{13.50}$$

Interestingly, this result is also obtained for the susceptibility of simple two-level tunneling systems with tunnel splitting Δ_0. This is a consequence of the high symmetry of the defect potential in cubic crystal KCl. For "high" temperatures $T > \Delta_0/k_B$, we may expand the hyperbolic tangent and equation (13.50) changes to the classical Langevin-Debye equation (13.47). The susceptibility in this range is proportional to T^{-1} and independent of the tunnel splitting.

Figure 13.16 shows the temperature dependence of the dielectric susceptibility χ_{Li} caused by only 6 ppm ^{6}Li or 4 ppm ^{7}Li in KCl crystals. First of all, it should be noted that the expected *isotope effect* occurs. The tunnel splitting is larger for ^{6}Li than for ^{7}Li, since the tunneling probability and thus the tunnel splitting increases with decreasing mass of the tunneling particles. The curve of the lighter isotope is therefore slightly shifted towards higher temperatures. As predicted, the overall effect increases proportionally to $1/\Delta_0$. The solid lines show the theoretical prediction (13.50) taking into account the different tunnel splitting of the two isotopes. The agreement between theory and experiment is almost perfect for both crystals, although the theory contains no free parameter. It is remarkable that the very small lithium concentration increases the dielectric constant of the sample at low temperatures by about 1%.

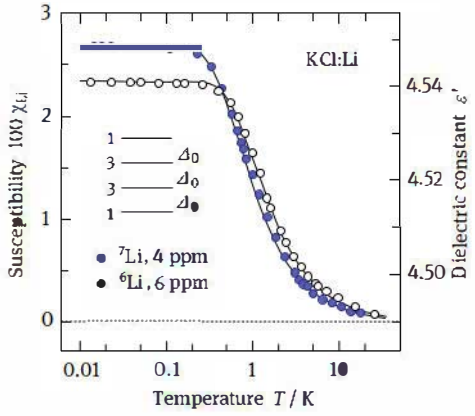


Fig. 13.16: Susceptibility χ_{Li} of Li-doped KCl crystals as a function of temperature. The solid curves show the theoretical predictions (13.50). The term diagram of the lithium defects and the degree of degeneration of the levels are also shown. The scale on the right side refers to the dielectric constant of the *entire sample*. (After C. Enss, Physica B**219 &220**, 239 (1996).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Relaxation Effects. After this short excursus on quantum effects at very low temperatures we turn back to classical dipoles. Interesting phenomena occur when an alternating field is applied instead of a static field, because then the dynamics of the reorientation processes influence the response of the system. According to equation (13.48), the static equilibrium polarization P_0 increases proportionally to the field. Since the orientation of the dipoles and thus the energy difference δE in the alternating field varies periodically, the population of the two energy levels changes constantly. The process approaching the equilibrium state, which in a similar form plays an important role in many areas of physics, is called **relaxation**. We have already seen a special case in connection with the electrical conductivity in Section 9.2. We use the relaxation equation (9.29) and instead of the distribution function $f(E, T)$ we formulate it in terms of the polarization P as the relaxation quantity:

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = -\frac{P(t) - \hat{P}(t)}{\tau} \,. \tag{13.51}$$

Here τ stands for the **relaxation time** and $\hat{P} = np^2\mathcal{E}/k_BT$ is the instantaneous equilibrium value of the polarization, which is determined by the momentary electric field strength and which would occur if the system had any amount of time. If, for example, an electric field is suddenly applied, an exponential increase of the polarization is expected until the new equilibrium value $\hat{P}$ is reached: $P(t) = \hat{P}[1 - \exp(-t/\tau)]$. The time τ is the average time that elapses between two reorientations of the dipoles which occurs via thermally activated processes. As shown schematically in Figure 13.14, a jump over the potential barrier of height V_a is required. Since formally the same situation exist as for diffusion of vacancies, we can equate the jump rate ν introduced in Section 5.1 with the relaxation rate τ^{-1} and find the expression

$$\tau = \tau_0 \, \mathrm{e}^{V_a/k_BT} \,. \tag{13.52}$$

For the *attempt frequency* τ_0^{-1} we use the Debye frequency ν_D, because the molecules vibrate with the frequency ν_D in the respective potential wells.

If we apply the alternating field $\mathcal{E}(t) = \mathcal{E}_0 \exp(-\mathrm{i}\omega t)$, the polarization will show the same periodic variation. This also applies to the instantaneous equilibrium value $\hat{P}$, since the equilibrium is also periodically modulated. For the polarization the ansatz $P(t) = P(\omega) \exp(-\mathrm{i}\omega t)$ and for the instantaneous equilibrium value the ansatz $\hat{P} = \hat{P}(0) \exp(-\mathrm{i}\omega t)$ is suitable. If we use these in (13.51) and consider the relationship $P(\omega) = \varepsilon_0 \chi(\omega) \mathcal{E}(t)$ between polarization and susceptibility, we obtain for the dipolar susceptibility

$$\chi_d(\omega) = \frac{\chi_d(0)}{1 - \mathrm{i}\omega\tau} \,. \tag{13.53}$$

To get the dielectric function we have to add the contributions of the ions and electrons: $\varepsilon(\omega) = [1 + \chi_i(\omega) + \chi_e(\omega) + \chi_d(\omega)]$. Since the contribution of the permanent dipoles is only important up to frequencies in the microwave range, we can assume χ_e and χ_i as

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

constant and obtain for microwave frequencies

$$\varepsilon(\omega) = 1 + \chi_i + \chi_e + \frac{\chi_d(0)}{1 - i\omega\tau} = \varepsilon_\infty + \frac{\varepsilon_{st} - \varepsilon_\infty}{1 - i\omega\tau} . \tag{13.54}$$

The limit values ε_∞ and ε_{st} are as defined in the previous section, albeit with modified cut-off frequencies. Instead of the resonance denominator, as it appears in equation (13.39), the typical relaxation denominator occurs. If we split the dielectric function again into the real and imaginary part, we obtain the two equations:

$$\varepsilon'(\omega) = \varepsilon_\infty + \frac{\varepsilon_{st} - \varepsilon_\infty}{1 + \omega^2\tau^2} , \tag{13.55}$$

$$\varepsilon''(\omega) = \frac{(\varepsilon_{st} - \varepsilon_\infty)\omega\tau}{1 + \omega^2\tau^2} . \tag{13.56}$$

In Figure 13.17 the two quantities ε' and ε'' are shown as a function of $\omega\tau$. The real part decreases steadily with $\omega\tau$ and does not exhibit a resonance behavior. Similar to a resonance, the imaginary part passes through a maximum which is, however, comparatively broad. The steepest drop of ε' and the maximum of ε'' occur at $\omega\tau = 1$, i.e. if the period of oscillation and the relaxation time are comparable. As can be seen from equation (13.56) or Figure 13.17, ε'' only decreases by a factor of 5 if the measuring frequency (or the relaxation time) deviates from the condition $\omega\tau = 1$ by one decade.

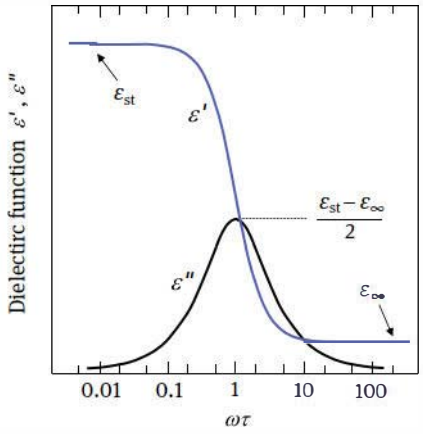


Fig. 13.17: Course of the real part (blue) and the imaginary part (black) of the dielectric function in case of relaxation processes as a function of $\omega\tau$. The scale of the x-axis is logarithmic, the y-axis is linear.

As an example, Figure 13.18 shows dielectric measurements on cesium cyanide. As mentioned in Section 3.2 and schematically indicated in Figure 3.10b, the CN^- ions are statistically oriented along the ⟨111⟩-directions. They carry an electric dipole moment of about 1.1×10^{-30} Asm pointing in one of these directions. Because of this disorder, such systems are often referred to as *orientational glass*. As shown in Figure 13.14, the ions have to overcome an energy barrier during reorientation, the height of which is about 0.14 eV. For the *loss angle* $\tan\delta = \varepsilon''/\varepsilon'$ depicted in Figure 13.18a one finds good qualitative agreement with the theoretical concepts that have been developed.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

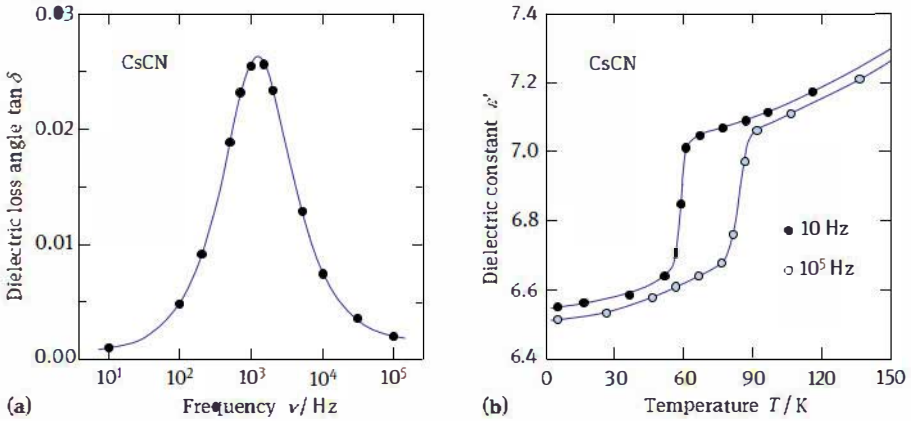


Fig. 13.18: **a)** Frequency dependence of the dielectric loss angle $\tan\delta = \varepsilon''/\varepsilon'$ of CsCN. **b)** Temperature dependence of the real part ε' of CsCN at 10 Hz and 10^5 Hz. (After J. Ortiz-Lopez et al, phys. stat. sol. (b) **199**, 245 (1997).)

A detailed analysis shows that the curve in Figure 13.18a is not completely symmetrical and is somewhat wider than the simple theory would suggest. Such deviations are common, since a distribution of relaxation times occurs due to crystal defects and the interaction of the electric dipoles with each other. From the position of the maximum of the imaginary part, the mean relaxation time can be determined directly via the condition $\omega\tau = 1$.

In many dielectric experiments the frequency is kept constant and the temperature is changed. Since according to equation (13.52) the relaxation time is strongly dependent on temperature, the product $\omega\tau$ can be varied over a wide range. Such measurements can be seen in Figure 13.18b, which shows the results of a measurement on CsCN at 10 Hz and 100 kHz. Clearly recognizable is the "step", which is caused by the relaxation of the CN ions. Please note that $\omega\tau$ decreases with increasing temperatures. Superimposed on the relaxation data is an overall increase of ε' with temperature, which is not caused by the dipole reorientation. It is also observed in CsBr, which exhibits no ions with dipole moment.

We will look at another example, namely the dielectric behavior of glycerol at the glass transition. Glycerol with a melting point of 291 K can easily be supercooled and thus be transformed into the glass state. The glass temperature is 188 K. Figure 13.19 shows the frequency dependence of the dielectric function during cooling at 227 K. A distinct absorption maximum occurs, which is also wider here than predicted by (13.56). Carrying out this measurement at different temperatures, one finds that the curve always looks very similar, but the maximum shifts with decreasing temperature to smaller frequencies. This is a result of the slower relaxation in the supercooled melt. Whereas at 290 K the maximum is observed at 10^8 Hz, it occurs at 10^{-4} Hz when the measurement is carried out at 184 K, i.e. below the glass transition. The dielectric loss

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

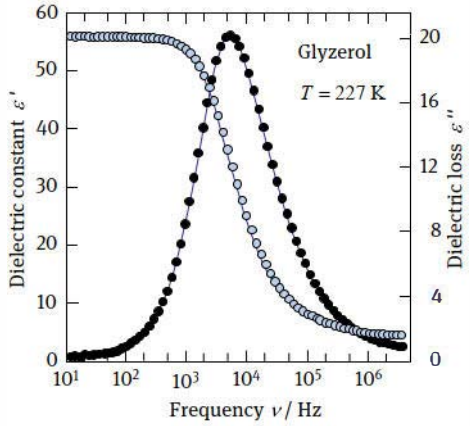


Fig. 13.19: Dielectric function of glycerol at 227 K as a function of frequency. The real part is drawn in light blue, the imaginary part in black. (After G. Kasper, A. Reiser, private communication.)

shown here is typical for the glass transition and is caused by structural rearrangements in the supercooled liquid phase, the so-called *α-process*. In addition, it should be noted that broad band *dielectric spectroscopy*, in which dielectric measurements are performed over a wide frequency range, plays a central role in the investigation of the still largely unknown processes at the glass transition.

To conclude the discussion of the relaxation effects, we remark on the frequency range within which these effects play an important role. The step-like transition of ε' and the maximum of ε'' occur when the condition $\omega\tau = 1$ is fulfilled. Since the relaxation time τ according to (13.52) depends strongly on temperature, the upper frequency, up to which relaxation effects are important, is mainly a question of the measuring temperature. Conversely, the experimentally available frequency range determines the temperature at which the investigations can be carried out successfully.

13.3.7 Ferroelectricity

In Section 12.3 we have learned about the ferromagnetic solids for which the occurrence of a spontaneous magnetization is characteristic. **Pyroelectric materials** show a similar behavior, where a *spontaneous electrical polarisation* already occurs without an external field. In general, however, the polarization is not noticeable, since the resulting surface charge is compensated by charges from the surrounding. However, it can be easily detected when the temperature of the sample changes.

Spontaneous polarization can only occur if the crystal has *one* polar axis. If there are several, the crystal is just **piezoelectric**, i.e. a mechanical deformation causes an electrical polarization. Since the absence of structural inversion symmetry is a prerequisite for the existence of polar axes, amorphous substances are not piezoelectric. Crystals are called ferroelectric if the spontaneous polarization can be reversed by a sufficiently strong field in the direction opposite to the polarization. There are, however,

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

pyroelectric crystals such as $LiNbO_3$ or $LiTaO_3$, which are of great technical importance, in which the energy barrier between the two opposing spontaneous polarizations is so large that the field required for reversing the polarization is greater than the field strength at which internal electrical breakdown occurs. These materials are therefore not classified as ferroelectrics. Similar to ferromagnetism, there are also substances that have *ferri-* and *antiferroelectric* properties.

If ferroelectric crystals are heated, the polarization disappears at the critical temperature T_c. Above this temperature the crystals are *paraelectric* and have a static dielectric constant which can be described by the *Curie-Weiss law* (12.25):

$$\varepsilon_{st} = \frac{C}{T - \Theta} \,. \tag{13.57}$$

Here C is a material-specific constant and Θ is the paraelectric Curie temperature. Figure 13.20 shows the temperature dependence of the static dielectric constant ε_{st} of $BaTiO_3$, which exhibits the expected behavior. The extremely high value of this constant is remarkable.

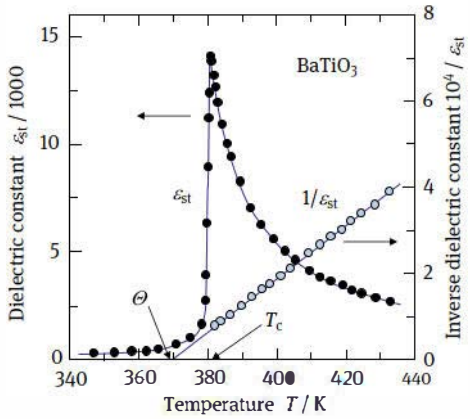


Fig. 13.20: Static dielectric constant ε_{st} of $BaTiO_3$ measured in the direction of the tetragonal c-axis. Additionally, the temperature dependence of ε_{st}^{-1} in the paraelectric phase is shown to illustrate the agreement with the Curie-Weiss law. (After W.J. Merz, Phys. Rev. **91**, 513 (1953).)

The naming indicates the great phenomenological similarity with ferromagnets, which goes beyond the effects discussed here. Therefore it is tempting to describe ferroelectrics like ferromagnets and to replace the magnetic dipoles by permanent electric ones in this theory. However, this approach is doomed to failure because the two phenomena are based on quite different microscopic causes.

The transition from the para- to the ferroelectric phase shows, compared to ferromagnetic transitions, a relatively complex behavior. While ferromagnetic transitions are always second order phase transitions, the ferroelectric ones can be first or second order depending on the substance. The nature of the transition relates to the order parameter characteristic (cf. Section 5.4) at T_c which either exhibits a jump or a continuous variation. We can use the spontaneous polarization which disappears above T_c and increases below T_c with decreasing temperature as order parameter.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

As an example of the behavior of ferroelectrics, Figure 13.21 shows the temperature dependence of the spontaneous polarization P_s of $BaTiO_3$, whose transition on cooling is accompanied by a structural phase transition. This is a first order transition from a cubic to a tetragonal phase. During cooling, the polarization jumps to a finite value at the transition temperature T_c and then increases further. As can be seen in the figure, in the case of $BaTiO_3$ at lower temperatures, two further jumps occur due to changes in direction of the polar axis caused by further structural changes. In addition, hysteresis effects are observed, which are typical for first order phase transitions, hut which we will not go into further here.

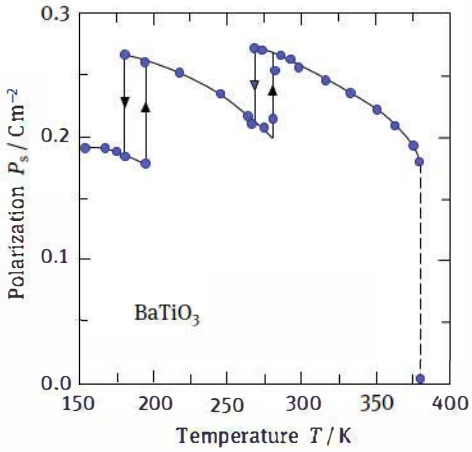


Fig. 13.21: Spontaneous polarization of $BaTiO_3$. At $T_c = 383$ K a discontinuous change of polarization occurs. The jumps at lower temperatures are a consequence of further structural phase transitions. (After W.J. Merz, Phys. Rev. **91**, 513 (1953); H.H. Wieder, Phys. Rev **99**, 1161 (1955).)

We will discuss briefly two mechanisms that lead to the formation of spontaneous polarization in ferroelectrics. The first category includes substances in which the ferroelectric transition is accompanied by an order-disorder transition (see Section 5.4). This means that above the phase transition, the regular arrangement of the electrical dipole moments is lost, but the dipoles themselves do not disappear. This substance class includes the ferroelectric potassium dihydrogen phosphate (KH_2PO_4), in which protons in hydrogen bonds (cf. Section 2.6) between the negatively charged PO_4 ions cause the dipole moments. As with ice, the hydrogen bond is associated with a dipole moment, the strength and direction of which are determined by the position of the proton relative to the bonding partners. Above the critical temperature of 123 K, the protons statistically occupy one of the two possible equilibrium positions between the adjacent PO_4 ions, whereby the mean value of the polarization disappears. When the temperature drops below the transition temperature, a certain orientation is preferred, so that a macroscopic polarization develops.

Table 13.3 shows for a number of ferroelectric crystals the Curie temperature T_c, the spontaneous polarization P_s and the temperature T_m at which the polarization was measured.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Tab. 13.3: Ferroelectric crystals. The Curie temperature T_c, the spontaneous polarization P_s and the measuring temperature T_m of some ferroelectrics are listed. (Various sources.)

	$BaTiO_3$	$PbTiO_3$	$LiTaO_3$	$LiNbO_3$	KH_2PO_4	KD_2PO_4
Curie temperature T_c (K)	383	765	883	1430	123	213
Polarization P_s (C/m^2)	0.26	0.30	0.50	0.71	0.047	0.048
Measuring temperature T_m (K)	300	300	300	300	100	180

In the second group of ferroelectrics, two sublattices shift against each other at the transition into the ferroelectric phase. Instead of ordering dipole moments that are already present they are generated by shifting ions. Typical representatives of this kind of ferroelectrics are ionic crystals with cubic perovskite structure,[10] to which $BaTiO_3$ also belongs. The temperature dependence of the dielectric constant and the spontaneous polarization near the phase transition has already been shown in Figures 13.20 and 13.21. At the critical temperature $T_c = 388\,K$, which is often somewhat reduced by impurities, the negatively charged oxygen ions are displaced permanently by 0.1 Å against the positive metal ions. This leads to the formation of a dipole moment of about 2×10^{-29} Asm per unit cell.

Displacement transitions can be described with the Clausius-Mossotti or Lyddane-Sachs-Teller relation. Let us first consider the Clausius-Mossotti relation (13.21), which we resolve for the static dielectric constant ε_{st}:

$$\varepsilon_{st} = \frac{1 + \frac{2}{3}\sum_i n_i\alpha_i}{1 - \frac{1}{3}\sum_i n_i\alpha_i} \,. \tag{13.58}$$

α_i is composed of the contributions of the electronic and the ionic polarizability. The summation takes into account that ferroelectrics consist of different types of ions, each of which contribute to the dielectric constant. The denominator is a consequence of the Lorentz field. In ferroelectrics, high values of $\sum n_i\alpha_i$ are expected because of the high ionic polarization and thus the denominator goes to zero and $\varepsilon_{st} \to \infty$. The increase of ε_{st} can already be caused by the reduction of the temperatures if $\sum n_i\alpha_i$ is already sufficiently close to the critical value. Then a permanent displacement and thus the spontaneous polarization occurs, because the displacing force acting on an ion due to the local electric field increases faster with the displacement than the linear elastic restoring force. This process is called a **polarization catastrophe**. Of course, displacement and polarization remain finite, because non-linear lattice forces additionally counteract the distortion at large displacement.

Since obviously small deviations from the critical value play a crucial role, we approximate $\sum n_i\alpha_i$ by $\frac{1}{3}\sum n_i\alpha_i = (1-\delta)$ and make the simple assumption that the small

10 The mineral perovskite ($CaTiO_3$) has a cubic structure and was described in Section 2.3.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

quantity δ changes proportionally to the reduced temperature, so that $\delta \propto (T - \Theta)$. As a consequence the Curie-Weiss law follows from (13.58):

$$\varepsilon_{st} \propto \frac{1}{(T - \Theta)} . \tag{13.59}$$

This temperature dependence was already visible in Figure 13.20 for $BaTiO_3$ and is found in many substances with a perovskite structure.

The increase of the static dielectric constant and the formation of a spontaneous polarization can be described in displacement ferroelectrics also in other ways. We want to illustrate this consideration, which is also transferable to non-ferroelectric solids, by the example of $SrTiO_3$. This material also has a perovskite structure. The temperature dependence of its dielectric constant is shown in Figure 13.22. Coming from high temperatures, the dielectric constant increases, as known from ferroelectrics, following the Curie-Weiss law. In strontium titanate, however, *no* ferroelectric transition occurs near the Curie temperature of about 35 K. The increase of the dielectric constant slows down and the $1/\varepsilon_{st}$-curve flattens out.

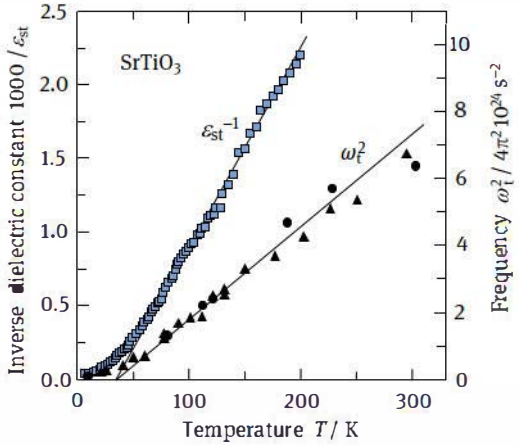


Fig. 13.22: Temperature dependence of ε_{st} and ω_t^2 of $SrTiO_3$. The phonon frequencies were measured by neutron and Raman scattering. (After T. Sakudo, H. Unoki, Phys. Rev. Lett. **26**, 851 (1971); Y. Yamada, G. Shirane, J. Phys. Soc. Japan **26**, 396 (1969).)

The static dielectric constant ε_{st} and the frequency ω_t of the transverse optical phonons, which in turn depends on the electrical polarization according to equation (13.36), are linked via the Lyddane-Sachs-Teller relationship. If we insert the Curie-Weiss law (13.57) into equation (13.43), we get

$$\omega_t^2 = \varepsilon_\infty \varepsilon_{st}^{-1} \omega_\ell^2 \propto (T - \Theta) . \tag{13.60}$$

Because ε_∞ and ω_ℓ are hardly dependent on temperature, ω_t^2 should vary proportionally to the reduced temperature $(T - \Theta)$. This prediction is confirmed by Figure 13.22. The decrease of ω_t during cooling is called *softening* of the transverse lattice vibrations. From this point of view, the spontaneous polarization of ferroelectrics could be described as

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

a "frozen" transverse vibration in which the oppositely charged ions were separated from each other.

13.3.8 Excitons

In this chapter we have so far discussed three mechanisms that cause the absorption of electromagnetic waves in insulators. Apart from special processes in substance classes like ferromagnetic materials, two other important absorption mechanisms exist. These are the absorption by impurities, which we have already learned about in connection with point defects in Section 10.2 and the absorption by excitons, which we will now discuss.

In the fundamental absorption mentioned before, each absorbed photon causes an electron to be lifted into the conduction band, leaving a hole in the valence band. So far we have treated electrons in the conduction band and holes in the valence band as independent particles. This changes, however, if the energy of the absorbed photons is slightly smaller than that of the band gap. Then a special, electrically neutral excitation occurs where the electron and the hole are not separated from one another but due to the Coulomb attraction form one entity called an **exciton**. A prerequisite for the generation of this excitation is that both charge carriers move at the same group velocity and therefore do not spatially separate. Since in optical absorption the transition of electrons from the valence band to the conduction band is almost vertical, this prerequisite for the formation of excitons is only fulfilled if the group velocity of the two types of charge carriers is zero. This is the case at the critical points (see Section 8.5) of the bands. Existing electron-hole pairs can be broken up by thermal energy or additional irradiation by electromagnetic waves (photoionization). Then free charge carriers with the above-mentioned properties are created. Of course an exciton can also decay by recombination. The hole is filled up by the recombining electron. In this process, either luminescence radiation with a frequency corresponding to the energy difference is emitted, or the recombination takes place without radiation by transferring the excitation energy directly to the lattice.

A distinction is made between the strongly and the weakly bound excitons. *Strongly bound* or **Frenkel excitons** mainly occur in molecular, noble gas or ionic crystals. Due to the relatively strong Coulomb interaction between an electron and a hole, the binding energy of this type of exciton is comparatively large, namely about 1 eV or even larger. Thus, the energies for exciton excitation and band-band transitions differ significantly, so that the Frenkel excitons can be investigated very well with optical methods, e.g. by measuring the optical absorption. In most excitons, the hole and the electron are located on one and the same atom or molecule and can therefore also be regarded as the excited state of a molecule. Excitons have the ability to move from atom to atom and thus diffuse in the crystal and transport their excitation energy through the solid.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Figure 13.23 shows the optical absorption of solid krypton, which has an energy gap of 11.7 eV. Even at significantly lower photon energies, distinct absorption maxima are observed, which are associated with the generation of excitons. Since the first maximum is found at 10.2 eV, the ground state energy of the exciton has a relatively high value of 1.5 eV.

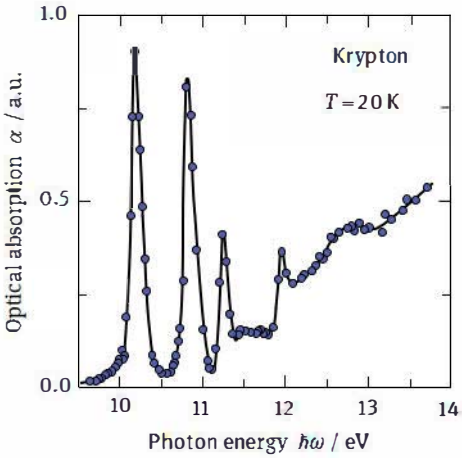


Fig. 13.23: Absorption spectrum of solid krypton. (After G. Baldini, Phys. Rev. **128**, 1562 (1962).)

Weakly bound or **Wannier-Mott excitons** are found in materials with a small band gap; they are therefore typical of semiconductors. Since the binding energy of the excitons must be smaller than the band gap, it follows from the uncertainty relation that the spatial distance between electron and the hole is relatively large. In germanium, for the example, the binding energy of the excitons is about 4 meV, the electron-hole spacing about 10 nm. The eigenstates of excitons can be represented in simplest approximation by a hydrogen atom model with a modified Rydberg formula:

$$E_\nu = E_g - \frac{\mu^* e^4}{32\pi^2\hbar^2\varepsilon_r^2\varepsilon_0^2}\,\frac{1}{\nu^2} + \frac{\hbar^2 K^2}{2(m_n^* + m_p^*)}\,. \tag{13.61}$$

Here E_g is the energy of the band gap, ν the main quantum number and μ^* the reduced mass given by $\mu^{*-1} = m_n^{*-1} + m_p^{*-1}$. The third term is the translational energy of the exciton with the wave vector **K**.

The strong influence of excitons on the optical absorption of semiconductors will be demonstrated by two examples, namely GaAs and Cu_2O. Figure 13.24a shows a measurement of the optical absorption of GaAs at low temperatures. The maximum which precedes the fundamental absorption and is caused by the generation of excitons can be clearly seen. However, the individual lines, which are expected from equation (13.61), cannot be resolved in this material. At room temperature, the thermal energy is greater than the ionization energy of the excitons and the optical transition

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

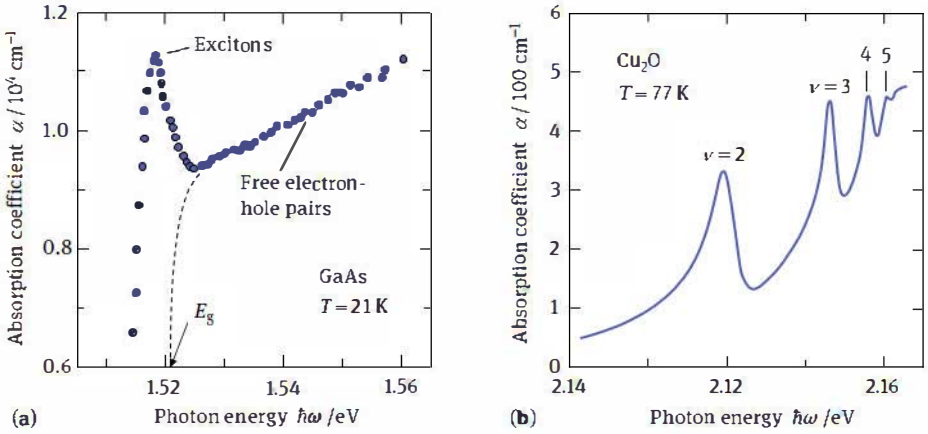


Fig. 13.24: Influence of excitons on the optical absorption of semiconductors. **a)** Absorption edge of GaAs, measured at 21 K. Even before the onset of fundamental absorption, a maximum occurs which can be assigned to excitons. (After M.D. Sturge, Phys. Rev. **127**, 768 (1962).) **b)** Optical absorption of Cu_2O at 77 K as a function of the irradiated photon energy. The fundamental absorption is preceded by a series of exciton lines. (After P.W. Baumeister, Phys. Rev. **121**, 359 (1961).)

occurs with the participation of phonons directly into the conduction band and the maximum disappears.

A substance on which the occurrence of individual exciton lines can be demonstrated particularly well is Cu_2O. In the measurement of the absorption coefficient shown in Figure 13.24b, the transitions to the exciton levels with the quantum numbers $\nu = 2$ to $\nu = 5$ are clearly visible. In experiments at helium temperature even states up to $\nu = 11$ could be observed. The transition to the ground state is missing in the picture. It would not be visible even if the energy axis was continued towards smaller values. This excitation is forbidden as electric dipole transition and occurs as quadrupole transition, which however gives rise to a much weaker absorption.

13.4 Optical Properties of Free Charge Carriers

In this section, we deal with the motion of electrons subject to electromagnetic waves in the partially filled bands of metals and in heavily doped semiconductors. In a sense, this is a continuation of the discussion we had on the electrical conductivity of metals in Section 9.2. As mentioned at the beginning of Section 13.2, intraband transitions are possible in partially occupied bands. These can be classically described as the acceleration of the conduction electrons by the electric field of the incident radiation.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

In deriving the related dielectric function, we assume a one-dimensional motion of a quasi-free conduction electron in a periodic field. The equation of motion is

$$m^*\ddot{u} + \frac{m^*\dot{u}}{\tau} = -e\mathcal{E}(t) \,. \tag{13.62}$$

It has the form of equation (13.31), if one takes into account that $\omega_0 = 0$ can be set, because in the free electron gas there are no restoring forces. Instead of the damping constant γ, which was used in (13.31), we have introduced the mean scattering time τ, which is derived from the electrical DC conductivity $\sigma_0 = ne^2\tau/m^*$. For a periodic field $\mathcal{E} = \mathcal{E}_0\exp[-i(\omega t - kx)]$ the solution has the simple form

$$\varepsilon(\omega) = 1 + n\alpha - \frac{ne^2}{\varepsilon_0 m^*}\,\frac{1}{\omega^2 + i\omega/\tau} \,. \tag{13.63}$$

We use the abbreviation $\varepsilon_\infty = (1 + n\alpha)$, which contains the contribution of the bound electrons, and introduce the **plasma frequency**

$$\omega_p^2 = \frac{ne^2}{\varepsilon_0\varepsilon_\infty m^*} \,, \tag{13.64}$$

whose meaning we will learn more about in the following. Now we separate the real and imaginary part and get

$$\varepsilon'(\omega) = \varepsilon_\infty\left(1 - \frac{\omega_p^2\tau^2}{1 + \omega^2\tau^2}\right) \,, \tag{13.65}$$

$$\varepsilon''(\omega) = \varepsilon_\infty\frac{\omega_p^2\tau}{\omega(1 + \omega^2\tau^2)} \,. \tag{13.66}$$

Since the average scattering time of electrons in metals of medium purity at room temperature is greater than 10^{-14} s, the condition $\omega\tau \gg 1$ applies, so that attenuation at optical frequencies does not play a significant role. For this case we can therefore set $\varepsilon''(\omega) \approx 0$ and for (13.65) we obtain

$$\varepsilon(\omega) = \varepsilon_\infty\left(1 - \frac{\omega_p^2}{\omega^2}\right) \,. \tag{13.67}$$

We will use this result in the next section to discuss the optical properties of metals and heavily doped semiconductors.

At low frequencies, i.e. in the infrared, the simplification made above no longer applies. In this frequency range we have $\omega\tau \ll 1$ and $\sigma_0 \gg \varepsilon_0\varepsilon_\infty\omega$. After a short calculation (see Problem 5 at the end of this chapter), we find in this case the **Hagen-Rubens relation**[11,12] for the refractive index n' or the extinction coefficient κ

$$n' \approx \kappa \approx \sqrt{\frac{\sigma_0}{2\varepsilon_0\omega}} \tag{13.68}$$

11 Carl Ernst Bessel Hagen, *1851 Königsberg, †1923 Solin near Munich
12 Heinrich Leopold Rubens, *1865 Wiesbaden, †1922 Berlin

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

and for the reflectivity R the expression

$$R \approx 1 - \sqrt{\frac{8\varepsilon_0\omega}{\sigma_0}} \,. \tag{13.69}$$

13.4.1 Electromagnetic Waves in Metals

The very simple dielectric function (13.65) already allows a broad understanding of the optical properties of metals. We start from the wave equation (13.44) for the electromagnetic waves and again use a plane wave as the solution. For the electric field strength of the plane wave we write $\mathcal{E} = \mathcal{E}_0 \exp[-i(\omega t - kx)]$. The choice of the x-direction does not limit the validity of the solution, since the free electron gas is isotropic. From that we obtain the well-known relation $\varepsilon(\omega)\,\omega^2 = c^2k^2$ and insert the dielectric function (13.65). Thus we find the dispersion relation for electromagnetic waves in metals

$$\varepsilon_\infty \omega^2 \left(1 - \frac{\omega_p^2}{\omega^2}\right) = c^2k^2 \,. \tag{13.70}$$

Depending on the frequency, two areas with completely different properties of the electron gas occur. Below the plasma frequency, i.e. for $\omega < \omega_p$ we have $k^2 < 0$ and k is therefore imaginary. Long wavelength electromagnetic waves therefore cannot propagate in metals. As we have already seen in the discussion of polar solids, total reflection occurs at $\varepsilon < 0$.

If $\omega > \omega_p$, then $\varepsilon(\omega) > 0$ and the wave equation describes the propagation of an electromagnetic wave called a **plasmon polariton**. The dispersion relation in this case is

$$\omega^2 = \omega_p^2 + \frac{c^2k^2}{\varepsilon_\infty} \,. \tag{13.71}$$

For small wave vectors the group velocity approaches zero and for large ones it approaches the value $v_g = c/\sqrt{\varepsilon_\infty}$, as known from isolators. Figure 13.25 shows the principle course of the dispersion curve. At low frequencies, metals are impermeable to electromagnetic radiation, but above the plasma frequency they are transparent.

Table 13.4 shows the plasma frequency and the corresponding wavelength λ_p for some electron densities under the assumption $\varepsilon_\infty = 1$. If one uses the numerical

Tab. 13.4: Plasma frequency ω_p and plasma wavelength λ_p at different electron densities n.

$n\,(\mathrm{m}^{-3})$	10^{28}	10^{24}	10^{20}	10^{16}	10^{12}
$\omega_p\,(\mathrm{s}^{-1})$	5.7×10^{15}	5.7×10^{13}	5.7×10^{11}	5.7×10^{9}	5.7×10^{7}
$\lambda_p\,(\mathrm{m})$	3.3×10^{-7}	3.3×10^{-5}	3.3×10^{-3}	3.3×10^{-1}	3.3×10^{1}

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

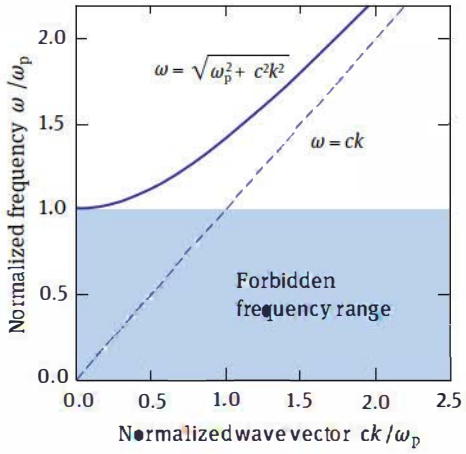


Fig. 13.25: Dispersion of plasmon polaritons in the free electron gas. Below the plasma frequency ω_p a forbidden frequency range occurs and electromagnetic waves cannot propagate there. In this schematic representation $\varepsilon_\infty = 1$ was assumed.

values for sodium in equation (13.64), one finds based on the small electron density that sodium is already transparent in the ultraviolet (see also Table 13.5). The same applies to other alkali metals. For transition metals, the situation is less clear since interband transitions play an important role there, so that the necessary theoretical considerations are somewhat more complex.

A comparison between the expected reflectivity within the model of free electrons and the actual data of aluminum can be seen in Figure 13.26. The curve for aluminum is based on the experimental results and theoretical considerations. A look at the reflective properties of polar crystals shows again the importance of the plasma frequency. In ionic crystals, a very high reflectivity is observed between the two frequencies ω_t and ω_ℓ which drops steeply above ω_ℓ (see Figure 13.12). This drop is connected with the zero crossing of the real part ε' of the dielectric function. For electron gases, the reflectivity disappears above the plasma frequency, where $\varepsilon' = 0$. This suggests that

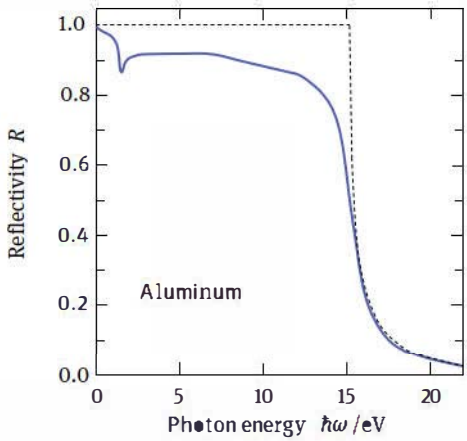


Fig. 13.26: Reflectivity of aluminum. Experimental data are represented by the blue curve and the dashed curve is calculated with equation (13.64) using the $\varepsilon_\infty = 1$ and $\hbar\omega_p = 15.3$ eV. (After H. Ehrenreich et al., Phys. Rev. **132**, 1918 (1963).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

longitudinal oscillations are associated with the plasma frequency. We will focus on the experimental observations of these oscillations in the following subsection. While the reflectivity of polar crystals decreases sharply at low frequencies, beyond the frequency of the transverse optical phonons, it remains very high for metals down to DC fields. The different behavior at low frequencies is due to the lack of shear stiffness of the electron gas, so that $\omega_t = 0$ can be assumed for metals. The "dip" in the reflectivity of aluminum at about 1.5 eV and the reduced reflectivity in the forbidden frequency range compared to the ideal value of $R = 1$ is due to the finite value of ε'' caused by interband transitions.

The dependence of the plasma frequency on the charge carrier concentration can be seen in Figure 13.27 using the example of the reflectivity of the semiconductor InSb. The high doping of the samples causes a high concentration of electrons in the conduction band, which can be treated like a free electron gas. The plasma frequency increases with the electron concentration and the reflection edge is shifted towards shorter wavelengths. This also applies to the minimum reflectivity, which occurs at $\varepsilon = 1$ and thus according to (13.70) at

$$\omega = \omega_p \sqrt{\frac{\varepsilon_\infty}{(\varepsilon_\infty - 1)}} \,. \tag{13.72}$$

The increase in reflectivity towards small wavelengths is due to the relatively large real part of the dielectric constant of indium antimonide at these frequencies. From the position of the minimum, the plasma frequency can be determined and thus with the help of equation (13.64) the effective mass of the electrons. Hence measurements of the reflectivity of semiconductors are often used to determine m^*.

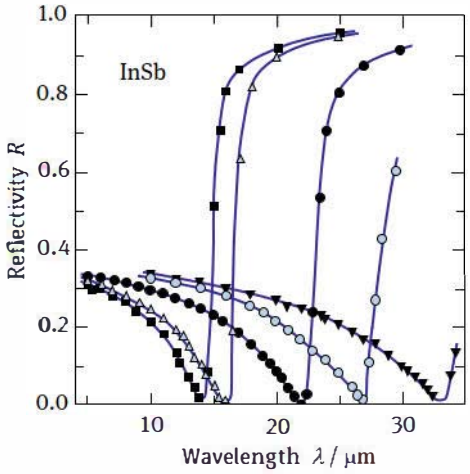


Fig. 13.27: Reflectivity of tellurium-doped indium antimonide at the plasma edge. With increasing doping the edge shifts to smaller wavelengths. The density of the charge carriers range from 3.5×10^{23} m^{-3} to 4.0×10^{24} m^{-3} (After W.G. Spitzer, H.Y. Fan, Phys. Rev. **106**, 882 (1957).)

As an example of how the variation of the plasma frequency by doping can be used to tune the reflection and absorption properties of materials, we will consider the

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

case of tin-doped In_2O_3 layers (better known as ITO for **i**ndium **t**in **o**xide), whose behavior is shown in Figure 13.28. By doping with tin at different concentrations, the absorption edge of In_2O_3 can be adjusted almost at will. In the example shown the plasma wavelength is just under 2 μm or just over 1 μm, depending on the electron density of 5×10^{26} m^{-3} to 1.3×10^{27} m^{-3}. Since in indium oxide the interband transitions only occur above 2.8 eV, the doped layer is optically transparent but is reflective in the infrared. Such layers are used to reduce the heat radiation losses of windows or sodium vapor lamps. In addition, such electrically conductive layers can be used as transparent electrodes, e.g. in liquid crystal displays (LCD) or solar cells.

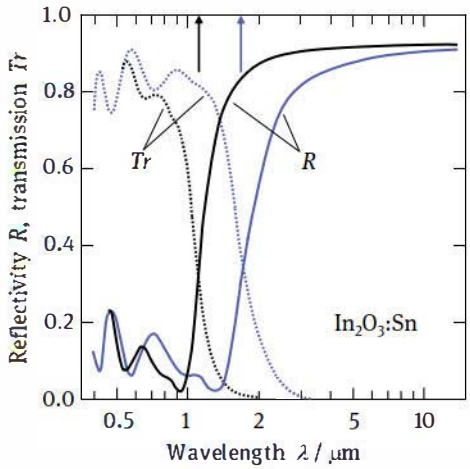


Fig. 13.28: Transmission (*Tr*) and reflectivity (*R*) of two 0.3 μm thick In_2O_3 layers with different tin doping. The arrows indicate the position of the plasma wavelengths for the two samples. The wave-like oscillation at short wavelengths is caused by interferences in the thin layers. (After G. Frank et al., Phys. Bl. **34**, 106 (1978).)

13.4.2 Plasmons

In a gas of free electrons, longitudinal oscillations are possible, but they do not couple to electromagnetic waves. The reason for the missing coupling is the same as for LO phonons: the electric fields of the two oscillations are perpendicular to each other. While there are no restoring forces when free electrons are transversely displaced, longitudinal oscillations do involve forces. In the limiting case $k \to 0$ their strength can be easily understood, because then all electrons oscillate uniformly against all ions. As shown in Fig. 13.29, with a displacement u a surface charge of density $\sigma_e = neu$ builds up at the boundary of the sample, which causes the electric field $\mathcal{E} = neu/\varepsilon_0\varepsilon_\infty$. The factor ε_∞ takes into account the polarizability of the core electrons. Thus the equation of motion of the free electrons without consideration of the attenuation can be written as

$$nm\ddot{u}(t) = -ne\mathcal{E}(t) = -\frac{n^2e^2u(t)}{\varepsilon_0\varepsilon_\infty} \tag{13.73}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

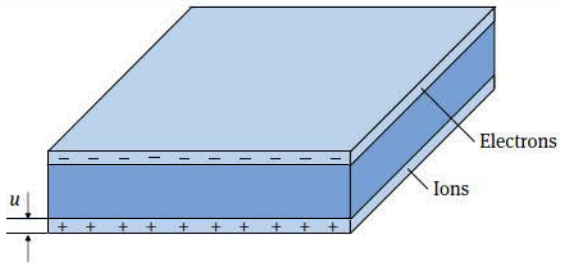


Fig. 13.29: Schematic representation of a plasma oscillation in the limiting case $k \to 0$. All electrons oscillate in phase against the rigid ionic lattice. The displacement u of the electrons generates the surface charge density $\sigma_e = neu$.

and with the plasma frequeny ω_p again given by equation (13.64) we have:

$$\ddot{u} + \omega_p^2 u = 0 \,. \tag{13.74}$$

This is the equation of motion of a harmonic oscillator with the eigenfrequency ω_p. The same oscillation frequency is obtained from the condition that the dielectric function (13.65) disappears at the frequency of the longitudinal eigen-oscillation, i.e. $\varepsilon(\omega_\ell = \omega_p) = 0$. As with the phonon polaritons, the frequency of the longitudinal oscillations represents a lower limit for the propagation of electromagnetic waves.

Similar to phonons, which represent a coherent motion of all atoms of the lattice, the **plasma oscillations** are coherent, collective excitations of all electrons of the Fermi gas. The amplitude of the harmonic oscillator that describes such a plasma oscillation is quantized. The excitations have the energy $\hbar\omega_p$ and are called **plasmons**. In addition, there are of course the one-electron excitations, which are based on the motion of single, independent electrons. Table 13.5 lists the energy and wavelength of plasmons in some metals.

Tab. 13.5: Plasmon energy $\hbar\omega_p$ and plasma wavelength λ_p of selected metals.

	Li	Na	K	Mg	Ag	Au	Cu	Al	Pt
Plasmon energy $\hbar\omega_p$ (eV)	7.12	5.71	3.72	10.6	9.6	8.55	7.39	15.3	5.15
Plasma wavelength λ_p (nm)	174	217	333	117	129	145	168	81.0	241

If one goes to finite wavelengths, i.e. to $k > 0$, the density of electrons is spatially modulated by the plasmons. Due to the gradient in charge density, the restoring forces acting on the electrons increase in this case and the frequency of the plasmons increases with the wave vector. For small values of the wave vector the dispersion relation can be approximated by

$$\omega \approx \omega_p \left(1 + \frac{3v_F^2}{10\omega_p^2} k^2 + \ldots\right) . \tag{13.75}$$

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Experimentally, plasmons can be studied by the reflection of high-energy photons or by electron scattering. For example, plasma oscillations can be excited in metal films by electrons penetrating through the film. Figure 13.30a shows the loss spectrum observed when 20 keV of electrons pass through a 258 nm thick aluminum film. It can be clearly seen that the scattered electrons undergo an energy loss which is an integer multiple of the plasmon energy $\hbar\omega_p \approx 15.3$ eV. This value corresponds to the photon energy at which the edge of the reflectivity of aluminum appears in Figure 13.26.

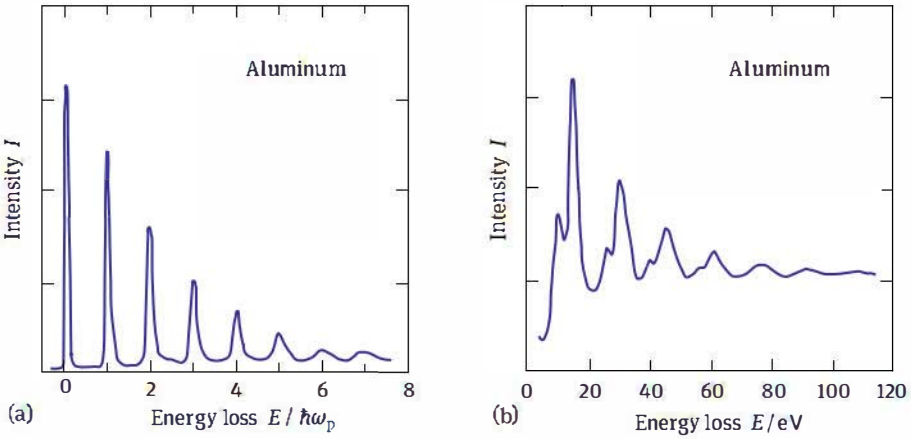


Fig. 13.30: Energy loss of high-energy electrons when passing through thin aluminum foils. **a)** The energy loss of 20 keV electrons is normalized to the plasmon energy $\hbar\omega_p = 15.3$ eV. (After L. Marton et al., Phys. Rev. 126, 182 (1962).) **b)** Energy loss of 2 keV electrons. In this experiment, surface plasmons were observed in addition to volume plasmons. (After C.J. Powell and J.B. Swan, Phys. Rev. 115, 869 (1959).)

In many of such scattering experiments, however, one does not get a clear signal for two reasons. Firstly, intraband transitions often occur at the same energy leading to a strong reduction of the plasmon lifetime and thus a strong broadening of the signal. Secondly, in addition to the **volume plasmons** discussed here, there are also **surface plasmons** in which the collective electron motion is bound to the surface of the sample. These excitations are also stable, but have a lower energy, because the accompanying electric fields partly run in vacuum and do not cause polarization there. In most experiments, both excitations are actually observed in the loss spectra. This can be seen in Figure 13.30b, which shows data also recorded on aluminum, but in this case the energy of the incident electrons was approximately one order of magnitude smaller than in Figure 13.30a.

Finally, we will briefly discuss the use of surface plasmons in sensor technology. In this application, we take advantage of the fact that surface plasmons can be used to detect changes on suitably prepared sample surfaces with extremely high sensitivity.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

The dispersion relation of surface plasmons can be calculated as follows: Staring from Maxwell's equations for plane waves at the interface between two uncharged materials, we obtain

$$\frac{\omega^2}{c^2} = \frac{1}{\varepsilon(\omega)} k^2 . \tag{13.76}$$

Here $\varepsilon(\omega)$ stands for the dielectric function, which in the case of a metal/air interface can be written as $\varepsilon^{-1} = \varepsilon_{\text{metal}}^{-1} + \varepsilon_{\text{air}}^{-1}$. Let us substitute for the dielectric function this expression in the above equation with $\varepsilon_{\text{air}} = 1$ and denote the wave number of the surface plasmons with k_{sp} and we find

$$k_{\text{sp}} = \frac{\omega}{c} \sqrt{\frac{\varepsilon_{\text{metal}}}{\varepsilon_{\text{metal}} + 1}} , \tag{13.77}$$

whereby according to equation (13.65) for the dielectric function of metals, the relation $\varepsilon_{\text{metal}} = \varepsilon_{\infty} \left(1 - \omega_{\text{p}}^2/\omega^2\right)$ holds.

Figure 13.31a shows the dependence of the frequency on the wave number $k_{\text{sp}}(\omega)$. Since $k_{\text{sp}}(\omega)$ does not intersect the curve for the propagation of light in vacuum, the straight line $k_{\text{vac}}(\omega)$, an electromagnetic wave striking a metal film cannot satisfy energy and momentum conservation at the same time and thus cannot excite surface plasmons. However, this is possible for a thin gold or silver film with a thickness of about 50 nm in a certain geometry. As shown in Figure 13.31b, this film is deposited on a glass prism and illuminated by a collimated light beam through the prism. Due to the refractive index of glass of about 1.5, the speed of light in glass is lower than in vacuum, i.e. the wave vector is larger. If a suitable angle of incidence is chosen, the horizontal component $k_x(\alpha, \omega) = n(\omega/c) \sin\alpha$ of the wave vector corresponds to

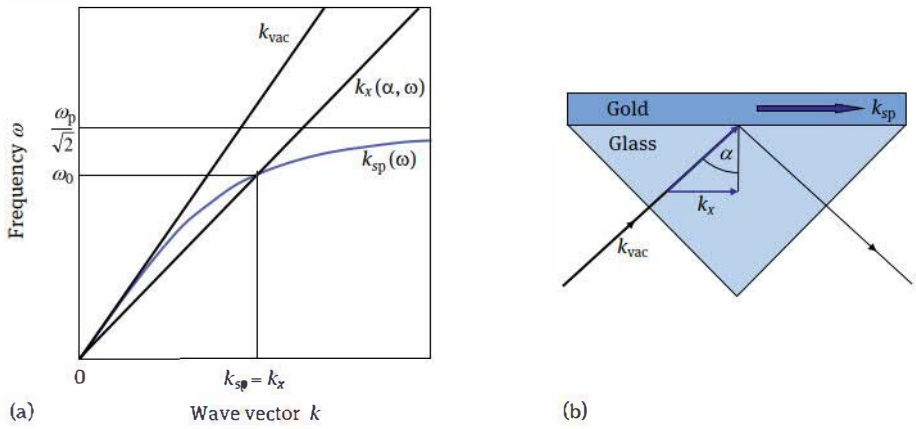


Fig. 13.31: Surface plasmons. **a)** The dispersion curve of the surface plasmons is drawn in blue. **b)** Kretschmann geometry for the detection of surface plasmons. By the excitation of surface plasmons (thick arrow), energy is very effectively extracted from the incident light. (After V. Temnov and U. Woggon, Physik Journal **9**, 45 (2010).)

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

the wave vector k_{sp} of the surface plasmons, i.e. $k_x(\alpha, \omega_0) = k_{sp}(\omega_0)$. This condition is fulfilled at the intersection of the two curves in Figure 13.31a, so that surface plasmons can be generated at a specific angle.

The *Kretschmann geometry* sketched in Figure 13.31b is often used in experiments. Since the thickness of the metal film is smaller than the wavelength of light, the electric field also runs within the film. If the condition $k_x = k_{sp}$ is fulfilled under a certain angle, a very effective excitation of surface plasmons occurs. This means that the conduction electrons in the metal film very effectively absorb energy from the light, so that the intensity of the reflected light decreases strongly. For the technical application of this effect it is crucial that the plasmon generation only takes place in a very narrow angular range $\Delta\alpha$. Small changes in plasmon frequency caused by external influences therefore cause strong changes in the intensity of the reflected light. It follows that molecules adsorbed to the metal layer have a strong influence on the reflected light intensity and can therefore be detected with high sensitivity. Surface plasmons are therefore used in biosensor technology, among other things. Since for the detection of biological reactions often only small amounts of material are available, a high detection sensitivity is a prerequisite for the successful application of this technique.

13.5 Exercises and Problems

1. Polarizability. A KCl-crystal (density $\varrho = 1980\,\mathrm{kg/m^3}$) is subjected to a static electric field with a strength of $\mathcal{E} = 1\,\mathrm{kV/m}$ and an electromagnetic field in the optical frequency range with $\mathcal{E} = 1\,\mathrm{MV/m}$. Calculate the polarizability and the moment of the generated dipoles.

2. Dipole Moment. Assume a 1 cm thick germanium disc with density $\varrho = 5320\,\mathrm{kg/m^3}$ and the dielectric constant $\varepsilon = 16.6$ between two capacitor plates.
(a) Determine the polarizability of the germanium atoms.
(b) Calculate the local field when a voltage of 50 V is applied between the plates.
(c) Calculate the dipole moment of the germanium atoms.

3. Optical Phonons. The static dielectric constant of NaCl is $\varepsilon_{st} = 5.9$, the refractive index $n' = 1.55$. The reflectivity R has a minimum at $\lambda = 30.6\,\mu\mathrm{m}$. Calculate the frequency of the optical phonons. Assume that the absorption can be neglected and that at minimum $\varepsilon' \approx 1$.

4. Dipole Moment of Hydrogen Chloride. Figure 13.15 shows the temperature dependence of the static dielectric constant of liquid and solid hydrogen chloride. Use this figure and the value of the density $\varrho(T = 98\,\mathrm{K}) = 1.48\,\mathrm{g/cm^3}$ to determine the dipole moment of HCl.

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

5. Hagen-Rubens Law. In the long-wavelength infrared range, most metals fulfill the limiting conditions $\omega\tau \ll 1$ and $\sigma_0 \gg \varepsilon_0\omega$, where σ_0 stands for direct current conductivity.

(a) Calculate the refractive index n' and the extinction coefficient κ for this limiting case.

(b) Show that the reflectivity in this limiting case can be described by the Hagen-Rubens law (13.69).

(c) Determine the reflectivity of silver (density $\varrho_{Ag} = 10.49\,\mathrm{kg/cm^3}$) and aluminum (density $\varrho_{Ag} = 2.70\,\mathrm{kg/cm^3}$) at a wavelength of 200μm.

(d) Check whether the requirements for the Hagen-Rubens Law are fulfilled.

6. Plasma Edge. Indium-doped tin oxide (ITO) is a technologically important material because of its transparency in the visible wavelength range and its high electrical conductivity.

(a) Up to which wavelength is an ITO layer with 4×10^{27} free charge carriers per $\mathrm{m^3}$ transparent? ($m^* = m$, $\varepsilon_\infty = 3.84$)

(b) At which wavelength is the reflectivity minimum?

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Index

Abrikosov lattice, 484
Absorption,
– optical,
  – amorphous semiconductors, 401, 405
  – color centers, 128
  – excitons, 574
  – fundamental absorption processes, 550
  – semiconductor, 376
  – solar cells, 426
– ultrasound,
  – amorphous solids, 235
  – crystals, 230
  – superconductor, 462, 493
Acceptors, 388
Acoustic phonons, 177
Activation energy,
– amorphous semiconductors, 404
– charge transport, 137
– diffusion, 133
– table, 127
Adiabatic approximation, 159, 170, 255
Alkali metals, binding energy, 26
Allotropy, 24
Alloy,
– entropy of mixing, 37
– miscibility gap, 36
– production, 36
Alternating current conductivity, 322
Amorphous solids,
– defects, 151, 405
– glass production, 42
– pair correlation function, 72
– semiconductors, 400
– specific heat, 215
– structure, 71
– structure determination, 106
– thermal transport, 250
– ultrasound absorption, 235
– vibrational spectrum, 213
Andreev reflexion, 467
Anharmonicity, 223
Anisotropic properties,
– crystals, 33
– ferromagnetism, 509
– high-temperature superconductors, 487
Antiferromagnetism, 528
Atomic force microscope, 120
Atomic scattering factor, 101
Atomic shape factor, 101
Atomic structure factor, 97, 101
Atomic transport,
– diffusion of vacancies, 132
– interstitial diffusion, 134
– vacancy diffusion, 136
Attempt frequency,
– diffusion, 133

Ballistic
– electrons, 331
– phonons, 239
Band bending,
– p-n junction, 411
Band discontinuity, 420, 434
Band ferromagnetism, 518
Band gap,
– amorphous semiconductors, 401
– quasi-free electrons, 284
– semiconductor, 376
  – table, 376
Band structure,
– metal, 291
– tight-binding model, 284
– calculation, 273
– experimental determination, 296
– insulator, 291
– quasi-free electron model, 277
– semiconductor, 376
– transition metals, 268
– two-dimensional solids, 299
Basis, 48
BCS theory,
– critical current, 469
– critical field, 469
– energy gap, 459
– excitation spectrum, 460
– groundstate, 456
– state at $T \neq 0$, 461
– transition temperature, 461
Belly orbit, 350

https://doi.org/10.1515/9783110666502-014

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Binding energy, 7
– covelent bond, 23
– Ionic crystals, 13
– metals, 26
– table, 7, 29, 127
– Van der Waals crystals, 11
Bipolar transistor, 429
Bloch
– function, 274, 276
– oscillation, 311
– theorem, 276
– wall, 526
Bloch's $T^{3/2}$ law, 526
Bloch-Grüneisen law, 329
Boltzmann
– distribution, 133
– equation, 318
– statistics, 384
Bonding types, 5
Born-Haber cycle, 14
Born-Oppenheimer approximation, 159
Bose-Einstein statistics, 207
Boson peak, 214
Boundary conditions
– fixed, 201
– periodic, 200
Bragg
– condition, 96
– reflexion, 97
Bravais
– lattice, 53
– mesh, 70
Bridgman-Stockbarger technique, 35
Brillouin
– function, 503
– scattering, 194, 197
– zone, 87, 292
– zone, reduction, 277
Buckminsterfullerene, 26
Burgers vector, 142

Carbon nanotubes, 24, 68, 299, 336
Casimir regime, 245
Cesium chloride structure, 62
Chalcogenide glasses, 152
Channels, 331
Charge carrier density,
– doped semiconductors, 391
– intrinsic semiconductors, 383
Charge compensation, 137
Charge neutrality, 127
Charge transport, 320
– amorphous semiconductors, 402
– electron gas, 316
– graphene, 301
– in bands, 312
– ionic conductor, 136
– magnetic field influence, 363
– one-dimensional conductors, 331
Charging energy, 338
Chemical potential, 39, 262
Clausius-Mossottl relation, 547
Clusters, 55
Coherence length, 446, 480, 491
– table, 446
Coherence volume, 491
Color centers, 128
Compensation, 392
Compensation effect, 395
Compressibility, 16
Compression modulus
– table, 17
Condensation energy, 449, 458
Conductance quantum, 332
Conduction band, 291, 376, 387
– graphene, 300
Conduction channel, 331
Conductivity, electrical,
– doped semiconductors, 397
– Hall effect, 361
– ionic crystals, 136
– metals, 316, 322
– metals,
  – table, 343
Conductivity, thermal,
– amorphous solids, 250
– crystal, 240
– metals, 340
– metals,
  – table, 343
Cooper pair, 451
– size, 456
Coordination number, 61, 151
Coordination shell, 74
Correlation effects, 517
Correlation energy, 406
Coulomb
– blockade, 339

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

– oscillations, 339
– potential, screened, 269
Covalent bond, 5, 17, 23
Critical
– current density, 468
– magnetic field, 441, 447, 468, 482, 486
  – table, 438, 485
– shear stress, 140
Crystal classes, 58
Crystal field, 390, 505
Crystal growing, 33
Crystal lattice, 48, 61
Crystal momentum, 184
Crystal plane, 90
Crystal structure,
– direct imaging, 79
Crystal systems,
– table, 52
Cubic lattice
– body-centered cubic lattice, 62
– face-centered cubic lattice, 63
– simple cubic lattice, 61
Cuprate superconductors, 489
Curie law, 502
Curie temperature,
– ferrimagnetic, 528
  – table, 528
– ferromagnetic, 511
– paramagnetic, 513
Curie-Weiss law, 513
Cycle processes, 14
Cyclotron frequency, 347
Cyclotron mass, 347
Cyclotron resonance,
– metals, 346
– semiconductor, 380
Czochralski-Kyropoulos technique, 35

Dangling bond, 151
De Haas-van Alphen effect, 358
Debye
– approximation, 205
– density of states, 206
– frequency, 206
– temperature, 207
  – table, 208
Debye-Scherrer method, 115
Debye-Waller factor, 188
Defect scattering,
– electrons, 322
– phonons, 230
Defect structures,
– amorphous semiconductors, 405
– amorphous solids, 151
– crystals, 123
Degenerate semiconductors, 384
Degree of order, 155
Demagnetization field, 441, 500
Density of states,
– electrons,
  – amorphous semiconductors, 401
  – metal, 296
  – Nanotubes, 303
  – semiconductors, 384
  – superconductor, 460
– free electrons,
  – in magnetic field, 356
  – one-dimensional, 261
  – three-dimensional, 258
  – two-dimensional, 259
– phonons,
  – amorphous solids, 214
  – Debye approximation, 205
  – definition, 202
  – low-dimensional systems, 210
– spin waves, 522
– tunneling systems, 218
Depletion zone, 412, 419
Depolarisation field, 545
Devices, 424
Diamagnetism,
– atomic, 500
– conduction electrons, 501
– ideal, 441
Diamond structure, 63
Dielectric constant, 542
– defect contribution, 564
– table, 556
Dielectric function,
– definition, 542
– electron gas, 549
– ionic crystals, 555
– metals, 576
– orientation polarization, 566
Dielectric properties, 541
Diffraction, 79

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Diffraction experiments
– elastic scattering, 81
Diffusion constant, 134
Diffusion current, p-n junction, 414
Diffusion length, 134, 385
Diffusion voltage, 411, 414
Diffusion,
– interstitials, 134
– vacancies, 132
Dipole moment,
– electrical, 543, 561, 565
– oscillating, 178
Dipole-dipole interaction, 9
Dirac-Weyl equation, 301
Dislocation line, 141
Dislocations, 141
Dispersion curves, experimental, 190
Dispersion relation,
– conduction electrons, 310
– lattice vibrations, 182
– metal, 577
– phonon polariton, 557
– plasmon polariton, 577, 581
– semiconductors, 381
– sound waves, 166
– spin waves, antiferromagnets, 531
– spin waves, ferromagnets, 522
– superconductor, 459
Domains,
– ferromagnetic, 526
– giant magnetoresistance, 532
Dominant phonon approximation, 241
Donors, 387
Doped semiconductors, 387
Doping superlattice, 423
Drift velocity, 316, 361, 400
Drude model, 316, 322
Dulong-Petit law, 199, 208
Dynamic mass, 309
Dynamical matrix, 182

Edge dislocation, 141
Effective electron mass,
– dynamic mass, 309
– semiconductor electrons, 380, 422
  – table, 381
– thermal mass, 267
– tightly bound electrons, 288
Effective mass approximation, 309
Einstein
– model, 199
– specific heat, 210
Einstein relation, 136
Elastic constants, 162
– table, 164
Elastic continuum, 159
Elastic properties, 159
Elastic wave, 165
Elasticity tensor, 162
Electrical conductivity,
– amorphous semiconductors, 402
– Boltzmann equation, 320
– doped semiconductors, 397
– ionic crystals, 136
– metals, 316
  – temperature dependence, 327
Electron affinity, 14
Electron gas,
– density of states, 257
– exchange interaction, 517
– magnetic field, 346
– metallic bond, 26
– optical properties, 575
– screening effects, 269
– specific heat, 265
– table, 264
Electron gas, two-dimensional,
– quantum dot contact, 333
– Quantum Hall effect, 364
– superlattice, 422
Electron mass,
– metals, table, 268
Electron orbit, 350
Electron-defect scattering, 322
Electron-electron
– interaction, 27, 271, 325, 330
Electron-phonon scattering, 323, 342
Electronic polarizability, 548
Electrons in magnetic fields, 346
Elementary cell, 50
Empty lattice, 277
Energy bands, 279, 290
Energy gap, 279, 282, 291
– amorphous semiconductors, 401
– semiconductor, 376
– superconductor, 459
  – experimental evidence, 462
  – table, 462

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Energy gap function, 492
Entropy,
– alloys, 37
– superconductor, 450
Equation of motion,
– electrons, 307
– lattice atoms, 180
Equation of state, 223, 225
Esaki diode, 424
Eutectic point, 41
Eutectic system, 41
Evjen cells, 15
Ewald sphere, 185
Exchange constant, 514, 535
Exchange energy, 528
Exchange hole, 518
Exchange integral, 19
Exchange interaction, 509, 516
– band ferromagnetism, 519
– ferromagnetic insulators, 514
– free electron gas, 517
– Heisenberg model, 515
– magnons, 523
– metallic binding, 27
– RKKY interaction, 516
– spin glasses, 535
– super exchange, 516
Excitons, 573
Expansion coefficient,
–crystals, 225
Extended defects,
– dislocations, 141
– grain boundaries, 149
Extinction coefficient, 543, 559, 576
Extremal orbit, 349
Extrinsic conduction, 393

F-center, 128
Fermi
– energy, 261
– function, 343
– gas, 256, 262
– level, 385
– liquid, 326
– momentum, 263
– sphere, 265, 317
– surface, 265, 292
– temperature, 263
– velocity, 263
– wave vector, 263
Fermi-Dirac statistics, 262
Ferrimagnetism, 527
Ferroelectricity, 568
Ferromagnetism, 509
– band model, 518
– exchange interaction, 514
– Heisenberg model, 515
– meanfield approximation, 510
Field current, 414
Flux line lattice, 484
Flux quantisation, 471
Flux tubes, 484
Fluxon, 472
Forbidden frequency range, 578
Force component, 171
Force constant tensor, 181
Fourier law, 240
Fractional quantum Hall effect, 370
Frank-Read source, 148
Freeze-out regime, 395
Frenkel defect, 131
Frenkel excitons, 573
Frequency gap, 179, 557
Friedels rule, 104
Frustration, 536
Fullerene, 26
Fundamental absorption processes, 550

Γ-point, 88
Gap function, 492
Generation current, 414
Giant magnetoresistance, 532
Ginzburg-Landau
– coherence length, 480, 491
– equations, 480
– parameter, 481
– theory, 479
Glass production, 42
Glass transition, 42
Glass transition temperature, 42
Glasses, 42
Graphene, 24, 299, 371
– band structure, 300
Graphite, 24
Groupvelocity, 174
Grüneisen
– parameter, 224
– relation, 225

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

H_2-molecule, 21
H_2^+ molecular ion, 17
Hagen-Rubens relation, 576
Hall constant, 363
– table, 363
Hall effect, 361
– metals, 361
– seminconductor, 396
Hardening, 147
Harmonic approximation, 180
Heat pulse experiments, 239
Heat transport, 238
Heavy fermion system UPt_3,
– energy gap, 495
– specific heat, 494
– superconductivity, 493
– ultrasound damping, 494
Heisenberg model, 515, 516
Heteropolar bond, 13
Heterostructures, 333, 420
Hexagonal close-packing of identical spheres, 65
High-temperature superconductors, 487
– critical magnetic field, 489
– electrical resistance, 489
– energy gap, 492
– specific heat, 490
Hole mass, semiconductor
– table, 383
Hole orbit, 350
Holes, 313
– concept, 314
– density, 384
– heavy, 382
– light, 382
– Luttinger liquid, 335
– semiconductor, 380
Hooke's law, 138, 162
Hopping conductivity, 403
Hubbard energy, 406
Hybrid orbital, 24
Hydrogen bond, 5
Hydrogen bridge bond, 29
Hydrogen molecular ion, 17
Hydrogen molecule, 21

Ideal diamagnet, 441
Impurities, 132
Impurity atoms, 70
Impurity conductivity, 391
Impurity diffusion, 136
Inelastic scattering, 182
Infrared absorption, 178, 196
Infrared active, 178
Inhomogeneous semiconductor, 410
Interband transitions, 377, 548, 578
Interface energy, superconductivity, 482
International notation, 51, 58
Interstitials, 131
Intrinsic conduction regime, 394
Intrinsic conductivity
– amorphous semiconductors, 402
– crystalline semiconductor, 383
Intrinsic semiconductors, 376
Inversion layer, 430
Inversion symmetry, 51, 105, 196, 568
Ion spacing
– table, 17
Ionic bond, 5, 13
Ionic conductivity, 136
Ionic crystals, 5
– dielectric constant, 554
– optical eigen modes, 552
– binding energy, 13
Ionic polarization, 551
Ionic radius
– table, 29
Ionization energy, donors
– table, 390
Iron-containing superconductors, 487
Ising model, 536
Isotope effect, superconductivity, 451
Isotope scattering, phonons, 246
Isotropic materials, 163

Josephson
– AC effect, 476
– DC effect, 475
– effect, 474
– equation, 475
– frequence, 476
– junctions in magnetic fields, 477

k-space, 86
Klitzing constant, 365
Kramers-Kronig relationen, 542
Kretschmann geometry, 584

Lamé constants, 163

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Landau
– cylinders, 354
– diamagnetism, 359, 501
– levels, 351
– tubes, 354
Landau-Rumer damping, 233
Landauer-Büttiker formalism, 367
Langevin function, 503
Langevin-Debye equation, 561
Larmor diamagnetism, 500
Lattice anharmonicity, 223
Lattice constant, 50
Lattice dynamics, 159, 170, 182
Lattice energy, 7
Lattice plane, 90
Lattice potential, 180
Lattice vector, 49
Lattice vibrations, 170
– amorphous solids, 213
– monatomic basis, 171
– polyatomic basis, 175
Laue procedure, 116
Law of mass action, 385
LCAO, 18
LEED, 118
Lennard-Jones potential, 8, 10
Level inversion, 131
Light emitting diode, 428
Light scattering, 193
Linear chain, 170, 176
Linear combination of atomic orbitals, 18
Liquid crystals, 47
Local electrical field, 543
London equations, 443
London penetration depth, 444
Long-range order, 45
Lorentz
– approximation, 544
– field, 546
– relation, 546
Lorenz number, 342
– table, 343
Luttinger liquid, 334
Lyddane-Sachs-Teller relation, 556

Macroscopic wave function, 471
Madelung
– energy, 16
– constant, 15
Magnetic domains, 526, 532
Magnetization
– antiferromagnet, 529
– De Haas-van Alphen effect, 358
– definition, 499
– domains, 526
– ferromagnetism, 511
– giant magnetoresistance, 532
– nuclear spin contribution, 507
– paramagnetic, 503
– spontaneous, 509
– type I superconductor, 441
Magnetoresistance, 532
Magnons
– antiferromagnetic, 530
– ferromagnetic, 522
  – dispersion relation, 523
  – thermodynmics, 524
Majority charge carriers, 411
Matthiessen's rule, 328
Mean field approximation, 510
Mean field constant, 510
Mean free path, 229
Mechanical
– deformation, 160
– properties, 138
– strength, 138
– stress, 160
Meissner phase, 482
Meissner-Ochsenfeld effect, 439
Melting temperature
– table, 7
Mermin-Wagner theorem, 299
Mesh, 70
Metallic bond, 5, 26
– binding energy, 26
Metal-semiconductor junction, 418
Metal-isolator transition, 272
Miller indices, 90
Minibands, 423
Minority charge carriers, 411, 414
Mixed crystal, 36
Mobility edge, 401
Mobility,
– Drude model, 316
– Hall effect, 364
– ionic condutors, 136
– semiconductor,
  – table, 398

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Molecule crystals, 9
MOSFET, 430
Mott's two-current model, 534
Mott-Wannier excitons, 574

N-processes, 242
Nanotubes, 68, 299, 336
Nearest neighbors, 61
Neck orbit, 350
Néel temperature, 529
– table, 532
Neutron scattering,
– elastic, 81
– inelastic, 186
Noble gas crystals, 9
– material parameters
  – table, 10
Nordheim rule, 154
Normal mode, 182, 228
Normal process,
– electrons, 324
– phonons, 242
Nuclear magnetism, 507

Ohm's law, 316
One-dimensional conductor, 330, 334, 367
One-dimensional heat transport, 247
One-electron approximation, 255, 274, 285, 325, 519
Optical absorption, 128
– dielectic materials, 548
– dopent atoms, 390
– excitons, 573
– semiconductor, 376
– semiconductors, amorphous, 401
Optical phonons, 177
– ionic crystals, 552
– table, 556
Optical properties of metals, 575
Order parameter, 479, 569
Order-disorder, 43
Order-disorder transition, 45, 154, 570
Orientation polarization, 561
Oscillator model,
– ionic polarization, 551
– optical transitions, 548
Overlap integral, 19

Packing density, 61
Pair correlation function, 72
Pair interaction, 12, 453
Pair state, 453
– occupation probability, 458
Pair wave function, 453
Paraelectricity, 569
Paramagnetism, 502, 507, 512, 529
Pauli paramagnetism, 507
Penetration depth in superconductors,
– Ginzburg-Landau, 480
– London, 444
– Pippard, 446
– table, 446
Penetration depth,
– em waves in metals, 348
– microwaves in semiconductors, 380
Periodic boundary conditions, 200
Periodic potential, electrons, 273
Periodic zone scheme, 284
Perovskite structure, 62
Pertubative potential, 285
Phase diagram,
– alloy, 39
– alloys, 36
– spin glass, 535
– superconductor, 495
Phase problem, 104
Phase transition,
– ferroelectric, 569
– ferromagnet, 513, 517
– first-order, 42, 155
– order-disorder, 154
– second-order, 155
– superconductor, 449, 479, 491, 495
Phase velocity, 174
Phonon
– amplitude, 185
– ballistic, 239
– density of states, 199
– momentum, 184
– number, 212
– spontaneous decay, 234
Phonon branch,
– acoustic, 177
– longitudinal, 179
– optical, 177
– transverse, 179
Phonon polaritons, 556

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Phonon scattering, 188
– by conduction electrons, 323, 329
– by defects, 230, 245
– by electrons in semiconductors, 399
– conduction electrons, 342
– from the surface, 245
– selection rules, 196, 234
Phonon-phonon interaction, 228
Phonon-phonon scattering, 230, 242
Photodiode, 427
Photoelectron spectroscopy, 297
Photoemssion spectroscopy, 296
Piezoelectricity, 569
Plasma frequency, 576
Plasma oscillations, 581
Plasmon, 580, 581
Plasmon energy,
–table, 581
Plasmon polariton, 577
Plastic deformation, 139, 146
p-n junction,, 410
– diffusion voltage, 411
– external voltage, 415
– forward direction, 415
– reverse direction, 415
– space-charge density, 412
Point contacts, 333
Point defects, 123
Point groups, 58
Point symmetry, 51
Polariton, 556, 577
Polarizability
– atomic, 544
– electronic, 548
  – table, 549
– ionic, 551
Polarization, 547
– electronic, 548
– ionic, 551
– orientational direction, 561
– spontaneous electrical, 568
  – table, 571
– static, 561
Polarization catastrophe, 571
Polycrystalline solids, 33
Population inversion, 432
Potential function, 180
Powder method, 115
Primitive unit cell, 51
Pyroelectricity, 568

Quantum dot, 261
Quantum dot contact, 337
Quantum Hall effect, 364
– edge channels, 367
– fractional, 370
– graphene, 371
– resistance plateau, 365
Quasi momentum, 184, 309
– crystals, 229
Quasi momentum conservation,
– amorphous solids, 214
– crystal, 184
Quasi particles,
– electrons, 326
– phonons, 185
– superconductivity, 459
Quasi-Fermi level, 415, 432
Quasicrystals, 55

Raman scattering,
– amorphous solids, 214
– crystal, 194
Rayleigh scattering, 194, 230, 246
Real space, 86
Reciprocal lattice, 84
Recombination, 131, 414, 426, 428
Recombination current, 414
Reduced zone scheme, 277, 284
Reflection, 51
Refractive index, 543
Relaxation absorption, 237
Relaxation time, 316, 327, 538, 565
Relaxation time approximation, 319
Repulsion, interaction potential, 8
Residual resistance, 154
Resonant interaction, 235, 251
Response functions, 542
Reststrahlen method, 561
RKKY interaction, 516
Rotary crystal method, 114
Rotary inversion, 52
Rotation axis, 51
Ruderman-Kittel oscillations, 271
Rutherford scattering, 399

Saturation effect, 237
Saturation magnetization, 504

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Saturation regime, 395
Scattering amplitude, 82, 83
Scattering cross section, 230
Scattering density distribution, 82
Scattering experiments, 79
– amorphous solids, 106
– experimental methods, 111
– inelastic scattering, 182
– surfaces, 103
Scattering length, 101
Scattering process,
– energy conversation, 184
– momentum conservation, 184
Scattering vector, 83
Schottky
– barrier, 419
– defekt, 124
– diode, 419
– model, 413
Schrödinger equation,
– covalent bond, 18
– magnetic field, 351
– periodic potential, 275
Screened Coulomb potential, 271, 326
Screening effects,
– core potential, 518
– dopant atoms, 388
– electron-electron interaction, 325
– electrostatic, 269
– magnetic fields, 443
– Thomas-Fermi, 271
Screening length, 270
Selection rules, 229
Self-diffusion, 136
Semi-metals, 292
Semiclassical equation, 309
Semiconductor, 375
– band gap, 376
– band structure, 376
– devices, 424
– direct, 376
– Heterostructures, 419
– impurity conductivity, 391
– indirect, 376
– inhomogeneous, 410
– intrinsic, 383
– intrinsic conductivity, 383
Semiconductor laser, 432
Semiconductors
– acceptors, 388
– amorphous, 400
– donors, 387
– Fermi level, 391
Shear deformation, 162
Shear stress, 140
– critical, 140
Shear waves, 167
Short-range order, 45
Shubnikov phase, 482
Shubnikov-de Haas oscillations, 361, 365
Silicon, amorphous, 408
Simple-cubic lattice,
– empty lattice, 277
Single crystal production , 33
Skin effect, 348
Small-angle scattering, 108
Solar cell, 409, 425
Sommerfeld
– coefficient, 266
– expansion, 264
– theory, 256, 317
Sound waves, 165
sp^2 hybridisation, 24
sp^3 hybridisation, 23
Space groups, 60, 70
Space-charge capacitance, 417
Space-charge density, 412
Space-charge zone, 415
Specific heat,
– Debye theory, 205
– Einstein model, 199
– glasses, 215
– high-temperature superconductors, 491
– lattice, 198
– low-dimensional systems, 210
– magnetic field dependence, 358
– metals, 265
  – table, 268
– superconductor, 463
– superconductor UPt_3, 494
Spin
– degeneracy, 332, 367
– magnetism, 359, 504, 507, 519
– susceptibility, 507
Spin glass, 535
Spin waves, 522
Spin-charge separation, 335

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Spin-orbit splitting, 383
– table, 383
Spins, unpaired, 153, 405
SQUID, 479
Stacking faults, 150
Static polarization, 561
Stokes process, 194
Stoner
– criterion, 520
– excitation, 524
– parameter, 519
Strain, 161
Stress tensor, 160
Stress, mechanical, 138, 160
Structural defects, 123
Structural relaxation, 406
Structure amplitude, 98
Structure determination, 79
Structure factor, 97
Structure of crystals, 48
Super structure lines, 157
Superconductivity, 437
– BCS state, 461
– BCS theory, 456
– coherence length, 446, 480, 491
– coherence volume, 491
– Cooper pairs, 453
– critical current, 468, 475
– critical magnetic field, 441, 447, 482
– critical temperature
  – table, 439, 488
– energy gap, 459, 461, 492
– entropy, 450
– Ginzburg-Landau theory, 479
– intermediate state, 442
– isotope effect, 451
– Josephson effect, 474
– macroscopic wave function, 471
– penetration depth, 444, 480
– specific heat, 463, 490, 494
  – table, 463
– thermodynamic, 447
– transition temperature, 437
  – table, 438
– tunnel junction spectroscopy, 464
– ultrasound absorption, 464, 493
Superconductor,
– critical magnetic field
  – table, 438
– critical temperature
  – table, 438
– energy gap
  – table, 462
– technical application, 496
– type-I, 482
– type-II, 482
Superlattice, 421
Superstructures, 70
Surface, 70
– direct imaging, 119
– phonon scattering, 245
– plasmon sensor, 582
– reconstruction, 70
– relaxation, 70
– structure determination, 117
Surface plasmons, 582
Surface scattering experiments, 103
Susceptibility
– antiferromagnetic, 530
– diamagnetic, 441, 501
  – table, 501
– paramagnetic, 503, 507, 513, 519
– spin glasses, 537

T^3 - law, specific heat, 208
Tensile stress, 139, 161
Tensile-strain diagram, 139
Tetrahedral bond, 23
Thermal conductivity, 240
– amorphous solids, 250
– dielectric, 240
– metals, 340
– one-dimensional systeme, 247
Thermal expansion, 223
Thermal expansion coefficient,
– amorphous solid, 227
Thermally activated process, 133, 403, 565
Thomas-Fermi screening length, 270, 326
Three phonon processes, 228
Three-axis spectrometer, 186
Tight-binding model, 284
Tightly bound electrons, 284
Time-of-flight spectrometers, 187
Tomonaga-Luttinger liquid, 334
Transistor,
– bipolar, 429
– field effect, 430

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use

Transition temperature, 437
– table, 438, 439, 485
Translation vector, 49
Translational symmetry, 48, 49
Transmission electron microscope, 119
Tunnel junction spectroscopy, 464
Tunneling current, 467, 476
Two-dimensional
– electron gas, 333, 422, 431
Two-dimensional hexagonal systems, 299
Two-dimensional system, 202, 259, 356
Two-level system, 217, 235, 562
Two-particle wave function, 22, 453

U-processes, 242
Ultrasonic absorption,
– crystals, 229
Ultrasonic measurement technique, 169
Ultrasound absorption,
– amorphous solids, 235
Ultrasound damping,
– heavy fermion system UPt_3, 494
– superconductor, 464
Ultrasound, resonant interaction, 235
Umklapp process,
– electrons, 324
– phonons, 242
Unconventional superconductors, 487
Unit cell, 50

Vacancy
– concentration, 125
– diffusion, 136
– ionic crystals, 127
Valence alternating pair, 153
Valence band, 291, 376
Valence band,
– graphene, 300
Van der Waals Bond, 9
Van der Waals force, 5, 9
Van-Hove singularity, 203, 261, 356
Van-Vleck paramagnetism, 509
Variable range hopping, 404
Vibration spectrum, amorphous solids, 213
Voigt notation, 162
Volume plasmons, 582

Wannier function, 277
Wave packet, 307
Weak link, 474
Whisker, 146
Wiedemann-Franz law, 342
Wigner-Seitz cell, 67
Work function, 256, 298, 418

X-raydiffration, 81

YBCO, 488

Zener breakdown, 425
Zener Diode, 425
Zero-point energy, 7, 11
– lattice, 212, 224
– noble gas crystals, 12
Zigzag structure, 68
Zinc blende structure, 63
Zone scheme, 284

EBSCOhost - printed on 2/13/2023 7:45 PM via . All use subject to https://www.ebsco.com/terms-of-use