Graduate Texts in Physics

Marius Grundmann

The Physics of Semiconductors

An Introduction Including Nanophysics and Applications

Third Edition

Springer

Graduate Texts in Physics

Series editors

Kurt H. Becker, Brooklyn, USA
Sadri Hassani, Normal, USA
Bill Munro, Kanagawa, Japan
Richard Needs, Cambridge, UK
William T. Rhodes, Boca Raton, USA
Susan Scott, Acton, Australia
H. Eugene Stanley, Boston, USA
Martin Stutzmann, Garching, Germany
Andreas Wipf, Jena, Germany

Graduate Texts in Physics

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

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

Marius Grundmann

The Physics of Semiconductors

An Introduction Including Nanophysics and Applications

Third Edition

Springer

Marius Grundmann
Institut für Experimentelle Physik II
Universität Leipzig
Leipzig
Germany

ISSN 1868-4513 ISSN 1868-4521 (electronic)
Graduate Texts in Physics
ISBN 978-3-319-23879-1 ISBN 978-3-319-23880-7 (eBook)
DOI 10.1007/978-3-319-23880-7

Library of Congress Control Number: 2015954622

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2006, 2010, 2016
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

To Michelle,
Sophia Charlotte
and Isabella Rose

Preface

Semiconductor electronics is commonplace in every household. Semiconductor devices have enabled economically reasonable fiber-based optical communication, optical storage, and high-frequency amplification and have recently revolutionized photography, display technology, and lighting. By now solar energy harvesting with photovoltaics contributes a significant portion to the energy mix. Along with these tremendous technological developments, semiconductors have changed the way we work, communicate, entertain, and think. The technological progress of semiconductor materials and devices is evolving continuously with a large worldwide effort in human and monetary capital. For students, semiconductors offer a rich and exciting field with a great tradition, offering diverse fundamental and applied topics [1] and a bright future.

This book introduces students to semiconductor physics and semiconductor devices. It brings them to the point where they can specialize and enter supervised laboratory research. It is based on the two-semester semiconductor physics course taught at Universität Leipzig in its Master of Science physics curriculum. Since the book can be followed with little or no pre-existing knowledge in solid-state physics and quantum mechanics, it is also suitable for undergraduate students. For the interested reader several additional topics are included in the book that can be covered in subsequent, more specialized courses. The material is selected to provide a balance between aspects of solid-state and semiconductor physics, the concepts of various semiconductor devices and modern applications in electronics and photonics.

The first semester contains the fundamentals of semiconductor physics (Part I, Chaps. 1–10) and selected topics from Part II (Chaps. 11–20). Besides important aspects of solid-state physics such as crystal structure, lattice vibrations, and band structure, semiconductor specifics such as technologically relevant materials and their properties, doping and electronic defects, recombination, surfaces, and hetero- and nanostructures are discussed. Semiconductors with electric polarization and magnetization are introduced. The emphasis is put on inorganic semiconductors, but a brief introduction to organic semiconductors is given in Chap. 17. Dielectric structures (Chap. 19) serve as mirrors, cavities, and microcavities and are a vital

part of many semiconductor devices. Other chapters give introduction to carbon-based nanostructures and transparent conductive oxides (TCOs). The third part (Part III, Chaps. 21–24) is dedicated to semiconductor applications and devices that are taught in the second semester of the course. After a general and detailed discussion of various diode types, their applications in electrical circuits, photodetectors, solar cells, light-emitting diodes, and lasers are treated. Finally, bipolar and field-effect transistors including thin-film transistors are discussed.

In the present text of the third edition, a few errors and misprints of the second edition have been corrected. Several topics have been extended and are treated in more depth, e.g., double donors and double acceptors, negative-U centers, Boltzmann transport equation, ionic conductivity, hopping conductivity, impact ionization, thermopower, polarons, intra-band transitions, amorphous semiconductors, disorder effects, heteroepitaxy on mismatched, curved and patterned substrates, and noise. A chapter on semiconductor surfaces has been added.

The list of references has been augmented by almost 400 quotations with respect to the list in the second edition. All references now include title and complete page numbers. The references have been selected to (i) cover important historical and milestone papers, (ii) direct to reviews and topical books for further reading and (iii) give access to current literature and up-to-date research. In Fig. 1, the original papers within the more than 1800 references in this book are shown by year. Roughly three phases of semiconductor physics and technology can be seen. Before the realization of the first transistor in 1947, only a few publications are noteworthy. Then an intense phase of understanding the physics of semiconductors and developing semiconductor technology and devices based on bulk semiconductors (mostly Ge, Si, GaAs) followed. At the end of the 1970s, a new era began with the advent of quantum wells and heterostructures, and later nanostructures (nanotubes, nanowires, and quantum dots) and new materials (e.g., organic semiconductors, nitrides or graphene). Also several very recent references to emerging topics such as 2D materials, topological insulators or novel amorphous semiconductors are given.

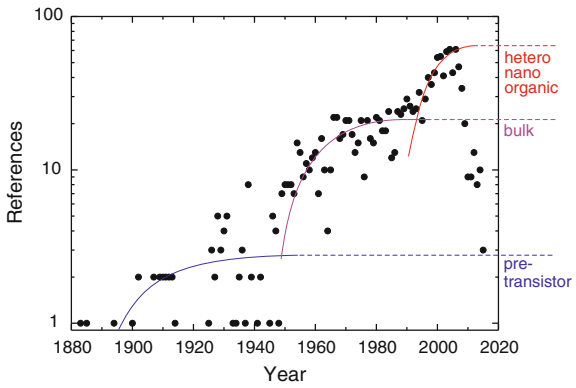


Fig. 1 Histogram of references in this book

I would like to thank many colleagues for their various contributions to this book, in alphabetical order (if no affiliation is given, at the time at Universität Leipzig): Gabriele Benndorf, Klaus Bente, Rolf Böttcher, Matthias Brandt, Christian Czekalla, Christof Peter Dietrich, Pablo Esquinazi, Heiko Frenzel, Volker Gottschalch, Helena Franke (née Hilmer), Axel Hoffmann (TU Berlin), Alois Krost† (Otto-von-Guericke Universität Magdeburg), Michael Lorenz, Stefan Müller, Thomas Nobis, Rainer Pickenhain, Hans-Joachim Queisser (Max-Planck-Institut für Festkörperforschung, Stuttgart), Bernd Rauschenbach (Leibniz-Institut für Oberflächenmodifizierung, Leipzig), Bernd Rheinländer, Heidemarie Schmidt, Mathias Schmidt, Rüdiger Schmidt-Grund, Matthias Schubert, Jan Sellmann, Oliver Stier (TU Berlin), Chris Sturm, Florian Tendille (CNRS-CRHEA), Gerald Wagner, Eicke Weber (UC Berkeley), Holger von Wenckstern, Michael Ziese, and Gregor Zimmermann. This book has benefitted from their comments, proof reading, experimental data, and graphic material. Also, numerous helpful comments from my students on my lectures and previous editions of this book are gratefully acknowledged.

I am also indebted to many other colleagues, in particular to (in alphabetical order) Gerhard Abstreiter, Zhores Alferov, Martin Allen, Levon Asryan, Günther Bauer, Manfred Bayer, Friedhelm Bechstedt, Dieter Bimberg, Otto Breitenstein, Len Brillson, Fernando Briones, Immanuel Broser†, Jean-Michel Chauveau, Jürgen Christen, Philippe De Mierry, Steve Durbin, Laurence Eaves, Klaus Ellmer, Guy Feuillet, Elvira Fortunato, Ulrich Gösele†, Alfred Forchel, Manus Hayne, Frank Heinrichsdorff, Fritz Henneberger†, Detlev Heitmann, Robert Heitz†, Evamarie Hey-Hawkins, Detlef Hommel, Evgeni Kaidashev, Eli Kapon, Nils Kirstaedter, Claus Klingshirn, Fred Koch†, Jörg Kotthaus, Nikolai Ledentsov, Peter Littlewood, Dave Look, Axel Lorke, Anupam Madhukar, Ingrid Mertig, Bruno Meyer†, David Mowbray, Hisao Nakashima, Jörg Neugebauer, Michael Oestreich, Louis Piper, Mats-Erik Pistol, Fred Pollak†, Volker Riede, Bernd Rosenow, Hiroyuki Sakaki, Lars Samuelson, Darrell Schlom, Vitali Shchukin, Maurice Skolnick, Robert Suris, Volker Türck, Konrad Unger†, Victor Ustinov, Leonid Vorob'jev, Richard Warburton, Alexander Weber, Peter Werner, Wolf Widdra, Ulrike Woggon, Roland Zimmermann, Arthur Zrenner, Alex Zunger, and Jesús Zúñiga-Pérez, with whom I have worked closely, had enjoyable discussions with and who have posed questions that stimulated me. It is my distinct privilege and joy that this list becomes longer as I pursue studies in semiconductor physics but sadly the number of †-symbols increases too rapidly from edition to edition.

Leipzig Marius Grundmann

Contents

1 Introduction 1
- 1.1 Timetable and Key Achievements 2
- 1.2 Nobel Prize Winners 12
- 1.3 General Information 21

Part I Fundamentals

2 Bonds 25
- 2.1 Introduction 25
- 2.2 Covalent Bonds 25
  - 2.2.1 Electron-Pair Bond 26
  - 2.2.2 sp^3 Bonds. 26
  - 2.2.3 sp^2 Bonds. 29
- 2.3 Ionic Bonds 32
- 2.4 Mixed Bonds 33
- 2.5 Metallic Bonding 36
- 2.6 van-der-Waals Bonds. 37
- 2.7 Hamilton Operator of the Solid. 38

3 Crystals 41
- 3.1 Introduction 41
- 3.2 Crystal Structure 42
- 3.3 Lattice 43
  - 3.3.1 Unit Cell 43
  - 3.3.2 Point Group 43
  - 3.3.3 Space Group. 45
  - 3.3.4 2D Bravais Lattices 46
  - 3.3.5 3D Bravais Lattices 46
  - 3.3.6 Polycrystalline Semiconductors 51
  - 3.3.7 Amorphous Semiconductors 51

3.4 Important Crystal Structures . . . 53
3.4.1 Rocksalt Structure . . . 53
3.4.2 CsCl Structure . . . 53
3.4.3 Diamond Structure . . . 54
3.4.4 Zincblende Structure . . . 55
3.4.5 Wurtzite Structure . . . 56
3.4.6 Chalcopyrite Structure . . . 57
3.4.7 Spinel Structure . . . 59
3.4.8 Fluorite Structure . . . 61
3.4.9 Delafossite Structure . . . 61
3.4.10 Perovskite Structure . . . 61
3.4.11 NiAs Structure . . . 62
3.4.12 Further Structures . . . 63
3.5 Polytypism . . . 64
3.6 Reciprocal Lattice . . . 66
3.6.1 Reciprocal Lattice Vectors . . . 66
3.6.2 Miller Indices . . . 67
3.6.3 Brillouin Zone . . . 70
3.7 Alloys . . . 72
3.7.1 Random Alloys . . . 72
3.7.2 Phase Diagram . . . 74
3.7.3 Virtual Crystal Approximation . . . 77
3.7.4 Lattice Parameter . . . 77
3.7.5 Ordering . . . 77

4 Defects . . . 81
4.1 Introduction . . . 81
4.2 Point Defects . . . 81
4.2.1 Point Defect Types . . . 81
4.2.2 Thermodynamics . . . 83
4.2.3 Diffusion . . . 85
4.2.4 Dopant Distribution . . . 88
4.2.5 Large Concentration Effects . . . 92
4.3 Dislocations . . . 96
4.3.1 Dislocation Types . . . 96
4.3.2 Visualization of Dislocations by Etching . . . 100
4.3.3 Impurity Hardening . . . 101
4.4 Extended Defects . . . 103
4.4.1 Micro-cracks . . . 103
4.4.2 Stacking Faults . . . 103
4.4.3 Grain Boundaries . . . 105
4.4.4 Antiphase and Inversion Domains . . . 106
4.5 Disorder . . . 110

5 Mechanical Properties 111
5.1 Introduction 111
5.2 Lattice Vibrations 111
5.2.1 Monoatomic Linear Chain 112
5.2.2 Diatomic Linear Chain 115
5.2.3 Lattice Vibrations of a Three-Dimensional Crystal 118
5.2.4 Density of States 121
5.2.5 Phonons 123
5.2.6 Localized Vibrational Modes 124
5.2.7 Phonons in Alloys 127
5.2.8 Disorder 128
5.2.9 Electric Field Created by Optical Phonons 129
5.3 Elasticity 132
5.3.1 Thermal Expansion 132
5.3.2 Stress–Strain Relation 133
5.3.3 Biaxial Strain 137
5.3.4 Three-Dimensional Strain 139
5.3.5 Substrate Bending 141
5.3.6 Scrolling 143
5.4 Plasticity 145
5.4.1 Critical Thickness 145
5.4.2 Cleaving 150
5.4.3 Wafer Breakage 151

6 Band Structure 153
6.1 Introduction 153
6.2 Electrons in a Periodic Potential 154
6.2.1 Bloch's Theorem 154
6.2.2 Free-Electron Dispersion 155
6.2.3 Non-Vanishing Potential 156
6.2.4 Kramer's Degeneracy 159
6.2.5 Symmetry Considerations 160
6.3 Band Structures of Selected Semiconductors 162
6.3.1 Silicon 163
6.3.2 Germanium 163
6.3.3 GaAs 163
6.3.4 GaP 163
6.3.5 GaN 163
6.3.6 Lead Salts 163
6.3.7 MgO, ZnO, CdO 164
6.3.8 Chalcopyrites 165
6.3.9 Spinels 166
6.3.10 Delafossites 166
6.3.11 Perovskites 167

6.4 Systematics of Semiconductor Band Gaps 168
6.5 Alloy Semiconductors 170
6.6 Amorphous Semiconductors 173
6.7 Temperature Dependence of the Band Gap 174
6.8 Isotope Dependence of the Band Gap 176
6.9 Electron Dispersion 177
6.9.1 Equation of Electron Motion 177
6.9.2 Effective Mass of Electrons 178
6.9.3 Nonparabolicity of Electron Mass 181
6.10 Holes 183
6.10.1 Hole Concept 183
6.10.2 Hole Dispersion Relation 184
6.10.3 Valence-Band Fine Structure 188
6.10.4 Band Inversion 189
6.11 Strain Effects on the Band Structure 190
6.11.1 Strain Effect on Band Edges 191
6.11.2 Strain Effect on Effective Masses 193
6.11.3 Interaction with a Localized Level 194
6.12 Density of States 195
6.12.1 General Band Structure 195
6.12.2 Amorphous Semiconductors 196
6.12.3 Free-Electron Gas 199

7 Electronic Defect States 203
7.1 Introduction 203
7.2 Carrier Concentration 204
7.3 Intrinsic Conduction 207
7.4 Doping 209
7.4.1 Concept 209
7.4.2 Doping Principles 210
7.5 Shallow Defects 211
7.5.1 Donors 211
7.5.2 Acceptors 220
7.5.3 Compensation 223
7.5.4 Multiple Impurities 226
7.5.5 Amphoteric Impurities 228
7.5.6 Autodoping 229
7.5.7 High Doping 230
7.6 Quasi-Fermi Levels 234
7.7 Deep Levels 235
7.7.1 Charge States 236
7.7.2 Double Donors 238
7.7.3 Double Acceptors 240
7.7.4 Jahn–Teller Effect 241
7.7.5 Negative-*U* Center 242

7.7.6 DX Center . . . 245
7.7.7 EL2 Defect . . . 247
7.7.8 Semi-insulating Semiconductors . . . 247
7.7.9 Isoelectronic Impurities . . . 249
7.7.10 Surface States . . . 251
7.8 Hydrogen in Semiconductors . . . 251

8 Transport . . . 255
8.1 Introduction . . . 255
8.2 Conductivity . . . 256
8.3 Low-Field Transport . . . 258
8.3.1 Mobility . . . 258
8.3.2 Microscopic Scattering Processes . . . 259
8.3.3 Ionized Impurity Scattering . . . 260
8.3.4 Deformation Potential Scattering . . . 261
8.3.5 Piezoelectric Potential Scattering . . . 262
8.3.6 Polar Optical Scattering . . . 262
8.3.7 Dislocation Scattering . . . 263
8.3.8 Grain Boundary Scattering . . . 263
8.3.9 Temperature Dependence . . . 264
8.3.10 Doping Dependence . . . 266
8.3.11 Piezoresistivity . . . 267
8.4 High-Field Transport . . . 268
8.4.1 Drift-Saturation Velocity . . . 269
8.4.2 Negative Differential Resistivity . . . 269
8.4.3 Velocity Overshoot . . . 270
8.4.4 Impact Ionization . . . 271
8.5 High-Frequency Transport . . . 275
8.6 Polarons . . . 275
8.6.1 Large Polarons . . . 275
8.6.2 Small Polarons . . . 277
8.7 Hopping Transport . . . 278
8.8 Transport in Amorphous Semiconductors . . . 280
8.9 Ionic Transport . . . 281
8.10 Diffusion . . . 282
8.11 Continuity Equation . . . 284
8.12 Heat Conduction . . . 284
8.13 Coupled Heat and Charge Transport . . . 286
8.13.1 Thermopower and Seebeck Effect . . . 286
8.13.2 Peltier Effect . . . 289

9 Optical Properties . . . 291
9.1 Spectral Regions and Overview . . . 291
9.2 Complex Dielectric Function . . . 292
9.3 Reflection and Diffraction . . . 293

9.4 Absorption . . . 295
9.5 Electron–Photon Interaction . . . 297
9.6 Band–Band Transitions . . . 299
9.6.1 Joint Density of States . . . 299
9.6.2 Direct Transitions . . . 300
9.6.3 Indirect Transitions . . . 304
9.6.4 Urbach Tail . . . 307
9.6.5 Amorphous Semiconductors . . . 308
9.6.6 Excitons . . . 309
9.6.7 Phonon Broadening . . . 313
9.6.8 Exciton Polariton . . . 313
9.6.9 Bound-Exciton Absorption . . . 317
9.6.10 Biexcitons . . . 319
9.6.11 Trions . . . 319
9.6.12 Band Gap Renormalization . . . 320
9.6.13 Electron–Hole Droplets . . . 321
9.6.14 Two-Photon Absorption . . . 322
9.7 Impurity Absorption . . . 323
9.7.1 Shallow Levels . . . 323
9.7.2 Deep Levels . . . 326
9.8 Absorption in the Presence of Free Charge Carriers . . . 328
9.8.1 Free-Carrier Absorption . . . 328
9.8.2 Burstein–Moss Shift . . . 332
9.8.3 Inter-Valenceband Transitions . . . 334
9.8.4 Inter-Valley Transitions . . . 335
9.8.5 Intra-Band Transitions . . . 336
9.9 Lattice Absorption . . . 337
9.9.1 Dielectric Constant . . . 337
9.9.2 Reststrahlenbande . . . 338
9.9.3 Polaritons . . . 339
9.9.4 Phonon–Plasmon Coupling . . . 340

10 Recombination . . . 343
10.1 Introduction . . . 343
10.2 Band–Band Recombination . . . 344
10.2.1 Spontaneous Emission . . . 344
10.2.2 Absorption . . . 346
10.2.3 Stimulated Emission . . . 347
10.2.4 Net Recombination Rate . . . 347
10.2.5 Recombination Dynamics . . . 349
10.2.6 Lasing . . . 350
10.3 Exciton Recombination . . . 351
10.3.1 Free Excitons . . . 351
10.3.2 Bound Excitons . . . 351
10.3.3 Alloy Broadening . . . 358

10.4 Phonon Replica . . . 362
10.5 Self-Absorption . . . 365
10.6 Donor–Acceptor Pair Transitions . . . 366
10.7 Inner-Impurity Recombination . . . 367
10.8 Auger Recombination . . . 368
10.9 Band–Impurity Recombination . . . 370
10.9.1 Shockley–Read–Hall Kinetics . . . 370
10.9.2 Multilevel Traps . . . 374
10.10 ABC Model . . . 374
10.11 Field Effect . . . 375
10.11.1 Thermally Activated Emission . . . 375
10.11.2 Direct Tunneling . . . 376
10.11.3 Assisted Tunneling . . . 376
10.12 Recombination at Extended Defects . . . 377
10.12.1 Surfaces . . . 377
10.12.2 Grain Boundaries . . . 378
10.12.3 Dislocations . . . 378
10.13 Excess-Carrier Profiles . . . 379
10.13.1 Generation at Surface . . . 379
10.13.2 Generation in the Bulk . . . 380

Part II Selected Topics

11 Surfaces . . . 385
11.1 Introduction . . . 385
11.2 Surface Crystallography . . . 386
11.3 Surface Energy . . . 387
11.4 Surface Reconstruction . . . 388
11.5 Surface Morphology . . . 390
11.6 Surface Physical Properties . . . 392
11.6.1 Surface Phonons . . . 392
11.6.2 Surface Plasmons . . . 393
11.6.3 Electronic Surface States . . . 395

12 Heterostructures . . . 399
12.1 Introduction . . . 399
12.2 Heteroepitaxy . . . 400
12.2.1 Growth Methods . . . 400
12.2.2 Substrates . . . 400
12.2.3 Growth Modes . . . 403
12.2.4 Heterosubstrates . . . 405
12.2.5 Patterned Substrates . . . 410
12.2.6 Pseudomorphic Structures . . . 412
12.2.7 Plastic Relaxation . . . 413
12.2.8 Surfactants . . . 414

12.3 Energy Levels in Heterostructures . . . 415
12.3.1 Band Lineup in Heterostructures . . . 415
12.3.2 Quantum Wells . . . 417
12.3.3 Superlattices . . . 424
12.3.4 Single Heterointerface Between Doped Materials . . . 424
12.4 Recombination in Quantum Wells . . . 427
12.4.1 Thickness Dependence . . . 427
12.4.2 Broadening Effects . . . 428
12.4.3 Quantum Confined Stark Effect . . . 431
12.5 Isotope Superlattices . . . 433
12.6 Wafer Bonding . . . 433

13 External Fields . . . 437
13.1 Electric Fields . . . 437
13.1.1 Bulk Material . . . 437
13.1.2 Quantum Wells . . . 439
13.2 Magnetic Fields . . . 441
13.2.1 Hall Effect . . . 441
13.2.2 Free-Carrier Absorption . . . 448
13.2.3 Energy Levels in Bulk Crystals . . . 449
13.2.4 Energy Levels in a 2DEG . . . 451
13.2.5 Shubnikov–de Haas Oscillations . . . 452
13.3 Quantum Hall Effect . . . 453
13.3.1 Integral QHE . . . 455
13.3.2 Fractional QHE . . . 458
13.3.3 Weiss Oscillations . . . 460

14 Nanostructures . . . 461
14.1 Introduction . . . 461
14.2 Quantum Wires . . . 462
14.2.1 V-Groove Quantum Wires . . . 462
14.2.2 Cleaved-Edge Overgrowth Quantum Wires . . . 465
14.2.3 Nanowhiskers . . . 465
14.2.4 Nanobelts . . . 467
14.2.5 Quantization in Two-Dimensional Potential Wells . . . 468
14.3 Quantum Dots . . . 471
14.3.1 Quantization in Three-Dimensional Potential Wells . . . 471
14.3.2 Electrical and Transport Properties . . . 473
14.3.3 Self-Assembled Preparation . . . 477
14.3.4 Optical Properties . . . 482

15 Polarized Semiconductors . . . 489
15.1 Introduction . . . 489
15.2 Spontaneous Polarization . . . 489

15.3 Ferroelectricity 490
15.3.1 Materials 492
15.3.2 Soft Phonon Mode 492
15.3.3 Phase Transition 493
15.3.4 Domains. 496
15.3.5 Optical Properties 497
15.4 Piezoelectricity 498
15.4.1 Piezoelectric Effect 498
15.4.2 Zincblende Crystals 498
15.4.3 Wurtzite Crystals. 499
15.4.4 Piezoelectric Effects in Nanostructures 503

16 Magnetic Semiconductors 505
16.1 Introduction 505
16.2 Magnetic Semiconductors. 505
16.3 Diluted Magnetic Semiconductors 507
16.4 Spintronics 511
16.4.1 Spin Transistor 512
16.4.2 Spin LED. 512

17 Organic Semiconductors 515
17.1 Introduction 515
17.2 Materials 515
17.2.1 Small Organic Molecules, Polymers. 515
17.2.2 Organic Semiconductor Crystals 516
17.3 Electronic Structure 519
17.4 Doping 520
17.5 Transport Properties 521
17.6 Optical Properties 522

18 Graphene and Carbon Nanotubes 529
18.1 Graphene 529
18.1.1 Structure 529
18.1.2 Band Structure 529
18.1.3 Electrical Properties 531
18.1.4 Other Two-Dimensional Crystals 536
18.2 Carbon Nanotubes. 536
18.2.1 Structure 536
18.2.2 Band Structure 538
18.2.3 Optical Properties 541
18.2.4 Other Anorganic Nanotubes 541

19 Dielectric Structures 545
19.1 Photonic Band Gap Materials 545
19.1.1 Introduction 545
19.1.2 General 1D Scattering Theory 546

19.1.3 Transmission of an *N*-Period Potential 548
19.1.4 The Quarter-Wave Stack 550
19.1.5 Formation of a 3D Band Structure 552
19.1.6 Disorder 558
19.1.7 Defect Modes 558
19.1.8 Coupling to an Electronic Resonance 562
19.2 Microscopic Resonators 565
19.2.1 Microdiscs 565
19.2.2 Purcell Effect 567
19.2.3 Deformed Resonators 568
19.2.4 Hexagonal Cavities 571

20 Transparent Conductive Oxide Semiconductors 575
20.1 Introduction 575
20.2 Materials 575
20.3 Properties 577

Part III Applications

21 Diodes 583
21.1 Introduction 583
21.2 Metal–Semiconductor Contacts 583
21.2.1 Band Diagram in Equilibrium 584
21.2.2 Space-Charge Region 589
21.2.3 Schottky Effect 593
21.2.4 Capacitance 594
21.2.5 Current–Voltage Characteristic 598
21.2.6 Ohmic Contacts 611
21.2.7 Metal Contacts to Organic Semiconductors 613
21.3 Metal–Insulator–Semiconductor Diodes 615
21.3.1 Band Diagram for Ideal MIS Diode 616
21.3.2 Space-Charge Region 619
21.3.3 Capacitance 623
21.3.4 Nonideal MIS Diode 624
21.4 Bipolar Diodes 626
21.4.1 Band Diagram 626
21.4.2 Space-Charge Region 626
21.4.3 Capacitance 632
21.4.4 Current–Voltage Characteristics 633
21.4.5 Breakdown 644
21.4.6 Heterostructure Diodes 649
21.4.7 Organic Semiconductor Diodes 650
21.5 Applications and Special Diode Devices 652
21.5.1 Rectification 653
21.5.2 Frequency Mixing 655

21.5.3 Voltage Regulator 656
21.5.4 Zener Diodes 658
21.5.5 Varactors 658
21.5.6 Fast-Recovery Diodes 660
21.5.7 Step-Recovery Diodes 661
21.5.8 Pin-Diodes 662
21.5.9 Tunneling Diodes 663
21.5.10 Backward Diodes 666
21.5.11 Gunn Diodes 667

22 Light-to-Electricity Conversion 669
22.1 Photocatalysis 669
22.2 Photoconductors 670
22.2.1 Introduction 670
22.2.2 Photoconductivity Detectors 671
22.2.3 Electrophotography 674
22.2.4 QWIPs 675
22.2.5 Blocked Impurity-Band Detectors 678
22.3 Photodiodes 680
22.3.1 Introduction 680
22.3.2 pn Photodiodes 681
22.3.3 Pin Photodiodes 684
22.3.4 Position-Sensing Detector 686
22.3.5 MSM Photodiodes 686
22.3.6 Avalanche Photodiodes 692
22.3.7 Traveling-Wave Photodetectors 695
22.3.8 Charge Coupled Devices 698
22.3.9 Photodiode Arrays 705
22.4 Solar Cells 707
22.4.1 Solar Radiation 708
22.4.2 Ideal Solar Cells 709
22.4.3 Real Solar Cells 714
22.4.4 Design Refinements 715
22.4.5 Modules 716
22.4.6 Solar-Cell Types 718
22.4.7 Commercial Issues 721

23 Electricity-to-Light Conversion 725
23.1 Radiometric and Photometric Quantities 725
23.1.1 Radiometric Quantities 725
23.1.2 Photometric Quantities 725
23.2 Scintillators 727
23.2.1 CIE Chromaticity Diagram 727
23.2.2 Display Applications 730

23.2.3 Radiation Detection 731
23.2.4 Luminescence Mechanisms 732
23.3 Light-Emitting Diodes 733
23.3.1 Introduction 733
23.3.2 Spectral Ranges 734
23.3.3 Efficiencies 735
23.3.4 Device Design 736
23.3.5 White LEDs 743
23.3.6 Quantum Dot LEDs 746
23.3.7 Organic LEDs 747
23.4 Lasers 749
23.4.1 Introduction 749
23.4.2 Applications 750
23.4.3 Gain 752
23.4.4 Optical Mode 755
23.4.5 Loss Mechanisms 762
23.4.6 Threshold 763
23.4.7 Spontaneous Emission Factor 765
23.4.8 Output Power 766
23.4.9 Temperature Dependence 768
23.4.10 Mode Spectrum 769
23.4.11 Longitudinal Single-Mode Lasers 770
23.4.12 Tunability 771
23.4.13 Dynamics and Modulation 774
23.4.14 Surface-Emitting Lasers 778
23.4.15 Optically Pumped Semiconductor Lasers 782
23.4.16 Quantum Cascade Lasers 783
23.4.17 Hot-Hole Lasers 784
23.5 Semiconductor Optical Amplifiers 785

24 Transistors 787
24.1 Introduction 787
24.2 Bipolar Transistors 788
24.2.1 Carrier Density and Currents 790
24.2.2 Current Amplification 792
24.2.3 Ebers–Moll Model 794
24.2.4 Current–Voltage Characteristics 796
24.2.5 Basic Circuits 798
24.2.6 High-Frequency Properties 800
24.2.7 Heterojunction Bipolar Transistors 800
24.2.8 Light-Emitting Transistors 801
24.3 Field-Effect Transistors 802
24.4 JFET and MESFET 804
24.4.1 General Principle 804
24.4.2 Static Characteristics 805

24.4.3 Normally on and Normally Off FETs . . . 809
24.4.4 Field-Dependent Mobility . . . 809
24.4.5 High-Frequency Properties . . . 812
24.5 MOSFETs . . . 812
24.5.1 Operation Principle . . . 812
24.5.2 Current–Voltage Characteristics . . . 814
24.5.3 MOSFET Types . . . 818
24.5.4 Complementary MOS . . . 818
24.5.5 Large-Scale Integration . . . 821
24.5.6 Tunneling FETs . . . 829
24.5.7 Nonvolatile Memories . . . 830
24.5.8 Heterojunction FETs . . . 833
24.6 Thin-Film Transistors . . . 837
24.6.1 Annealing of Amorphous Silicon . . . 837
24.6.2 TFT Devices . . . 838
24.6.3 OFETs . . . 839

Appendix A: Tensors . . . 843

Appendix B: Point and Space Groups . . . 847

Appendix C: Kramers–Kronig Relations . . . 851

Appendix D: Oscillator Strength . . . 853

Appendix E: Quantum Statistics . . . 859

Appendix F: Kronig–Penney Model . . . 865

Appendix G: The $\mathbf{k} \cdot \mathbf{p}$ Perturbation Theory . . . 869

Appendix H: Effective-Mass Theory . . . 873

Appendix I: Boltzmann Transport Theory . . . 875

Appendix J: Noise . . . 881

References . . . 891

Index . . . 967

Abbreviations

0D	Zero-dimensional
1D	One-dimensional
2D	Two-dimensional
2DEG	Two-dimensional electron gas
3D	Three-dimensional
A	Atomic mass ($A = 12$ for ^{12}C)
AAAS	American Association for the Advancement of Science
AB	Antibonding (position)
ac	Alternating current
ACS	American Chemical Society
AFM	Atomic force microscopy
AHE	Anomalous Hall effect
AIP	American Institute of Physics
AM	Air mass
APD	Antiphase domain, avalanche photodiode
APS	American Physical Society
AR	Antireflection
ARPES	Angle-resolved photoemission spectroscopy
ASE	Amplified spontaneous emission
AVS	American Vacuum Society (The Science & Technology Society)
bc	Body-centered
bcc	Body-centered cubic
BC	Bond center (position)
BD	Blu-ray™ disc
BEC	Bose–Einstein condensation
BGR	Band gap renormalization
BIA	Bulk inversion asymmetry
BJT	Bipolar junction transistor
CAS	Calorimetric absorption spectroscopy
CCD	Charge coupled device

CD	Compact disc
CEO	Cleaved-edge overgrowth
CIE	Commission Internationale de l'Éclairage
CIGS	$Cu(In,Ga)Se_2$ material
CIS	$CuInSe_2$ material
CL	Cathodoluminescence
CMOS	Complementary metal–oxide–semiconductor
CMY	Cyan–magenta–yellow (color system)
CNT	Carbon nanotube
COD	Catastrophical optical damage
CPU	Central processing unit
CRT	Cathode ray tube
CSL	Coincident site lattice
CVD	Chemical vapor deposition
cw	Continuous wave
CZ	Czochralski (growth)
DAP	Donor–acceptor pair
DBR	Distributed Bragg reflector
dc	Direct current
DF	Dielectric function
DFB	Distributed feedback
DH(S)	Double heterostructure
DLTS	Deep level transient spectroscopy
DMS	Diluted magnetic semiconductor
DOS	Density of states
DPSS	Diode-pumped solid-state (laser)
DRAM	Dynamic random access memory
DVD	Digital versatile disc
EA	Electron affinity
EBL	Electron blocking layer
EEPROM	Electrically erasable programmable read-only memory
EHL	Electron–hole liquid
EIL	Electron injection layer
EL	Electroluminescence
ELA	Excimer laser annealing
ELO	Epitaxial lateral overgrowth
EMA	Effective mass approximation
EML	Emission layer
EPR	Electron paramagnetic resonance
EPROM	Erasable programmable read-only memory
ESF	Extrinsic stacking fault
ESR	Electron spin resonance
ETL	Electron transport layer
EXAFS	Extended X-ray absorption fine structure
F_4-TCNQ	2,3,5,6-tetrafluoro-7,7,8,8-tetracyano-quinodimethane

fc	Face-centered
fcc	Face-centered cubic
FeRAM	Ferroelectric random access memory
FET	Field-effect transistor
FIB	Focused ion beam
FIR	Far infrared
FKO	Franz–Keldysh oscillation
FLG	Few layer graphene
FPA	Focal plane array
FQHE	Fractional quantum Hall effect
FWHM	Full width at half-maximum
FZ	Float-zone (growth)
Gb	Gigabit
GIZO	GaInZnO
GLAD	Glancing-angle deposition
GRINSCH	Graded-index separate confinement heterostructure
GSMBE	Gas-source molecular beam epitaxy
GST	$Ge_2Sb_2Te_5$
HBL	Hole blocking layer
HBT	Heterobipolar transistor
Hcp	Hexagonally close packed
HCSEL	Horizontal cavity surface-emitting laser
HEMT	High electron mobility transistor
HIGFET	Heterojunction insulating gate FET
HIL	Hole injection layer
HJFET	Heterojunction FET
hh	Heavy hole
HOPG	Highly ordered pyrolitic graphite
HOMO	Highest occupied molecular orbital
HR	High reflection
HRTEM	High-resolution transmission electron microscopy
HTL	Hole transport layer
HWHM	Half-width at half-maximum
IBM	International Business Machines Corporation
IC	Integrated circuit
IDB	Inversion domain boundary
IE	Ionization energy
IEEE	Institute of Electrical and Electronics Engineers[1]
IF	Intermediate frequency
IPAP	Institute of Pure and Applied Physics, Tokyo
IQHE	Integral quantum Hall effect
IR	Infrared

[1]Pronounced Eye-triple-E,

ISF	Intrinsic stacking fault
ITO	Indium tin oxide
JDOS	Joint density of states
JFET	Junction field-effect transistor
KKR	Kramers–Kronig relation
KOH	Potassium hydroxide
KTP	$KTiOPO_4$ material
LA	Longitudinal acoustic (phonon)
LCD	Liquid crystal display
LDA	Local density approximation
LEC	Liquid encapsulated Czochralski (growth)
LED	Light-emitting diode
lh	Light hole
LO	Longitudinal optical (phonon), local oscillator
LPE	Liquid phase epitaxy
LPCVD	Low-pressure chemical vapor deposition
LPP	Longitudinal phonon plasmon (mode)
LST	Lyddane–Sachs–Teller (relation)
LT	Low temperature
LUMO	Lowest unoccupied molecular orbital
LVM	Local vibrational mode
MBE	Molecular beam epitaxy
MEMS	Micro-electro-mechanical system
MESFET	Metal–semiconductor field-effect transistor
MHEMT	Metamorphic HEMT
MIGS	Midgap (surface) states
MILC	Metal-induced lateral crystallization
MIOS	Metal–insulator–oxide–semiconductor
MIR	Mid-infrared
MIS	Metal–insulator–semiconductor
MIT	Massachusetts Institute of Technology
ML	Monolayer
MLC	Multi-level cell
MMIC	Millimeter-wave integrated circuit
MO	Master oscillator
MODFET	Modulation-doped FET
MOMBE	Metalorganic molecular beam epitaxy
MOPA	Master oscillator power amplifier
MOS	Metal–oxide–semiconductor
MOSFET	Metal–oxide–semiconductor field-effect transistor
MOVPE	Metalorganic chemical vapor deposition
MQW	Multiple quantum well
MRAM	Magnetic random access memory
MRS	Materials Research Society
MS	Metal–semiconductor (diode)

MSA	Mobility spectral analysis
MSM	Metal–semiconductor–metal (diode)
MTJ	Magneto-tunneling junction
MWNT	Multi-walled (carbon) nanotube
NDR	Negative differential resistance
NEP	Noise equivalent power
NIR	Near infrared
NMOS	N-channel metal–oxide–semiconductor (transistor)
NTSC	National television standard colors
OLED	Organic light-emitting diode
OMC	Organic molecular crystals
ONO	Oxide/nitride/oxide
OPSL	Optically pumped semiconductor laser
PA	Power amplifier
PBG	Photonic band gap
pc	Primitive cubic
PCM	Phase change memory
PFM	Piezoresponse force microscopy
PHEMT	Pseudomorphic HEMT
PL	Photoluminescence
PLD	Pulsed laser deposition
PLE	Photoluminescence excitation (spectroscopy)
PMC	Programmable metallization cell
PMMA	Poly-methyl methacrylate
PMOS	P-channel metal–oxide–semiconductor (transistor)
PPC	Persistent photoconductivity
PPLN	Perodically poled lithium niobate
PV	Photovoltaic(s)
PWM	Pulsewidth modulation
PZT	$PbTi_xZr_{1-x}O_3$ material
QCL	Quantum cascade laser
QCSE	Quantum confined Stark effect
QD	Quantum dot
QHE	Quantum Hall effect
QW	Quantum well
QWIP	Quantum-well intersubband photodetector
QWR	Quantum wire
RAM	Random access memory
RAS	Reflection anisotropy spectroscopy
RF	Radio frequency
RFID	Radio frequency identification
RGB	Red–green–blue (color system)
RHEED	Reflection high-energy electron diffraction
RIE	Reactive ion etching
RKKY	Ruderman–Kittel–Kasuya–Yoshida (interaction)

rms	Root mean square
ROM	Read-only memory
RRAM	Resistance random access memory
SAGB	Small-angle grain boundary
SAM	Separate absorption and amplification (structure)
sc	Simple cubic
SCH	Separate confinement heterostructure
SdH	Shubnikov–de Haas (oscillation)
SEL	Surface-emitting laser
SEM	Scanning electron microscopy
SET	Single-electron transistor, single-electron tunneling
SGDBR	Sampled grating distributed Bragg reflector
SHG	Second-harmonic generation
si	Semi-insulating
SIA	Semiconductor Industry Association
SIA	Structural inversion asymmetry
SIMS	Secondary ion mass spectrometry
SL	Superlattice
SLC	Single-level cell
SLG	Single layer graphene
SNR	Signal-to-noise ratio
s-o	Spin-orbit (or split-off)
SOA	Semiconductor optical amplifier
SPD	Spectral power distribution
SPICE	Simulation Program with Integrated Circuit Emphasis
SPIE	Society of Photographic Instrumentation Engineers
SPP	Surface plasmon polariton
SPS	Short-period superlattice
sRGB	Standard RGB
SRH	Shockley–Read–Hall (kinetics)
SSR	Side-mode suppression ratio
STEM	Scanning transmission electron microscopy
STM	Scanning tunneling microscopy
SWNT	Single-walled (carbon) nanotube
TA	Transverse acoustic (phonon)
TAS	Thermal admittance spectroscopy
TCO	Transparent conductive oxide
TE	Transverse electric (polarization)
TED	Transferred electron device
TEGFET	Two-dimensional electron gas FET
TEM	Transmission electron microscopy
TES	Two-electron satellite
TF	Thermionic field emission
TFET	Transparent FET, tunneling FET
TFT	Thin-film transistor

TM	Transverse magnetic (polarization)
TMAH	Tetramethyl-ammonium-hydroxide
TMR	Tunnel-magnetoresistance
TO	Transverse optical (phonon)
TOD	Turn-on delay (time)
TPA	Two-photon absorption
TSO	Transparent semiconducting oxide
UHV	Ultrahigh vacuum
UV	Ultraviolet
VCA	Virtual crystal approximation
VCO	Voltage-controlled oscillator
VCSEL	Vertical-cavity surface-emitting laser
VFF	Valence force field
VGF	Vertical gradient freeze (growth)
VIS	Visible
VLSI	Very large scale integration
WGM	Whispering gallery mode
WKB	Wentzel–Kramer–Brillouin (approximation or method)
WS	Wigner–Seitz (cell)
YSZ	Yttria-stabilized zirconia (ZrO_2)
X	Exciton
XSTM	Cross-sectional STM
XX	Biexciton
Z	Atomic number ($Z = 2$ for helium)
ZnPc	Zinc-phthalocyanine

Symbols

α	Madelung constant, disorder parameter, linewidth enhancement factor
$\alpha(\omega)$	Absorption coefficient
α_{m}	Mirror loss
α_{n}	Electron ionization coefficient
α_{p}	Hole ionization coefficient
α_{T}	Base transport factor
β	Used as abbreviation for $e/(k_{\mathrm{B}}T)$, spontaneous emission coefficient, two-photon absorption coefficient
γ	Broadening parameter, Grüneisen parameter, emitter efficiency
Γ	Broadening parameter
$\gamma_1, \gamma_2, \gamma_3$	Luttinger parameters
δ_{ij}	Kronecker symbol
Δ_0	Spin-orbit splitting
$\epsilon(\omega)$	Dielectric function
ϵ_0	Permittivity of vacuum
ϵ_{i}	Dielectric constant of insulator
ϵ_{r}	Relative dielectric function
ϵ_{s}	Dielectric function of semiconductor ($= \epsilon_{\mathrm{r}}\epsilon_0$)
ϵ_S	Gain compression coefficient
ϵ_{ij}	Components of strain tensor
η	Quantum efficiency
η_{d}	Differential quantum efficiency
η_{ext}	External quantum efficiency
η_{int}	Internal quantum efficiency
η_{w}	Wall-plug efficiency
θ	Angle
Θ_{D}	Debye temperature
Θ_{B}	Typical phonon energy parameter
κ	Imaginary part of index of refraction, heat conductivity
λ	Wavelength

λ_p	Plasma wavelength
μ	Mobility
μ_0	Magnetic susceptibility of vacuum
μ_h	Hole mobility
μ_H	Hall mobility
μ_n	Electron mobility
ν	Frequency
Π	Peltier coefficient
π_{ij}	Components of piezoresistivity tensor
ρ	Mass density, charge density, resistivity
ρ_{ij}	Components of resistivity tensor
ρ_P	Polarization charge density (per volume)
σ	Standard deviation, conductivity, stress, effective mass ratio
Σn	Grain boundary type
σ_{ij}	Components of stress tensor, conductivity tensor
σ_n	Electron capture cross section
σ_p	Hole capture cross section
σ_P	Polarization charge density (per area)
τ	Lifetime, time constant
τ_n	Electron (minority carrier) lifetime
τ_{nr}	Non-radiative lifetime
τ_p	Hole (minority carrier) lifetime
τ_r	Radiative lifetime
Θ_D	Debye temperature
ϕ	Phase
Φ_0	Photon flux
χ	Electron affinity, electric susceptibility
χ_{ex}	Light extraction efficiency
χ_{sc}	Electron affinity (of semiconductor)
$\chi(r)$	Envelope wavefunction
ψ	Angle
$\Psi(r)$	Wavefunction
ω	Angular frequency
ω_{LO}	Longitudinal optical phonon frequency
ω_p	Plasma frequency
ω_{TO}	Transverse optical phonon frequency
Ω	Interaction parameter
a	Hydrostatic deformation potential, lattice constant
$\mathbf{a}$	Accelaration
A	Area, band-impurity recombination coefficient
A, A	Vector potential
A^*	Richardson constant
a_0	(Cubic) lattice constant
b	Bowing parameter, deformation potential

$\mathbf{b}$	Burger's vector
B	Bimolecular recombination coefficient, bandwidth
$\mathbf{B}$, B	Magnetic induction field
c	Velocity of light in vacuum, lattice constant (along c-axis)
C	Capacitance, spring constant, Auger recombination coefficient
C_n, C_p	Auger recombination coefficient
C_{ij}, C_{ijkl}	Elastic constants
d	Distance, shear deformation potential
d_i	Insulator thickness
D	Density of states, diffusion coefficient
$\mathbf{D}$, D	Electric Displacement field
D^*	Detectivity
$D_\mathrm{e}(E)$	Electron density of states
$D_\mathrm{h}(E)$	Hole density of states
D_n	Electron diffusion coefficient
D_p	Hole diffusion coefficient
e	Elementary charge
e_i	Strain components (Voigt notation)
E	Energy, electric field
$\mathbf{E}$, E	Electric field
E_A	Energy of acceptor level
E_A^b	Acceptor ionization energy
E_C	Energy of conduction-band edge
E_xc	Exchange interaction energy
E_D	Energy of donor level, Dirac energy (in graphene)
E_D^b	Donor ionization energy
E_F	Fermi energy
$E_{\mathrm{F_n}}$	Electron quasi-Fermi energy
$E_{\mathrm{F_p}}$	Hole quasi-Fermi energy
E_g	Band gap
E_i	Intrinsic Fermi level
E_m	Maximum electric field
E_P	Energy parameter
E_V	Energy of valence-band edge
E_vac	Energy of vacuum level
E_X	Exciton energy
E_X^b	Exciton binding energy
E_XX	Biexciton energy
f	Oscillator strength
F	Free energy
$\mathbf{F}$, F	Force
$F(M)$	Excess noise factor
F_B	Schottky barrier height
f_e	Fermi–Dirac distribution function (for electrons)

f_i	Ionicity
f_h	Fermi–Dirac distribution function for holes ($= 1 - f_\mathrm{e}$)
F_n	Electron quasi-Fermi energy
F_p	Hole quasi-Fermi energy
F_P	Purcell factor
g	Degeneracy, g-factor, gain
G	Free enthalpy, generation rate
$\mathbf{G}$, G	Reciprocal lattice vector
g_m	Transconductance
G_th	Thermal generation rate
h	Planck constant
H	Enthalpy, Hamiltonian
$\mathbf{H}$, H	Magnetic field
$\hbar$	$h/(2\pi)$
i	Imaginary number
I	Current, light intensity
I_D	Drain current
I_s	Saturation current
I_sc	Short circuit current
I_thr	Threshold current
j	Current density, orbital momentum
j_s	Saturation current density
j_thr	Threshold current density
k	Wavenumber, Boltzmann constant
$\mathbf{k}$	Wavevector
k_B	Boltzmann constant
k_F	Fermi wavevector
l	Angular orbital momentum
L	Length of line element
$\mathbf{L}$	Line vector (of dislocation)
L_D	Diffusion length
L_z	Quantum-well thickness
m	Mass
m_0	Free electron mass
m^*	Effective mass
m^*_{ij}	Effective mass tensor
m_c	Cyclotron mass
M	Mass, multiplication factor
$\mathbf{M}$, M	Magnetization
m_e	Effective electron mass
m_h	Effective hole mass
m_j	Magnetic quantum number
m_l	Longitudinal mass
m_r	Reduced mass

m_t	Transverse mass
n	Electron concentration (in conduction band), ideality factor
$\mathbf{n}$	Normal vector
$N(E)$	Number of states
n^*	Complex index of refraction ($= n_r + i\kappa$)
N_A	Acceptor concentration
N_c	Critical doping concentration
N_C	Conduction-band edge density of states
N_D	Donor concentration
n_i	Intrinsic electron concentration
n_{if}	Ideality factor due to image force effect
n_r	Index of refraction (real part)
n_s	Sheet electron density
N_t	Trap concentration
n_{tr}	Transparency electron concentration
n_{thr}	Threshold electron concentration
N_V	Valence-band edge density of states
p	Pressure, free hole density
$\mathbf{p}$, p	Momentum
P	Power
$\mathbf{P}$, P	Electric polarization
p_{cv}	Momentum matrix element
p_i	Intrinsic hole concentration
q	Charge
$\mathbf{q}$, q	Heat flow
Q	Charge, quality factor, impurity-bound exciton localization energy
r	Radius
r	Spatial coordinate
R	Resistance, radius, recombination rate
R_λ	Responsivity
$\mathbf{R}$	Vector of direct lattice
r_H	Hall factor
R_H	Hall coefficient
s	spin
S	Entropy, Seebeck coefficient, total spin, surface index, surface recombination velocity
S_{ij}, S_{ijkl}	Stiffness coefficients
t	Time
t_{ox}	(Gate) oxide thickness
T	temperature
u	Displacement, cell-internal parameter
U	energy
u_{nk}	Bloch function
v, v	Velocity

V	Volume, voltage, potential
$V(\lambda)$	(Standardized) sensitivity of human eye
V_a	Unit-cell volume
V_A	Early voltage
V_{bi}	Built-in voltage
v_g	Group velocity
V_G	Gate voltage
V_{oc}	Open circuit voltage
V_P	Pinch-off voltage
v_s	Velocity of sound, drift-saturation velocity
V_{SD}	Source-drain voltage
v_{th}	Thermal velocity
w	Depletion-layer width, width
w_B	Base width (bipolar transistor)
W_m	(Metal) work function
X	Electronegativity
Y	Young's modulus, CIE brightness parameter
Z	Partition sum, atomic order number

Physical Constants

Constant	Symbol	Numerical value	Unit
Speed of light in vacuum	c_0	2.99792458×10^{8}	ms^{-1}
Permeability of vacuum	μ_0	$4\pi \times 10^{-7}$	NA^{-2}
Permittivity of vacuum	$\epsilon_0 = (\mu_0 c_0^2)^{-1}$	$8.854187817 \times 10^{-12}$	Fm^{-1}
Elementary charge	e	$1.60217733 \times 10^{-19}$	C
Electron mass	m_e	$9.1093897 \times 10^{-31}$	kg
Planck constant	h	$6.6260755 \times 10^{-34}$	J s
	$\hbar = h/(2\pi)$	$1.05457266 \times 10^{-34}$	J s
	$\hbar$	6.582122×10^{-16}	eV s
Boltzmann constant	k_B	1.380658×10^{-23}	JK^{-1}
		8.617385×10^{-5}	meVK^{-1}
von-Klitzing constant	R_H	25812.8056	Ω
Rydberg constant		13.6056981	eV
Bohr radius	a_B	$5.29177249 \times 10^{-11}$	m
Avogadro constant	N_A	6.0221418×10^{23}	mol^{-1}

Chapter 1
Introduction

The proper conduct of science lies in the pursuit of Nature's puzzles, wherever they may lead.

J.M. Bishop [2]

Abstract Important dates and events in the history of semiconductors are chronologically listed, from the early days (Volta, Seebeck and Faraday) to the latest achievements like the blue and white LED. Many known and not so well known scientists are mentioned. Also a list of semiconductor related Nobel prizes and their winners is given.

The historic development of semiconductor physics and technology began in the second half of the 19th century. Interesting discussions of the history of the physics and chemistry of semiconductors can be found in [3–5]. The development of crystal growth is covered in [6]. The history of semiconductor industry can be followed in [7, 8]. In 1947, the commercial realization of the transistor was the impetus to a fast-paced development that created the electronics and photonics industries. Products founded on the basis of semiconductor devices such as computers (CPUs, memories), optical-storage media (lasers for CD, DVD), communication infrastructure (lasers and photodetectors for optical-fiber technology, high frequency electronics for mobile communication), displays (thin film transistors, LEDs), projection (laser diodes) and general lighting (LEDs) are commonplace. Thus, fundamental research on semiconductors and semiconductor physics and its offspring in the form of devices has contributed largely to the development of modern civilization and culture.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_1

1.1 Timetable and Key Achievements

In this section early important milestones in semiconductor physics and technology are listed.

1782
A. Volta—coins the phrase 'semicoibente' (semi-insulating) which was translated then into English as 'semiconducting' [9].

1821
T.J. Seebeck—discovery of thermopower (electrical phenomena upon temperature difference) in metals and PbS, FeS_2, $CuFeS_2$ [10].

1833
M. Faraday—discovery of the temperature dependence of the conductivity of Ag_2S (sulphuret of silver, negative dR/dT) [11].

1834
J. Peltier—discovery of the Peltier effect (cooling by current) [12].

1873
W. Smith—discovery of photoconductivity in selenium [13]. Early work on photoconductivity in Se is reviewed in [14, 15].

1874
F. Braun[1]—discovery of rectification in metal–sulfide semiconductor contacts [17], e.g. for $CuFeS_2$ and PbS. The current through a metal–semiconductor contact is nonlinear (as compared to that through a metal, Fig. 1.1), i.e. a deviation from Ohm's law. Braun's structure is similar to a MSM diode.

1876
W.G. Adams and R.E. Day—discovery of the photovoltaic effect in selenium [18].

W. Siemens—large response from selenium photoconductor [19], made by winding two thin platinum wires to the surface of a sheet of mica, and then covering the surface with a thin film of molten selenium. Resistance ratio between dark and illuminated by sunlight was larger than ten [19] and measured to 14.8 in [20].

1879
E.H. Hall—measurement of the transverse potential difference in a thin gold leaf on glass [21, 22]. Experiments were continued by his mentor H.A. Rowland [23]. A detailed account of the discovery of the Hall efect is given in [24, 25].

1883
Ch. Fritts—first solar cell, based on an gold/selenium rectifier [20]. The efficiency was below 1 %.

[1] F. Braun made his discoveries on metal–semiconductor contacts in Leipzig while a teacher at the Thomasschule zu Leipzig [16]. He conducted his famous work on vacuum tubes later as a professor in Strasbourg, France.

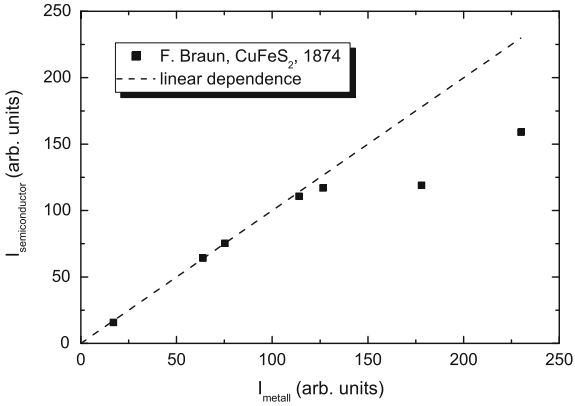


Fig. 1.1 Current through a silver–$CuFeS_2$–silver structure as a function of the current through the metal only, 1874. Data points are for different applied voltages. Experimental data from [17]

1901
J.C. Bose—point contact detector for electromagnetic waves based on galena (PbS) [26]. At the time, the term semiconductor was not introduced yet and Bose speaks about 'substances of a certain class (...) presenting a decreasing resistance to the passage of the electric current with an increasing impressed electromotive force'.

1906
G.W. Pickard—rectifier based on point contact (cat's whisker) diode on silicon [27-29]. Erroneously, the rectifying effect was attributed to a thermal effect, however, the drawing of the 'thermo-junction' (TJ in Fig. 1.2) developed into the circuit symbol for a diode (cmp. Fig. 21.61a).

1907
H.J. Round—discovery of electroluminescence investigating yellow and blue light emission from SiC [30].

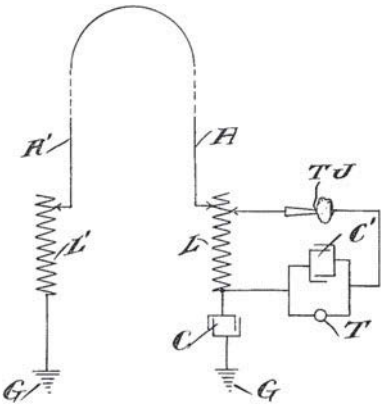


Fig. 1.2 Circuit diagram for a radio receiver with a point-contact diode (TJ). Adapted from [27]

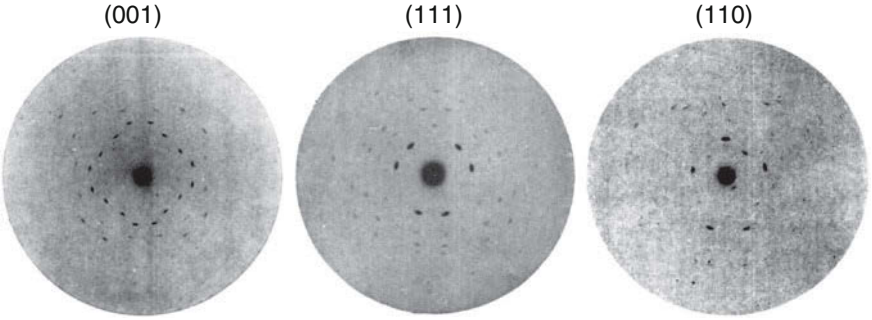


Fig. 1.3 Laue images of 'regular' (cubic) ZnS along three major crystallographic directions, directly visualizing their 4-, 3- and 2-fold symmetry. Adapted from [41]

K. Bädeker—preparation of metal (e.g. Cd, Cu) oxides and sulfides and also CuI from metal layers using a vapor phase transport method [31].[2] CuI is reported transparent (∼200 nm thick films) with a specific resistivity of $\rho = 4.5 \times 10^{-2}\,\Omega\,\text{cm}$, the first transparent conductor.[3] Also CdO (films of thickness 100–200 nm) is reported to be highly conductive, $\rho = 1.2 \times 10^{-3}\,\Omega\,\text{cm}$, and orange-yellow in color, the first reported TCO (transparent conductive oxide).

1909
K. Bädeker—discovery of doping. Controlled variation of the conductivity of CuI by dipping into iodine solutions (e.g. in chloroform) of different concentrations [34].

1910
W.H. Eccles—negative differential resistance of contacts with galena (PbS), construction of crystal oscillators [38].[4]

1911
The term 'Halbleiter' (semiconductor) is introduced for the first time by J. Weiss [39] and J. Königsberger and J. Weiss [40]. Königsberger preferred the term 'Variabler Leiter' (variable conductor).

1912
M. von Laue—X-ray diffraction of bulk crystals including ZnS (Fig. 1.3) [41, 42].

[2] This work was conducted as Habilitation in the Physics Institute of Universität Leipzig. Bädeker became subsequently professor in Jena and fell in WW I. His scientific contribution to semiconductor physics is discussed in [32, 33].

[3] CuI is actually a p-type transparent conductor; at that time the positive sign of the Hall effect [34, 35] could not be interpreted as hole conduction yet.

[4] Historical remarks on Eccles' contributions to radio technology can be found in [36, 37].

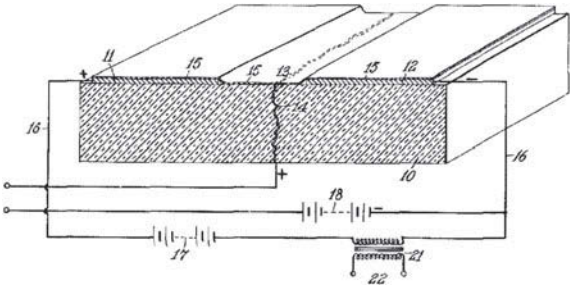


Fig. 1.4 Sketch of a field-effect transistor, 1926. From [46]

1925
J.E. Lilienfeld[5]—proposal of the metal-semiconductor field-effect transistor (MESFET) [46], with suggested copper sulfide thin film channel and aluminum gate.[6] (Fig. 1.4). Lilienfeld was also awarded patents for a depletion mode MOSFET [48] with proposed copper sulfide, copper oxide or lead oxide channel and current amplification with nppn- and pnnp-transistors [49]. Due to the lack of other publications of Lilienfeld on transistors, it is under discussion whether Lilienfeld just patented ideas or also build working devices with mounting evidence for the latter [44, 47, 50].

1927
A. Schleede, H. Buggisch—synthesis of pure, stoichiometric PbS, influence of sulphur excess and impurities [51].
A. Schleede, E. Körner—activation of luminescence of ZnS [52, 53].

1928
F. Bloch—quantum mechanics of electrons in a crystal lattice, 'Bloch functions' [54].
O.V. Losev—description of the light emitting diode[7] (SiC) [58]; light emission was observed in forward direction and close to breakdown (Fig. 1.5a). Also current modulation of LED light output was reported (Fig. 1.5b) [58].

1929
R. Peierls—explanation of positive (anomalous) Hall effect with unoccupied electron states [59, 60].

[5] After obtaining his PhD in 1905 from the Friedrich-Wilhelms-Universität Berlin, Julius Edgar Lilienfeld joined the Physics Department of Universität Leipzig and worked on gas liquefaction and with Lord Zeppelin on hydrogen-filled blimps. In 1910 he became professor at the Universität Leipzig where he mainly researched on X-rays and vacuum tubes [43]. To the surprise of his colleagues he left in 1926 to join a US industrial laboratory [44, 45].

[6] In [44] it is suggested that the device works as a npn transistor, in [47] it is suggested to be a JFET.

[7] The historic role of Losev regarding the invention of the LED and oscillators is discussed in [55–57].

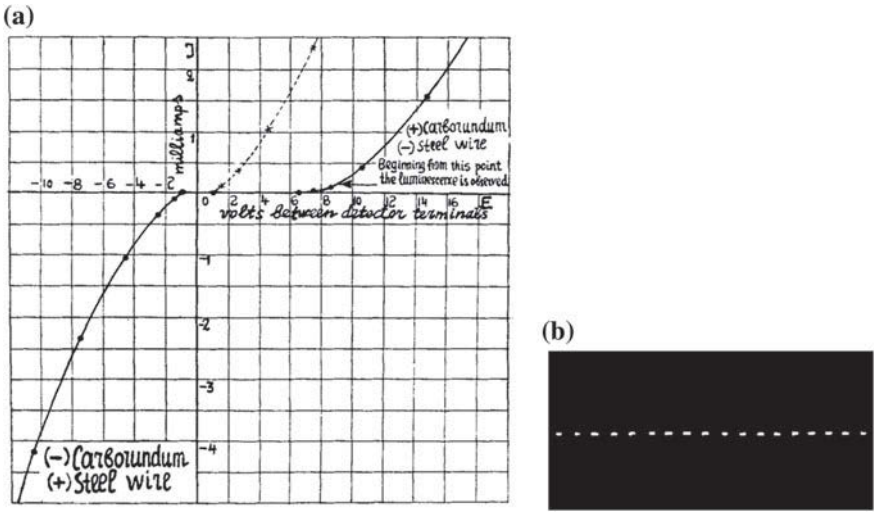


Fig. 1.5 (**a**) I–V characteristic of SiC/steel wire light emitting diode. The *dotted curve* is the flipped curve for negative voltage (3rd quadrant). (**b**) Recording of current modulated (at 500 Hz) LED on moving photographic plate. Adapted from [58]

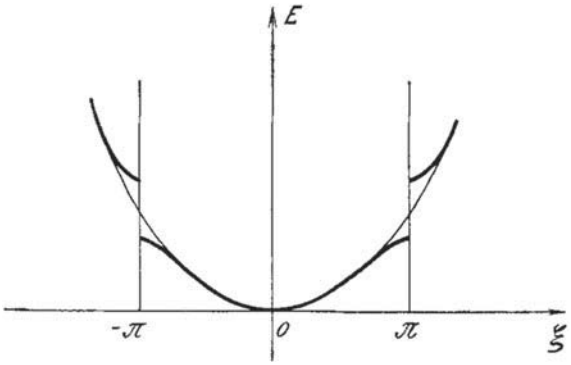


Fig. 1.6 First band structure calculation ($\xi = k\,a$). Adapted from [62]

1930
R. Peierls—first calculation of a band structure and band gap[8] (Fig. 1.6) [62].
1931
W. Heisenberg—theory of hole (‘Löcher’) states [63].
R. de L. Kronig and W.G. Penney—properties of periodic potentials in solids [64].
A.H. Wilson[9]—development of band-structure theory [67, 68].

[8]Peierls performed this work at suggestion of W. Pauli at ETH Zürich. The mathematical problem of Schrödinger’s equation with a sinusoidal potential had been already treated by M.J.O Strutt in 1928 [61].

[9]Wilson was theoretical physicist in Cambridge, who spent a sabbatical with Heisenberg in Leipzig and applied the brand new field of quantum mechanics to issues of electrical conduction,

1933
C. Wagner—excess ('Elektronenüberschuss-Leitung', n-type) and defect ('Elektronen-Defektleitung', p-type) conduction [69]. Anion deficiency in ZnO causes conducting behavior [70].

1934
C. Zener—Zener tunneling [71].

1936
J. Frenkel—description of excitons [72].

1938
B. Davydov—theoretical prediction of rectification at pn-junction [73] and in Cu_2O [74].
W. Schottky—theory of the boundary layer in metal–semiconductor contacts [75], being the basis for Schottky contacts and field-effect transistors.
N.F. Mott—metal–semiconductor rectifier theory [76, 77].
R. Hilsch and R.W. Pohl—three-electrode crystal (KBr) [78].

1940
R.S. Ohl—Silicon-based photoeffect (solar cell, Fig. 1.7) [79] from a pn-junction formed within a slab of polycrystalline Si fabricated with directed solidification due to different distribution coefficients of p- and n-dopants (boron and phosphorus, cmp. Fig. 4.6b) (J. Scaff and H. Theurer) [80, 81].

1941
R.S. Ohl—Silicon rectifier with point contact [82, 83] (Fig. 1.8), building on work from G.W. Pickard (1906) and using metallurgically refined and intentionally doped silicon (J. Scaff and H. Theurer) [80].

1942
K. Clusius, E. Holz and H. Welker—rectification in germanium [84].

1945
H. Welker—patents for JFET and MESFET [85].

1947
W. Shockley, J. Bardeen and W. Brattain fabricate the first transistor in the AT&T

(Footnote 9 continued)
first in metals and then in semiconductors. When he returned to Cambridge, Wilson urged that attention be paid to germanium but, as he expressed it long afterward, 'the silence was deafening' in response. He was told that devoting attention to semiconductors, those messy entities, was likely to blight his career among physicists. He ignored these warnings and in 1939 brought out his famous book 'Semiconductors and Metals' [65] which explained semiconductor properties, including the much-doubted phenomenon of intrinsic semiconductivity, in terms of electronic energy bands. His academic career seems indeed to have been blighted, because despite his great intellectual distinction, he was not promoted in Cambridge (he remained an assistant professor year after year) [66]. Compare the remark of W. Pauli (p. 205).

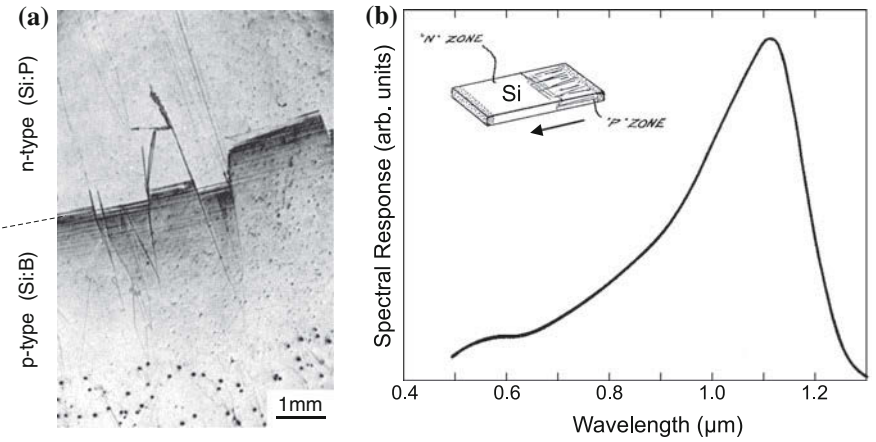


Fig. 1.7 (**a**) Optical image of directionally solidified silicon. The lower part contains predominantly boron, the upper part contains predominantly phosphorous. First the growth is porous and subsequently columnar. Adapted from [80]. (**b**) Spectral response of silicon pn-junction photoelement, 1940. The *inset* depicts schematically a Si slab with built-in pn-junction formed during directed solidification as shown in panel (**a**). The *arrow* denotes the direction of solidification (cmp. Fig. 4.6). Adapted from [79]

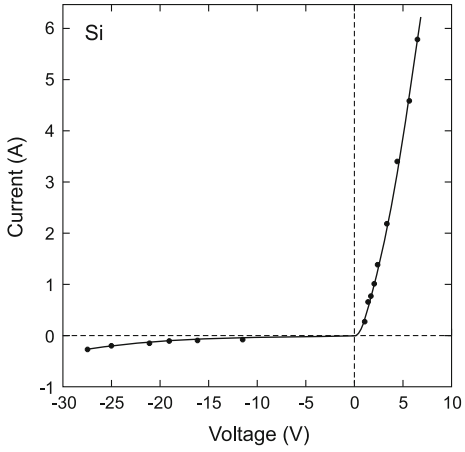


Fig. 1.8 Characteristics of a silicon rectifier, 1941. Adapted from [82]

Bell Laboratories, Holmdel, NJ in an effort to improve hearing aids [86].[10] Strictly speaking the structure was a point-contact transistor. A 50-μm wide slit was cut with a razor blade into gold foil over a plastic (insulating) triangle and pressed with a spring on n-type germanium (Fig. 1.9a) [87]. The surface region of the germanium is p-type due to surface states and represents an inversion layer. The two gold con-

[10] Subsequently, AT&T, under pressure from the US Justice Department's antitrust division, licensed the transistor for $25,000. This action initiated the rise of companies like Texas Instruments, Sony and Fairchild.

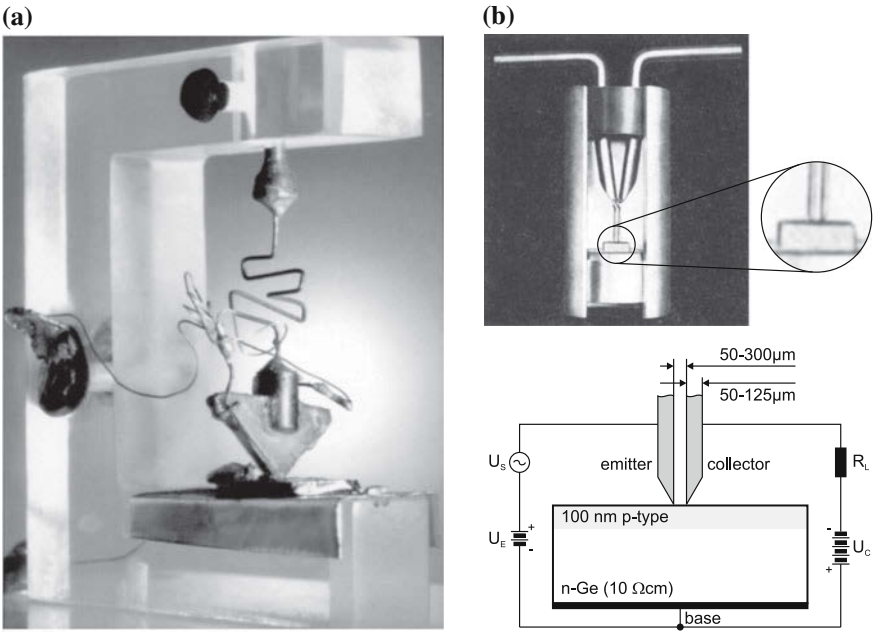


Fig. 1.9 (**a**) The first transistor, 1947 (length of side of wedge: 32 mm). (**b**) Cutaway model of a 1948 point contact transistor ('Type A') based on n-type bulk Ge ($n = 5 \times 10^{14}\,\mathrm{cm}^{-3}$) and common base circuit diagram. The surface region (∼100 nm depth) of the Ge is p-type due to surface states and represents an inversion layer. The two wires are made from phosphor bronze. Adapted from [88]

tacts form emitter and collector, the large-area back contact of the germanium the base contact [88]. For the first time, amplification was observed [89]. Later models use two close point contacts made from wires with their tips cut into wedge shape (Fig. 1.9b) [88].[11] More details about the history and development of the semiconductor transistor can be found in [90], written on the occasion of the 50th anniversary of its invention.

1948
W. Shockley—invention of the bipolar (junction) transistor [91].

1952
H. Welker—fabrication of III–V compound semiconductors[12] [94–97]
W. Shockley—description of today's version of the (J)FET [98].

[11] The setup of Fig. 1.9b represents a common base circuit. In a modern bipolar transistor, current amplification in this case is close to unity (Sect. 24.2.2). In the 1948 germanium transistor, the reversely biased collector contact is influenced by the emitter current such that current amplification $\partial I_C/\partial I_E$ for constant U_C was up to 2–3. Due to the collector voltage being much larger than the emitter voltage, a power gain of ∼125 was reported [88].

[12] An early concept for III–V semiconductors was developed in [92, 93].

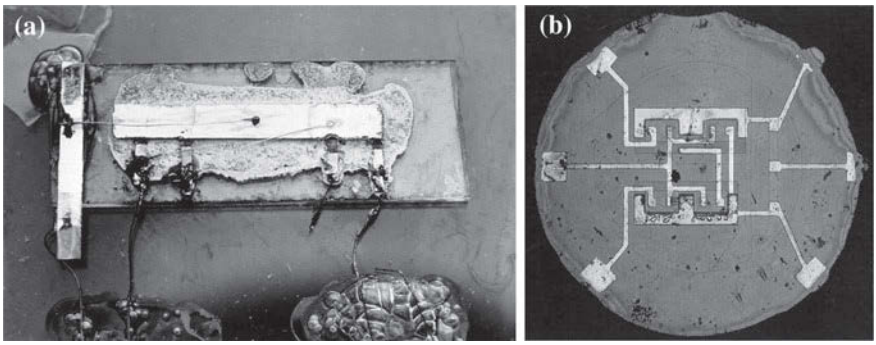


Fig. 1.10 (**a**) The first integrated circuit, 1958 (germanium, $11 \times 1.7\,\text{mm}^2$). (**b**) The first planar integrated circuit, 1959 (silicon, diameter: 1.5 mm)

1953
G.C. Dacey and I.M. Ross—first realization of a JFET [99].
D.M. Chapin, C.S. Fuller and G.L. Pearson—invention of the silicon solar cell at Bell Laboratories [100]. A single $2\,\text{cm}^2$ photovoltaic cell from Si, Si:As with an ultra-thin layer of Si:B, with about 6 % efficiency generated 5 mW of electrical power.[13] Previously existing solar cells based on selenium had very low efficiency (<0.5 %).

1958
J.S. Kilby made the first integrated circuit at Texas Instruments. The simple 1.3 MHz RC-oscillator consisted of one transistor, three resistors and a capacitor on a $11 \times 1.7\,\text{mm}^2$ Ge platelet (Fig. 1.10a). J.S. Kilby filed in 1959 for a US patent for miniaturized electronic circuits [101]. At practically the same time R.N. Noyce from Fairchild Semiconductors, the predecessor of INTEL, invented the integrated circuit on silicon using planar technology [102]. A detailed and (very) critical view on the invention of the integrated circuit can be found in [103].

Figure 1.10b shows a flip-flop with four bipolar transistors and five resistors. Initially, the invention of the integrated circuit[14] met scepticism because of concerns regarding yield and the achievable quality of the transistors and the other components (such as resistors and capacitors).

[13] A solar cell with 1 W power cost $300 in 1956 ($3 in 2004). Initially, 'solar batteries' were only used for toys and were looking for an application. H. Ziegler proposed the use in satellites in the 'space race' of the late 1950s.

[14] The two patents led to a decade-long legal battle between Fairchild Semiconductors and Texas Instruments. Eventually, the US Court of Customs and Patent Appeals upheld R.N. Noyce's claims on interconnection techniques but gave J.S. Kilby and Texas Instruments credit for building the first working integrated circuit.

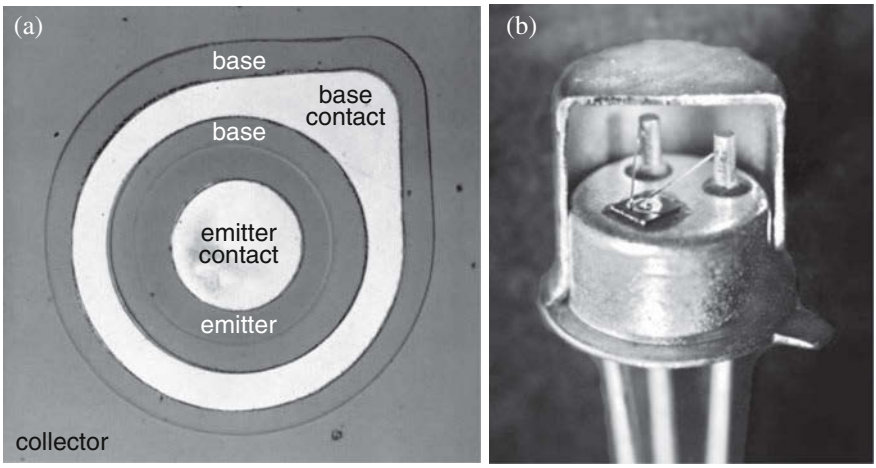


Fig. 1.11 (**a**) Optical image of planar pnp silicon transistor (2N1613 [110]), 1959. The contacts are Al surfaces (not bonded). (**b**) Housing of such transistor cut open

1959
J. Hoerni[15] and R. Noyce—first realization of a planar transistor (in silicon) (Fig. 1.11) [105–109].

1960
D. Kahng and M.M. Atalla—first realization of a MOSFET [111, 112].

1962
The first semiconductor laser on GaAs basis at 77 K at GE [113, 114] (Fig. 1.12) and at IBM [115] and MIT [116].
First visible laser diode [117].[16]

1963
Proposal of a double heterostructure laser (DH laser) by Zh.I. Alferov [120] and H. Kroemer [121].
J.B. Gunn—discovery of the Gunn effect, the spontaneous microwave oscillations in GaAs and InP at sufficiently large applied electric field (due to negative differential resistance) [122].

1966
C.A. Mead—proposal of the MESFET ('Schottky Barrier Gate FET') [123].

[15]The Swiss born Jean Hoerni also contributed $12 000 for the building of the first school in the Karakoram Mountain area in Pakistan and has continued to build schools in Pakistan and Afghanistan as described in [104].

[16]Remarks on the discovery and further development of the laser diode can be found in [118, 119].

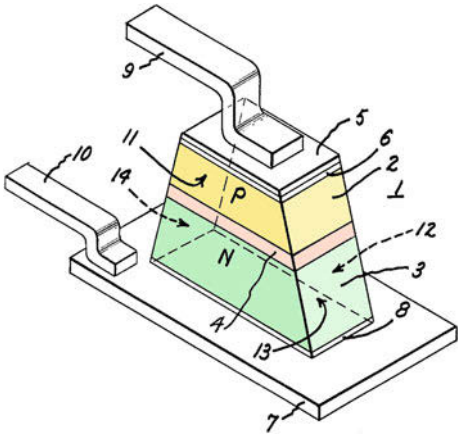


Fig. 1.12 Schematics of GaAs-based laser diode. The active layer is highlighted in *red*. Adapted from [114]

1967
Zh.I. Alferov—report of the first DH laser on the basis of GaAsP at 77 K [124, 125].
W.W. Hooper and W.I. Lehrer—first realization of a MESFET [126].

1968
DH laser on the basis of GaAs/AlGaAs at room temperature, independently developed by Zh.I. Alferov [127] and I. Hayashi [128].
GaP:N LEDs with yellow-green emission (550 nm) and 0.3 % efficiency [129].

1968
SiC blue LED with efficiency of 0.005 % [130].

1975
First monolithic microwave integrated circuit (MMIC) (Fig. 1.13) [131]

1992
S. Nakamura—growth of high-quality group-III–nitride thin films [132], blue nitride heterostructure LED with efficiency exceeding 10 % (1995) [133] (Fig. 1.14a). Later the white LED was built by combining a blue LED with yellow phosphors (Fig. 1.14b, c).

1.2 Nobel Prize Winners

Several Nobel Prizes[17] have been awarded for discoveries and inventions in the field of semiconductor physics (Fig. 1.15).
1909
Karl Ferdinand Braun
'in recognition of his contributions to the development of wireless telegraphy'

[17] www.nobel.se.

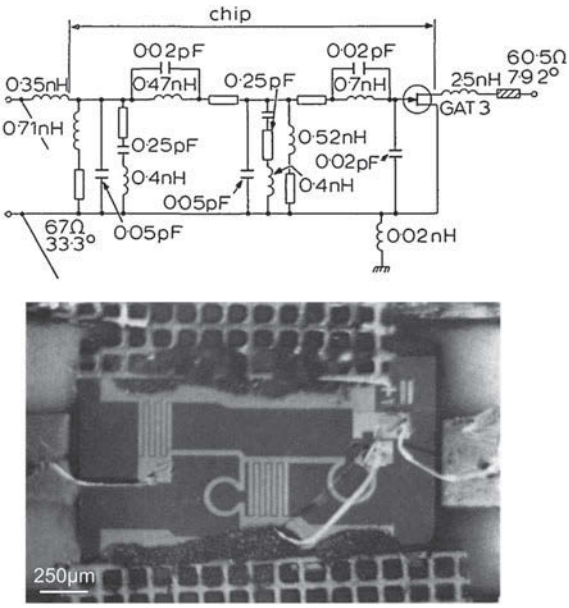


Fig. 1.13 Equivalent circuit and optical image of first monolithic microwave integrated circuit (exhibiting gain (4.5 ± 0.9 dB) in the frequency range 7.0–11.7 GHz). Adapted from [131]

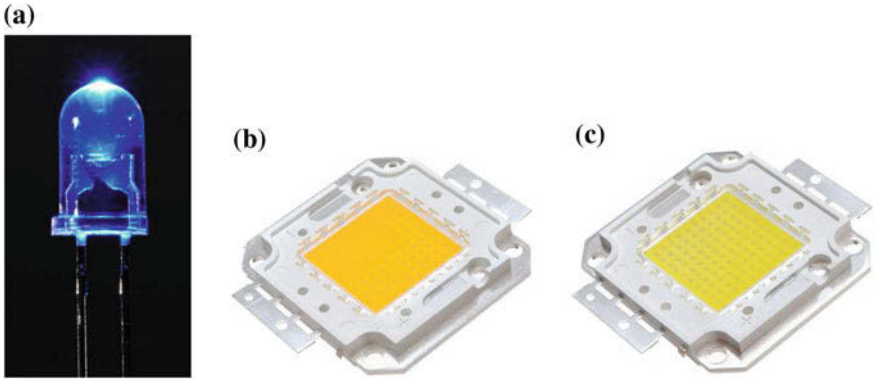


Fig. 1.14 (**a**) Blue LED (standard housing). 50 W, 4000 lm (**b**) warm white and (**c**) cold white LED ($45 \times 45\,\text{mm}^2$)

1914
Max von Laue 'for his discovery of the diffraction of X-rays by crystals'

1915
Sir William Henry Bragg
William Lawrence Bragg
'for their services in the analysis of crystal structure by means of X-rays'

1946
Percy Williams Bridgman
'for the invention of an apparatus to produce extremely high pressures, and for the discoveries he made therewith in the field of high pressure physics'

1953
William Bradford Shockley
John Bardeen
Walter Houser Brattain
'for their researches on semiconductors and their discovery of the transistor effect'

1973
Leo Esaki
'for his experimental discoveries regarding tunneling phenomena in semiconductors'

1985
Klaus von Klitzing
'for the discovery of the quantized Hall effect'

1998
Robert B. Laughlin
Horst L. Störmer
Daniel C. Tsui
'for their discovery of a new form of quantum fluid with fractionally charged excitations'

2000
Zhores I. Alferov
Herbert Kroemer
'for developing semiconductor heterostructures used in high-speed and optoelectronics'
Jack St. Clair Kilby
'for his part in the invention of the integrated circuit'

2009
Willard S. Boyle
George E. Smith
'for the invention of an imaging semiconductor circuit—the CCD sensor'

1909
Karl Ferdinand Braun
(1850–1918)

1914
Max von Laue
(1879–1960)

1915
Sir William Henry Bragg
(1862–1942)

1915
William Laurence Bragg
(1890–1971)

1946
Percy Williams Bridgman
(1882–1961)

1953
William B. Shockley
(1910–1989)

1953
John Bardeen
(1908–1991)

1953
Walter Hauser Brattain
(1902–1987)

1973
Leo Esaki
(*1925)

1985
Klaus von Klitzing
(*1943)

1998
Robert B. Laughlin
(*1930)

1998
Horst L. Störmer
(*1949)

1998
Daniel C. Tsui
(*1939)

2000
Zhores I. Alferov
(*1938)

2000
Herbert Kroemer
(*1928)

2000
Jack St. Clair Kilby
(1923–2005)

Fig. 1.15 Winners of Nobel Prize in Physics and year of award with great importance for semiconductor physics

2009
Willard S. Boyle
(1924–2011)

2009
George E. Smith
(*1930)

2010
Andre Geim
*1958

2010
Konstantin Novoselov
*1974

2014
Isamu Akasaki
(*1929)

2014
Hiroshi Amano
(*1960)

2014
Shuji Nakamura
(*1954)

Fig. 1.15 (continued)

2010
Andre Geim
Konstantin Novoselov
'for groundbreaking experiments regarding the two-dimensional material graphene'

2014
Isamu Akasaki
Hiroshi Amano
Shuji Nakamura
'for the invention of efficient blue light-emitting diodes which has enabled bright and energy-saving white light sources'.

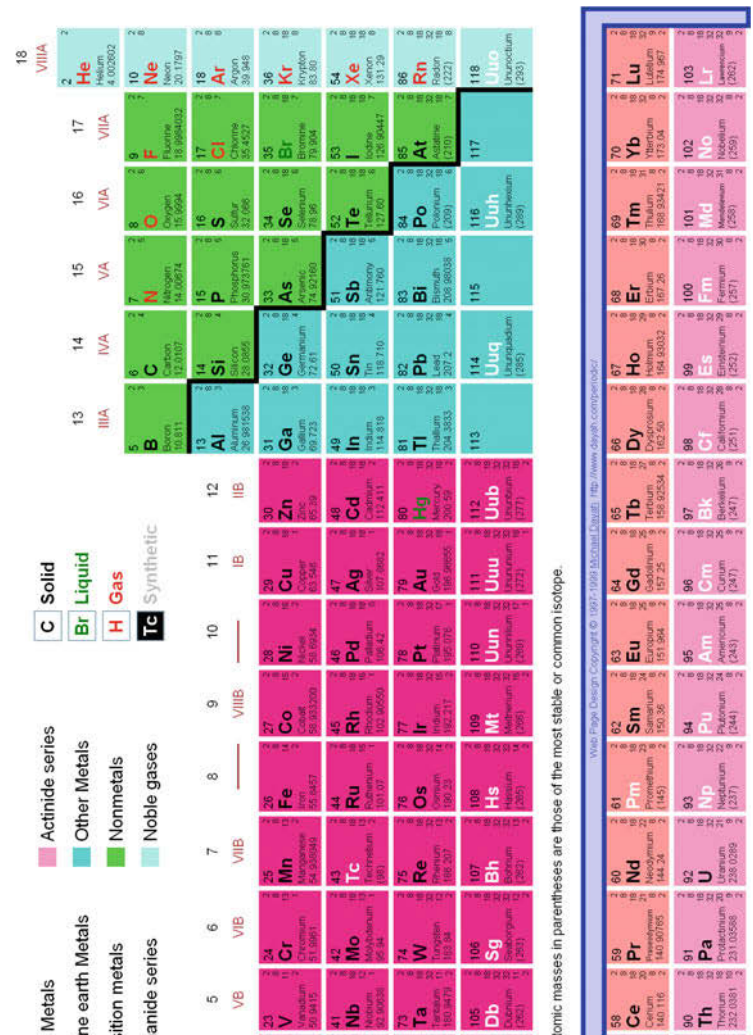


Fig. 1.16 Periodic table of elements. From [147]

Table 1.1 Physical properties of various semiconductors at room temperature

	S	a_0 (nm)	E_g (eV)	m_e^*	m_h^*	ϵ_0	n_r	μ_e (cm^2/Vs)	μ_h (cm^2/Vs)
C	d	0.3567	5.45 (Γ)			5.5	2.42	2200	1600
Si	d	0.5431	1.124 (X)	0.98 (m_l) 0.19 (m_t)	0.16 (m_{lh}) 0.5 (m_{hh})	11.7	3.44	1350	480
Ge	d	0.5646	0.67 (L)	1.58 (m_l) 0.08 (m_t)	0.04 (m_{lh}) 0.3 (m_{hh})	16.3	4.00	3900	1900
α-Sn	d	0.64892	0.08 (Γ)	0.02				2000	1000
3C-SiC	zb	0.436	2.4			9.7	2.7	1000	50
4H-SiC	w	0.3073 (a) 1.005 (c)	3.26			9.6	2.7		120
6H-SiC	w	0.30806 (a) 1.5117 (c)	3.101			10.2	2.7	1140	850
AlN	w	0.3111 (a) 0.4978 (c)	6.2			8.5	3.32		
AlP	zb	0.54625	2.43 (X)	0.13		9.8	3.0	80	
AlAs	zb	0.56605	2.16 (X)	0.5	0.49 (m_{lh}) 1.06 (m_{hh})	12		1000	80
AlSb	zb	0.61335	1.52 X)	0.11	0.39	11	3.4	200	300
GaN	w	0.3189 (a) 0.5185 (c)	3.4 (Γ)	0.2	0.8	12	2.4	1500	
GaP	zb	0.54506	2.26 (Γ)	0.13	0.67	10	3.37	300	150
GaAs	zb	0.56533	1.42 (Γ)	0.067	0.12 (m_{lh}) 0.5 (m_{hh})	12.5	3.4	8500	400
GaSb	zb	0.60954	0.72 (Γ)	0.045	0.39	15	3.9	5000	1000

(continued)

Table 1.1 (continued)

	S	a_0 (nm)	E_g (eV)	m_e^*	m_h^*	ϵ_0	n_r	μ_e (cm^2/Vs)	μ_h (cm^2/Vs)
InN	w	0.3533 (a)	0.69 (Γ)						
		0.5693 (c)							
InP	zb	0.58686	1.35 (Γ)	0.07	0.4	12.1	3.37	4000	600
InAs	zb	0.60584	0.36 (Γ)	0.028	0.33	12.5	3.42	22 600	200
InSb	zb	0.64788	0.18 (Γ)	0.013	0.18	18	3.75	100 000	1700
ZnO	w	0.325 (a)	3.4 (Γ)	0.24	0.59	6.5	2.2	220	
		0.5206 (c)							
ZnS	zb	0.54109	3.6 (Γ)	0.3		8.3	2.4	110	
ZnSe	zb	0.56686	2.58 (Γ)	0.17		8.1	2.89	600	
ZnTe	zb	0.61037	2.25 (Γ)	0.15		9.7	3.56		
CdO	rs	0.47	2.16						
CdS	w	0.416 (a)	2.42 (Γ)	0.2	0.7	8.9	2.5	250	
		0.6756 (c)							
CdSe	zb	0.650	1.73 (Γ)	0.13	0.4	10.6		650	
CdTe	zb	0.64816	1.50 (Γ)	0.11	0.35	10.9	2.75	1050	100
MgO	rs	0.421	7.3						
HgS	zb	0.5852	2.0 (Γ)					50	
HgSe	zb	0.6084	−0.15 (Γ)	0.045		25		18 500	
HgTe	zb	0.64616	−0.15 (Γ)	0.029	0.3	20	3.7	22 000	100
PbS	rs	0.5936	0.37 (L)	0.1	0.1	170	3.7	500	600
PbSe	rs	0.6147	0.26 (L)	0.07 (m_{lh})	0.06 (m_{lh})	250		1800	930
				0.039 (m_{hh})	0.03 (m_{hh})				

(continued)

Table 1.1 (continued)

	S	a_0 (nm)	E_g (eV)	m_e^*	m_h^*	ϵ_0	n_r	μ_e (cm^2/Vs)	μ_h (cm^2/Vs)
PbTe	rs	0.645	0.29 (L)	0.24 (m_{lh})	0.3 (m_{lh})	412		1400	1100
				0.02 (m_{hh})	0.02 (m_{hh})				
$ZnSiP_2$	ch	0.54 (a)	2.96 (Γ)	0.07					
		1.0441 (c)							
$ZnGeP_2$	ch	0.5465 (a)	2.34 (Γ)		0.5				
		1.0771 (c)							
$ZnSnP_2$	ch	0.5651 (a)	1.66 (Γ)						
		1.1302 (c)							
$CuInS_2$	ch	0.523 (a)	1.53 (Γ)						
		1.113 (c)							
$CuGaS_2$	ch	0.5347 (a)	2.5 (Γ)						
		1.0474 (c)							
$CuInSe_2$	ch	0.5784 (a)	1.0 (Γ)						
		1.162 (c)							
$CuGaSe_2$	ch	0.5614 (a)	1.7 (Γ)						
		1.103 (c)							

'S' denotes the crystal structure (*d* diamond, *w* wurtzite, *zb* zincblende, *ch* chalcopyrite, *rs* rocksalt)

1.3 General Information

In Fig. 1.16, the periodic table of elements is shown.

In Table 1.1 the physical properties of various semiconductors are summarized. Data on semiconductors can be found in [134–146].

Part I
Fundamentals

Chapter 2
Bonds

Protons give an atom its identity, electrons its personality.
B. Bryson [148]

Abstract A little bit of solid state physics... The schemes of covalent, ionic and mixed bonds are explained which are the basis for the atomic arrangement and crystal structures of semiconductors.

2.1 Introduction

The positively charged atomic nuclei and the electrons in the atomic shells of the atoms making up the semiconductor (or any other solid) are in a binding state. Several mechanisms can lead to such cohesiveness. First, we will discuss the homopolar, electron-pair or covalent bond, then the ionic bond and subsequently the mixed bond. We will only briefly touch on the metallic bond and the van-der-Waals bond. A classical book on bonds in semiconductors is [149, 150].

2.2 Covalent Bonds

Covalent bonds are formed due to quantum-mechanical forces. The prototype covalent bond is the bonding of the hydrogen molecule due to overlapping of the atomic shells. If several electron pairs are involved, directional bonds can be formed in various spatial directions, eventually making up a solid.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_2

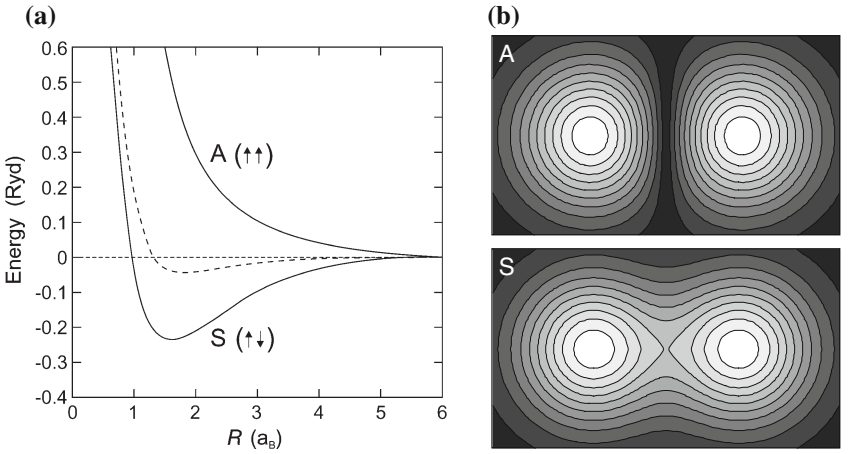


Fig. 2.1 Binding of the hydrogen molecule. (**a**) *Dashed line*: classical calculation (electrostatics), 'S', 'A': quantum-mechanical calculation taking into account Pauli's principle (S: symmetric orbital, antiparallel spins, A: antisymmetric orbital, parallel spins). The distance of the nuclei (protons) is given in units of the Bohr radius $a_B = 0.053\,\text{nm}$, the energy is given in Rydberg units (13.6 eV). (**b**) Schematic contour plots of the probability distribution ($\Psi^*\Psi$) for the S and A states

2.2.1 Electron-Pair Bond

The covalent bond of two hydrogen atoms in a H_2 molecule can lead to a reduction of the total energy of the system, compared to two single (distant) atoms (Fig. 2.1). For fermions (electrons have spin 1/2) the two-particle wavefunction of the two (indistinguishable) electrons A and B must be antisymmetric, i.e. $\Psi(A, B) = -\Psi(B, A)$ (Pauli principle). The wavefunction of each electron has degrees of freedom in real space ($\mathbf{r}$) and spin (σ), $\Psi(A) = \Psi_\mathbf{r}(A)\,\Psi_\sigma(A)$. The two-particle wavefunction of the molecule is nonseparable and has the form $\Psi(A, B) = \Psi_\mathbf{r}(\mathbf{r_A}, \mathbf{r_B})\,\Psi_\sigma(\sigma_A, \sigma_B)$. The binding state has a wavefunction with a symmetric orbital and antiparallel spins, i.e. $\Psi_\mathbf{r}(\mathbf{r_A}, \mathbf{r_B}) = \Psi_\mathbf{r}(\mathbf{r_B}, \mathbf{r_A})$ and $\Psi_\sigma(\sigma_A, \sigma_B) = -\Psi_\sigma(\sigma_B, \sigma_A)$. The antisymmetric orbital with parallel spins is antibinding for all distances of the nuclei (protons).

2.2.2 sp^3 Bonds

Elements from group IV of the periodic system (C, Si, Ge, ...) have 4 electrons on the outer shell. Carbon has the electron configuration $1s^2 2s^2 2p^2$. For an octet configuration bonding to four other electrons would be optimal (Fig. 2.2). This occurs through the mechanism of sp^3 hybridization.[1] First, one electron of the $ns^2 np^2$ configuration

[1] It is debated in femtosecond chemistry whether the bond *really* forms in this way. However, it is a picture of overwhelming simplicity.

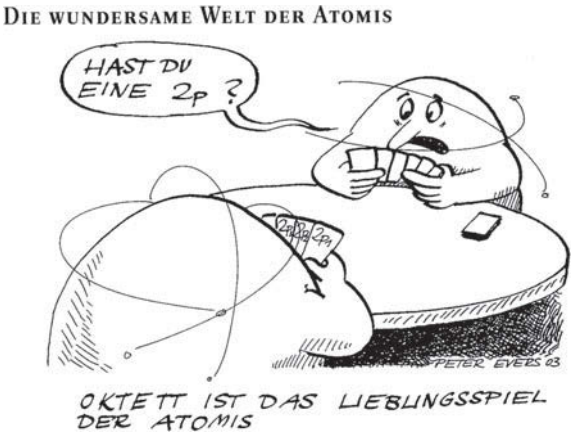


Fig. 2.2 Octet, the favorite card game of the 'Atomis' (trying to reach octet configuration in a bond by swapping wavefunctions). The bubble says: 'Do you have a 2p?'. Reprinted with permission from [151], © 2002 Wiley-VCH

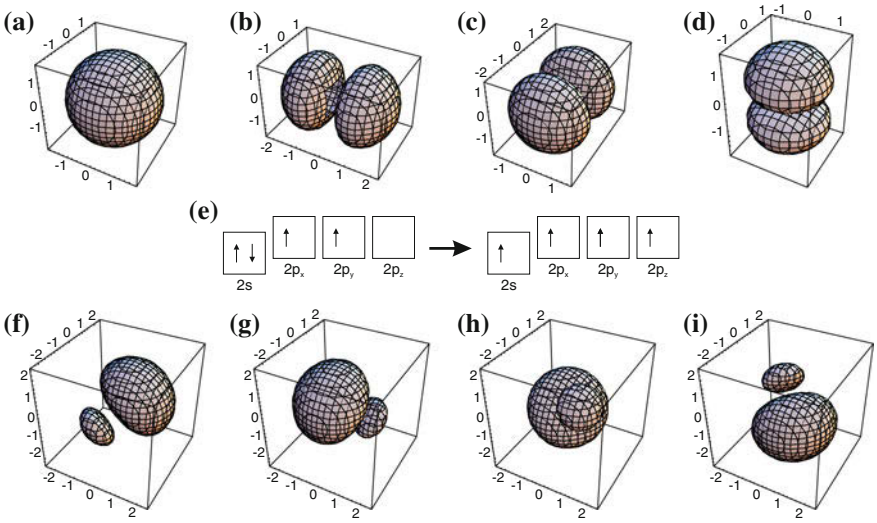


Fig. 2.3 (**a**) s orbital, (**b**, **c**, **d**) p_x, p_y and p_z orbital, (**e**) hybridization, (**f**, **g**, **h**, **i**) orbitals of the sp^3 hybridization: (**f**) $(s+p_x+p_y+p_z)/2$, (**g**) $(s+p_x-p_y-p_z)/2$, (**h**) $(s-p_x+p_y-p_z)/2$, (**i**) $(s-p_x-p_y+p_z)/2$

is brought into a p orbital, such that the outermost shell contains one s, p_x, p_y, and p_z orbital each (Fig. 2.3a–e). The energy necessary for this step is much less than regained in the subsequent formation of the covalent bonds. The four orbitals can be reconfigured into four other wavefunctions, the sp^3 hybrids (Fig. 2.3f–i), i.e.

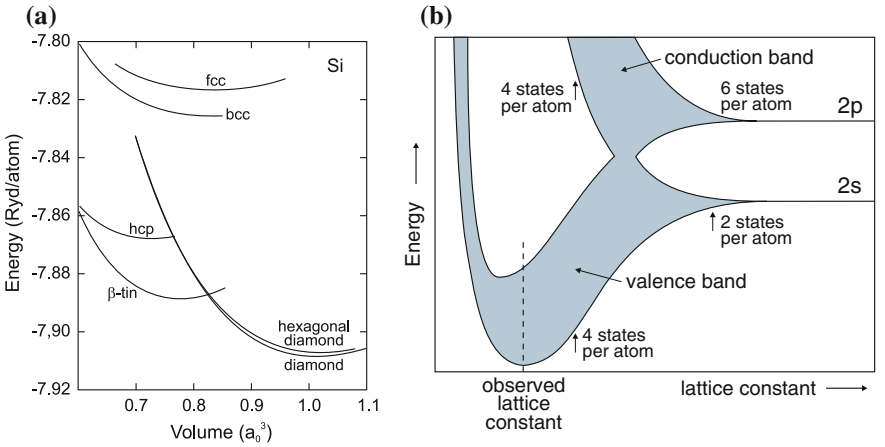


Fig. 2.4 (**a**) Energy per atom in silicon for various crystal structures. Adapted from [152]. (**b**) Electron energy levels in (diamond structure) carbon as a function of the distance of the atomic nuclei (schematic). Adapted from [153, 154]

$$\Psi_1 = (\mathrm{s} + \mathrm{p}_x + \mathrm{p}_y + \mathrm{p}_z)/2 \tag{2.1a}$$
$$\Psi_2 = (\mathrm{s} + \mathrm{p}_x - \mathrm{p}_y - \mathrm{p}_z)/2 \tag{2.1b}$$
$$\Psi_3 = (\mathrm{s} - \mathrm{p}_x + \mathrm{p}_y - \mathrm{p}_z)/2 \tag{2.1c}$$
$$\Psi_4 = (\mathrm{s} - \mathrm{p}_x - \mathrm{p}_y + \mathrm{p}_z)/2. \tag{2.1d}$$

These orbitals have a directed form along tetrahedral directions. The binding energy (per atom) of the covalent bond is large, for H–H 4.5 eV, for C–C 3.6 eV, for Si–Si 1.8 eV, and for Ge–Ge 1.6 eV. Such energy is, for *neutral* atoms, comparable to the ionic bond, discussed in the next section.

In Fig. 2.4a the energy of a crystal made up from silicon atoms is shown for various crystal structures[2] or phases (cf. Chap. 3). We note that the crystal energy of further silicon structures are discussed in [155]. The lattice constant with the lowest total energy determines the lattice spacing for each crystal structure. The thermodynamically stable configuration is the phase with the lowest overall energy for given external conditions.

The covalent bond of a group-IV atom to other group-IV atoms has a tetrahedral configuration with electron-pair bonds, similar to the hydrogen molecule bond. In Fig. 2.4b the energy states of the $n = 2$ shell for tetrahedrally bonded carbon (diamond, see Sect. 3.4.3) are shown as a function of the distance from the nuclei. First, the energetically sharp states become a band due to the overlap and coupling of the atomic wavefunctions (cf. Chap. 6). The mixing of the states leads to the formation of the filled lower valence band (binding states) and the empty upper conduction band (antibinding states).

[2]Hexagonal diamond is wurtzite structure with two identical atoms in the base.

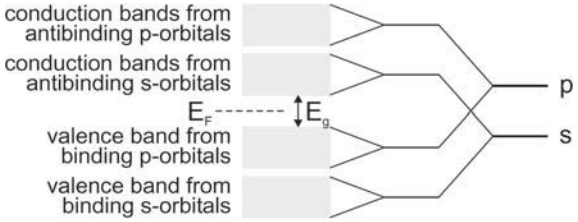


Fig. 2.5 Schematic of the origin of valence and conduction band from the atomic s and p orbitals. The band gap E_g and the position of the Fermi level E_F are indicated

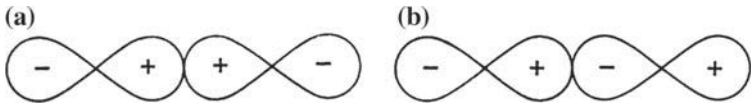


Fig. 2.6 Schematic representation of (**a**) bonding and (**b**) antibinding p orbitals. The signs denote the phase of the wavefunction

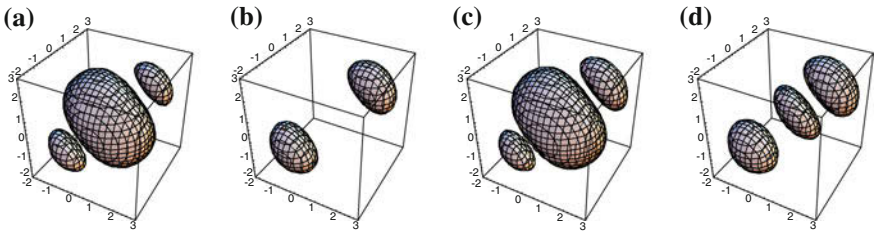


Fig. 2.7 Schematic representation of (**a**, **c**) bonding and (**b**, **d**) antibinding symmetric (**a**, **b**) and nonsymmetric (**c**, **d**) sp^3 orbitals

This principle is valid for most semiconductors and is shown schematically also in Fig. 2.5. The configuration of bonding and antibinding p orbitals is depicted schematically in Fig. 2.6. The bonding and antibinding sp^3 orbitals are depicted in Figs. 2.7a, b and 2.13. We note that the energy of the crystal does not only depend on the distance from the nuclei but also on their geometric arrangement (crystal structure).

Per carbon atom there are (in the second shell) four electrons and four unoccupied states, altogether eight. These are redistributed into four states (filled) per atoms in the valence band and four states per atom (empty) in the conduction band. Between the top of the valence band and the bottom of the conduction band there is an energy gap, later called the *band gap* (cf. Chap. 6).

2.2.3 sp^2 Bonds

Organic semiconductors (see Chap. 17) are made up from carbon compounds. While for inorganic semiconductors the covalent (or mixed, cf. Sect. 2.4) bond with sp^3

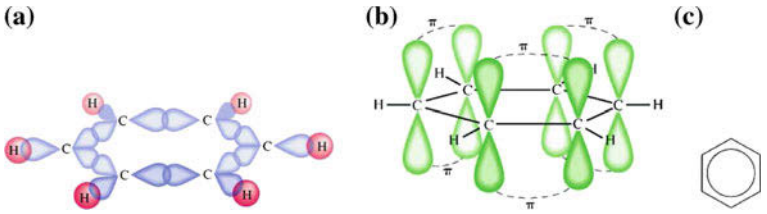


Fig. 2.8 Schematic representation of the (**a**) σ and (**b**) π bonds in benzene, (**c**) schematic symbol for benzene

hybridization is important, the organic compounds are based on the sp^2 hybridization. This bonding mechanism, which is present in graphite, is stronger than the sp^3-bond present in diamond. The prototype organic molecule is the benzene ring[3] (C_6H_6), shown in Fig. 2.8. The benzene ring is the building block for small organic molecules and polymers.

In the benzene molecule neighboring carbon atoms are bonded within the ring plane via the binding σ states of the sp^2 orbitals (Fig. 2.8a). The wavefunctions (Fig. 2.9) are given by (2.2a–c).

$$\Psi_1 = (s + \sqrt{2}\, p_x)/\sqrt{3} \tag{2.2a}$$

$$\Psi_2 = (s - \sqrt{1/2}\, p_x + \sqrt{3/2}\, p_y)/\sqrt{3} \tag{2.2b}$$

$$\Psi_3 = (s - \sqrt{1/2}\, p_x - \sqrt{3/2}\, p_y)/\sqrt{3}. \tag{2.2c}$$

The 'remaining' p_z orbitals do not directly take part in the binding (Fig. 2.8b) and form bonding (π, filled) and antibinding (π^*, empty) orbitals (see Fig. 2.10). The π and π^* states are delocalized over the ring. Between the highest populated molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) is typically an energy gap (Fig. 2.11). The antibinding σ^* orbitals are energetically above the π^* states.

[3] Supposedly, the chemist Friedrich August Kekulé von Stadonitz had a dream about dancing carbon molecules and thus came up with the ring-like molecule structure [156]. Kekulé remembered: 'During my stay in Ghent, I lived in elegant bachelor quarters in the main thoroughfare. My study, however, faced a narrow side-alley and no daylight penetrated it. . . . I was sitting writing on my textbook, but the work did not progress; my thoughts were elsewhere. I turned my chair to the fire and dozed. Again the atoms were gamboling before my eyes. This time the smaller groups kept modestly in the background. My mental eye, rendered more acute by the repeated visions of the kind, could now distinguish larger structures of manifold conformation; long rows sometimes more closely fitted together all twining and twisting in snake-like motion. But look! What was that? One of the snakes had seized hold of its own tail, and the form whirled mockingly before my eyes. As if by a flash of lightning I awoke; and this time also I spent the rest of the night in working out the consequences of the hypothesis.'

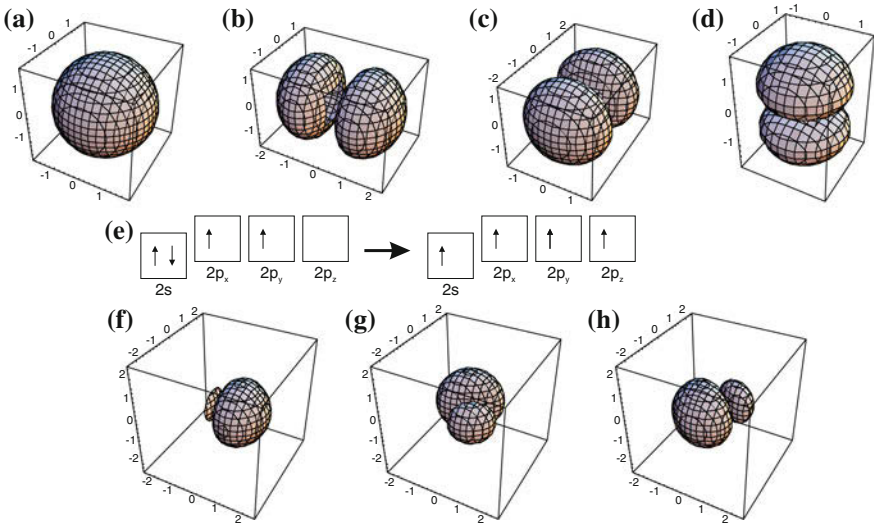


Fig. 2.9 (**a**) s orbital, (**b**, **c**, **d**) p_x, p_y and p_z orbital, (**e**) hybridization, (**f**, **g**, **h**) orbitals of the sp^2 hybridization: (**f**) $(s+\sqrt{2}p_x)/\sqrt{3}$, (**g**) $(s-\sqrt{1/2}p_x+\sqrt{3/2}p_y)/\sqrt{3}$, (**h**) $(s-\sqrt{1/2}p_x-\sqrt{3/2}p_y)/\sqrt{3}$

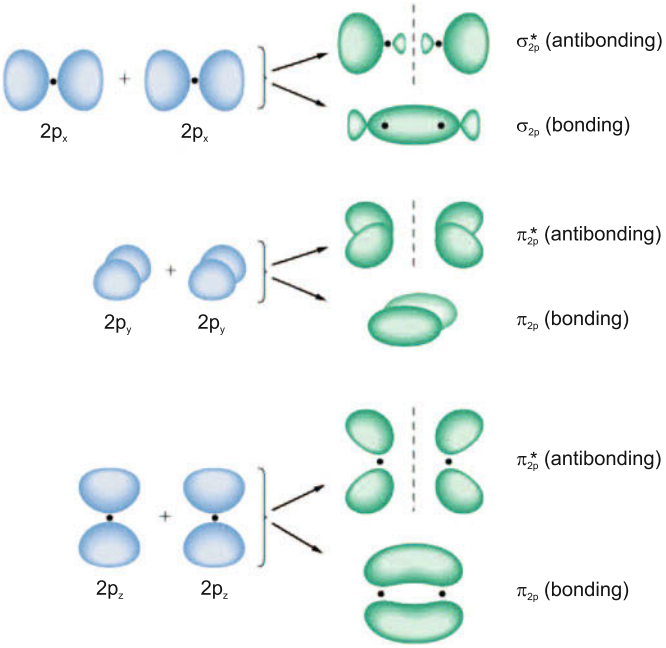


Fig. 2.10 Orbitals due to binding and antibinding configurations of various π orbitals

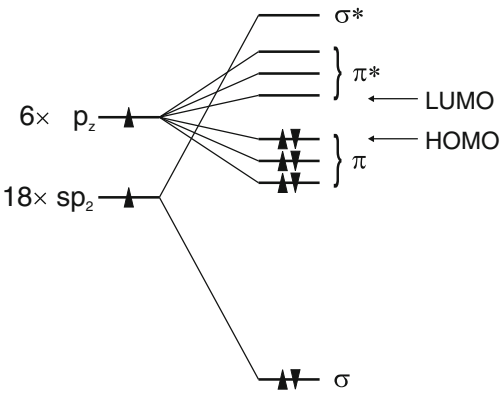


Fig. 2.11 Schematic energy terms of the benzene molecule

2.3 Ionic Bonds

Ionic crystals are made up from positively and negatively charged ions. The heteropolar or ionic bond is the consequence of the electrostatic attraction between the ions. However, the possibly repulsive character of next neighbors has to be considered.

For I–VII compounds, e.g. LiF or NaCl, the shells of the singly charged ions are complete: Li: $1s^2 2s^1 \rightarrow Li^+$: $1s^2$, F: $1s^2 2s^2 2p^5 \rightarrow F^-$: $1s^2 2s^2 2p^6$. Compared to ions in a gas, a Na–Cl pair in the crystal has a binding energy of 7.9 eV that mostly stems from the electrostatic energy (Madelung energy). Van-der-Waals forces (cf. Sect. 2.6) only contribute 1–2 %. The ionization energy of Na is 5.14 eV, the electron affinity of Cl is 3.61 eV. Thus the energy of the NaCl pair in the solid is 6.4 (= 7.9−5.1+3.6) eV smaller than in a gas of neutral atoms.

The interaction of two ions with distance vector $\mathbf{r}_{ij}$ is due to the Coulomb interaction

$$U_{ij}^{\mathrm{C}} = \frac{q_i\, q_j}{4\pi\epsilon_0}\,\frac{1}{r_{ij}} = \pm\frac{e^2}{4\pi\epsilon_0}\,\frac{1}{r_{ij}} \tag{2.3}$$

and a repulsive contribution due to the overlap of (complete) shells. This contribution is typically approximated by a radially symmetric core potential

$$U_{ij}^{\mathrm{core}} = \lambda\, \exp(-\lambda/\rho) \tag{2.4}$$

that only acts on next neighbors. λ describes the strength of this interaction and ρ parameterizes its range.

The distance of ions is denoted as $r_{ij} = p_{ij} R$, where R denotes the distance of next neighbors and the p_{ij} are suitable coefficients. The electrostatic interaction of an ion with *all* its neighbors is then written as

$$U_{ij}^{\mathrm{C}} = -\alpha\,\frac{e^2}{4\pi\epsilon_0}\,\frac{1}{R}, \tag{2.5}$$

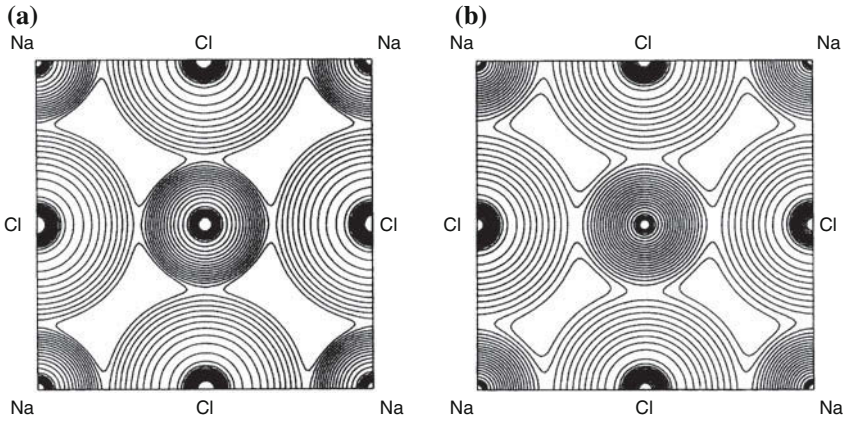


Fig. 2.12 (**a**) Experimental and (**b**) theoretical charge distribution in the (100) plane of NaCl. The lowest contour in the interstitial region corresponds to a charge density of $7\,\mathrm{e/nm^3}$ and adjacent contours differ by $\sqrt{2}$. Differences are mainly due to the fact that the X-ray experiments have been made at room temperature. Adapted from [157]

where α is the Madelung constant. For an attractive interaction (as in a solid), α is positive. It is given (calculated for the i-th ion) as

$$\alpha = \sum_{ij} \frac{\pm 1}{p_{ij}}. \tag{2.6}$$

For a one-dimensional chain $\alpha = 2\ln 2$. For the rocksalt (NaCl) structure (cf. Sect. 3.4.1) it is $\alpha \approx 1.7476$, for the CsCl structure (see Sect. 3.4.2) it is $\alpha \approx 1.7627$, and for the zincblende structure (see Sect. 3.4.4) it is $\alpha \approx 1.6381$. This shows that ionic compounds prefer the NaCl or CsCl structure. The charge distribution for NaCl is shown in Fig. 2.12. For tetragonal and orthorhombic structures, the Madelung constant has been calculated in [158].

2.4 Mixed Bonds

The group-IV crystals are of perfectly covalent nature, the I–VII are almost exclusively ionically bonded. For III–V (e.g. GaAs, InP) and II–VI compounds (e.g. CdS, ZnO) we have a mixed case.

The (screened) Coulomb potentials of the A and B atoms (in the AB compound) shall be denoted V_A and V_B. The origin of the coordinate system is in the center of the A and B atom (i.e. for the zincblende structure (cf. Sect. 3.4.4) at $(1/8, 1/8, 1/8)a$. The valence electrons then see the potential

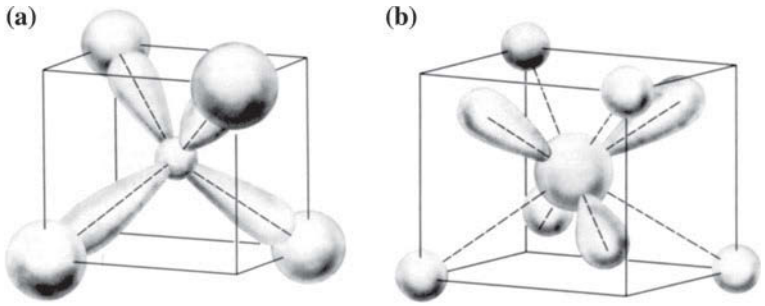


Fig. 2.13 Schematic representation of (**a**) bonding and (**b**) antibinding sp^3 orbitals. From [149]

$$V_{\text{crystal}} = \sum_{\alpha} V_A(\mathbf{r} - \mathbf{r}_\alpha) + \sum_{\beta} V_B(\mathbf{r} - \mathbf{r}_\beta), \tag{2.7}$$

where the sum α (β) runs over all A (B) atoms. This potential can be split into a symmetric (V_c, covalent) and an antisymmetric (V_i, ionic) part (2.8b), i.e. $V_{\text{crystal}} = V_\text{c} + V_\text{i}$

$$V_\text{c} = \frac{1}{2}\left\{\sum_{\alpha} V_A(\mathbf{r} - \mathbf{r}_\alpha) + \sum_{\alpha} V_B(\mathbf{r} - \mathbf{r}_\alpha) + \sum_{\beta} V_B(\mathbf{r} - \mathbf{r}_\beta) + \sum_{\beta} V_A(\mathbf{r} - \mathbf{r}_\beta)\right\} \tag{2.8a}$$

$$V_\text{i} = \frac{1}{2}\left\{\sum_{\alpha} V_A(\mathbf{r} - \mathbf{r}_\alpha) - \sum_{\alpha} V_B(\mathbf{r} - \mathbf{r}_\alpha) + \sum_{\beta} V_B(\mathbf{r} - \mathbf{r}_\beta) - \sum_{\beta} V_A(\mathbf{r} - \mathbf{r}_\beta)\right\}. \tag{2.8b}$$

For homopolar bonds $V_\text{i} = 0$ and the splitting between bonding and antibinding states is E_h, which mainly depends on the bond length l_{AB} (and the related overlap of atomic wavefunctions). In a partially ionic bond the orbitals are not symmetric along A–B, but the center is shifted towards the more electronegative material (Figs. 2.7c, d and 2.13).

The band splitting[4] between the (highest) bonding and (lowest) antibinding state E_ba is then written as

[4]This energy should not be confused with the *band gap* ΔE_cv, the energy separation of the highest valence-band state and the lowest conduction-band state. The energy splitting E_ba is the energy separation between the centers of the valence and conduction bands. Mostly, the term E_g is used for ΔE_cv.

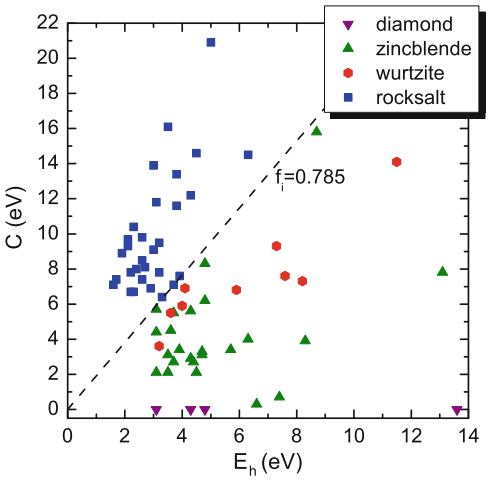


Fig. 2.14 Values of E_h and C for various $A^N B^{8-N}$ compounds. The *dashed line* $f_\mathrm{i} = 0.785$ separates 4-fold from 6-fold coordinated structures. Most data taken from [159]

$$E_\mathrm{ba} = E_\mathrm{h} + \mathrm{i}C, \tag{2.9}$$

where C denotes the band splitting due to the ionic part of the potential and depends only on $V_A - V_B$. C is proportional to the difference of the electronegativities X of the A and B atoms, $C(A, B) = 5.75(X_A - X_B)$. A material thus takes a point in the (E_h,C) plane (Fig. 2.14). The absolute value for the band splitting is given as $E_\mathrm{ba}^2 = E_\mathrm{h}^2 + C^2$.

The ionicity of the bond is described with the ionicity (after Phillips) f_i, defined as [160, 161]

$$f_\mathrm{i} = \frac{C^2}{E_h^2 + C^2}. \tag{2.10}$$

The covalent part is $1 - f_\mathrm{i}$. In Table 2.1 the ionicity is given for a number of binary compounds. The ionicity can also be interpreted as the angle $\tan(\phi) = C/E_h$ in the (E_h,C) diagram. The critical value of $f_\mathrm{i} = 0.785$ for the ionicity separates quite exactly (for about 70 compounds) the 4-fold (diamond, zincblende and wurtzite) from the 6-fold (rocksalt) coordinated substances ($f_\mathrm{i} = 0.785$ is indicated by a dashed line in Fig. 2.14).

For ionic compounds, an effective ionic charge e^* is defined connecting the displacement $\mathbf{u}$ of negative and positive ions and the resulting polarization $\mathbf{P} = (e^*/2a^3)\,\mathbf{u}$ [162]. Connected with the ionicity is the so-called s-parameter, describing the change of the charge upon change of bond length b from its equilibrium value b_0 [163]

$$e^*(b) = e^*(b_0)\left(\frac{b}{b_0}\right)^s \approx e_0^*\,(1 + s\,\epsilon), \tag{2.11}$$

ϵ being the strain of the bond length, $b/b_0 = 1 + \epsilon$. It seems justified to assume that $e^*(b_0)$ is always positive at the metal atom in III–V and II–VI compounds.

Table 2.1 Ionicity f_i (2.10) for various binary compounds

C	0.0	AlAs	0.27	BeO	0.60	CuCl	0.75
Si	0.0	BeS	0.29	ZnTe	0.61	CuF	0.77
Ge	0.0	AlP	0.31	ZnO	0.62	AgI	0.77
Sn	0.0	GaAs	0.31	ZnS	0.62	MgS	0.79
BAs	0.002	InSb	0.32	ZnSe	0.63	MgSe	0.79
BP	0.006	GaP	0.33	HgTe	0.65	CdO	0.79
BeTe	0.17	InAs	0.36	HgSe	0.68	HgS	0.79
SiC	0.18	InP	0.42	CdS	0.69	MgO	0.84
AlSb	0.25	AlN	0.45	CuI	0.69	AgBr	0.85
BN	0.26	GaN	0.50	CdSe	0.70	LiF	0.92
GaSb	0.26	MgTe	0.55	CdTe	0.72	NaCl	0.94
BeSe	0.26	InN	0.58	CuBr	0.74	RbF	0.96

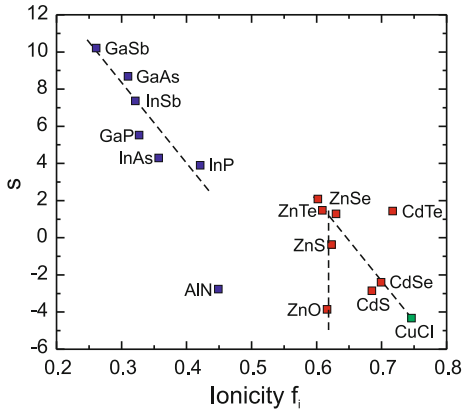


Fig. 2.15 s-Parameter as defined in (2.11) as a function of the ionicity f_i (2.10) for various compound semicondcutors. *Dashed lines* are guides to the eye. Data from [164], value for CuCl from [163]

The relation of s with the ionicity f_i is shown in Fig. 2.15 for various compound semiconductors.

2.5 Metallic Bonding

In a metal, the positively charged atomic cores are embedded in a more or less homogeneous sea of electrons. The valence electrons of the atoms become the conduction electrons of the metal. These are freely moveable and at $T = 0\,\mathrm{K}$ there is no energy gap between filled and empty states. The bonding is mediated by the energy reduction for the conduction electrons in the periodic potential of the solid compared to free atoms. This will be clearer when the band structure is discussed (Chap. 6). In transition metals the overlap of inner shells (d or f) can also contribute to the bonding.

2.6 van-der-Waals Bonds

The van-der-Waals bond is a dipole bond that leads to bonding in the noble-gas crystals (at low temperature). Ne, Ar, Kr and Xe crystallize in the densely packed fcc lattice (cf. Sect. 3.3.5). He^3 and He^4 represent an exception. They do not solidify at zero pressure at $T = 0\,K$ due to the large zero-point energy. This quantum-mechanical effect is especially strong for oscillators with small mass.

When two neutral atoms come near to each other (distance of the nuclei R), an attractive dipole–dipole interaction $-A\,R^{-6}$ arises (London interaction) the van-der-Waals interaction. The quantum-mechanical overlap of the (filled) shells leads to a strong repulsion $+B\,R^{-12}$. Altogether, a binding energy minimum results for the Lennard–Jones potential V_{LJ} (see Fig. 2.16)

$$V_{\mathrm{LJ}}(R) = -\frac{A}{R^6} + \frac{B}{R^{12}}. \tag{2.12}$$

The energy minimum $E_{\min} = -A^2/(2B)$ is at $R = (2B/A)^{1/6}$.

The origin of the attractive dipole–dipole interaction can be understood from a one-dimensional (1D) model as follows: Two atoms are modeled by their fixed positively charged nuclei in a distance R and their negatively charged electron shells that are polarizable, i.e. can be displaced along one direction x. Additionally, we assume (two identical) 1D harmonic oscillators for the electron motion at the positions 0 and R. Then, the Hamilton operator H_0 of the system without interaction (R is very large)

$$H_0 = \frac{1}{2m}\,p_1^2 + C\,x_1^2 + \frac{1}{2m}\,p_2^2 + C\,x_2^2. \tag{2.13}$$

The indices 1 and 2 denote the two electrons of atoms. x_1 and x_2 are the displacements of the electrons. Both harmonic oscillators have a resonance frequency $\omega_0 = \sqrt{C/m}$, and the zero-point energy is $\hbar\omega_0/2$.

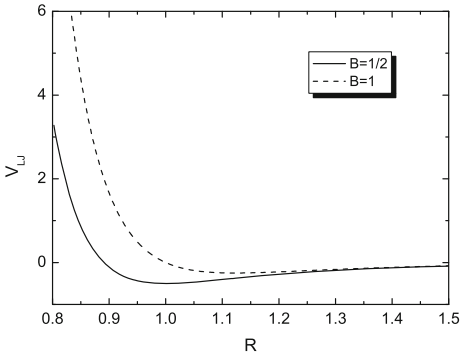


Fig. 2.16 Lennard–Jones potential (2.12) for $A = 1$ and two values of B

Taking into account the Coulomb interaction of the four charges, an additional term H_1 arises

$$H_1 = \frac{e^2}{R} + \frac{e^2}{R + x_1 + x_2} - \frac{e^2}{R + x_1} - \frac{e^2}{R - x_2} \approx -\frac{2e^2}{R^3}\, x_1\, x_2. \tag{2.14}$$

The approximation is valid for small amplitudes $x_i \ll R$. A separation of variables can be achieved by transformation to the normal modes

$$x_s = \frac{x_1 + x_2}{\sqrt{2}}, \, x_a = \frac{x_1 - x_2}{\sqrt{2}}. \tag{2.15}$$

Then we find

$$\begin{aligned} H &= H_0 + H_1 \\ &= \left[\frac{1}{2m}\, p_s^2 + \frac{1}{2}\left(C - \frac{2e^2}{R^3}\right) x_s^2\right] + \left[\frac{1}{2m}\, p_a^2 + \frac{1}{2}\left(C - \frac{2e^2}{R^3}\right) x_a^2\right]. \end{aligned} \tag{2.16}$$

This equation is the Hamiltonian of two decoupled harmonic oscillators with the normal frequencies

$$\omega_\pm = \sqrt{\left(C \pm \frac{2e^2}{R^3}\right)/m} \approx \omega_0 \left[1 \pm \frac{1}{2}\left(\frac{2e^2}{C\, R^3}\right) - \frac{1}{8}\left(\frac{2e^2}{C\, R^3}\right)^2 + \cdots\right]. \tag{2.17}$$

The coupled system thus has a lower (zero-point) energy than the uncoupled. The energy difference per atom is (in lowest order) proportional to R^{-6}.

$$\Delta U = \hbar\omega_0 - \frac{1}{2}\,(\omega_+ - \omega_-) \approx -\hbar\omega_0\, \frac{1}{8}\left(\frac{2e^2}{C\, R^3}\right)^2 = -\frac{A}{R^6}. \tag{2.18}$$

The interaction is a true quantum-mechanical effect, i.e. the reduction of the zero-point energy of coupled oscillators.

2.7 Hamilton Operator of the Solid

The total energy of the solid, including kinetic and potential terms, is

$$\begin{aligned} H = &\sum_i \frac{\mathbf{p}_i^2}{2m_i} + \sum_j \frac{\mathbf{P}_j^2}{2M_j} + \frac{1}{2}\sum_{j,j'} \frac{Z_j\, Z_{j'}\, e^2}{4\pi\epsilon_0\, |\mathbf{R}_j - \mathbf{R}_{j'}|} \\ &+ \frac{1}{2}\sum_{i,i'} \frac{e^2}{4\pi\epsilon_0\, |\mathbf{r}_i - \mathbf{r}_{i'}|} - \sum_{i,j} \frac{Z_j\, e^2}{4\pi\epsilon_0\, |\mathbf{R}_j - \mathbf{r}_i|}, \end{aligned} \tag{2.19}$$

where $\mathbf{r}_i$ and $\mathbf{R}_i$ are the position operators and $\mathbf{p}_i$ and $\mathbf{P}_i$ are the momentum operators of the electrons and nuclei, respectively. The first term is the kinetic energy of the electrons, the second term is the kinetic energy of the nuclei. The third term is the electrostatic interaction of the nuclei, the fourth term is the electrostatic interaction of the electrons. In the third and fourth terms the summation over the same indices is left out. The fifth term is the electrostatic interactions of electrons and nuclei.

In the following, the usual approximations in order to treat (2.19) are discussed. First, the nuclei and the electrons tightly bound to the nuclei (inner shells) are united with ion cores. The remaining electrons are the valence electrons.

The next approximation is the Born–Oppenheimer (or adiabatic) approximation. Since the ion cores are much heavier than the electrons (factor $\approx 10^3$) they move much slower. The frequencies of the ion vibrations are typically in the region of several tens of meV (phonons, cf. Sect. 5.2), the energy to excite an electron is typically 1 eV. Thus, the electrons always 'see' the momentary position of the ions, the ions, however, 'see' the electron motions averaged over many periods. Thus, the Hamiltonian (2.19) is split into three parts:

$$H = H_{\text{ions}}(\mathbf{R}_j) + H_{\text{e}}(\mathbf{r}_i, \mathbf{R}_{j_0}) + H_{\text{e-ion}}(\mathbf{r}_i, \delta\mathbf{R}_j). \tag{2.20}$$

The first term contains the ion cores with their potential and the time-averaged contribution of the electrons. The second term is the electron motion around the ion cores at their averaged positions $\mathbf{R}_{j_0}$. The third term is the Hamiltonian of the electron–phonon interaction that depends on the electron positions and the deviation of the ions from their average position $\delta\mathbf{R}_j = \mathbf{R}_j - \mathbf{R}_{j_0}$. The electron–phonon interaction is responsible for such effects as electrical resistance and superconductivity.

Chapter 3
Crystals

La science cristallographique ne consiste donc point à décrire scrupuleusement tous les accidens des formes cristallines; mais à spécifier, en décrivant ces formes, les rapports plus ou moins immédiats qu'elles ont entre elles.
Crystallographic science does not consist in the scrupulous description of all the accidents of crystalline form, but in specifying, by the description of these forms, the more or less close relationship they have with each other.
J.-B. Romé de l'Isle, 1783 [165]

Abstract A little bit of crystallography… The concepts of the direct and reciprocal lattice, point and space groups, unit and elementary cells and the Wigner–Seitz cell are laid out. The important crystal structures for semiconductors (diamond, sphalerite, wurtzite, chalcopyrite,…) are discussed in some detail. Also alloys and ordering are covered.

3.1 Introduction

The economically most important semiconductors have a relatively simple atomic arrangement and are highly symmetric. The symmetry of the atomic arrangement is the basis for the classification of the various crystal structures. Using group theory [166], basic and important conclusions can be drawn about the physical properties of the crystal, such as its elastic and electronic properties. The presence of highly symmetric planes is obvious from the crystal shape of the minerals and their cleavage behavior.

Polycrystalline semiconductors consist of grains of finite size that are structurally perfect but have various orientations. The grain boundaries are a lattice defect (see also Sect. 4.4.3). Amorphous semiconductors are disordered on the atomic scale, see Sect. 3.3.7.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_3

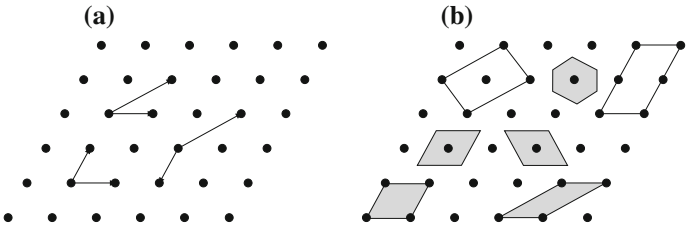


Fig. 3.1 (**a**) Two-dimensional lattice. It can be generated by various pairs of translation vectors. (**b**) Elementary cells of the lattice. Primitive elementary cells are *shaded*

3.2 Crystal Structure

A crystal is built up by the (quasi-) infinite periodic repetition of identical building blocks. This lattice (Bravais lattice) is generated by the three fundamental *translation* vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$. These three vectors may not lie in a common plane. The lattice (Fig. 3.1) is the set of all points $\mathbf{R}$

$$\mathbf{R} = n_1\,\mathbf{a}_1 + n_2\,\mathbf{a}_2 + n_3\,\mathbf{a}_3. \tag{3.1}$$

The crystal structure is made up by the lattice and the building block that is attached to each lattice point. This building block is called the *base* (Fig. 3.2). In the simplest case, e.g., for crystals like Cu, Fe or Al, this is just a single atom (monoatomic base). In the case of C (diamond), Si or Ge, it is a diatomic base with two identical atoms (e.g., Si–Si or Ge–Ge), in the case of compound semiconductors, such as GaAs or InP, it is a diatomic base with nonidentical atoms such as Ga–As or In–P. There exist far more involved structures, e.g., $NaCd_2$ where the smallest cubic cell contains 1192 atoms. In protein crystals, the base of the lattice can contain 10 000 atoms.

In summary: Crystal structure = Lattice × Base.

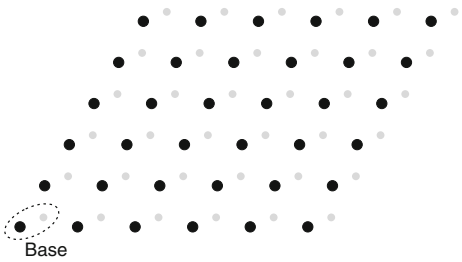


Fig. 3.2 Crystal structure, consisting of a lattice and a base

3.3 Lattice

As described in Sect. 3.2 the lattice is spanned by three vectors $\mathbf{a}_i$. The lattice symmetry is decisive for the physical properties of the semiconductor. It is described by the appropriate groups of the symmetry operations.

3.3.1 Unit Cell

The choice of the vectors $\mathbf{a}_i$ making up the lattice is not unique (Fig. 3.1). The volume that is enclosed in the parallelepiped spanned by the vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ is called the *elementary cell*. A *primitive* elementary cell is an elementary cell with the smallest possible volume (Fig. 3.1b). In each primitive elementary cell there is exactly *one* lattice point. The coordination number is the number of next-neighbor lattice points. A primitive cubic (pc) lattice, e.g., has a coordination number of 6.

The typically chosen primitive elementary cell is the *Wigner–Seitz* (WS) cell that reflects the symmetry of the lattice best. The Wigner–Seitz cell around a lattice point $\mathbf{R}_0$ contains all points that are closer to this lattice point than to any other lattice point. Since all points fulfill such a condition for some lattice point $\mathbf{R}_i$, the Wigner–Seitz cells fill the volume completely. The boundary of the Wigner–Seitz cell is made up by points that have the same distance to $\mathbf{R}_0$ and some other lattice point(s). The Wigner–Seitz cell around $\mathbf{R}_0$ is constructed by drawing lines from $\mathbf{R}_0$ to the next neighbors $\mathbf{R}_j$, taking the point at half distance and erecting a perpendicular plane at $(\mathbf{R}_j + \mathbf{R}_0)/2$. The WS cell is the smallest polyhedra circumscribed by these planes. A two-dimensional construction is shown in Fig. 3.3.

3.3.2 Point Group

Besides the translation there are other operations under which the lattice is invariant, i.e. the lattice is imaged into itself. These are:

Identity. The neutral element of any point group is the identity that does not change the crystal. It is denoted as 1 (E) in international (Schönfließ) notation.

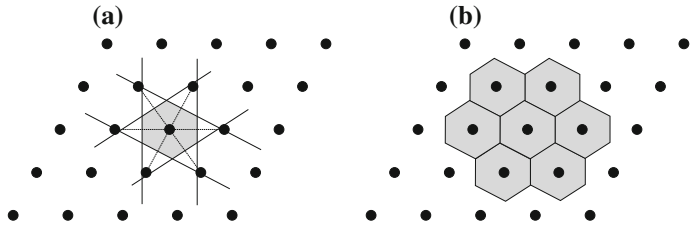


Fig. 3.3 (**a**) Construction of a two-dimensional Wigner–Seitz cell, (**b**) filling of space with WS cells

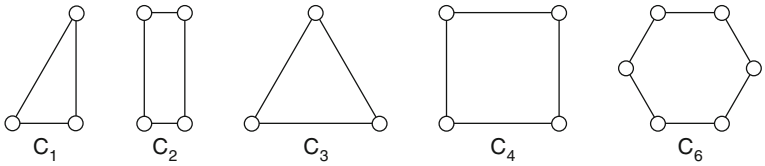


Fig. 3.4 Two-dimensional objects with perpendicular rotation axis C_n. Note that the *circles* do not exhibit σ_h symmetry with respect to the paper plane, i.e. they are different on the *top* and *bottom side*

Rotation. The rotation around an axis may have a rotation angle of 2π, $2\pi/2$, $2\pi/3$, $2\pi/4$ or $2\pi/6$ or their integer multiples. The axis is then called n = 1-, 2-, 3-, 4- or 6-fold, respectively,[1] and denoted as n (international notation) or C_n (Schönfließ). Objects with C_n symmetry are depicted in Fig. 3.4.

Mirror operation with respect to a plane through a lattice point. Different mirror planes are discerned (Fig. 3.5) (after Schönfließ) σ_h: a mirror plane perpendicular to a rotational axis, σ_v: a mirror plane that contains a rotational axis, and σ_d: a mirror plane that contains a rotational axis and bisects the angle between two C_2 axes. The international notation is $\bar{2}$.

Inversion. All points around the inversion center $\mathbf{r}$ are replaced by $-\mathbf{r}$. The inversion is denoted $\bar{1}$ (i) in international (Schönfließ) notation.

Improper rotation. The improper rotation S_n is a rotation C_n followed immediately by the inversion operation i denoted as $\bar{n}$ in international notation. There are $\bar{3}$, $\bar{4}$ and $\bar{6}$ and their powers. Only the combined operation $\bar{n}$ is a symmetry operation, while the individual operations C_n and i alone are not symmetry operations. In the Schönfließ notation the improper rotation is defined as $S_n = \sigma_\mathrm{h} C_n$, with σ_h being a mirror operation with a plane perpendicular to the axis of the C_n rotation, denoted as S_n. There are S_3, S_4 and S_6 and $\bar{3} = S_6^5$, $\bar{4} = S_4^3$ and $\bar{6} = S_3^5$. For successive applications, the S_n yield previously known operations, e.g., $S_4^2 = C_2$, $S_4^4 = E$, $S_6^2 = C_3$, $S_6^3 = i$, $S_3^2 = C_3^2$, $S_3^3 = \sigma_\mathrm{h}$, $S_3^4 = C_3$, $S_3^6 = E$. We note that formally S_1 is the inversion i and S_2 is the mirror symmetry σ. Objects with S_n symmetry are schematically shown in Fig. 3.6.

These symmetry operations form 32 *point groups*. These groups are shown (with their different notations and elements) in Table B.2. The highest symmetry is the cubic symmetry $O_h = O \times i$. The tetraeder group T_d (methane molecule) is a subgroup of O_h, lacking the inversion operation.

Important for surface symmetries, there are ten two-dimensional point groups (Sect. 11.2 and Table B.1).

[1] 5-fold periodic symmetry is geometrically impossible. However, *quasicrystals* with aperiodic five-fold symmetry exist [167, 168], some of them possibly being semiconducting [169, 170].

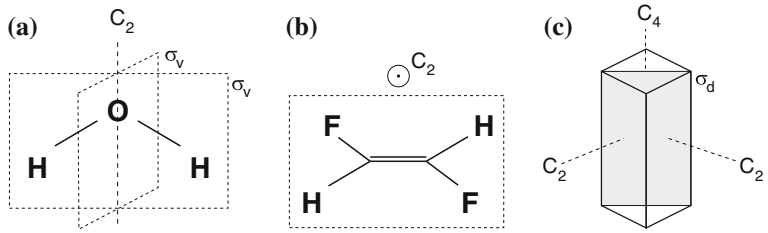


Fig. 3.5 Mirror planes: (**a**) σ_v (at H_2O molecule), (**b**) σ_h (at F_2H_2 molecule) and (**c**) σ_d

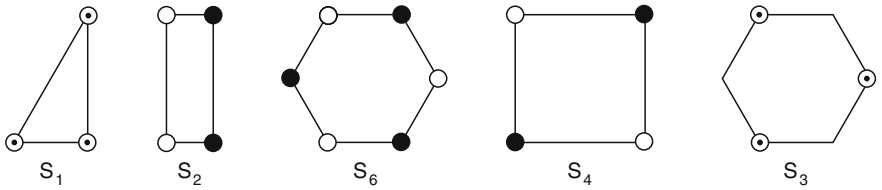


Fig. 3.6 Two-dimensional objects with perpendicular improper rotation axis S_n. Note that the *white* and *black circles* do not exhibit σ_h symmetry with respect to the paper plane, i.e. they are *white on the top* and *black on the bottom*. The *circles with a dot in the center* exhibit σ_h symmetry, i.e. they look the same from *top* and *bottom*

3.3.3 Space Group

The space group is formed by the combination of the elements of the point group with translations. The combination of a translation along a rotational axis with a rotation around this axis creates a screw axis n_m. In Fig. 3.7a, a so-called 4_1 screw axis is shown. The first index n indicates the rotation angle, i.e. $2\pi/n$, the second index indicates the translation, i.e. $c\,m/n$, c being the periodicity along the axis. There are eleven crystallographically allowed screw rotations.[2]

The combination of the mirror operation at a plane that contains a rotational axis with a translation along this axis creates a glide reflection (Fig. 3.7b). For an axial glide (or b-glide) the translation is parallel to the reflection plane. A diagonal glide (or d-glide) involves translation in two or three directions. A third type of glide is the diamond glide (or d-glide). There are 230 different space groups, listed in Appendix B. A detailed treatment can be found in [171].[3]

Important for surface symmetries, there are 17 two-dimensional space groups (Sect. 11.2).

[2] $2_1, 3_1, 3_2, 4_1, 4_2, 4_3, 6_1, 6_2, 6_3, 6_4, 6_5$.

[3] One should in particular consider the pitfalls pointed out in Appendix 10 of this reference.

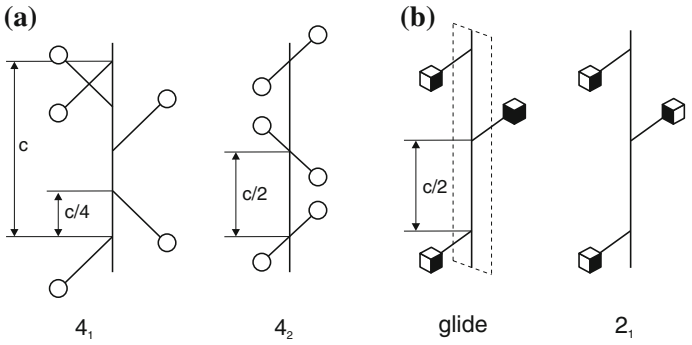


Fig. 3.7 (**a**) Schematic drawing of a 4_1 and 4_2 screw axis. (**b**) Schematic drawing of an axial glide reflection. The mirror plane is shown with *dashed outline*. Opposite faces of the cube have opposite color. For comparison a 2_1 screw axis is shown

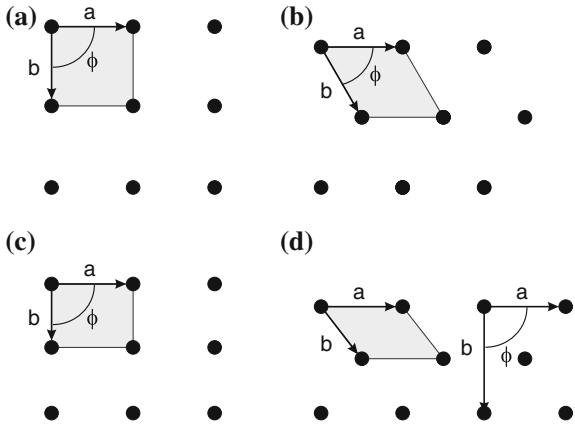


Fig. 3.8 The two-dimensional Bravais lattices with the primitive unit cells: (**a**) *square lattice* ($a = b$, $\phi = 90°$), (**b**) *hexagonal lattice* ($a = b$, $\phi = 60°$), (**c**) *rectangular lattice* ($a \neq b$, $\phi = 90°$), (**d**) *centered-rectangular lattice* ($a \neq b$, $\phi = 90°$, for the (nonprimitive) rectangular unit cell shown on the *right*)

3.3.4 2D Bravais Lattices

There are five two-dimensional (2D) Bravais lattices (Fig. 3.8) which are distinct and fill all (2D) space. These are very important for the description of symmetries at surfaces. The 2D Bravais lattices are the square, hexagonal, rectangular and centered-rectangular lattice.

3.3.5 3D Bravais Lattices

In three dimensions, the operations of the point group results in fourteen 3D Bravais lattices (Fig. 3.9), that are categorized into seven crystal classes (trigonal, monoclinic,

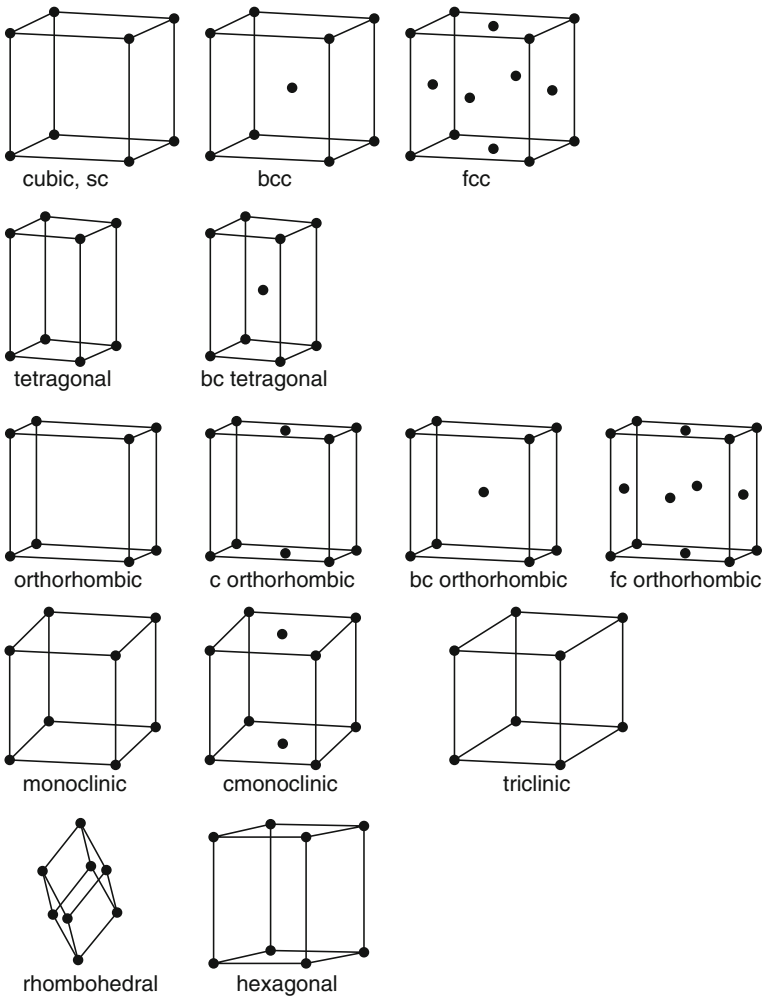


Fig. 3.9 The 14 three-dimensional Bravais lattices: cubic (sc: simple cubic, bcc: body-centered cubic, fcc: face-centered cubic), tetragonal (simple and body-centered), orthorhombic (simple, centered, body-centered and face-centered), monoclinic (simple and centered), triclinic, rhombohedral and hexagonal

rhombic, tetragonal, cubic, rhombohedral and hexagonal). These classes are discerned by the conditions for the lengths and the mutual angles of the vectors that span the lattice (Table 3.1). Some classes have several members. The cubic crystal can have a simple (sc), face-centered (fcc) or body-centered (bcc) lattice.

In the following, some of the most important lattices, in particular those most relevant to semiconductors, will be treated in some more detail.

Table 3.1 Conditions for lengths and angles for the 7 crystal classes

System	#	Lattice symbol	Conditions for the usual unit cell
Triclinic	1		None
Monoclinic	2	s, c	$\alpha = \gamma = 90°$ or $\alpha = \beta = 90°$
Orthorhombic	4	s, c, bc, fc	$\alpha = \beta = \gamma = 90°$
Tetragonal	2	s, bc	$a = b$, $\alpha = \beta = \gamma = 90°$
Cubic	3	s, bc, fc	$a = b = c$, $\alpha = \beta = \gamma = 90°$
Trigonal	1		$a = b$, $\alpha = \beta = 90°$, $\gamma = 120°$
(Rhombohedral)	1		$a = b = c$, $\alpha = \beta = \gamma$
Hexagonal	1		$a = b$, $\alpha = \beta = 90°$, $\gamma = 120°$

Note that only the positive conditions are listed. The rhombohedral system is a special case of the trigonal class. Conditions for the trigonal and hexagonal classes are the same, however, trigonal symmetry includes a single C_3 or S_6 axis, while hexagonal symmetry includes a single C_6 or S_6^5 axis

Cubic fcc and bcc Lattices

The primitive translation vectors for the cubic face-centered (fcc) and the cubic body-centered (bcc) lattice are shown in Figs. 3.10 and 3.11, respectively. Many metals crystallize in these lattices, e.g., copper (fcc) and tungsten (bcc).

In the fcc lattice, one lattice point sits in the center of each of the six faces of the usual cubic unit cell. The vectors spanning the primitive unit cell are

$$\begin{aligned} \mathbf{a}_1 &= \frac{a}{2}\,(1,1,0) \\ \mathbf{a}_2 &= \frac{a}{2}\,(0,1,1) \\ \mathbf{a}_3 &= \frac{a}{2}\,(1,0,1)\,. \end{aligned} \tag{3.2}$$

In the bcc lattice, one extra lattice point sits at the intersection of the three body diagonals at $(\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3)/2$. The vectors spanning the primitive unit cell are

$$\begin{aligned} \mathbf{a}_1 &= \frac{a}{2}\,(1,1,-1) \\ \mathbf{a}_2 &= \frac{a}{2}\,(-1,1,1) \\ \mathbf{a}_3 &= \frac{a}{2}\,(1,-1,1)\,. \end{aligned} \tag{3.3}$$

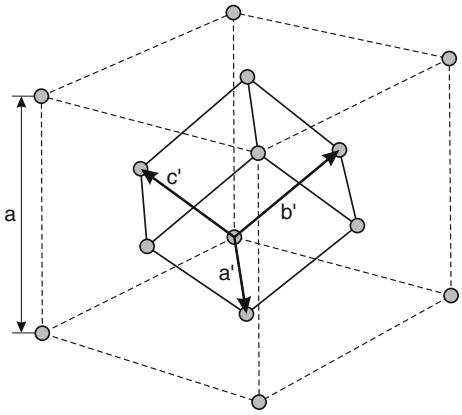


Fig. 3.10 Primitive translations of the fcc lattice. These vectors connect the origin with the face-center points. The primitive unit cell is the rhombohedron spanned by these vectors. The primitive translations $\mathbf{a}'$, $\mathbf{b}'$ and $\mathbf{c}'$ are given in (3.2). The angle between the vectors is 60°

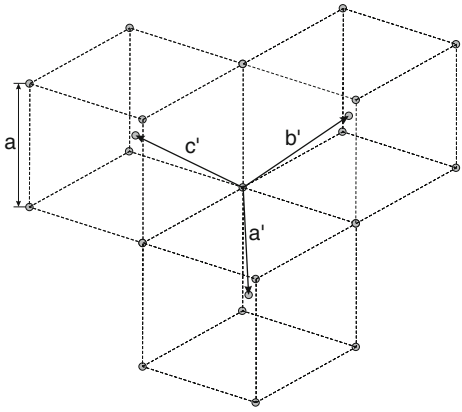


Fig. 3.11 Primitive translations of the bcc lattice. These vectors connect the origin with the lattice points in the cube centers. The primitive unit cell is the rhombohedron spanned by these vectors. The primitive translations $\mathbf{a}'$, $\mathbf{b}'$ and $\mathbf{c}'$ are given in (3.3). The angle between the vectors is ≈70.5°

Hexagonally Close Packed Lattice (hcp)

The 2D hexagonal Bravais lattice fills a plane with spheres (or circles) with maximum filling factor. There are two ways to fill space with spheres and highest filling factor. One is the fcc lattice, the other is the hexagonally close packed (hcp) structure. Both have a filling factor of 74 %.

For the hcp, we start with a hexagonally arranged layer of spheres (A), see Fig. 3.12. Each sphere has six next-neighbor spheres. This could, e.g., be a plane in the fcc perpendicular to the body diagonal. The next plane B is put on top in such a way that each new sphere touches three spheres of the previous layer. The third plane can now be added in two different ways: If the spheres of the third layer are vertically on top of the spheres of layer A, a plane A' identical to A has been created that is shifted from A along the stacking direction (normally called the c-axis) by

$$c_{\mathrm{hcp}} = \sqrt{8/3}\, a \approx 1.633\, a. \tag{3.4}$$

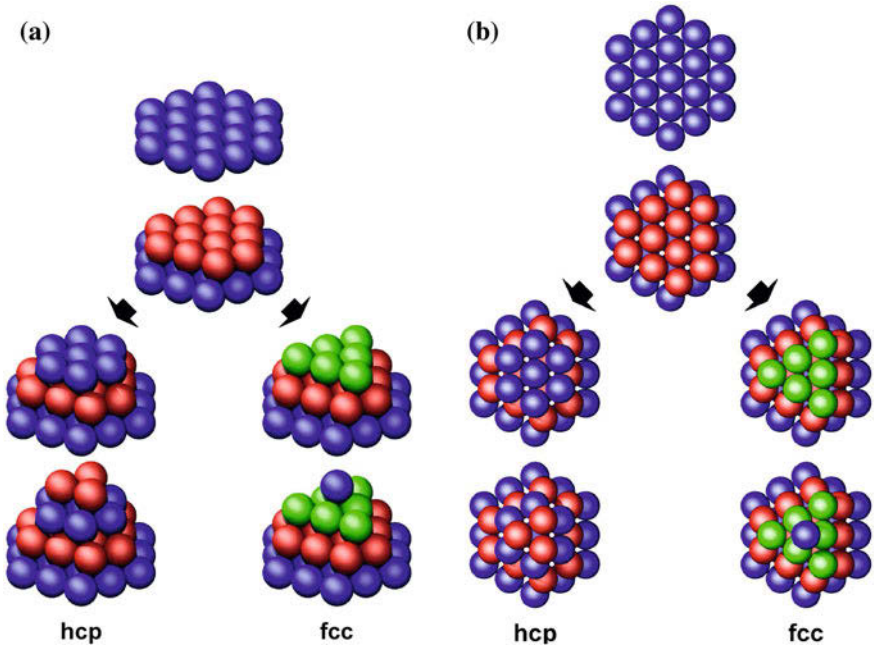


Fig. 3.12 Structure of the (**a**) hcp and (**b**) fcc lattice. For hcp the stacking (along the c-axis) is ABABAB. . ., for fcc (along the body diagonal) it is ABCABCABC. . .

The vectors spanning the primitive unit cell are

$$\begin{aligned} \mathbf{a}_1 &= \frac{a}{2}\left(1, -\sqrt{3}, 0\right) \\ \mathbf{a}_2 &= \frac{a}{2}\left(1, \sqrt{3}, 0\right) \\ \mathbf{a}_3 &= c\ (0, 0, 1)\,. \end{aligned} \tag{3.5}$$

The hcp stacking order is ABABAB. . . for hcp, the coordination number is 12. In the fcc structure, the third layer is put on the thus far unfilled positions and forms a new layer C. Only the forth layer is again identical to A and is shifted by

$$c_{\text{fcc}} = \sqrt{6}\,a \approx 2.45\,a. \tag{3.6}$$

The fcc stacking order is ABCABCABC. . .

In the hexagonal plane of the fcc lattice (which will later be called a {111} plane) the distance between lattice points is $a = a_0/\sqrt{2}$, where a_0 is the cubic lattice constant. Thus $c = \sqrt{3}\,a_0$, just what is expected for the body diagonal.

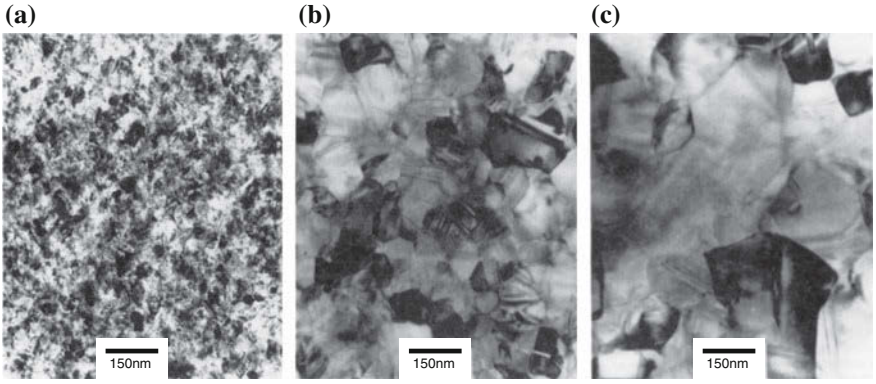


Fig. 3.13 Transmission electron micrographs of polycrystalline silicon (poly-Si). (**a**) As-deposited material from low-pressure chemical vapor deposition (LPCVD) at about 620 °C, grain size is about 30 nm. (**b**) After conventional processing (annealing at 1150 °C), average grain size is about 100 nm. (**c**) After annealing in HCl that provides enhanced point defect injection (and thus increased possibility to form larger grains), average grain size is about 250 nm. Adapted from [172]

For real materials with hexagonal lattice the ratio c/a deviates from the ideal value given in (3.4). Helium comes very close to the ideal value, for Mg it is 1.623, for Zn 1.861. Many hcp metals exhibit a phase transition to fcc at higher temperatures.

3.3.6 Polycrystalline Semiconductors

A polycrystalline material consists of crystal grains that are randomly oriented with respect to each other. Between two grains a (large-angle) grain boundary (see also Sect. 4.4.3) exists. An important parameter is the grain size and its distribution. It can be influenced via processing steps such as annealing. Polycrystalline semiconductors are used in cheap, large-area applications such as solar cells (e.g., polysilicon, $CuInSe_2$) or thin-film transistors (poly-Si) or as n-conducting contact material in MOS diodes (poly-Si) as shown in Fig. 3.13 (see also Fig. 21.28). Polycrystalline material can be fabricated from amorphous material using annealing procedures as discussed in Sect. 24.6.1 for silicon.

3.3.7 Amorphous Semiconductors

An amorphous material lacks the long-range order of the direct lattice. It is disordered on the atomic scale. Historically, amorphous Se (a-Se) has been investigated first; since the 1950s amorphous chalcogenides and a-Ge [173] and since the late 1960s a-Si [174] are researched. The field of amorphous oxides started in the mid 1950s

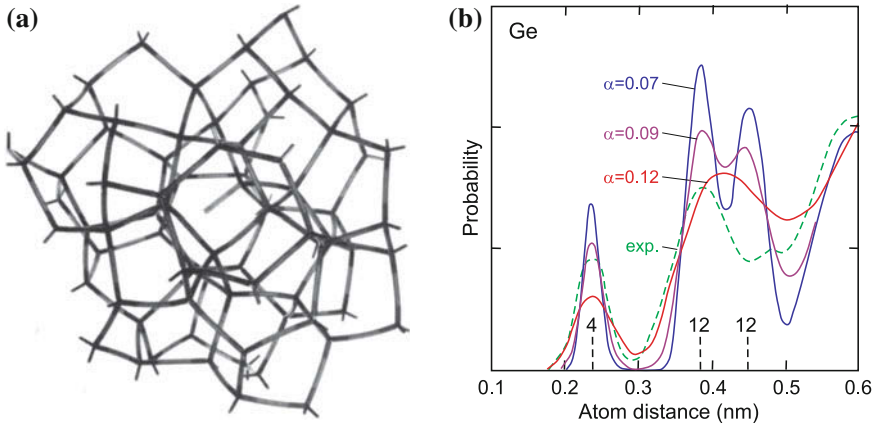


Fig. 3.14 (**a**) A continuous random network model of amorphous silicon containing a dangling bond in the center of the figure. Reprinted with permission from [178]. (**b**) Calculated radial atomic distribution functions of amorphous Ge (*solid lines*) for three different values of the disorder parameter α (3.7) as labeled and experimental result (*dashed line*). The positions of next, second-next and third-next neighbors are indicted by *vertical bars* with numbers of their multiplicity (4, 12, and 12). Adapted from [177]

with vanadates [175] and is currently very active with mixed-metal-based oxides [176] (cmp. Chap. 20).

The local quantum mechanics provides almost rigorous requirements for the bond length to next neighbors. The constraints for the bond angle are less strict. Covalently bonded atoms arrange in an open network with correlations up to the third and fourth neighbors. The short-range order is responsible for the observation of semiconductor properties such as an optical absorption edge and also thermally activated conductivity. In Fig. 3.14a a model of a continuous random network (with a bond-angle distortion of less than about 20 %) of a-Si is depicted. The diameter d_{SR} of the short-range order region is related to the disorder parameter α via [177]

$$d_{\mathrm{SR}} = \frac{a}{2\,\alpha}, \tag{3.7}$$

where a is the next-neighbor interatomic distance. For a diamond structure it is related to the lattice constant by $a = \sqrt{3}\,a_0/4$.

Typically, a significant number of dangling bonds exists. Bonds try to pair but if an odd number of broken bonds exists locally, an unsaturated, dangling bond remains. This configuration can be passivated by a hydrogen atom. Thus, the hydrogenation of amorphous semiconductors is very important, in particular for a-Si. A hydrogen atom can also break an overlong (and therefore weak) bond, saturate one side and eventually leave a dangling bond.

Amorphous material can be (re-)crystallized into crystalline, mostly polycrystalline material upon annealing. This is technologically very important for a-Si (see Sect. 24.6.1).

3.4 Important Crystal Structures

Now the crystal structures that are important for semiconductor physics will be discussed. These are mainly the rocksalt (PbS, CdO, ...), diamond (C, Si, Ge), zincblende (GaAs, InP, ...) and wurtzite (GaN, ZnO, ...) structures.

3.4.1 Rocksalt Structure

The rocksalt (rs, NaCl, space group 225, $Fm\bar{3}m$) structure (Fig. 3.15a) consists of a fcc lattice with the period a and a diatomic base in which the Cl atom is positioned at (0, 0, 0) and the Na atom at (1/2, 1/2, 1/2)a with a distance $\sqrt{3}\,a/2$. Materials that crystallize (under normal conditions) in the rocksalt lattice are, e.g., KCl, KBr, PbS (galena), PbSe, PbTe, AgBr, MgO, CdO, MnO. AlN, GaN and InN undergo, under high pressure, a phase transition from the wurtzite into the rocksalt structure.

3.4.2 CsCl Structure

The CsCl structure (space group 221, $Pm\bar{3}m$) (Fig. 3.15b) consists of a simple cubic lattice. Similar as for the rocksalt structure, the base consists of different atoms

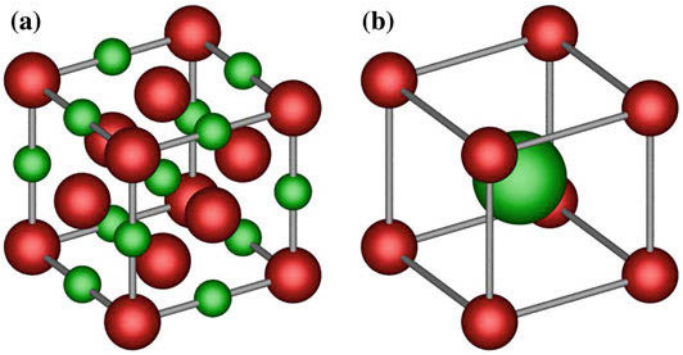


Fig. 3.15 (**a**) Rocksalt (NaCl) structure, (**b**) CsCl structure

at (0, 0, 0) and (1/2, 1/2, 1/2)a. Typical crystals with CsCl-structure are TlBr, TlI, CuZn (β-brass), AlNi.

3.4.3 Diamond Structure

The diamond structure (C, space group 227, Fd$\bar{3}$m) (Fig. 3.16a) has the fcc lattice. The base consists of two identical atoms at (0, 0, 0) and (1/4, 1/4, 1/4)a. Each atom has a tetrahedral configuration. The packing density is only about 0.34. The ABC-type stacking along the [111]-direction is visualized in Fig. 3.17a. The materials that crystallize in the diamond lattice are C, Ge, Si and α-Sn. Silicon as the most important semiconductor is particularly well researched [179].

The diamond structure (point group O_h) has an inversion center, located between the two atoms of the base, i.e. at (1/8, 1/8, 1/8)a. The radii of the wavefunctions for various group-IV elements increases with the order number (Table 3.2), and accordingly the lattice constant increases.

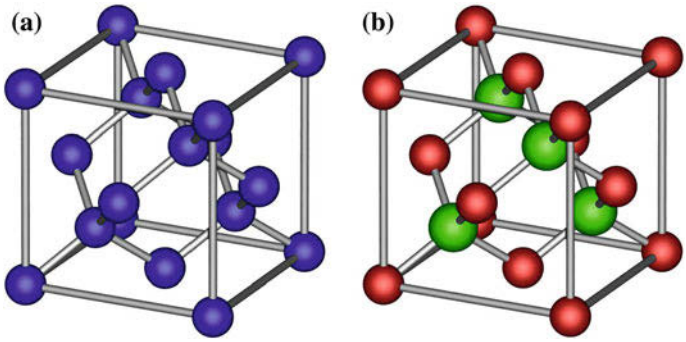


Fig. 3.16 (**a**) Diamond structure and (**b**) zincblende structure (*red spheres*: A atoms, *green spheres*: B atoms). The tetragonal bonds are indicated

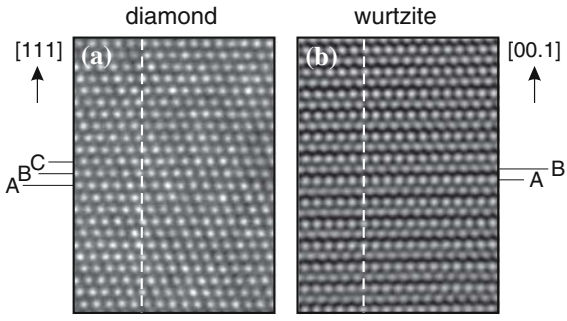


Fig. 3.17 HRTEM images of (**a**) diamond structure (Si, {110} cross section) and (**b**) wurtzite structure (GaN, ⟨10.0⟩ azimuth). The ABC and AB stacking is indicated

Table 3.2 Radii of the wavefunctions in the diamond structure, r_s and r_p are related to s^1p^3, r_d to $s^1p^2d^1$ and lattice constant a_0

	r_s (nm)	r_p (nm)	r_d (nm)	a_0 (nm)
C	0.121	0.121	0.851	0.3567
Si	0.175	0.213	0.489	0.5431
Ge	0.176	0.214	0.625	0.5646

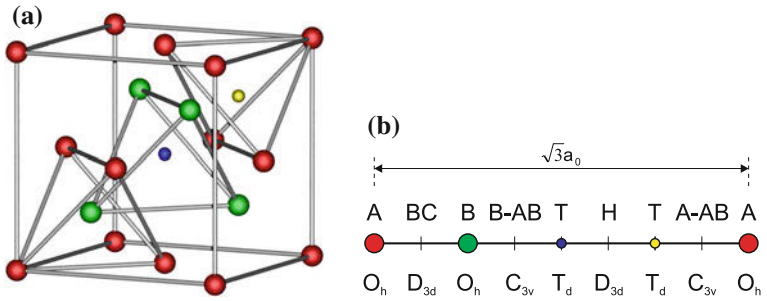


Fig. 3.18 (**a**) Unit cell of the zincblende structure with the indication of tetragonal symmetries. The position of the small *yellow* (*blue*) *sphere* is the tetrahedrally configured unoccupied positions of the A (B) sublattice, denoted with 'T' in part (**b**). (**b**) Line along [111] in the zincblende structure. The positions of the A and B atoms are denoted by *red* and *green circles* as labeled. Other positions are called the bond center ('BC'), antibonding ('AB') relative to A and B atoms ('A–AB', 'B–AB'), hexagonal ('H') and tetrahedral position ('T', *blue* and *yellow circles*)

In Fig. 3.18a the unit cell with tetragonal symmetry of three places along the ⟨111⟩ direction is shown. In Fig. 3.18b the arrangement of atoms along ⟨111⟩ is depicted. The symmetry along this line is at least C_{3v}. At the atoms sites it is O_h. The bond center (BC) and the hexagonal (H) position are a center of inversion and have D_{3d} symmetry. The unoccupied 'T' positions have T_d symmetry. High-pressure phases of silicon are already found in indentation experiments [180].

We note that α-Sn has little current importance. The diamond structure α–Sn phase is stable below 13.2 °C. The addition of Ge inhibits the retransformation to metallic tin up to higher temperatures (e.g., 60 °C for 0.75 weight percent Ge). The properties of gray tin are reviewed in [181].

3.4.4 Zincblende Structure

The zincblende (sphalerite,[4] ZnS, space group 216, $F\bar{4}3m$) structure (Fig. 3.16b) has a fcc lattice with a diatomic base. The metal (A) atom is at (0, 0, 0) and the

[4]Zincblende technically means the material ZnS which occurs in sphalerite (cubic) and wurtzite (hexagonal) phase. However, in the literature the term 'zincblende' for the sphalerite structure is common and used throughout this book.

nonmetal (B) atom is at $(1/4, 1/4, 1/4)a$. Thus the cation and anion sublattices are shifted with respect to each other by a quarter of the body diagonal of the fcc lattice. The atoms are tetrahedrally coordinated, a Zn atom is bonded to four S atoms and vice versa. However, no inversion center is present any longer (point group T_d). In the zincblende structure the stacking order of diatomic planes along the body diagonal is aAbBcCaAbBcC...

Many important compound semiconductors, such as GaAs, InAs, AlAs, InP, GaP and their alloys (cf. Sect. 3.7), but also the II–VI compounds ZnS, ZnSe, ZnTe, HgTe and CdTe and halides, including AgI, CuF, CuCl, CuBr, and CuI, crystallize in the zincblende structure.

Four-fold coordinated materials (zincblende and wurtzite) typically undergo a phase transition into 6-fold coordinated structures upon hydrostatic pressure [182]. For GaAs under pressure see [183].

3.4.5 *Wurtzite Structure*

The wurtzite structure (ZnS, space group 186, $P6_3mc$) is also called the hexagonal ZnS structure (because ZnS has both modifications). It consists (Fig. 3.19) of a hcp lattice with a diatomic base. The c/a ratio typically deviates from the ideal value $\zeta_0 = \sqrt{8/3} \approx 1.633$ (3.4) as listed in Table 3.3. The c-axis is a 6_3 screw axis.

The Zn atom is located at (0, 0, 0), the S atom at $(0, 0, \sqrt{3/8})a$. This corresponds to a shift of $3/8\,c$ along the c-axis. This factor is called the cell-internal parameter u. For real wurtzite crystals u deviates from its ideal value $u_0 = 3/8 = 0.375$, e.g., for group-III nitrides $u > u_0$. The diatomic planes have a stacking order of aAbBaAbB... along the [00.1]-direction as visualized in Fig. 3.17b.

In Fig. 3.20 the different local structural environment of the atoms in the zincblende and wurtzite structure is shown.

Many important semiconductors with large band gap crystallize in the wurtzite structure, such as GaN, AlN, InN, [185] ZnO, [186] SiC, [187], CdS und CdSe.

Table 3.3 c/a ratio of various wurtzite semiconductors

Material	ξ (%)	Material	ξ(%)	Material	ξ (%)	Material	ξ (%)
AlN	−2.02	CdS	−0.61	CuBr	0.43	BeO	−0.61
GaN	−0.49	CdSe	−0.18	CuCl	0.55	ZnO	−1.9
InN	−1.35	CdTe	0.25	CuI	0.74	6H-SiC	0.49
ZnS	0.25	MgS	−0.80	AgI	0.12	BN	0.74
ZnSe	0.06	MgSe	−0.67	ZnTe	0.74	MgTe	−0.67

Listed is $\xi = (c/a - \zeta_0)/\zeta_0$. Data based on [184]

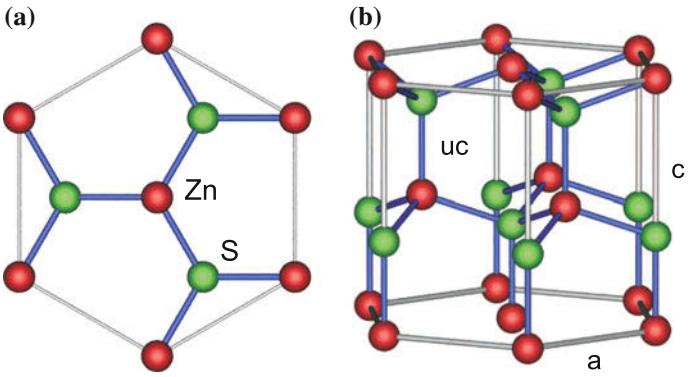


Fig. 3.19 (**a**) *Top view* (along the *c*-axis) and (**b**) *side view* of the wurtzite structure with the tetragonal bonds indicated. The *top* (*bottom*) surface of the depicted structure is termed the Zn-face, (00.1) (O-face, $(00.\bar{1})$)

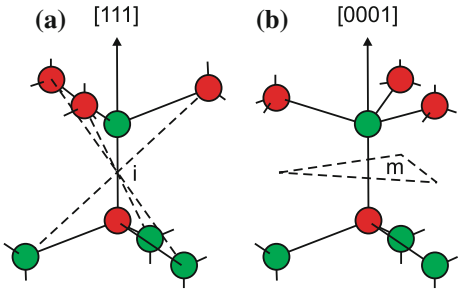


Fig. 3.20 Comparison of the tetragonal bonds in the (**a**) zincblende and (**b**) wurtzite structure (*i*: inversion center, *m*: symmetry plane)

3.4.6 Chalcopyrite Structure

The chalcopyrite [188] (ABC_2, named after 'fool's gold' $CuFeS_2$, space group 122, $I\bar{4}2d$) structure is relevant for I–III–VI_2 (with chalcogenide anions) and II–IV–V_2 (with pnictide anions) semiconductors such as, e.g., $(Cu,Ag)(Al,Ga,In)(S,Se,Te)_2$ and $(Mg,Zn,Cd)(Si,Ge,Sn)(As,P,Sb)_2$. A nonmetallic anion atom ('C') is tetrahedrally bonded to two different types of cation atoms ('A' and 'B') as shown in Fig. 3.21. The local surrounding of each anion is identical, two of both the A and B atoms. The structure is tetragonal. The aspect ratio $\eta = c/(2a)$ deviates from its ideal value 1; typically $\eta < 1$ [189, 190].

If the C atom is in the tetrahedral center of the two A and two B atoms, the bond lengths R_{AC} and R_{BC} of the A–C and B–C bonds, respectively, are equal. Since the ideal A–C and B–C bond lengths d_{AC} and d_{BC} are typically unequal, this structure is strained. The common atom C is therefore displaced along [100] and [010] such

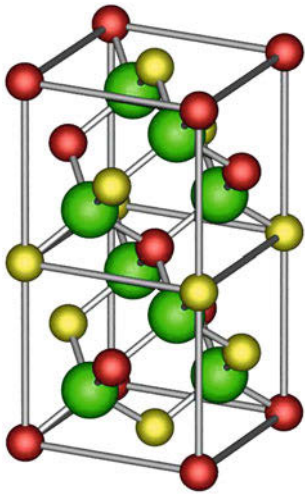


Fig. 3.21 Chalcopyrite structure, *red* and *yellow spheres* denote the metal species. The bigger *green spheres* represent the nonmetal anion

that it is closer (if $d_{AC} < d_{BC}$) to the pair of A atoms and further away from the B atoms. The displacement parameter is

$$u = \frac{1}{4} + \frac{R_{AC}^2 - R_{BC}^2}{a^2} \tag{3.8}$$

and it deviates from the ideal value $u_0 = 1/4$ for the zincblende structure as listed in Table 3.4 for a number of chalcopyrite compounds. In the chalcopyrite structure

$$R_{AC} = a\sqrt{u^2 + \frac{1+\eta^2}{16}} \tag{3.9a}$$

$$R_{BC} = a\sqrt{\left(u - \frac{1}{2}\right)^2 + \frac{1+\eta^2}{16}}. \tag{3.9b}$$

The minimization of the microscopic strain yields (in first order) [191]

$$u \cong \frac{1}{4} + \frac{3}{8}\frac{d_{AC}^2 - d_{BC}^2}{d_{AC}^2 + d_{BC}^2}. \tag{3.10}$$

Compounds with $u > u_c$, $u_c = 0.265$ being a critical displacement parameter, (or $u < 1/2 - u_c = 0.235$) are stable with regard to cation disorder [190]. In Fig. 3.22 the correlation of the calculated value for u according to (3.10) and the experimental values is shown.

Table 3.4 Lattice nonideality parameters η and u (from (3.10)) of various chalcopyrite compounds and their experimentally observed disorder stability (+/− indicates compound with/without order–disorder (D–O) transition, respectively)

	η	u	D–O		η	u	D–O
$CuGaSe_2$	0.983	0.264	+	$ZnSiAs_2$	0.97	0.271	−
$CuInSe_2$	1.004	0.237	+	$ZnGeAs_2$	0.983	0.264	+
$AgGaSe_2$	0.897	0.287	−	$CdSiAs_2$	0.92	0.294	−
$AgInSe_2$	0.96	0.261	+	$CdGeAs_2$	0.943	0.287	
$CuGaS_2$	0.98	0.264		$ZnSiP_2$	0.967	0.272	−
$CuInS_2$	1.008	0.236	+	$ZnGeP_2$	0.98	0.264	+
$AgGaS_2$	0.895	0.288	−	$CdSiP_2$	0.92	0.296	−
$AgInS_2$	0.955	0.262		$CdGeP_2$	0.939	0.288	−

Data from [190]

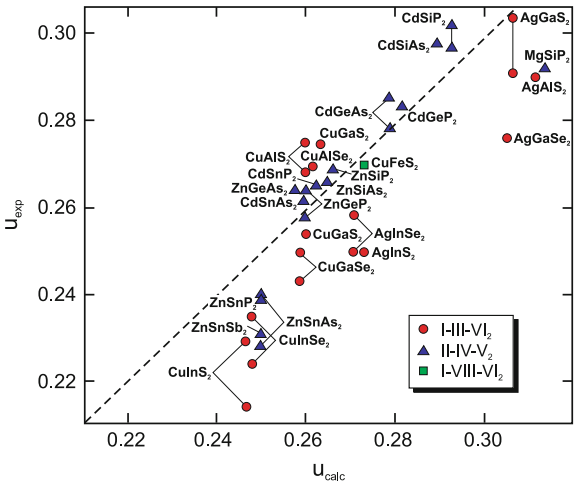


Fig. 3.22 Experimental values u_{exp} of the displacement parameter for various chalcopyrites versus the calculated value u_{calc} according to (3.10). The *dashed line* indicates $u_{exp} = u_{calc}$. Adapted from [191]

3.4.7 Spinel Structure

A large variety of ternary compounds of type $A^{II}B^{III}{}_2C^{VI}{}_4$ crystallize in the cubic spinel structure (spinel, $MgAl_2O_4$, space group 227, Fd3m). Typical elements are A: Mg, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd, Sn, B: Al, Ga, In, Ti, V, Cr, Mn, Fe, Co, Rh, Ni, and C: O, S, Se, Te.

As an example $ZnGa_2O_4$ (zinc gallate) has received attention as interfacial layer in ZnO/GaAs epitaxy [192], luminescent material [193], and as ferromagnetic semiconductor [194]. Also the (unwanted) appearance of zinc gallate inclusions is in

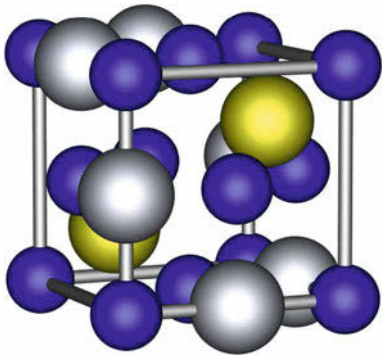


Fig. 3.23 Spinel AB_2C_4 crystal structure, the cations are depicted as *yellow* (A atoms) and *silver* (B atoms) *spheres*, the anions (C atoms) as *blue spheres*

competition with the formation of highly doped wurtzite ZnO:Ga; in the (normal) spinel cubic structure of $ZnGa_2O_4$ the Zn^{2+} ions still occupy tetrahedral sites, but the Ga^{3+} ions occupy octahedral sites instead of tetrahedral sites in doped wurtzite ZnO:Ga. (Sc,Al)MgO_4 (SCAM) is available as substrate material. Also $A^{VI}B^{II}{}_2C^{VI}{}_4$ compounds exist, e.g., GeB_2O_4 (with B = Mg, Fe, Co, Ni)

The anion atoms (C^{2-}) sit on a fcc lattice. The A atoms fill 1/8 of all tetraeder spaces and the B atoms fill half of all octaeder places (Fig. 3.23). Often the cations are charged A^{2+} and B^{3+}, e.g., in $ZnAl_2O_4$, $MgCr_2O_4$ or $ZnCo_2O_4$. Also A^{6+} and B^{1-} exists, e.g., in WNa_2O_4.

The cubic lattice constant is denoted as a. In real spinels, the anions deviate from the ideal fcc array which is accounted for by the parameter u, measuring the displacement of anions along the [111]-direction [195]; if the A-site cation is at (0,0,0), an anion is at (u, u, u). The cation-anion distances are given by [196]

$$R_{AC} = a\sqrt{3\,(u - 1/8)}, \tag{3.11a}$$

$$R_{BC} = a\sqrt{3\,u^2 - 2\,u - 3/8}. \tag{3.11b}$$

The ideal value is $u = 1/4$; examples are $u = 0.2624$ for $MgAl_2O_4$, $u = 0.2617$ for $ZnGa_2O_4$ and $u = 0.2409$ for $SiFe_2O_4$ [196].

In the inverted spinel structure, for $A^{II}B^{III}{}_2C^{VI}{}_4$ compounds, the cations are distributed like $B(AB)C_4$, i.e. the B cations occupy tetraeder and octaeder places, e.g., in $Mg^{2+}(Mg^{2+}Ti^{4+})O_4^{2-}$ or $Fe^{3+}(Ni^{2+}Fe^{3+})O_4^{2-}$. Examples are magnetite (Fe_3O_4), a material with high spin polarization, or $MgFe_2O_4$. Also $A^{VI}B^{II}{}_2C^{VI}{}_4$ compounds exist in this structure, e.g., SnB_2O_4 (with B = Mg, Mn, Co, Zn), TiB_2O_4 (with B = Mg, Mn, Fe, Co, Zn), and VB_2O_4 (with B = Mg, Co, Zn).

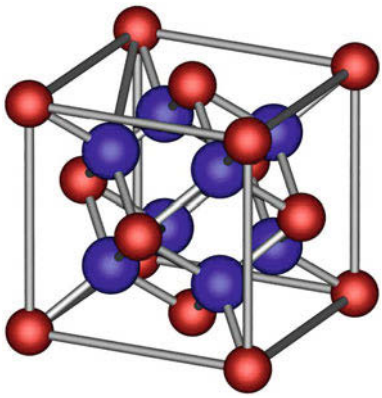


Fig. 3.24 Fluorite crystal structure, the cations are depicted as *red spheres*, the anions as *blue spheres*

3.4.8 Fluorite Structure

Named after the minerale fluorite (CaF_2, space group 225, Fm3m), this structure for binary ionic compounds occurs when the cation valence is twice the anion valence, e.g., for (cubic) ZrO_2 (zirconia) or HfO_2. The lattice is fcc with a triatomic base. At (0,0,0) is the cation (e.g., Zr^{4+}), the anions (e.g., O^{2-}) are at (1/4, 1/4, 1/4) a (as in the zincblende structure) and (3/4, 3/4, 3/4) a (Fig. 3.24). The anion atom positions are on a simple cubic lattice with lattice constant $a/2$. Zirconia can crystallize in various phases [197], the most prominent being the monoclinic, tetragonal and cubic phases. The cubic phase can be extrinsically stabilized using yttrium [198, 199] (YSZ, yttria-stabilized zirconia). Hafnium oxide has the remarkable property that the HfO_2/Si interface is stable and allows the fabrication of transistor gate oxides with high dielectric constant (see Sect. 24.5.5).

3.4.9 Delafossite Structure

The I–III–O_2 materials crystallize in the trigonal delafossite ($CuFeO_2$, space group 166, $R\bar{3}m$) structure (Fig. 3.25). This structure is also called caswellsilverite ($NaCrS_2$). In Table 3.5 the lattice parameters of some delafossite compounds are given. The (Cu, Ag) (Al, Ga, In)O_2 materials are transparent conductive oxides (TCO). We note that Pt and Pd as group-I component create metal-like compounds because of the d^9 configuration as opposed to the d^{10} configuration of Cu and Ag.

3.4.10 Perovskite Structure

The perovskite structure for ABO: 3 materials (calcium titanate, $CaTiO_3$, space group 62, Pnma) (Fig. 3.26) is relevant for ferroelectric semiconductors (cf. Sect. 15.3). It is cubic with the Ca (or Ba, Sr) ions (charge state 2+) on the corners of the

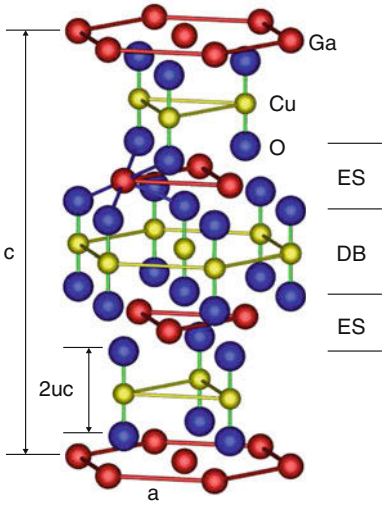


Fig. 3.25 Hexagonal unit cell of delafossite $CuGaO_2$. Oxygen atoms are bonded to the Cu in a dumbbell ('DB') configuration. In the edge-sharing ('ES') layer the Ga atoms are octahedrally configured as GaO_6

Table 3.5 Lattice parameters a, c, and u of some delafossite compounds

	a (nm)	c (nm)	u (nm)
$CuAlO_2$	0.2858	1.6958	0.1099
$CuGaO_2$	0.2980	1.7100	0.1073*
$CuInO_2$	0.3292	1.7388	0.1056*

Theoretical values are shown with asterisk. Data from [200]

cube, the O ions (2−) on the face centers and the Ti (4+) in the body center. The lattice is simple cubic, the base is Ca at (0, 0, 0), O at (1/2, 1/2, 0), (1/2, 0, 1/2) and (0, 1/2, 1/2) and Ti at (1/2, 1/2, 1/2). The ferroelectric polarization is typically evoked by a shift of the negatively and positively charged ions relative to each other. $LaAlO_3$ (lanthanium aluminate) is available as substrate material (space group 226, $Fm\bar{3}c$ [201]). Perovskites are also important for high temperature superconductivity.

3.4.11 NiAs Structure

The NiAs structure (space group 194, $P6_3/mmc$) (Fig. 3.27) is relevant for magnetic semiconductors, such as MnAs, and also occurs in the formation of Ni/GaAs Schottky contacts [202]. The structure is hexagonal. The arsenic atoms form a hcp structure and are trigonal prismatically configured with six nearest metal atoms. The metal atoms form hcp planes and fill all octahedral holes of the As lattice. For a cubic close

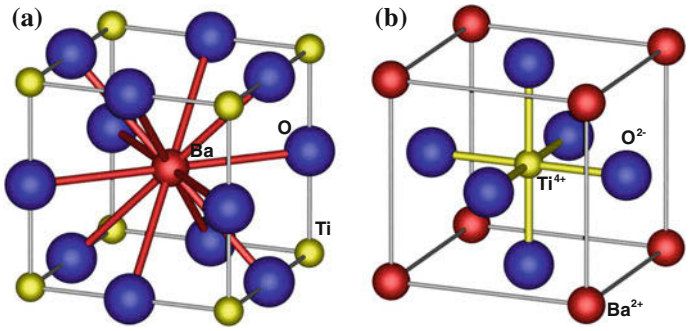


Fig. 3.26 Perovskite structure ($BaTiO_3$), (**a**) A cell with 12-fold (cuboctahedrally) configured Ba, (**b**) B cell with octahedrally configured Ti

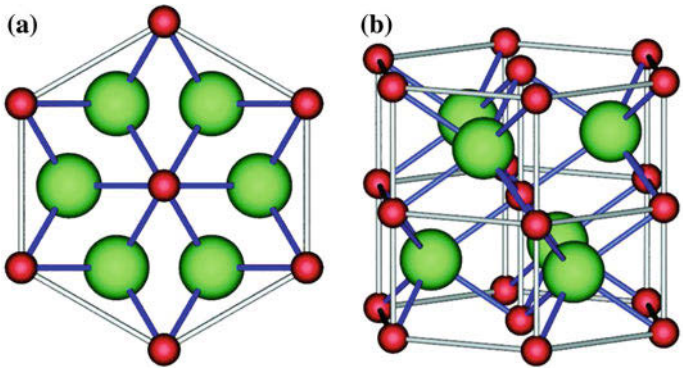


Fig. 3.27 NiAs structure, metal atoms: *dark grey*, chalcogenide atoms: *light grey*

packed, i.e. fcc, structure this would correspond to the rocksalt crystal. The stacking is ABACABAC… (A: Ni, B,C: As).

3.4.12 Further Structures

There are many other crystal structures that have relevance for semiconductor materials. Among them are the

- corundum structure (Al_2O_3, space group 167, $R\bar{3}c$) occurring, e.g., for sapphire substrates used in epitaxy or for gallium oxide α-Ga_2O_3 (Ga_2O_3 is a multiphase material [203])
- bixbyite structure (In_2O_3, δ-Ga_2O_3, space group 206, $Ia\bar{3}$) (see Fig. 20.3)
- β-Ga_2O_3 monoclinic structure (space group 12, C2/m)
- quartz (SiO_2) structures, α-quartz (space group 154, $P3_221$) and β-quartz (space group 180, $P6_222$)

Space does not permit to discuss these and other structures in more detail here. The reader should refer to textbooks on crystallography, e.g., [204–206], and space groups [171, 207]. A good source for information and images of crystal structures on the web is [208].

3.5 Polytypism

In polytype materials the stacking order is not only hcp or fcc but takes different sequences, such as, e.g., ACBCABAC as the smallest unit cell along the stacking direction. A typical example is SiC, for which in addition to hcp and fcc 45 other stacking sequences are known. The largest primitive unit cell of SiC [187] contains 594 layers. Some of the smaller polytypes are sketched in Fig. 3.28. In Fig. 3.29 cubic

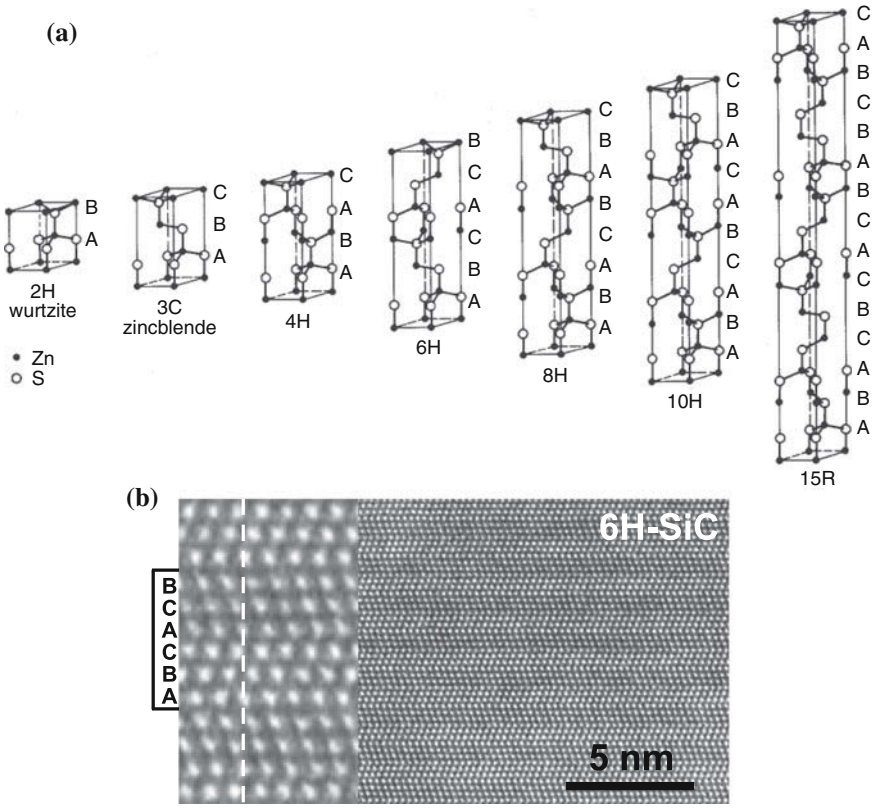


Fig. 3.28 (**a**) Polytypes of the zincblende and wurtzite lattice (found in SiC), the letters A, B and C denote the three possible positions of the diatomic layers (see Fig. 3.12). (**b**) High resolution TEM image of 6H-SiC. For the enlarged view on the *left*, the unit cell and the stacking sequence are indicated. Adapted from [209]

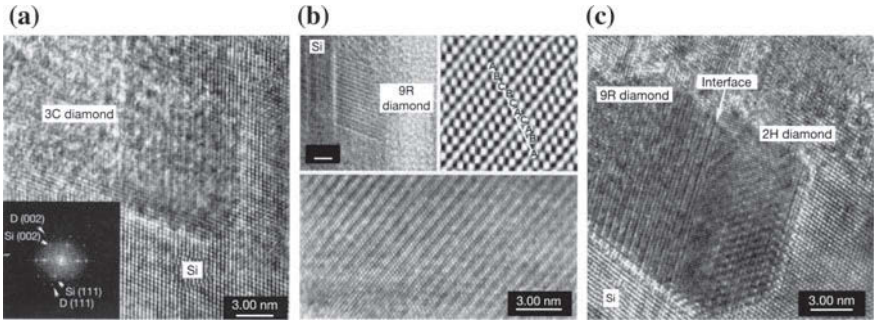


Fig. 3.29 Polytypes of diamond found in crystallites (metastable phases in silicon). (**a**) cubic type (3C) with stacking ABC, *inset* shows a diffractogram and the alignment of the C and Si lattice, (**b**) rhombohedral 9R crystallite with ABCBCACABA stacking, (**c**) 9R phase with interface to a hexagonal 2H (AB stacking) phase. Reprinted with permission from Nature [210], © 2001 Macmillan Magazines Limited

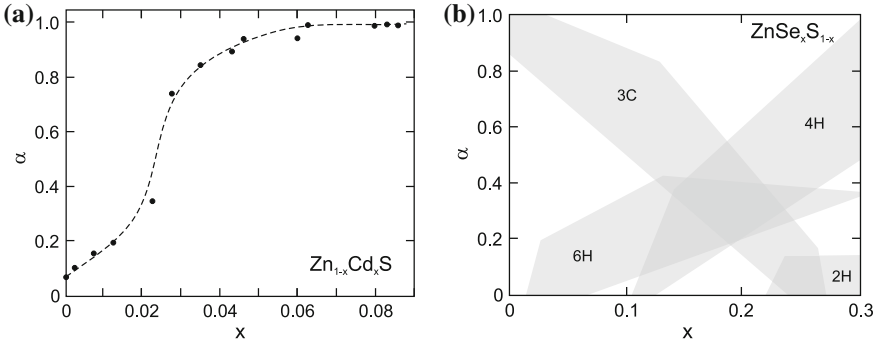


Fig. 3.30 (**a**) Hexagonality index α (3.12) of $Zn_{1-x}Cd_xS$ for various ternary compositions. *Dashed line* is a guide to the eye. (**b**) Regions of different polytypes in $ZnSe_xS_{1-x}$. Adapted from [211]

diamond crystallites and metastable hexagonal and orthorhombic phases (in silicon) are shown.

For the ternary alloy (cf. Sect. 3.7) $Zn_{1-x}Cd_xS$ the numbers n_h of diatomic layers with hexagonal stacking (AB) and n_c of layers with cubic stacking (ABC) have been investigated. CdS has wurtzite structure and ZnS mostly zincblende structure. The hexagonality index α as defined in (3.12) is shown in Fig. 3.30 for $Zn_{1-x}Cd_xS$

$$\alpha = \frac{n_h}{n_h + n_c}. \tag{3.12}$$

3.6 Reciprocal Lattice

The reciprocal lattice is of utmost importance for the description and investigation of periodic structures, in particular for X-ray diffraction [212], surface electron diffraction, phonons and the band structure. It is the quasi-Fourier transformation of the crystal lattice. The crystal lattice is also called the *direct* lattice, in order to distinguish it from the reciprocal lattice.

3.6.1 Reciprocal Lattice Vectors

When $\mathcal{R}$ denotes the set of vectors of the direct lattice, the set $\mathcal{G}$ of the reciprocal lattice vectors is given by the condition[5]

$$\exp(\mathrm{i}\,\mathbf{G}\cdot\mathbf{R}) = 1 \tag{3.13}$$

for all $\mathbf{R}\in\mathcal{R}$ and $\mathbf{G}\in\mathcal{G}$. Therefore, for all vectors $\mathbf{r}$ and a reciprocal lattice vector $\mathbf{G}$

$$\exp(\mathrm{i}\,\mathbf{G}\cdot(\mathbf{r}+\mathbf{R})) = \exp(\mathrm{i}\,\mathbf{G}\cdot\mathbf{r})\,. \tag{3.14}$$

Each Bravais lattice has a certain reciprocal lattice. The reciprocal lattice is also a Bravais lattice, since when $\mathbf{G}_1$ and $\mathbf{G}_2$ are two reciprocal lattice vectors, then this is obviously true also for $\mathbf{G}_1+\mathbf{G}_2$. For the primitive translation vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ of the direct lattice, the vectors $\mathbf{b}_1$, $\mathbf{b}_2$ and $\mathbf{b}_3$ that span the reciprocal lattice can be given directly as

$$\mathbf{b}_1 = \frac{2\pi}{V_a}(\mathbf{a}_2\times\mathbf{a}_3) \tag{3.15a}$$

$$\mathbf{b}_2 = \frac{2\pi}{V_a}(\mathbf{a}_3\times\mathbf{a}_1) \tag{3.15b}$$

$$\mathbf{b}_3 = \frac{2\pi}{V_a}(\mathbf{a}_1\times\mathbf{a}_2), \tag{3.15c}$$

where $V_a = \mathbf{a}_1\cdot(\mathbf{a}_2\times\mathbf{a}_3)$ is the volume of the unit cell spanned by the vectors $\mathbf{a}_i$. The volume of the unit cell in the reciprocal space is $V_a^* = (2\pi)^3/V_a$.

The vectors $\mathbf{b}_i$ fulfill the conditions

$$\mathbf{a}_i\cdot\mathbf{b}_j = 2\pi\,\delta_{ij}. \tag{3.16}$$

Thus, it is clear that (3.13) is fulfilled. For an arbitrary reciprocal lattice vector $\mathbf{G} = k_1\mathbf{b}_1 + k_2\mathbf{b}_2 + k_3\mathbf{b}_3$ and a vector $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ in direct space

[5] The dot product $\mathbf{a}\cdot\mathbf{b}$ of two vectors shall also be denoted as $\mathbf{ab}$.

we find

$$\mathbf{G} \cdot \mathbf{R} = 2\pi \ (n_1 k_1 + n_2 k_2 + n_3 k_3) \,. \tag{3.17}$$

The number in brackets is an integer. Additionally, we note that the reciprocal lattice of the reciprocal lattice is again the direct lattice. The reciprocal lattice of the fcc is bcc and vice versa. The reciprocal lattice of hcp is hcp (rotated by 30° with respect to the direct lattice).

For later, we note two important theorems. A (sufficiently well behaved) function $f(\mathbf{r})$ that is periodic with the lattice, i.e. $f(\mathbf{r}) = f(\mathbf{r} + \mathbf{R})$ can be expanded into a Fourier series with the reciprocal lattice vectors according to

$$f(\mathbf{r}) = \sum a_{\mathbf{G}} \ \exp(\mathrm{i}\,\mathbf{G} \cdot \mathbf{r}), \tag{3.18}$$

where $a_{\mathbf{G}}$ denotes the Fourier component of the reciprocal lattice vector $\mathbf{G}$, $a_{\mathbf{G}} = \int_V f(\mathbf{r}) \exp(-\mathrm{i}\,\mathbf{G} \cdot \mathbf{r})\, \mathrm{d}^3\mathbf{r}$. If $f(\mathbf{r})$ is lattice periodic, the integral given in (3.19) is zero unless $\mathbf{G}$ is a reciprocal lattice vector.

$$\int_V f(\mathbf{r}) \ \exp(-\mathrm{i}\,\mathbf{G} \cdot \mathbf{r}) \ \mathrm{d}^3\mathbf{r} = \begin{cases} a_{\mathbf{G}} \\ 0, \mathbf{G} \notin \mathcal{G} \end{cases} . \tag{3.19}$$

3.6.2 Miller Indices

A lattice plane is the set of all lattice points in a plane spanned by two independent lattice vectors $\mathbf{R_1}$ and $\mathbf{R_2}$. The lattice points on that plane form a two-dimensional Bravais lattice. The entire lattice can be generated by shifting the lattice plane along its normal $\mathbf{n} = (\mathbf{R_1} \times \mathbf{R_2})/|\mathbf{R_1} \times \mathbf{R_2}|$. The plane belongs to the reciprocal lattice vector $\mathbf{G_n} = 2\pi\mathbf{n}/d$, d being the distance between planes.

This correspondence between reciprocal lattice vectors and sets of planes allows the orientation of planes to be described in a simple manner. The shortest reciprocal lattice vector perpendicular to the plane is used. The coordinates with respect to the primitive translation vectors of the reciprocal space $\mathbf{b}_i$ form a triplet of integer numbers and are called Miller indices [213].

The plane described by $\mathbf{G_n} \cdot \mathbf{r} = A$ fulfills the condition for a suitable value of A. The plane intersects the axes $\mathbf{a}_i$ at the points $x_1\,\mathbf{a}_1$, $x_2\,\mathbf{a}_2$ and $x_3\,\mathbf{a}_3$. Thus we find $\mathbf{G_n}\, x_i\, \mathbf{a}_i = A$ for all i. From (3.17) follows $\mathbf{G_n} \cdot \mathbf{a}_1 = 2\pi\, h$, $\mathbf{G_n} \cdot \mathbf{a}_2 = 2\pi\, k$ and $\mathbf{G_n} \cdot \mathbf{a}_3 = 2\pi\, l$, where h, k and l are integers. The triplet of integer numbers (hkl), the reciprocal values of the axis intersections in the direct lattice, are the Miller indices. An example is shown in Fig. 3.31.

Planes are denoted as (hkl) with parentheses. The (outward) normal direction is denoted with $[hkl]$ (square brackets). A set of equivalent planes is denoted with curly brackets as $\{hkl\}$. For example, in the simple cubic lattice (100), (010), (001),

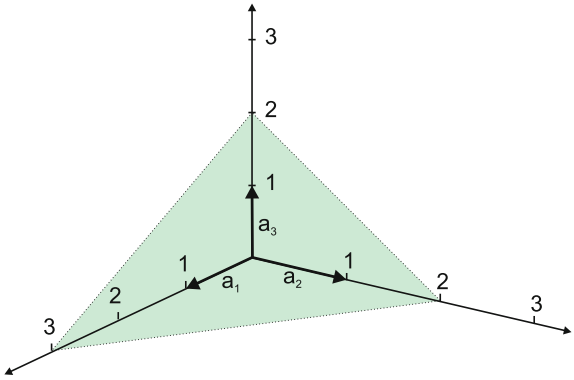


Fig. 3.31 The plane intersects the axes at 3, 2, and 2. The inverse of these numbers is 1/3, 1/2, and 1/2. The smallest integer numbers of this ratio form the Miller indices (233)

(−100), (0−10) are (00−1) equivalent and are denoted by {100}. (−100) can also be written as $(\bar{1}00)$. A set of equivalent directions is denoted with $\langle hkl \rangle$.

In a cubic lattice the faces of the cubic unit cell are {001} and the planes perpendicular to the area (body) diagonals are {110} ({111}) (Fig. 3.32a). In the zincblende lattice the {111} planes consist of diatomic planes with Zn and S atoms. It depends on the direction whether the metal or the nonmetal is on top. These two cases are denoted by A and B. We follow the convention that the (111) plane is (111)A and the metal is on top (as in Fig. 3.16b). For each change of sign the type changes from A to B and vice versa, e.g., (111)A, $(1\bar{1}\,1)$B and $(\bar{1}\,\bar{1}\,\bar{1})$B. In Fig. 3.32b the in-plane directions for the (001), (110) and (111) planes are visualized.

In the wurtzite lattice the Miller indices are denoted as $[hklm]$ (Fig. 3.33). Within the (0001) plane three indices hkl are used that are related to the three vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ (see Fig. 3.33a) rotated with respect to each other by 120°. Of course, the four indices are not independent and $l = -(h + k)$. The third (redundant) index can be denoted as a dot. The c-axis [0001] is then denoted as [00.1]. Wurtzite (and trigonal, e.g., sapphire) substrates are available typically with a ($\langle 11.0 \rangle$), r ($\langle 01.2 \rangle$), m ($\langle 01.0 \rangle$) and c ([00.1]) surface orientations (Fig. 3.33b).

The distance of lattice planes $d = 2\pi/|\mathbf{G}|$ can be expressed via the Miller indices for cubic (3.20a), tetragonal (3.20b) and hexagonal (3.20c) crystals as

$$d^{\mathrm{c}}_{hkl} = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \tag{3.20a}$$

$$d^{\mathrm{t}}_{hkl} = \frac{a}{\sqrt{h^2 + k^2 + l^2\,(a/c)^2}} \tag{3.20b}$$

$$d^{\mathrm{h}}_{hkl} = \frac{a}{\sqrt{4\,(h^2 + hk + k^2)/3 + l^2\,(a/c)^2}} \tag{3.20c}$$

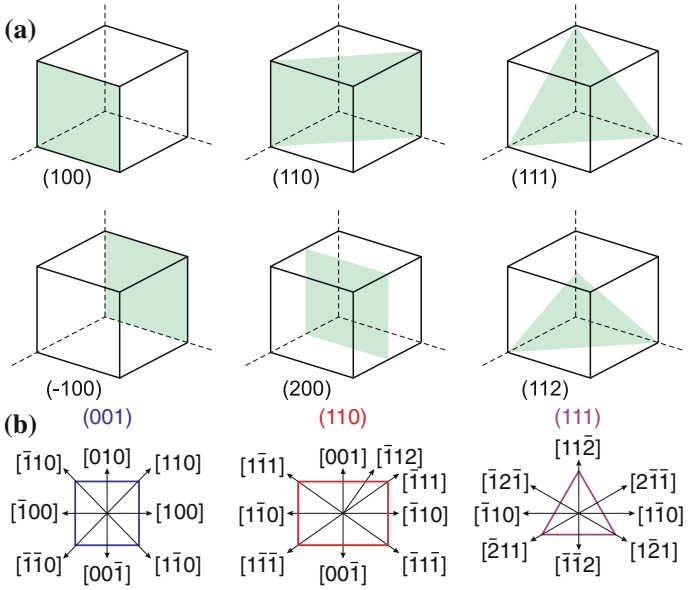


Fig. 3.32 (**a**) Miller indices of important planes for the simple cubic (and fcc, bcc) lattice. (**b**) Directions within three low index planes of cubic crystals

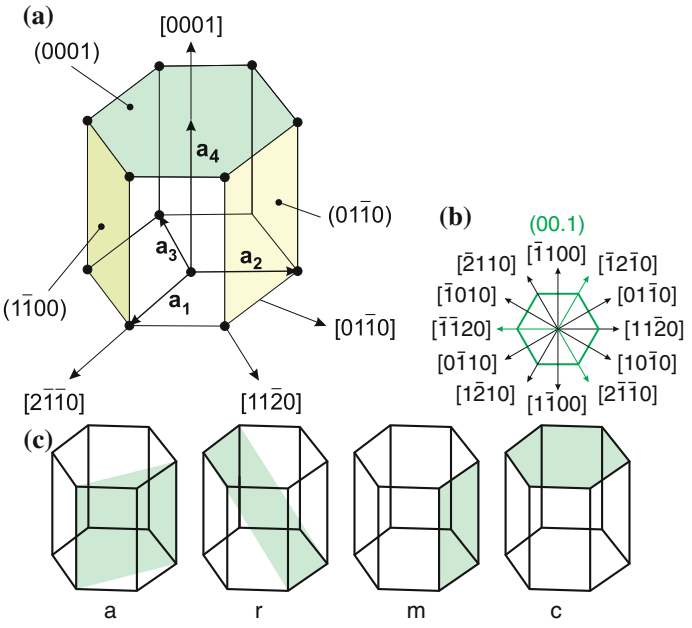


Fig. 3.33 (**a**, **b**) Miller indices for the wurtzite (or hcp) structure. (**c**) Orientation of the a-, r-, m-, and c-plane in the wurtzite structure

Table 3.6 High symmetry points and directions from Γ-point in the Brillouin zone of the fcc lattice

Point	**k**	Direction	$(0 \leq \lambda \leq 1)$
Γ	(0,0,0)		
X	$\frac{2\pi}{a}$ (0, 1, 0)	Δ	$\frac{2\pi}{a}$ (0, λ, 0)
K	$\frac{3\pi}{2a}$ (1, 1, 0)	Σ	$\frac{3\pi}{2a}$ (λ, λ, 0)
L	$\frac{\pi}{a}$ (1, 1, 1)	Λ	$\frac{\pi}{a}$ (λ, λ, λ)
W	$\frac{2\pi}{a}$ (1, 1/2, 0)		
U	$\frac{2\pi}{a}$ (1, 1/4, 1/4)		

Table 3.7 High symmetry points and directions from Γ-point in the Brillouin zone of the hcp lattice

Point	**k**	Direction	$(0 \leq \lambda \leq 1)$
Γ	(0,0,0)		
A	2π (0, 0, $\frac{1}{2c}$)	Δ	$\frac{\pi}{c}$ (0, 0, λ)
L	2π (0, $\frac{1}{\sqrt{3}a}$, $\frac{1}{2c}$)		
M	2π (0, $\frac{1}{\sqrt{3}a}$, 0)	Σ	$\frac{2\pi}{\sqrt{3}a}$ (0, λ, 0)
H	2π ($-\frac{1}{3a}$, $\frac{1}{\sqrt{3}a}$, $\frac{1}{2c}$)		
K	2π ($-\frac{1}{3a}$, $\frac{1}{\sqrt{3}a}$, 0)	T	2π ($-\frac{\lambda}{3a}$, $\frac{\lambda}{\sqrt{3}a}$, 0)

Useful formulas for the angle θ between a (hk.l)-plane and the [00.1]-direction in the cubic, tetragonal and wurtzite structures are:

$$\cos\theta^{\mathrm{c}} = \frac{l}{\sqrt{h^2 + k^2 + l^2}} \tag{3.21a}$$

$$\cos\theta^{\mathrm{t}} = \frac{l}{\sqrt{l^2 + \frac{c^2}{a^2}(h^2 + k^2)}} \tag{3.21b}$$

$$\cos\theta^{\mathrm{h}} = \frac{l}{\sqrt{l^2 + \frac{4}{3}\frac{c^2}{a^2}(h^2 + hk + k^2)}}. \tag{3.21c}$$

3.6.3 Brillouin Zone

The Wigner–Seitz cell in reciprocal space is called the (first) Brillouin zone. In Fig. 3.34, the Brillouin zones for the most important lattices are shown. Certain points in the Brillouin zone are labeled with dedicated letters. The Γ point always denotes $\mathbf{k} = 0$ (zone center). Certain paths in the Brillouin zone are labeled with dedicated Greek symbols.

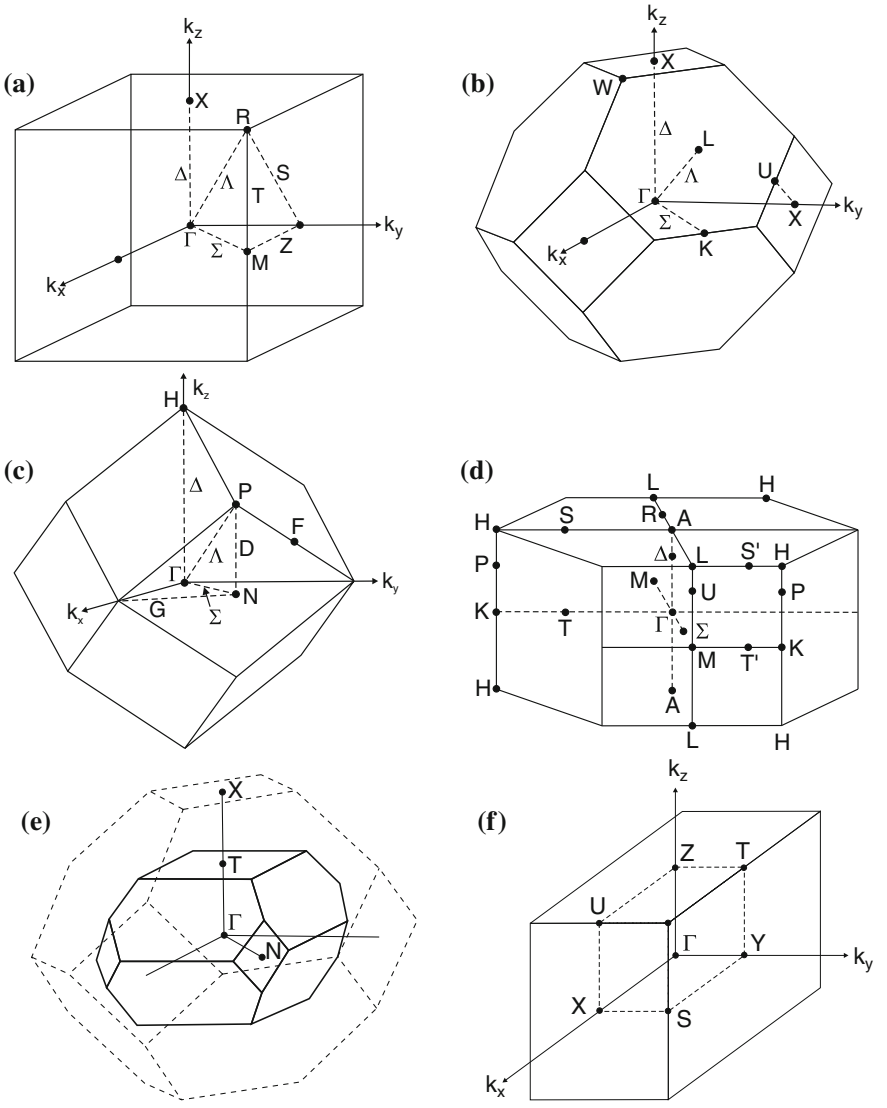


Fig. 3.34 Brillouin zones and special **k** points for the (**a**) primitive cubic (pc), (**b**) fcc, (**c**) bcc, and (**d**) hcp lattice. (**e**) Brillouin zone for chalcopyrite structure with fcc Brillouin zone shown as *dashed outline*. (**f**) Brillouin zone for orthorhombic lattice with one quadrant shown with *dashed lines*.

In the Brillouin zone of the fcc lattice (Si, Ge, GaAs, …) the X point denotes the point at the zone boundary in ⟨001⟩-directions (at distance $2\,\pi/a$ from Γ), K for ⟨110⟩-directions (at distance $3\,\pi/\sqrt{2}\,a$ from Γ) and L for the ⟨111⟩-directions (at distance $\sqrt{3}\,\pi/a$ from Γ) (see Table 3.6). The straight paths from Γ to X, K, and L are denoted as Δ, Σ, and Λ, respectively. High symmetry points and directions of the Brillouin zone of the hcp lattice are given in Table 3.7.

3.7 Alloys

When different semiconductors are mixed various cases can occur:

- The semiconductors are not miscible and have a so-called miscibility gap. They will tend to form clusters that build up the crystal. The formation of defects is probable.
- They form an ordered (periodic) structure that is called a superlattice.
- They form a random alloy.

3.7.1 Random Alloys

Alloys for which the probability to find an atom at a given lattice site is given by the fraction of these atoms (i.e. the stoichiometry), independent of the surrounding, are called *random* alloys. Deviations from the random population of sites is termed *clustering*.

For a Ge_xSi_{1-x} alloy this means that any given atom site has the probability x to have a Ge atom and $1-x$ to have a Si atom. The probability p_n that a Si atom has n next-neighbor Ge atoms is

$$p_n = \binom{4}{n} x^n\,(1-x)^{4-n}, \tag{3.22}$$

and is depicted in Fig. 3.35 as a function of the alloy composition. The symmetry of the Si atom is listed in Table 3.8. If it is surrounded by four of the same atoms (either Ge or Si), the symmetry is T_d. If one atom is different from the other three next neighbors, the symmetry is reduced to C_{3v} since one bond is singled out. For two atoms each the symmetry is lowest (C_{2v}).

In an alloy from binary compound semiconductors such as $Al_xGa_{1-x}As$ the mixing of the Al and Ga metal atoms occurs only on the metal (fcc) sublattice. Each As atom is bonded to four metal atoms. The probability that it is surrounded by n Al atoms is given by (3.22). The local symmetry of the As atom is also given by Table 3.8. For $AlAs_xP_{1-x}$ the mixing occurs on the nonmetal (anion) sublattice. If the alloy contains three atom species it is called a *ternary* alloy. In Fig. 3.36 the $(1\bar{1}0)$ surface

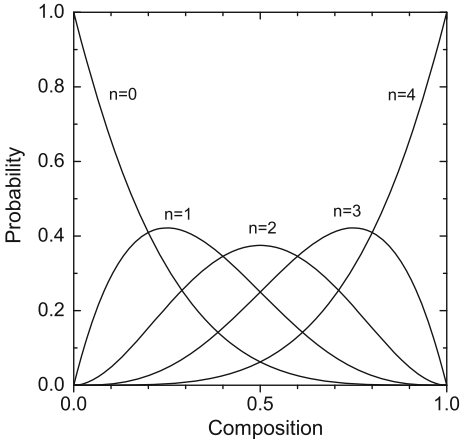


Fig. 3.35 Probability that a Si atom has n next-neighbor Ge atoms in a random Ge_xSi_{1-x} alloy

Table 3.8 Probability p_n (3.22) and symmetry of an A atom being surrounded by n B atoms in a tetrahedrally configured B_xA_{1-x} random alloy

n	p_n	Symmetry
0	x^4	T_d
1	$4\,x^3\,(1-x)$	C_{3v}
2	$6\,x^2\,(1-x)^2$	C_{2v}
3	$4\,x\,(1-x)^3$	C_{3v}
4	$(1-x)^4$	T_d

(UHV cleave) of an $In_{0.05}Ga_{0.95}As$ alloy is shown. Indium atoms in the first layer show up as brighter round dots [214]. Along the [001]-direction the positions are uncorrelated, along [110] an anti-correlation is found, corresponding to an effective repulsive pair interaction energy of 0.1 eV for the nearest neighbor In–In pairs along the [110]-direction due to strain effects [215].

If the binary end components have different crystal structure, the alloy shows a transition (or compositional transition range) from one structure to the other at a particular concentration. An example is the alloy between wurtzite ZnO and rocksalt MgO. $Mg_xZn_{1-x}O$ alloy thin films exhibits wurtzite structure up to about $x = 0.5$ and rocksalt structure for $x > 0.6$ [216] (cmp. Fig. 3.39).

If the alloy contains four atom species it is called *quaternary*. A quaternary zincblende alloy can have the mixing of three atom species on one sublattice, such as $Al_xGa_yIn_{1-x-y}As$ or $GaAs_xP_ySb_{1-x-y}$ or the mixing of two atom species on both of the two sublattices, such as $In_xGa_{1-x}As_yN_{1-y}$.

The random placement of different atoms on the (sub)lattice in an alloy represents a perturbation of the ideal lattice and causes additional scattering (alloy scattering). In the context of cluster formation, the probability of an atom having a direct neighbor of the same kind on its sublattice is important. Given a $A_xB_{1-x}C$ alloy, the probability p_S to find a single A atom surrounded by B atoms is given by (3.23a). The probability

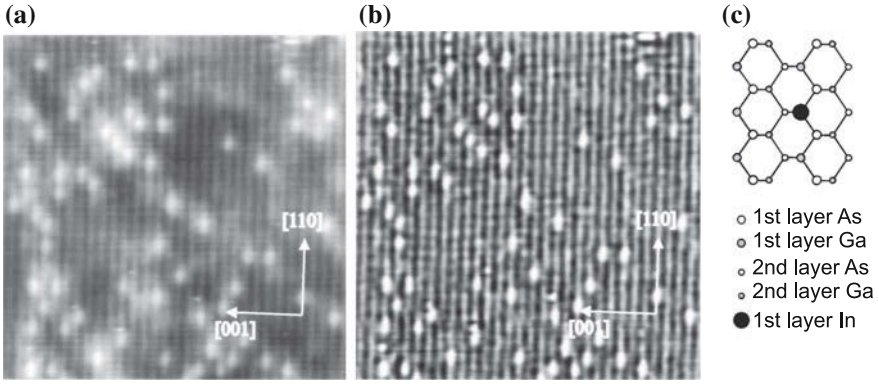


Fig. 3.36 (**a**) STM empty-state image ($17.5 \times 17.5\,\mathrm{nm}^2$) of the $(1\bar{1}0)$ surface of an MBE-grown $In_{0.05}Ga_{0.95}As$ alloy on GaAs, (**b**) curvature-enhanced image. (**c**) Schematic atomic arrangement of the first and second atomic layer. Adapted from [215]

p_{D^1} to find a cluster of two neighbored A atoms surrounded by B atoms is given by (3.23b).

$$p_S = (1-x)^{12} \tag{3.23a}$$

$$p_{D^1} = 12\,x\,(1-x)^{18}. \tag{3.23b}$$

These formulas are valid for fcc and hcp lattices. For larger clusters [217, 218], probabilities in fcc and hcp structures differ.

3.7.2 Phase Diagram

The mixture A_xB_{1-x} with average composition x between two materials A and B can result in a single phase (alloy), a two-phase system (phase separation) or a metastable system. The molar free enthalpy ΔG of the mixed system is approximated by

$$\Delta G = \Omega\,x\,(1-x) + kT\,[x\,\ln(x) + (1-x)\,\ln(1-x)]\,. \tag{3.24}$$

The first term on the right-hand side of (3.24) is the (regular solution) enthalpy of mixing with the interaction parameter Ω, which can depend on x. The second term is the ideal configurational entropy based on a random distribution of the atoms. The function is shown for various ratios of kT/Ω in Fig. 3.37a. In an equilibrium phase diagram (see Fig. 3.37b) the system is above the binodal curve in one phase (miscible). On the binodal line $T_b(x)$ in the (x, T) diagram the A- and B-rich disordered phases have equal chemical potentials, i.e. $\partial G/\partial x = 0$. For Ω independent of x the temperature T_b is given by (3.25a). A critical point is at the maximum temperature

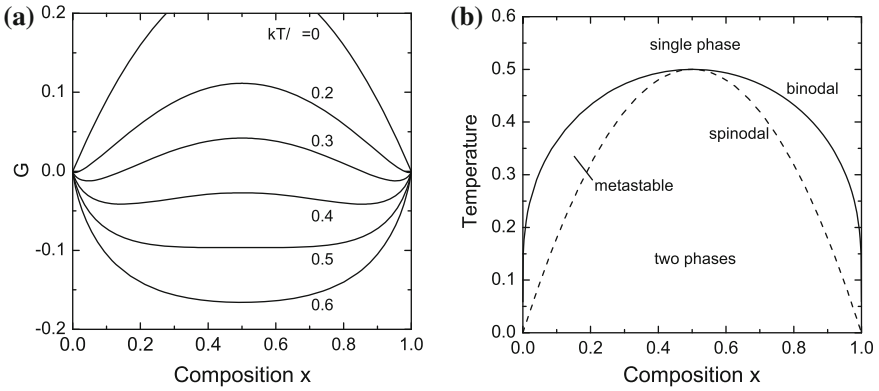


Fig. 3.37 (**a**) Free enthalpy ΔG of mixed binary system (3.24) in units of Ω for $\Omega = \text{const.}$ and various values of kT/Ω as labeled. (**b**) Schematic phase diagram for binary mixture. The temperature is given in units of Ω/k. The *solid* (*dashed*) denotes the binodal (spinodal) line

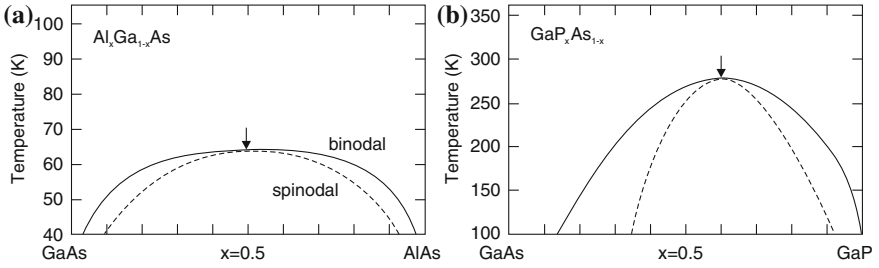


Fig. 3.38 Calculated phase diagrams for (**a**) $Al_xGa_{1-x}As$ and (**b**) GaP_xAs_{1-x}. The binodal (spinodal) curve is shown as *solid* (*dashed*) line. Adapted from [219]

T_{mg} and concentration x_{mg} of the miscibility gap. For Ω independent of x it is given by $T_{\text{mg}} = \Omega/2$ and $x_{\text{mg}} = 1/2$. In the region below the spinodal boundary, the system is immiscible and phases immediately segregate (by spinodal decomposition). On the spinodal line $T_{\text{sp}}(x)$ the condition $\partial^2 G/\partial x^2 = 0$ is fulfilled. For Ω independent of x the temperature T_{sp} is given by (3.25b). The region between the binodal and spinodal curves is the metastable region, i.e. the system is stable to small fluctuations of concentration or temperature but not for larger ones.

$$kT_{\text{b}}(x) = \Omega \frac{2x-1}{\ln(x) - \ln(1-x)} \tag{3.25a}$$

$$kT_{\text{sp}}(x) = 2\Omega\, x\,(1-x). \tag{3.25b}$$

In Fig. 3.38 calculated diagrams for GaAs-AlAs and GaAs-GaP [219] are shown. The arrows denote the critical point. These parameters and the interaction parameters for a number of ternary alloys are given in Table 3.9. For example, for $Al_xGa_{1-x}As$

Table 3.9 Calculated interaction parameter $\Omega(x)$ (at $T = 800\,\mathrm{K}$, 1 kcal/mol= 43.39 meV), miscibility-gap temperature T_{mg} and concentration x_{mg} for various ternary alloys

Alloy	T_{mg} (K)	x_{mg}	$\Omega(0)$ (kcal/mol)	$\Omega(0.5)$ (kcal/mol)	$\Omega(1)$ (kcal/mol)
$Al_xGa_{1-x}As$	64	0.51	0.30	0.30	0.30
GaP_xAs_{1-x}	277	0.603	0.53	0.86	1.07
$Ga_xIn_{1-x}P$	961	0.676	2.92	3.07	4.60
$GaSb_xAs_{1-x}$	1080	0.405	4.51	3.96	3.78
$Hg_xCd_{1-x}Te$	84	0.40	0.45	0.80	0.31
$Zn_xHg_{1-x}Te$	455	0.56	2.13	1.88	2.15
$Zn_xCd_{1-x}As$	605	0.623	2.24	2.29	2.87
$In_xGa_{1-x}N$	1505	0.50	6.32	5.98	5.63

Data for (In,Ga)N from [220], other data from [219]

complete miscibility is possible for typical growth temperatures (>700 K), but for $In_xGa_{1-x}N$ the In solubility at a typical growth temperature of 1100 K is only 6 % [220].

The alloy system (Al,Ga,In)(As,P,Sb) always crystallizes in the zincblende structure and (Al,Ga,In)N always in the wurtzite structure. If the binary end components of a ternary alloy have a *different* crystal structure things become more complicated and the crystal phase has to be determined experimentally (and modelled) for each composition. As an example the energy of the wurtzite, hexagonal and rocksalt structure of $Mg_xZn_{1-x}O$ has been calculated [221] as depicted in Fig. 3.39 (cmp. Fig. 2.4 for silicon). The transition between wurtzite and rocksalt structure is predicted for $x = 0.33$.

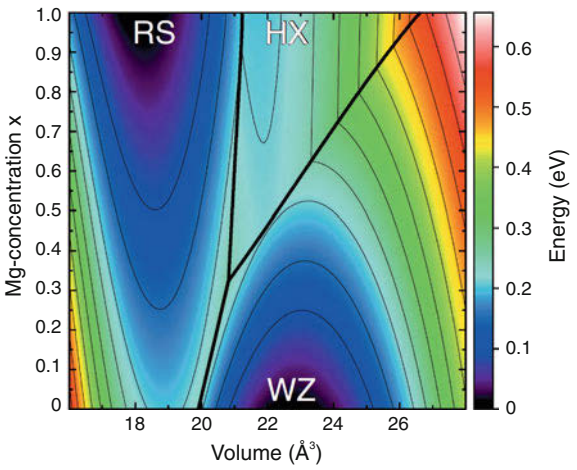


Fig. 3.39 Calculated energy versus volume of the formula unit for $Mg_xZn_{1-x}O$ in the wurtzite (WZ), hexagonal (HX) and rocksalt phase (RS). The separations between the three phase are denoted by *straight bold lines*. Adapted from [221]

3.7.3 Virtual Crystal Approximation

In the *virtual crystal approximation* (VCA) the disordered alloy AB_xC_{1-x} is replaced by an ordered binary compound AD with D being a 'pseudoatom' with properties that are configuration averaged over the properties of the B and C atoms, e.g., their masses or charges. Such an average is weighted with the ternary composition, e.g., the mass is $M_D = xM_B + (1-x)M_C$. For example, the A–D force constant would be taken as the weighted average over the A–B and A–C force constants.

3.7.4 Lattice Parameter

In the VCA for an alloy a new sort of effective atom is assumed that has an averaged bond length that depends linearly on the composition. Typically, Vegard's law (3.26), which predicts that the lattice constant of a ternary alloy $A_xB_{1-x}C$ depends linearly on the lattice constants of the binary alloys AC and BC, is indeed fulfilled

$$a_0(A_xB_{1-x}C) = a_0(BC) + x\,[a_0(AC) - a_0(BC)]. \tag{3.26}$$

In reality, the bond length of the AC and BC bonds changes rather little (Fig. 3.40a) such that the atoms in the alloy suffer a displacement from their average position and the lattice is deformed on the nanoscopic scale. In a lattice of the type $In_xGa_{1-x}As$ the anions suffer the largest displacement since their position adjusts to the local cation environment. For $In_xGa_{1-x}As$ a bimodal distribution, according to the As–Ga–As and As–In–As configurations, is observed (Fig. 3.40b). The cation–cation second-neighbor distances are fairly close to the VCA.

While the average lattice parameter in alloys changes linearly with composition, the cell-internal parameter u (for wurtzite structures, see Sect. 3.4.5) exhibits a nonlinear behavior as shown in Fig. 3.41. Therefore physical properties connected to u, such as the spontaneous polarization, will exhibit a bowing.

3.7.5 Ordering

Some alloys have the tendency for the formation of a superstructure [225]. Growth kinetics at the surface can lead to specific adatom incorporation leading to ordering. For example, in $In_{0.5}Ga_{0.5}P$ the In and Ga atoms can be ordered in subsequent (111) planes (CuPt structure) instead of being randomly mixed (Fig. 3.42). This impacts fundamental properties such as the phonon spectrum or the band gap. CuPt ordering on (111) and $(\bar{1}\bar{1}1)$ planes is called $CuPt_A$, on $(\bar{1}11)$ and $(1\bar{1}1)$ planes $CuPt_B$ ordering. In Fig. 3.43, a TEM image of a $Cd_{0.68}Zn_{0.32}Te$ epilayer is shown with simultaneous

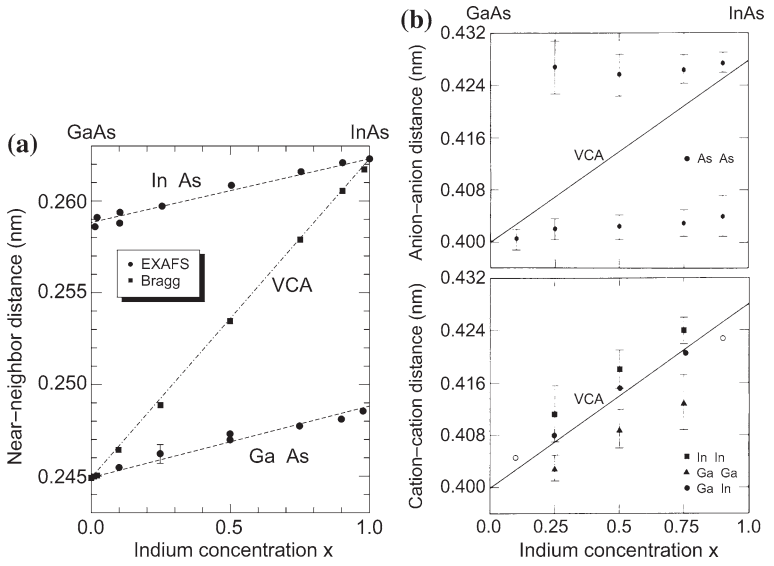


Fig. 3.40 (**a**) Near-neighbor distance ($\sqrt{3}\,a_0/4$) of $In_xGa_{1-x}As$ as measured by standard X-ray diffraction (Bragg reflection, *solid squares*) and VCA approximation (*dash-dotted line*). Near-neighbor Ga–As and In–As distances as determined by EXAFS (extended X-ray absorption fine structure, *solid squares*). *Dashed lines* are guides to the eye. Data from [222]. (**b**) Second-neighbor distances for $In_xGa_{1-x}As$ as determined from EXAFS, *top*: anion–anion distance (for As–As), *bottom*: cation–cation distance (for In–In, Ga–Ga, and Ga–In). *Solid lines* in both plots are the VCA ($a_0/\sqrt{2}$). Data from [223]

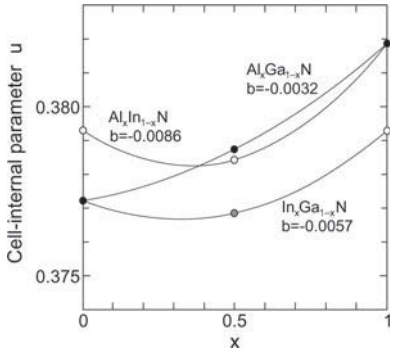


Fig. 3.41 Theoretical values ($T = 0\,K$) for the cell-internal parameter u as a function of the composition for group-III nitride alloys. The *solid lines* are quadratic curves (bowing parameter b is shown) through the points for $x = 0$, 0.5, and 1.0. Data from [224]

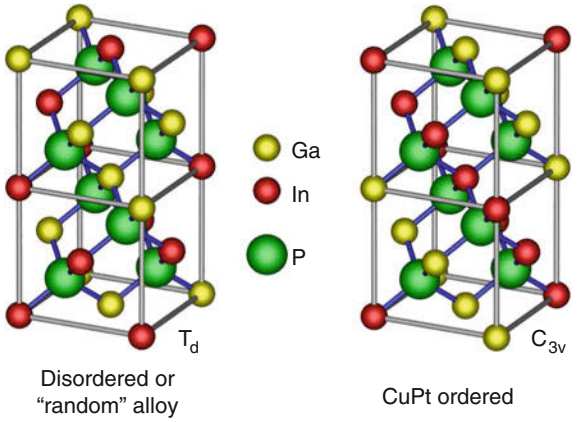


Fig. 3.42 CuPt-ordered ternary alloy $In_{0.5}Ga_{0.5}P$; the lattice symmetry is reduced from T_d to C_{3v}

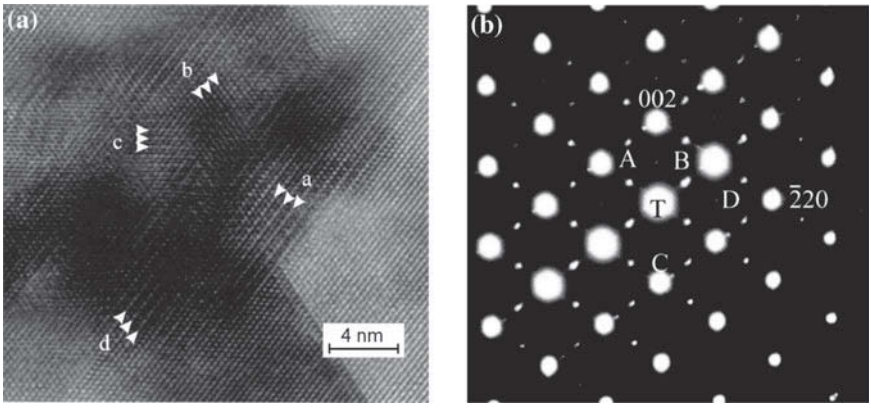


Fig. 3.43 (**a**) Cross-sectional transmission electron microscope image along the [110] zone axis of a $Cd_{0.68}Zn_{0.32}Te$ epilayer on GaAs showing ordered domains having a doublet periodicity on the {111} and {001} lattice planes. Two different {111} variants are labeled 'a' and 'b'. The doublet periodicity in the [001] is seen in the 'c' region. (**b**) Selected-area diffraction pattern along the [110] zone. Strong peaks are fundamental peaks of the zincblende crystal, weak peaks are due to CuPt ordering, labeled A and B, and CuAu-I ordering, labeled C and D. The latter are the weakest due to a small volume fraction of CuAu-ordered domains. Adapted from [226]

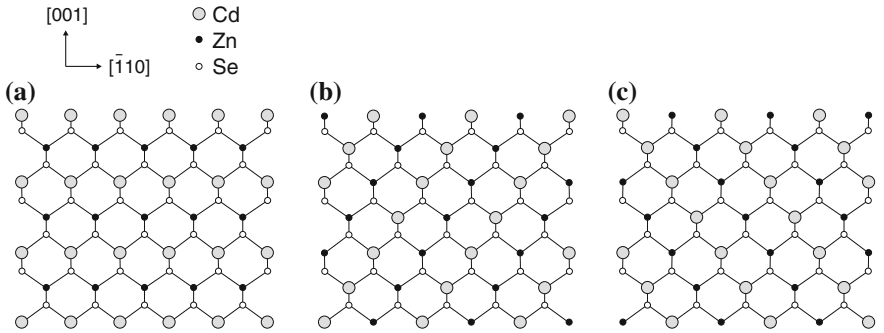


Fig. 3.44 Schematic diagrams of zincblende $Cd_xZn_{1-x}Te$ along [110] with (**a**) CuAu-I type ordering and (**b**,**c**) two types of the $CuPt_B$ type ordering. Doublet periodicity is along (**a**) [001] and [$\bar{1}$10], (**b**) [1$\bar{1}$1] and (**c**) [$\bar{1}$11]. Adapted from [226]

ordering in the CuPt structure (doublet periodicity along $\left[1\bar{1}1\right]$ and $\left[\bar{1}11\right]$) and in the CuAu-I structure[6] (doublet periodicity along [001] and [$\bar{1}$10]) (Fig. 3.44).

[6]The CuAu-I structure has tetragonal symmetry. There exists also the CuAu-II structure that is orthorhombic.

Chapter 4
Defects

Abstract No crystal is perfect. Various point defects and their thermodynamics, diffusion and distribution effects are discussed. Also dislocations and extended defects such as cracks, stacking faults, grain boundaries and antiphase domains are covered.

4.1 Introduction

In an ideal lattice each atom is at its designated position. Deviations from the ideal structure are called defects. In the following, we will briefly discuss the most common defects. The electrical activity of defects will be discussed in Sects. 7.5 and 7.7. For the creation (formation) of a defect a certain free enthalpy $G_{\mathrm{D}}^{\mathrm{f}}$ is necessary. At thermodynamical equilibrium a (point) defect density $\propto \exp\left(-G_{\mathrm{D}}^{\mathrm{f}}/kT\right)$ will always be present (cf. Sect. 4.2.2).

Point defects (Sect. 4.2) are deviations from the ideal structure involving essentially only one lattice point. The formation energy for line defects (Sect. 4.3) or area defects (Sect. 4.4) scales with $N^{1/3}$ and $N^{2/3}$, respectively, N being the number of atoms in the crystal. Therefore, these defects are not expected in thermodynamic equilibrium. However, the path into thermodynamical equilibrium might be so slow that these defects are metastable and must be considered quasi-frozen. There may also exist metastable point defects. By annealing the crystal, the thermodynamic equilibrium concentration might be re-established. The unavoidable two-dimensional defect of the bulk structure is the surface, discussed in Chap. 11.

4.2 Point Defects

4.2.1 Point Defect Types

The simplest point defect is a vacancy V, a missing atom at a given atomic position. If an atom is at a position that does not belong to the crystal structure an interstitial I

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_4

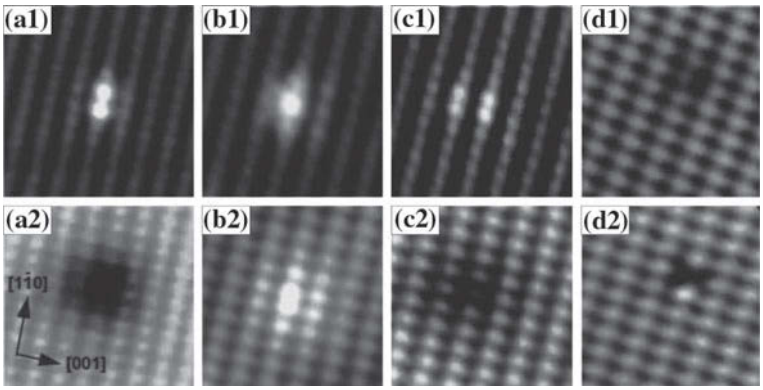


Fig. 4.1 Images of occupied (*upper frames*) and empty (*lower frames*) density of states of typical defects on Si-doped GaAs (110) surfaces. (**a1**, **a2**) show a Ga vacancy, (**b1**, **b2**) a Si_{Ga} donor, (**c1**, **c2**) a Si_{As} acceptor and (**d1**, **d2**) a Si_{Ga}–V_{Ga} complex. Adapted from [227]

(or Frenkel defect) is formed. Depending on the position of the interstitial different types are distinguished. An interstitial atom that has the same chemical species as the crystal is called 'self-interstitial'.

If an atom site is populated with an atom of different order number Z, an impurity is present. An impurity can also sit on interstitial position. If the number of valence electrons is the same as for the original (or correct) atom, then it is an isovalent impurity and quasi fits into the bonding scheme. If the valence is different, the impurity adds extra (negative or positive) charge to the crystal bonds, which is compensated by the extra, locally fixed charge in the nucleus. This mechanism will be discussed in detail in the context of doping (Chap. 7). If in an AB compound an A atom sits on the B site, the defect is called an antisite defect A_B.

A Ga vacancy, a silicon impurity atom on Ga- and As-site and a Si_{Ga}-vacancy complex at the (110) surface of Si doped GaAs are shown in Fig. 4.1 as observed with STM [227, 228]. Also antisite defects in GaAs can be observed with STM [229, 230].

A point defect is typically accompanied by a relaxation of the surrounding host atoms. As an example, we discuss the vacancy in Si (Fig. 4.2a). The missing atom leads to a lattice relaxation with the next neighbors moving some way into the void (Fig. 4.2b). The bond lengths of the next and second-next neighbor Si atoms around the neutral vacancy are shown in Fig. 4.2c. The lattice relaxation depends on the charge state of the point defect (Jahn–Teller effect) which is discussed in more detail in Sect. 7.7. In Fig. 4.2d the situation for the positively charged vacancy with one electron missing is shown. One of the two bonds is weakened since it lacks an electron. The distortion is therefore different from that for V^0. Also the (self-)interstitial is accompanied with a lattice relaxation as shown in Fig. 4.3 for a silicon interstitial at tetrahedral place. Self-interstitials in silicon and germanium are reviewed and compared in [231] for their various charge states.

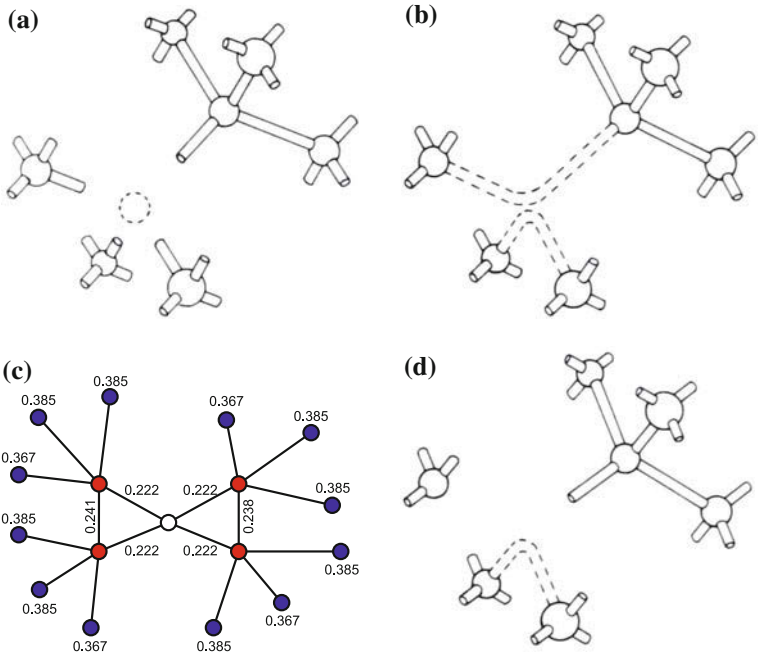


Fig. 4.2 (**a**) Schematic diamond lattice with vacancy, i.e. a missing Si atom without relaxation. (**b**) Si with neutral vacancy (V^0), lattice relaxation and formation of two new bonds. (**c**) Schematic diagram showing the (inward) relaxation of the neighbors around the neutral Si vacancy defect site (*empty circle*) calculated by an ab initio method. The distances of the outer shell of atoms (*red circles*) from the vacant site is labeled (in nm). The bond lengths of the two new bonds and the second-neighbor (*blue circles*) distance are also indicated. The bond length in bulk Si is 0.2352 nm, the second-neighbor distance 0.3840 nm. Adapted from [232]. (**d**) Si unit cell with positively charged vacancy (V^+). Parts (**a**, **b**, **d**) reprinted with permission from [233]

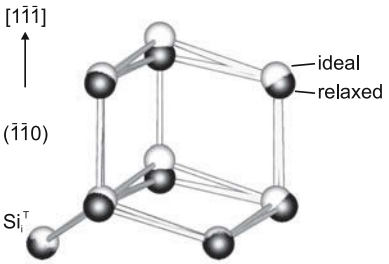


Fig. 4.3 Silicon tetrahedral interstitial Si_i^T and its next atoms in ideal (*white spheres*) and relaxed (*black spheres*) position. Adapted from [155]

4.2.2 Thermodynamics

For a given temperature, the free enthalpy G of a crystal (a closed system with regard to particle exchange)

$$G = H - TS \tag{4.1}$$

is minimum. H is the enthalpy and S the entropy. The enthalpy $H = E + pV$ is the thermodynamic potential for a system whose only external parameter is the volume V. It is used when the independent variables of the system are the entropy S and pressure p. The free enthalpy is used when the independent parameters are T and p. G_0 (H_0) is the free energy (enthalpy) of the perfect crystal. H^f is the formation enthalpy of an isolated defect. This could be, e.g., the enthalpy of a vacancy, created by bringing an atom from the (later) vacancy site to the surface, or an interstitial, created by bringing an atom from the surface to the interstitial site. In the limit that the n defects do not interact with each other, i.e. their concentration is sufficiently small, they can be considered independent and the enthalpy is given by

$$H = H_0 + n\,H^\mathrm{f}. \tag{4.2}$$

The increase of entropy due to increased disorder is split into the configurational disorder over the possible sites, denoted as S^d, and the formation entropy S_f due to localized vibrational modes. The total change ΔG of the free energy is

$$\Delta G = G - G_0 = n\,(H^\mathrm{f} - T\,S^\mathrm{f}) - T\,S^\mathrm{d} = n\,G^\mathrm{f} - T\,S^\mathrm{d}, \tag{4.3}$$

where $G^\mathrm{f} = H^\mathrm{f} - T\,S^\mathrm{f}$ denotes the free enthalpy of formation of a single isolated defect. In Table 4.1 experimental values for the formation entropy and enthalpy are given for several defects. Surprisingly, despite their fundamental importance in semiconductor defect physics, these numbers are not very well known and disputed in the literature.

The defect concentration is obtained by minimizing ΔG, i.e.

$$\frac{\partial \Delta G}{\partial n} = G^\mathrm{f} - T\,\frac{\partial S^\mathrm{d}}{\partial n} = 0. \tag{4.4}$$

The entropy S^d due to disorder is given as

$$S^\mathrm{d} = k_\mathrm{B} \ln W, \tag{4.5}$$

where W is the complexion number, usually the number of distinguishable ways to distribute n defects on N lattice sites

Table 4.1 Formation enthalpy H^f and entropy S^f of the interstitial (I) and vacancy (V) in Si and the Ga vacancy in GaAs

Material	Defect	H^f (eV)	S^f (k_B)
Si	I	3.2	4.1
Si	V	2.8	~ 1
GaAs	V_Ga	3.2	9.6

Data for Si from [234, 235], for GaAs from [236]

$$W = \binom{N}{n} = \frac{N!}{n!\,(N-n)!}. \tag{4.6}$$

With Stirling's formula $\ln x! \approx x(\ln x - 1)$ for large x we obtain

$$\frac{\partial S^{\mathrm{d}}}{\partial n} = k_{\mathrm{B}} \left[\frac{N}{n} \ln\left(\frac{N}{N-n}\right) + \ln\left(\frac{N-n}{n}\right) \right]. \tag{4.7}$$

If $n \ll N$, $\partial N/\partial n = 0$ and the right side of (4.7) reduces to $k_{\mathrm{B}} \ln(N/n)$. The condition (4.4) reads $G^{\mathrm{f}} + k_{\mathrm{B}} T \ln(n/N)$, or

$$\frac{n}{N} = \exp\left(-\frac{G^{\mathrm{f}}}{kT}\right). \tag{4.8}$$

In the case of several different defects i with a degeneracy Z_i, e.g. a spin degree of freedom or several equivalent configurations, (4.8) can be generalized to

$$\frac{n_i}{Z_i\, N} = \exp\left(-\frac{G_i^{\mathrm{f}}}{kT}\right). \tag{4.9}$$

In [237] the equilibrium concentration of interstitials C_I^{eq} in silicon has been given as

$$C_I^{\mathrm{eq}} = \left(1.0 \times 10^{27}\,\mathrm{cm}^{-3}\right) \exp\left(-\frac{3.8\,\mathrm{eV}}{kT}\right), \tag{4.10}$$

about $10^{14}\,\mathrm{cm}^{-3}$ at 1200 °C. The vacancy concentration has been investigated in [238]. Around a temperature of 1200 °C it is in the 10^{14}–$10^{15}\,\mathrm{cm}^{-3}$ range. Due to the reaction

$$0 \rightleftharpoons I + V, \tag{4.11}$$

a mass action law holds for the concentrations of interstitials and vacancies

$$C_I\, C_V = C_I^{\mathrm{eq}}\, C_V^{\mathrm{eq}}. \tag{4.12}$$

4.2.3 Diffusion

The diffusion of point defects is technologically very important, in particular for silicon as host material. Typically a dopant profile should be stable under following technological processing steps and also during device performance. Also defect annihilation is crucial after implantation processes. Diffusion of an interstitial I and a

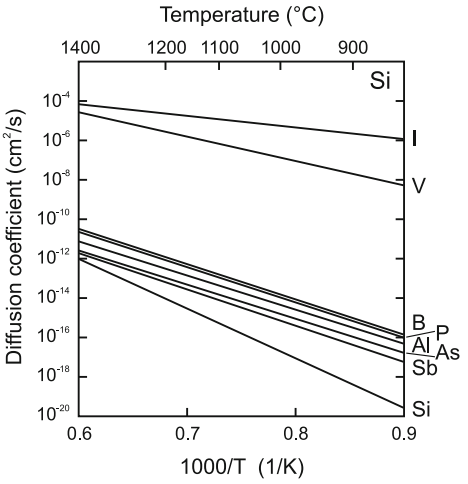


Fig. 4.4 The temperature dependent diffusion coefficient of Si interstitials I, vacancies V and various impurities in silicon. Also the self-diffusion coefficient, labeled with 'Si' is shown. Based on data from [239]

vacancy V to the same site is prerequisite for recombination of defects (so called *bulk process*) according to the scheme $I + V \rightarrow 0$. We note that the process $0 \rightarrow I + V$ is called *Frenkel pair process*.[1] Also the self-diffusion of silicon has been studied, e.g. using radioactively marked isotopes [237]. The diffusion of point defects including dopants in silicon has been reviewed in [239, 240]. Usually Fick's law is applied, stating how the flux J depends on the concentration gradient, for an interstitial it reads:

$$J_I = -D_I \nabla C_I, \tag{4.13}$$

D_I being the interstitial diffusion coefficient. For interstitials in Si it was found [237] that

$$D_I = 0.2 \exp\left(-\frac{1.2\,\text{eV}}{kT}\right) \text{cm}^2/\text{s}. \tag{4.14}$$

The diffusion of neutral vacancies occurs with [241]

$$D_V = 0.0012 \exp\left(-\frac{0.45\,\text{eV}}{kT}\right) \text{cm}^2/\text{s}. \tag{4.15}$$

The temperature dependent diffusion coefficients of point defects and dopants in silicon are shown in Fig. 4.4.

The self-diffusion coefficient of silicon has been determined from the annealing of isotope superlattices (Sect. 12.5) of sequence $^{28}\text{Si}_n/^{30}\text{Si}_n$, $n = 20$ to be [242]

[1] At higher temperatures a silicon atom can occasionally acquire sufficient energy from lattice vibrations to leave its lattice site and thus an interstitial and a vacancy are generated.

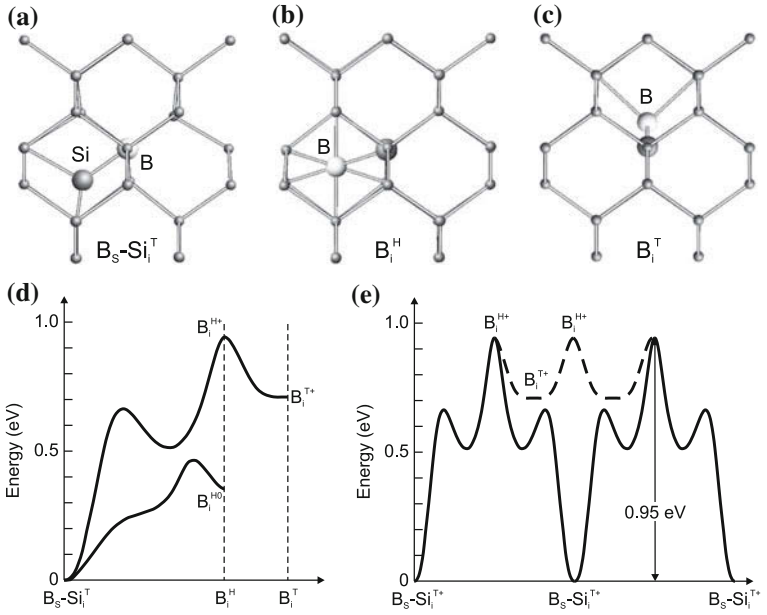


Fig. 4.5 Configurations of boron in Si: (**a**) Substitutional boron and Si self-interstitial at 'T' site (B_S–Si_i^T). Interstitial boron at (**b**) 'H' (B_i^H) and (**c**) 'T' site (B_i^T), each with the Si atoms on the Si lattice sites. The *large bright ball* represents the boron atom, *large* and *small dark balls* represent Si atoms. (**d**) Lowest energy barrier diffusion paths for positively charged and neutral B–Si states, total energy versus configuration. (**e**) Two diffusion pathways for positively charged B–Si, kick-out (*dashed line*) and pair diffusion (*solid line*); the activation energy is labeled. Adapted from [245]

$$D_{\mathrm{Si}}^{\mathrm{SD}} = \left[2175.4\ \exp\left(-\frac{4.95\,\mathrm{eV}}{kT}\right) + 0.0023\ \exp\left(-\frac{3.6\,\mathrm{eV}}{kT}\right)\right]\ \mathrm{cm}^2/\mathrm{s}, \quad (4.16)$$

the first (second) term being due to interstitial (vacancy) mechanism, dominant for temperatures larger (smaller) than 900 °C. The enthalpy in the exponent, e.g. $H_V = 3.6^{+0.3}_{-0.1}$ eV [242], consists of the formation and migration enthalpies,

$$H_V = H_V^{\mathrm{f}} + H_V^{\mathrm{m}}. \quad (4.17)$$

Using the experimental value $H_V^{\mathrm{f}} = 2.8 \pm 0.3$ eV [235] from Table 4.1, for the migration enthalpy a value around $H_V^{\mathrm{m}} \approx 0.8$ eV is obtained.

As an example for a dopant diffusion process that has been understood microscopically, we discuss here boron in silicon. In Fig. 4.5a the lowest-energy configuration of a boron-related defect in silicon is depicted, B_s–Si_i^T, i.e. boron on a substitutional site and a self-interstitial Si on the 'T' place with highest symmetry[2] (see Fig. 3.18).

[2]The positive charge state is stable, the neutral charge state is metastable since the defect is a negative-U center (see Sect. 7.7.5).

Due to its importance as an acceptor in Si, the configuration and diffusion of B in Si has found great interest [243–245]. The diffusion depends on the charge state of boron. The diffusion of positively charged boron has been suggested [245] to occur via the following route: The boron leaves its substitutional site and goes to the hexagonal site ('H') (Fig. 4.5b) with an activation energy of about 1 eV (Fig. 4.5d). It can then relax (∼0.1 eV) without barrier to the tetrahedral 'T' position (Fig. 4.5c). The direct migration B_s–$Si_i^{T+} \rightarrow B_i^{T+}$ has a higher activation energy of 1.12 eV and is thus unlikely. The boron atom can then diffuse through the crystal by going from 'H' to 'T' to 'H' and so on (Fig. 4.5e). However, long-range diffusion seems to be not possible in this way because the kick-in mechanism will bring back the boron to its stable configuration. The pair diffusion mechanism for neutral boron B_s–$Si_i^T \rightarrow B_i^H \rightarrow B_s$–$Si_i^T$ via the hexagonal site has an activation energy of about 0.5 eV (Fig. 4.5d) while the path via B_i^T has a larger 0.9 eV barrier. The concentration dependence of the diffusion mechanism has been discussed in [246].

Similarly, indium diffusion in silicon has been investigated suggesting a minimum energy In_s–$Si_i^T \rightarrow In_i^T \rightarrow In_s$–$Si_i^T$ diffusion pathway via the tetrahedral site with 0.8 eV activation energy [247]. Microscopic modeling has been reported also for diffusion of phosphorus [248].

4.2.4 Dopant Distribution

The introduction of impurities into a semiconductor (or other materials such as glasses) is termed *doping*. The unavoidable incorporation of impurities in the nominally pure (nominally undoped) material is called *unintentional* doping and leads to a residual or background impurity concentration. Several methods are used for doping and the creation of particular doping profiles (in depth or lateral). All doping profiles underly subsequent diffusion of dopants (Sect. 4.2.3).

Various methods of doping are used. A straightforward method of doping is the incorporation during crystal growth or epitaxy. For semiconductor wafers a homogeneous doping concentration is targeted, both laterally and along the rod from which the wafer is cut (Sect. 12.2.2). When a crystal is grown from melt, containing a concentration c_0 of the impurity, the concentration in the solid is given by ('normal freezing' case [249–251])[3]

$$c(x) = c_0\, k\, (1-x)^{k-1}, \tag{4.18}$$

where $c(x)$ is the impurity concentration in the crystal at the freezing interface, x is the frozen melt fraction (ratio of solid mass to total mass, $0 \le x \le 1$). k is the *distribution*

[3] Mass preservation of the impurities can be written at any time $c_m(1-x) + \int_0^x c(x')\,dx' = c_0$, where c_m is the (remaining) concentration in the melt. At the beginning $c_m(0) = c_0$. At the interface $c(x) = k\, c_m(x)$. Putting this into the mass preservation, building $c'(x)$ and solving the resulting differential equation $c' = c(1-k)/(1-x)$ with $c(0) = k\, c_0$ leads to (4.18).

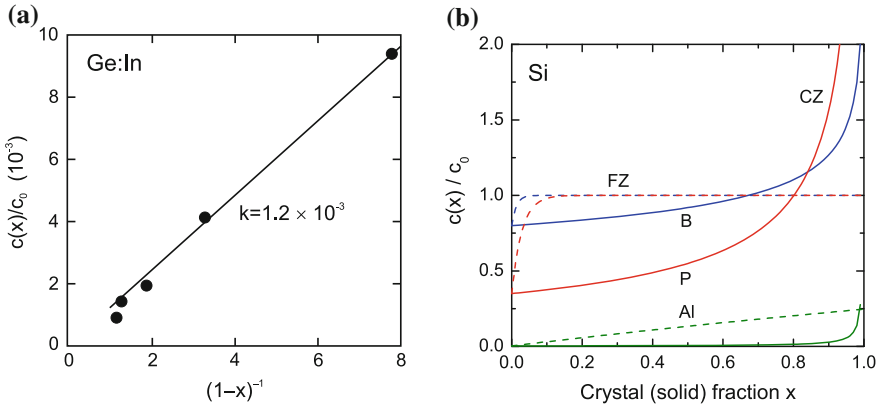


Fig. 4.6 (**a**) Relative concentration of indium along a CZ-grown germanium crystal. Absolute concentration is in the $10^{16}\,\mathrm{cm}^{-3}$ range. *Solid line* follows (4.19) with $k = 1.2 \times 10^{-3}$. *Symbols* are experimental data from [254]. (**b**) Impurity distribution (relative concentration $c(x)/c_0$) for CZ (4.18) (*solid lines*) and FZ (4.20) (*dashed lines*, $z = 0.01$) silicon crystals for B (*blue*), P (*red*), and Al (*green*). Distribution coefficients have been taken from Table 4.2. Note crossing of B and P lines and possibly associated change from p-type to n-type (cmp. Fig. 1.7)

coefficient (or segregation coefficient) which is the fraction of impurities that is built into the crystal at the liquid–solid interface. Since the melt volume reduces during the solidification, the impurity concentration rises over time. For small distribution coefficients (4.18) can be approximated to

$$c(x) \approx c_0 \frac{k}{1-x}\,, \tag{4.19}$$

An experimental example for Ge:In is shown in Fig. 4.6a.

In Table 4.2 the distribution coefficients for various impurities in Si, Ge and GaAs is given. The modification of distribution coefficients in SiGe alloys is discussed in [252]. Equilibrium values (k_{eq}) are obtained for 'slow' crystal growth. For finite growth rates, k becomes a function of the growth rate and is then called the *effective* distribution coefficient. For $k < 1$, $k_{\mathrm{eff}} > k_{\mathrm{eq}}$. k_{eff} approaches 1 for high growth rates, i.e. all impurities at the rapidly moving interface are incorporated.

Equation (4.18) applies to Czrochalski growth where the crystal is pulled out of the melt [257]. In float-zone (FZ) growth [257] a polycrystalline rod is transformed into a crystalline one while a RF-heated and liquid 'float' zone is moved through the crystal. In this case the impurity distribution is given by[4]

[4]When the float zone moves through the crystal, the change of mass of impurities $m_{\mathrm{m}} = c_{\mathrm{m}} z$ in the liquid is $m'_{\mathrm{m}} = c_0 - k c_{\mathrm{m}}$. The first term stems from the melting of the polycrystalline part, the second from the solidification of the crystal. Solving the resulting differential equation $c'_{\mathrm{m}} = (c_0 - k c_{\mathrm{m}})/z$ with $c_{\mathrm{m}}(0) = c_0$ and using $c(x) = k c_{\mathrm{m}}(x)$ yields (4.20).

Table 4.2 Equilibrium distribution coefficients (at melting point) of various impurities in silicon, germanium and GaAs

Impurity	Si	Ge	GaAs
C	0.07	>1.85	0.8
Si		5.5	0.1
Ge	0.33		0.03
N	7×10^{-4}		
O	≈1		0.3
B	0.8	12.2	
Al	2.8×10^{-3}	0.1	3
Ga	8×10^{-3}	0.087	
In	4×10^{-4}	1.2×10^{-3}	0.1
P	0.35	0.12	2
As	0.3	0.04	
Sb	0.023	3.3×10^{-3}	<0.02
S	10^{-5}	$> 5 \times 10^{-5}$	0.3
Fe	6.4×10^{-6}	3×10^{-5}	2×10^{-3}
Ni	$\approx 3 \times 10^{-5}$	2.3×10^{-6}	6×10^{-4}
Cu	8×10^{-4}	1.3×10^{-5}	2×10^{-3}
Ag	$\approx 1 \times 10^{-6}$	10^{-4}	0.1
Au	2.5×10^{-5}	1.5×10^{-5}	
Zn	2.5×10^{-5}	6×10^{-4}	0.1

Data for Si from [251, 253], for Ge from [146, 254–256] and for GaAs from [146]

$$c(x) = c_0 \left[1 - (1 - k) \exp\left(-\frac{k\,x}{z} \right) \right], \tag{4.20}$$

where x is the ratio of the crystal mass to the total mass, i.e. crystal, liquid and feed rod. z is the relative mass of the (liquid) float zone, i.e. the ratio of liquid mass to the total mass. The impurity distribution for CZ and FZ crystals is compared in Fig. 4.6b. Obviously the FZ process can create much more homogeneous profiles.[5]

Using epitaxy arbitrary doping profiles along the growth directions can be created by varying the impurity supply during growth. Impurities can be introduced through the surface of the material by diffusion from a solid or gas phase. In ion implantation [258] the impurity atoms are accelerated towards the semiconductor and deposited with a certain depth profile due to multiple scattering and energy loss events, depending on the acceleration voltage (increasing deposition depth with increasing voltage, Fig. 4.7a and ion mass (decreasing deposition depth with increasing mass, Fig. 4.7b. The depth profile is often investigated using secondary ion mass

[5]We note that during directed solidification of Si:(B,P) a pn-junction forms due to the different distribution coefficients of boron and phosphorus. This has been used in [79].

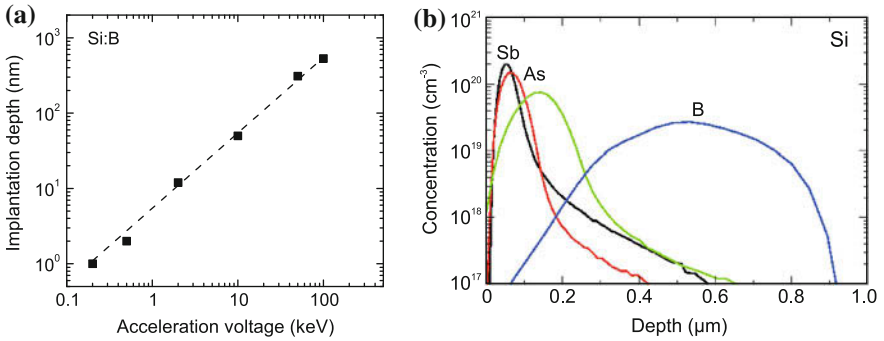


Fig. 4.7 (**a**) Depth of peak concentration of boron implanted in silicon for various acceleration voltages U. Data from various sources, for $U < 1$ keV from [264]. *Dashed line* is linear dependence. (**b**) Simulated depth profiles of impurity concentration for B, P, As, and Sb implanted into crystalline silicon with $U = 100$ keV and a dose of 10^{15} cm^{-2}. Adapted from [265]

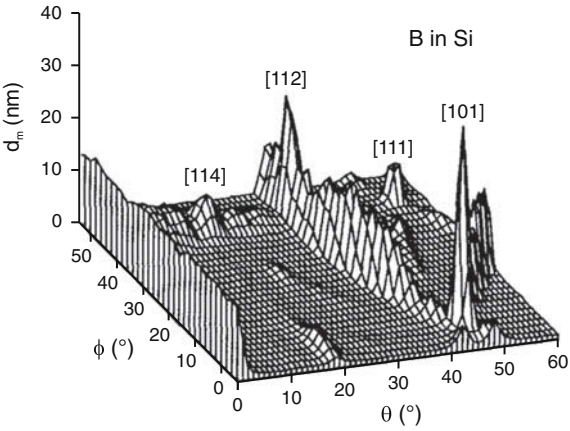


Fig. 4.8 Simulated mean path length as a function of implantation direction (azimuthal angle ϕ and polar angle θ) near [001] for 5 keV boron in silicon. The [001] channeling peak appears as a ridge at the left side of the plot ($\theta = 0$, any value of ϕ). Adapted from [261]

spectrometry (SIMS) [259]. The profile also depends on the matrix material whose stopping power depends on its density and atomic mass. While an implantation depth of about 50 nm is reached for boron in silicon ($A \approx 28$) for 10 keV, 20 keV are necessary in germanium ($A \approx 72.6$) [260]. The mean path length[6] d_m depends also on the crystallographic direction (channeling effects, Fig. 4.8) [261]. A simulation of the interaction of ions and solids can be performed using the SRIM software [262, 263].

[6]The mean path length is the distance integrated along the ion trajectory until its direction deviates by more than 4° from the incident direction.

4.2.5 Large Concentration Effects

Lattice Constant

At high doping concentration, a noticeable effect on the lattice constant a_0 is found. For silicon the atom density[7] is $N_{Si} = 5 \times 10^{22}\,cm^{-3}$. A doping level of $N = 10^{19}\,cm^{-3}$ corresponds thus to a dopant fraction of 0.02 %. Such crystal could also be considered a very dilute alloy. About each $(N_{Si}/N)^{1/3} \approx 17$th atom in a given direction is a dopant.

The effect of high doping on the lattice constant is due to different ionic radius of the dopant and the hydrostatic deformation potential of the band edge occupied by the free carriers [266]. In a linear approach, the effect is summarized in the coefficient β via

$$\beta = \frac{1}{N}\frac{\Delta a_0}{a_0}. \tag{4.21}$$

The effect due to charge carriers on β is negative (positive) for p-doping (n-doping). Experimental data for Si, Ge, GaAs and GaP are compiled in [267, 268] and theoretically discussed. The effect is in the order of $\beta = \pm(1 - 10) \times 10^{-24}\,cm^3$. For example, in the case of Si:B, the shrinkage of the lattice constant is mostly due to the charge carrier effect, for Si:P both effects almost cancel. In [269] it is shown that boron incorporation in silicon changes the lattice constant in various directions quite differently, e.g. d_{333} is shrunk by 0.4 % for a doping level of $10^{19}\,cm^{-3}$ while the {620} lattice constant remains constant.

Clustering

Point defects can cluster, i.e. several point defects aggregate at neighboring sites. An example the configuration of five nearby vacancies in silicon, the so-called V_5 cluster is shown in Fig. 4.9a. In [270] the ring-like hexavacancy in silicon is predicted a very stable defect. A large number of clustered vacancies is equivalent to a void. An example is depicted in Fig. 4.9b for an In_2O_3 crystal which has locally 'decomposed' into an indium particle and a void as revealed by TEM [271]. Also impurities can exhibit clustering (see Sect. 4.2.5).

Typically a random distribution of dopants in the host is assumed (cmp. Sect. 3.7.1 on random alloys). The introduction of several impurities can lead to pairing effects, e.g. described for Se and B, Ga, Al or In in silicon [273]. A high concentration of a single impurity makes the existence of clusters, i.e. two or more neighboring dopant atoms, more probable. This effect has been extensively studied for B in Si [274], showing that several boron atoms with interstitials I form thermodynamically stable clusters, e.g. B_3I_2. This cluster forms from B_2I and BI with only 0.2 eV activation barrier [275] as shown in Fig. 4.10. The formation is limited by diffusion of the smaller clusters to the same site. The number of free carriers (here holes) released

[7] Eight atoms per cubic unit cell of length $a_0 = 0.543\,nm$.

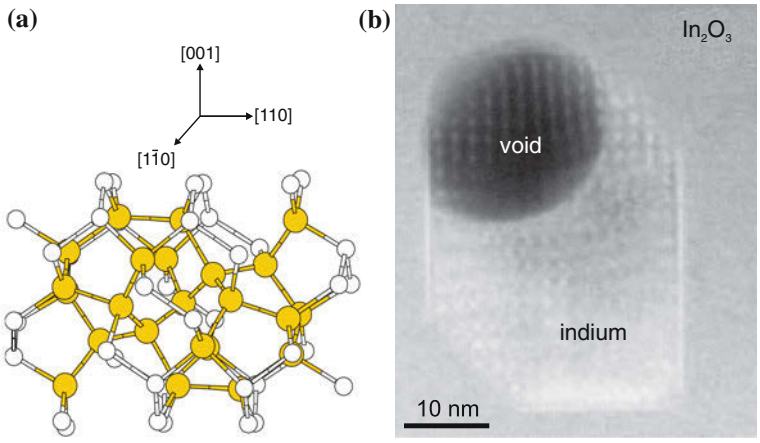


Fig. 4.9 (**a**) Predicted configuration of the V_5-cluster (five vacancies) in silicon. *Yellow spheres* indicate more distorted atoms than the rest of the lattice atoms (*white spheres*). Adapted from [272]. (**b**) Indium particle with adjacent void embedded in In_2O_3 (STEM image revealing Z-contrast in [001]). Adapted from [271]

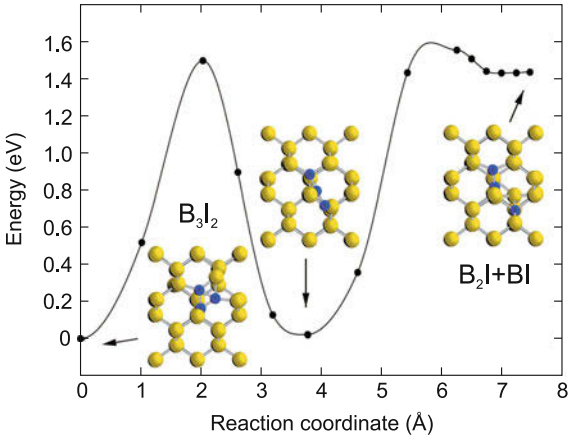


Fig. 4.10 Minimum energy path for the breakup of a B_3I_2 cluster into B_2I and BI. Silicon (boron) atoms are shown as *yellow* (*blue*) *spheres*. Adapted from [275]

from such cluster is smaller than the number of boron atoms since it forms a deep acceptor [274]. This autocompensation mechanism is thus limiting the maximum achievable free carrier concentration due to doping and is technologically unfavorable. Reactions between boron atoms and silicon self-interstitials often lead to boron clustering in the peak region of an implantation profile and require detailed optimization of the annealing process [276].

Solubility Limit

The steady-state impurity solubility can be defined as the maximum concentration of impurity atoms in a crystal allowing thermodynamic balance between the crystal and

Table 4.3 Maximum solubility N_s of some impurities in silicon

Impurity	N_s (10^{20} cm^{-3})
B	4
P	5
As	4
Sb	0.7
Al	0.13
Cu	1.4×10^{-2}
Au	1.2×10^{-3}
Fe	3×10^{-4}

Data for B, P, As, Sb from [278], other data from [281]

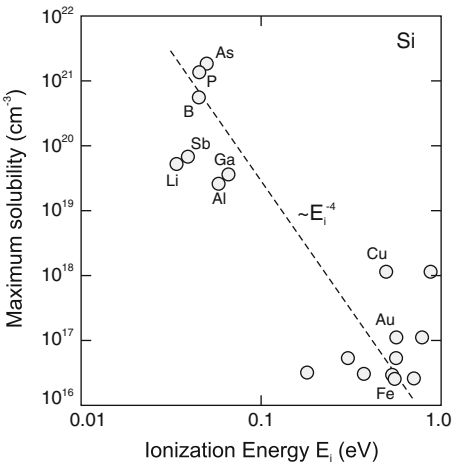


Fig. 4.11 Solubility limit for various impurities in silicon versus their ionization energy. Adapted from [279]

another phase, e.g. a liquid phase, an extended defect or a precipitate. Precipitates are small inclusions of a second phase in a crystal, exhibiting a high concentration of 'gathered' impurities that cannot be solved in the crystal. Solubility limits for impurities in silicon have been first determined in [277] with a bulk of subsequent research [278] due to its practical relevance in device fabrication. The solubility limits for a few impurities in silicon are listed in Table 4.3. It is related to the ionization energy of the defect (cmp. Sect. 7.4) as shown in Fig. 4.11.

The temperature dependence of the solubility for a few dopants is depicted in Fig. 4.12a. The solubility depends also on the present strain [280]. The simple empirical relation $x_s = 0.1\,k$ (Fig. 4.12b) between the maximum molar solubility x_s and the distribution coefficient k in silicon and germanium has been pointed out in [281].

A typical example for the formation of precipitates is Fe in InP, used for compensation of shallow donors in order to produce semi-insulating material (Sect. 7.7.8). The solubility of Fe in InP is fairly low, about 10^{17} cm^{-3} at growth temperature [282]. In Fig. 4.13 a high-resolution TEM image of a precipitate in InP doped with

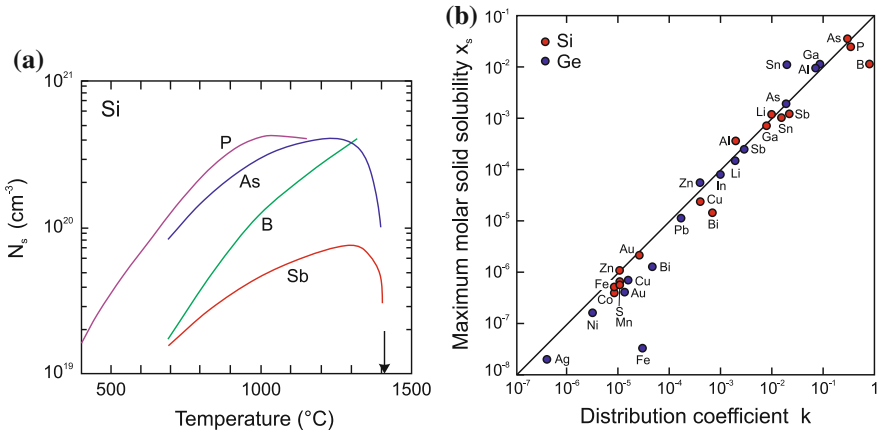


Fig. 4.12 (**a**) Steady-state solubility of impurities (P, As, B and Sb as labeled) in silicon. *Solid lines* are theoretical model matching various experimental data. *Arrow* denotes the melting point of silicon (1410°). Adapted from [278]. (**b**) Maximum molar solid solubility x_s versus the distribution coefficient for various impurities in crystalline silicon and germanium. *Solid line* follows $x_s = 0.1\,k$. Adapted from [281]

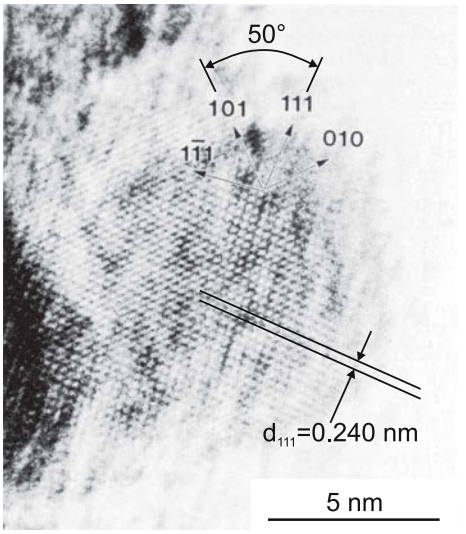


Fig. 4.13 High resolution TEM image of a FeP precipitate in iron-doped InP. Adapted from [283]

$3 \times 10^{18}\,\mathrm{cm}^{-3}$ Fe is shown. The precipitate exhibits a lattice constant of $d_{111} = 0.240\,\mathrm{nm}$ in [111]-direction, much different from that of InP ($d_{111}^{\mathrm{InP}} = 0.339\,\mathrm{nm}$). The angle between the [101] and [111] direction is 50° instead of 35° for InP. This is consistent with orthorhombic FeP [283]. Typically FeP and FeP_2 precipitates are found in highly Fe-doped InP [284].

4.3 Dislocations

Dislocations are line defects along which the crystal lattice is shifted by a certain amount. The vector along the dislocation line is called line vector **L**. A closed path around the dislocation core differs from that in an ideal crystal. The difference vector is called the Burger's vector **b**. Dislocations for which the Burger's vector is a vector of the lattice are called *full* dislocations. In contrast, dislocations with Burger's vectors that are not translation vectors of the lattice are called *partial* dislocations. The history of dislocation theory is described in [285].

Since the energy of a dislocation is proportional to $\mathbf{b}^2$, only dislocations with the shortest Burger's vector are stable. The plane spanned by **L** and **b** is called the *glide* plane. In Fig. 4.14 a high-resolution image of the atoms around a dislocation and the phase and amplitude of the (111) reflection are shown. The phase corresponds to the atomic columns, the amplitude to the displacement of the atoms at the dislocation core (see also Fig. 5.34).

4.3.1 Dislocation Types

Edge Dislocations

For an *edge* dislocation (Fig. 4.15a) **b** and **L** are perpendicular to each other. An extra half-plane spanned by **L** and $\mathbf{b} \times \mathbf{L}$ is inserted (Fig. 4.16).

Screw Dislocations

For a *screw* dislocation (Fig. 4.15b) **b** and **L** are collinear. The solid has been cut along a half-plane up to the dislocation line, shifted along **L** by the amount **b** and reattached.

Around the intersection of a screw dislocation with a surface, the epitaxial growth occurs, typically in the form of a growth spiral that images the lattice planes around the defect.

60° Dislocations

The most important dislocations in the zincblende lattice (Fig. 4.17) have the line vector along $\langle 110 \rangle$. With the Burger's vector $a/2\,\langle 110 \rangle$ three different types of dislocations can be formed: edge, screw and 60° dislocations. The vicinity of the core of the latter is shown in more detail in Fig. 4.17d. We note that the atomistic structure of 60° dislocations is different for **L** along [110] and [−1 10]; depending on whether the cations or anions are in the core, they are labeled α or β dislocations.

Misfit Dislocations

When materials with different lattice constants are grown on top of each other, the strain can plastically relax via the formation of misfit dislocations. A typical network of such dislocations is shown in Fig. 4.18 for SiGe on Si.

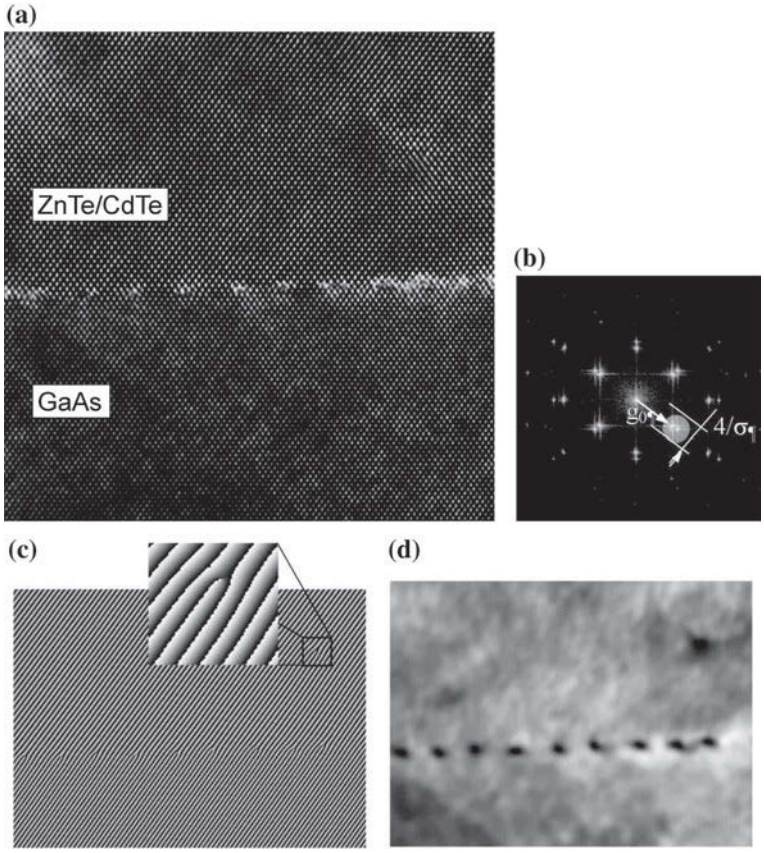


Fig. 4.14 (**a**) High-resolution transmission electron microscopy image (HRTEM) in the ⟨110⟩ projection of a network of misfit dislocations at a GaAs/CdTe/ZnTe interface. Substrate: GaAs (001), 2° off ⟨110⟩, ZnTe buffer layer is 2 monolayers thick. (**b**) Fourier transform with round mask around the (111) Bragg reflection. (**c**) Phase and (**d**) amplitude images for the mask from (**b**). From [286]

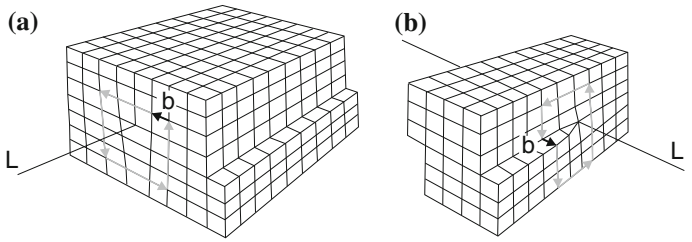


Fig. 4.15 Model of (**a**) an edge and (**b**) a screw dislocation. The line vector **L** and the Burger's vector **b** are indicated

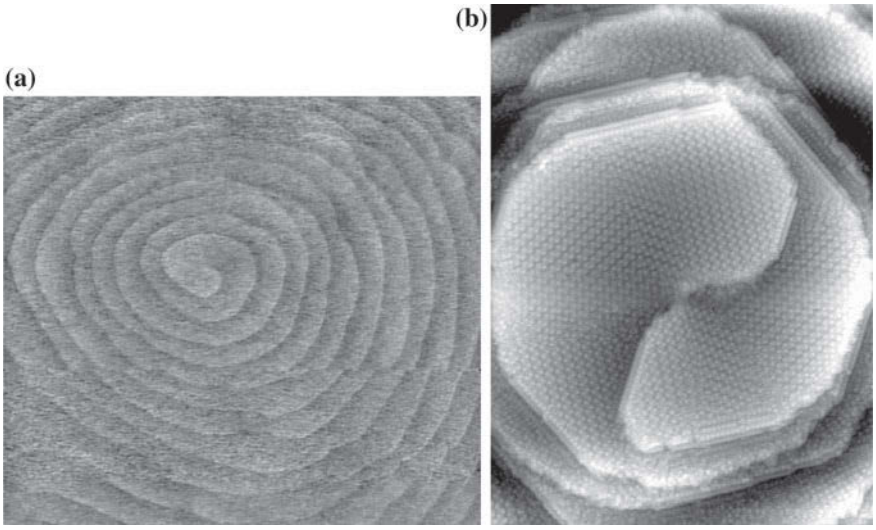


Fig. 4.16 (**a**) Atomic force microscopy image of growth spiral around a screw dislocation on a silicon surface; image width: 4 μm. (**b**) STM image (width: 75 nm) of a screw-type dislocation with a Burgers vector of [000-1] on the N-face of GaN. The reconstruction is c(6 × 12). The c(6 × 12) row directions correspond to $\langle \bar{1}100 \rangle$. Reprinted with permission from [287], © 1998 AVS

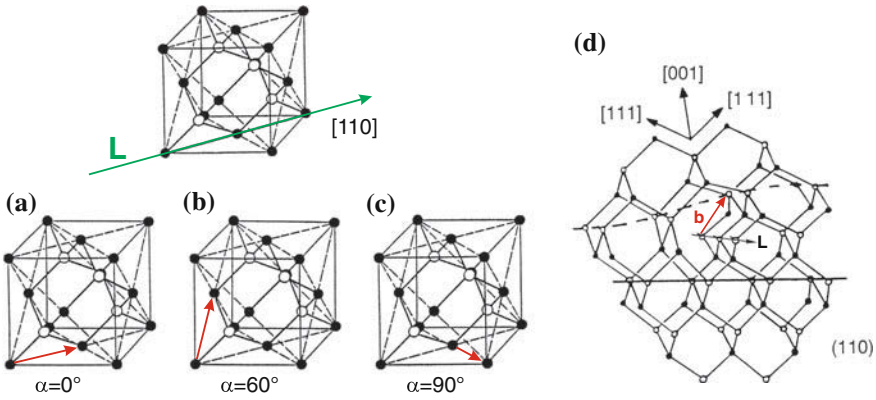


Fig. 4.17 Dislocations in the zincblende structure. The line vector is along [100]. The Burger's vector $a/2\,\langle 110 \rangle$ can create an (**a**) edge dislocation, a (**b**) screw dislocation, and (**c**) a 60° dislocation. (**d**) Atomistic structure of a 60° dislocationeps

Partial Dislocations

Partial dislocations, i.e. the Burger's vector is not a lattice vector, must necessarily border a two-dimensional defect, usually a stacking fault (Sect. 4.4.2). A typical partial dislocation in diamond or zincblende material is the Shockley partial dislocation (or just Shockley partial) with Burger's vector $b = (a_0/6)\,\langle 112 \rangle$. Another important partial is the Frank partial with $b = (a_0/3)\,\langle 111 \rangle$. A perfect dislocation

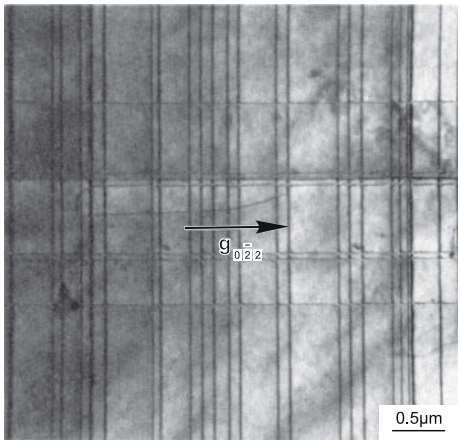


Fig. 4.18 Plan-view transmission electron microscopy image of a network of ⟨110⟩ dislocation lines in InGaAs on InP (001) with a lattice mismatch of about 0.1 %. The TEM diffraction vector is $\mathbf{g} = [2\bar{2}0]$. Adapted from [288]

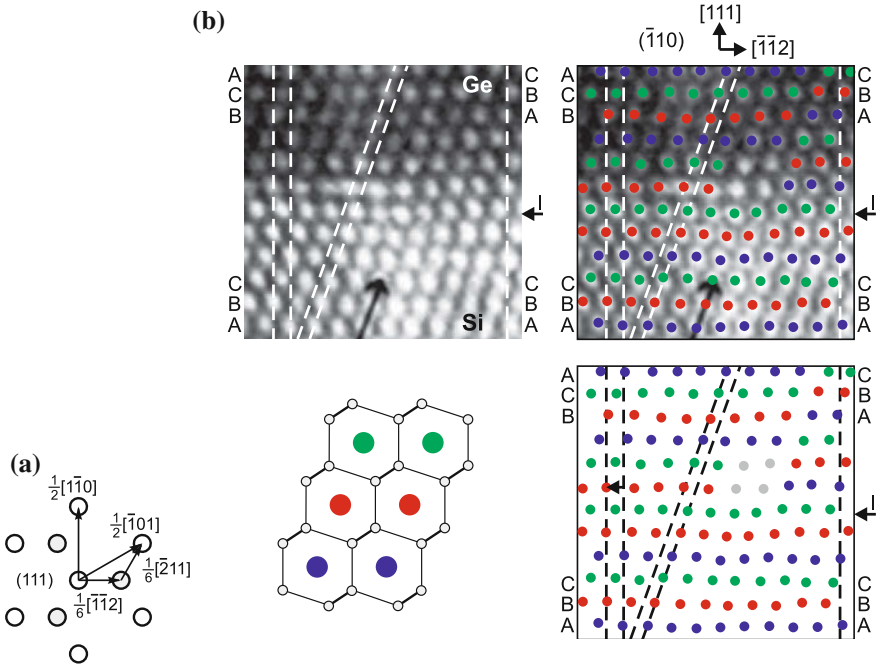


Fig. 4.19 (**a**) Graphical representation of the dislocation reaction of (4.22). (**b**) TEM image of the interface of a Ge/Si heterostructure with a $\left[\bar{2}11\right]/6$ Shockley partial dislocation. The image is overlayed with empty rod positions (as schematically shown in the *lower left* part of the figure) colored according to the stacking position (A: *blue*, B: *red*, C: *green*). The *arrows* labeled 'I' denote the position of the interface. Based on [289]

can be dissociated into two partials. This is energetically favorable. As an example we consider the reaction (Fig. 4.19a)

$$\frac{1}{2}\left[\bar{1}01\right] \rightarrow \frac{1}{6}\left[\bar{1}\bar{1}2\right] + \frac{1}{6}\left[\bar{2}11\right]. \tag{4.22}$$

The length of the full dislocation is $a_0/\sqrt{2}$. The length of the Shockley partial is $a_0/\sqrt{6}$. Thus the energy $E = G\,b^2$ of the full dislocation is $E_1 = Ga_0^2/2$ and the sum of the energies of the partials is smaller, $E_2 = 2Ga_0^2/6 = Ga_0^2/3$. In Fig. 4.19b a TEM image of a Ge/Si interface with a Shockley partial is shown.

4.3.2 Visualization of Dislocations by Etching

Defects can be made visible using etching techniques. This is particularly popular for finding dislocations. Many etches are anisotropic, i.e. the etch velocity varies for different crystal directions. As an example the result of etching a silicon sphere in molten KOH and a germanium sphere in a HNO_3/HF solution are shown in Fig. 4.20. The remaining bodies exhibit those planes with low etching velocity. The etch velocity of various etch solutions has been investigated in detail in particular for silicon (Fig. 4.21).

In a planar geometry, etch pits indicate the presence of dislocations, as shown in Fig. 4.22 for Ge of different orientation. The anisotropic etch prepares {111} planes. The dislocation core is at the intersection of the planes. In Fig. 4.23 hexagonal etch pits stretched along $[1\bar{1}0]$ are developed by molten KOH [292, 293]. The sides of the base are along [110], ⟨130⟩ and ⟨310⟩. The depth and width of the pits increases with increasing etching time. On the $(00\bar{1})$ surface, the orientation of the pits is rotated by 90° because of the polar [111]-axis of the zincblende structure [292]. Such etch

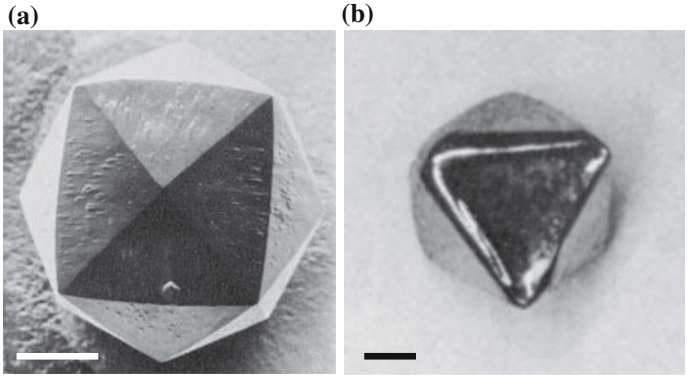


Fig. 4.20 (**a**) Resulting shape of Si sphere ('Lösungskörper') after 3 h at 100 °C in molten KOH. (**b**) Resulting shape of Ge sphere after etching in HNO_3:HF:CH_3COOH, 35:30:35 weight percent. The octaedric form indicates {111} faces. Markers are 1 mm. Adapted from [290]

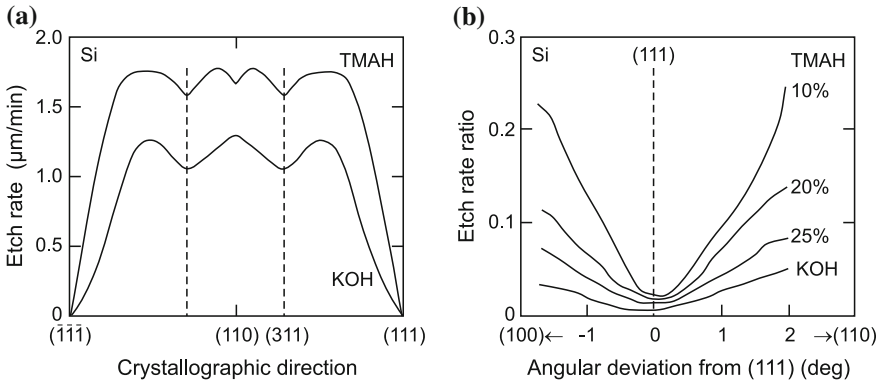


Fig. 4.21 (**a**) Etch rate of silicon for tetramethyl-ammonium-hydroxide (TMAH) water solution (25 %) at 86 °C and 40 % KOH at 70 °C as a function of crystallographic direction. (**b**) Detail of the anisotropy around the (111) direction for TMAH solutions with three different concentrations and 40 % KOH, all at 86 °C. Adapted from [291]

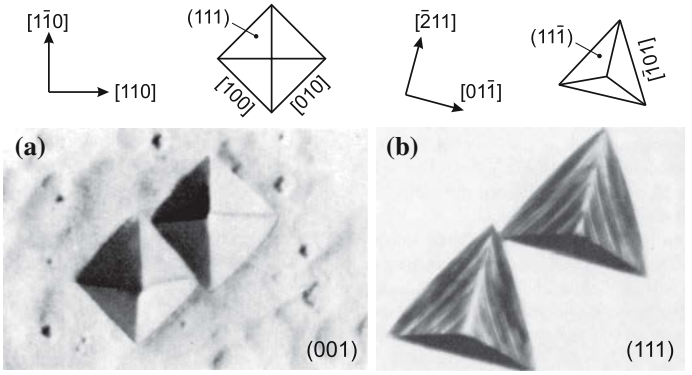


Fig. 4.22 Etch pits on germanium with (**a**) (001) and (**b**) (111) surface orientation. In both cases {111} facets are prepared by the etch. As etch in (**b**) a HNO_3/HF/CH_3COOH solution with $AgNO_3$ additive has been used. Width of the triangular etch pits is about 100 µm. Adapted from [296]

pit develops at a dislocation with Burger's vector $a/2\,[011]$ (inclined to the (001) surface) [294]. Other types of etch pits indicate dislocations with other Burger's vectors [294, 295]. Recipies how to wet chemically etch various semiconductors can be found in [290, 296–299]. Other etching techniques include dry processes such as plasma etching or reactive ion etching (RIE) [300–303].

4.3.3 Impurity Hardening

It has been found that the addition of impurities can lead to a substantial reduction of the dislocation density. This effect is known as impurity hardening and is caused by a hardening of the lattice due to an increase of the so-called critical resolved

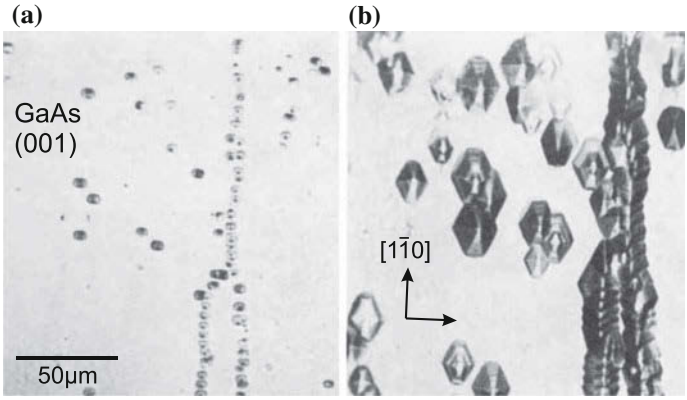


Fig. 4.23 Etch pits on GaAs (001) after (**a**) 3 min and (**b**) 10 min etch time in molten KOH at 300 °C. Adapted from [292]

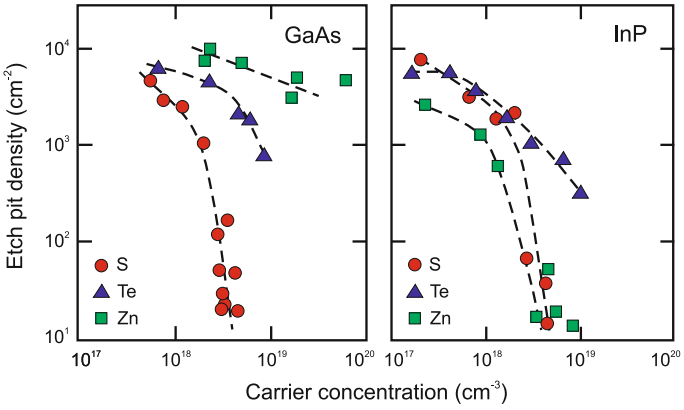


Fig. 4.24 Dislocation density (as revealed by etch pits) for GaAs and InP as a function of the carrier concentration for various concentrations of impurities (S, Te, and Zn). Adapted from [305]

shear stress [304]. In Fig. 4.24 the dependence of the dislocation density in GaAs and InP is shown as a function of the carrier density that is induced by the incorporation of (electrically active) group-II or group-VI atoms (acceptors or donors, cf. Sect. 7.5). The high carrier concentration is unwanted when semi-insulating substrates (cf. Sect. 7.7.8) or low optical absorption (cf. Sect. 9.8.1) are needed. Thus the incorporation of isovalent impurities, such as In, Ga or Sb in GaAs and Sb, Ga or As in InP, has been investigated and found to be remarkably effective. Material containing such impurities in high concentration ($>10^{19}\,\mathrm{cm}^{-3}$) must be considered a low-concentration alloy. The lattice constant is thus slightly changed, which can cause problems in the subsequent (lattice-mismatched) epitaxy of pure layers.

4.4 Extended Defects

4.4.1 Micro-cracks

If the stress in a material becomes too big to be accommodated by dislocations, cracks may form to release strain energy.[8] In Fig. 4.25 an example is shown. In this case, micro-cracks have formed in a bulk mercury indium telluride crystal upon incorporation of residual stress and thermal stress during cooling of the material from growth temperature (about 1000 K) to room temperature. See also Fig. 12.19 for micro-cracks in an epitaxial layer.

4.4.2 Stacking Faults

The ideal stacking of (111) planes in the zincblende structure, ABCABC. . ., can be disturbed in various ways and creates area defects. If one plane is missing, i.e. the stacking is ABC*AC*ABC, an *intrinsic* stacking fault is present. If an additional plane is present, the defect is called an *extrinsic* stacking fault, i.e. ABCAB*A*CABC. An extended stacking fault in which the order of stacking is reversed is called a twin lamella, e.g. ABCABC*BAC B*ABCABC. If two regions have inverted stacking order they are called twins and their joint interface is called a twin boundary, e.g. . . .ABCABCAB*C*BACBACBA. . . (Fig. 4.28). The various types of stacking faults are shown in Fig. 4.26. In Fig. 4.27 a cross-sectional image of stacking faults in GaAs on Si is shown. They block each other and thus partially annihilate with increasing thickness.

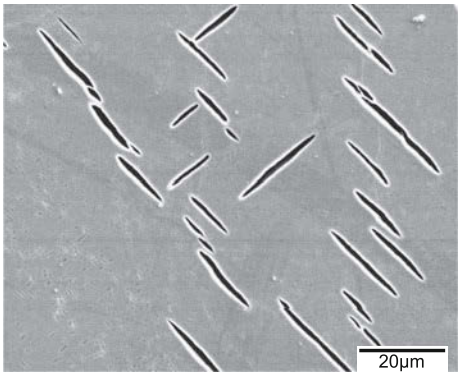


Fig. 4.25 Micro-cracks in a mercury indium telluride crystal. Adapted from [306]

[8] We note that in elasticity theory a *continuous* deformation is assumed. Obviously the separation (fracture) into two unstrained blocks is the lowest strain energy state of a stressed piece of material.

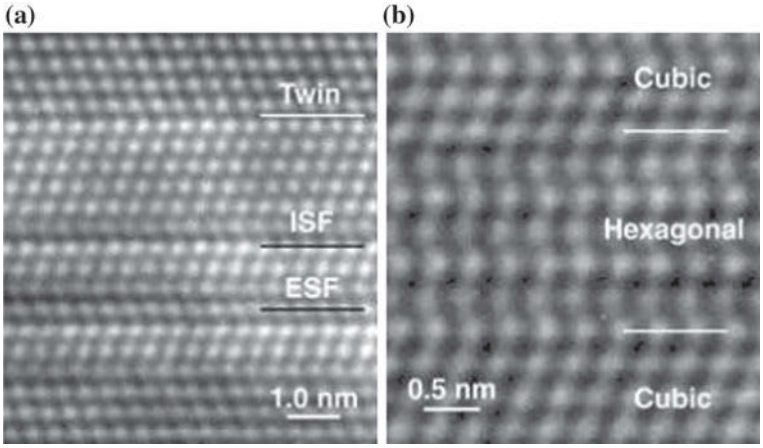


Fig. 4.26 HRTEM images of (**a**) thin-film silicon with intrinsic (labeled 'ISF') and extrinsic ('ESF') stacking faults and twin boundary ('Twin'). (**b**) Six monolayer thick hexagonal (wurtzite) CdTe layer in cubic (zincblende) CdTe. Stacking order (from bottom to top) is: ABCABABABABC... Reprinted with permission from [307]

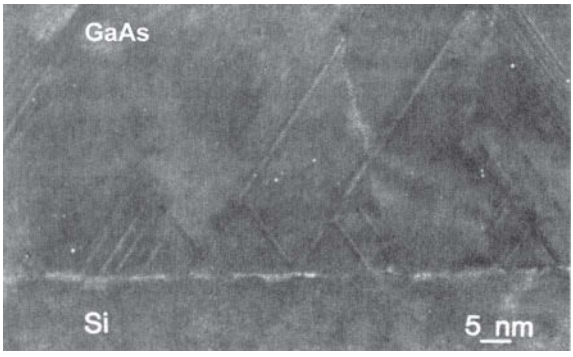


Fig. 4.27 Cross-sectional TEM image showing stacking faults in heteroepitaxial GaAs on Si. Adapted from [308]

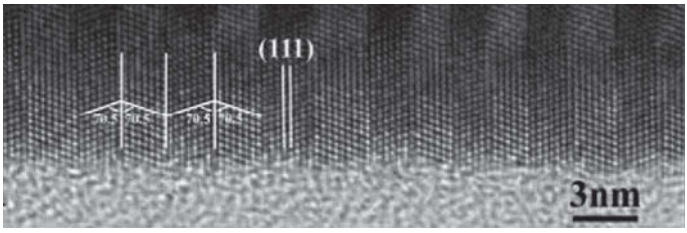


Fig. 4.28 High resolution TEM image of ZnS nanowire exhibiting periodical twin structures. Adapted from [309]

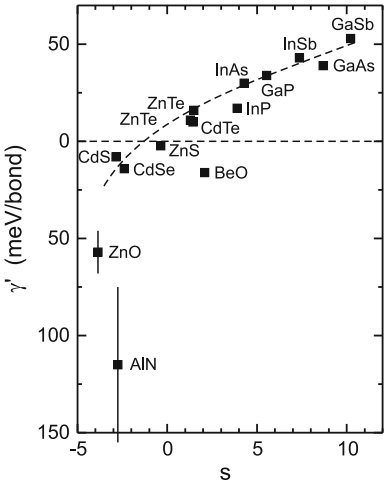


Fig. 4.29 Reduced stacking fault energy (stacking fault energy per bond) γ' for various compound semiconductors plotted as a function of the s-parameter. *Dashed line* is guide to the eye. Data from [164]

A stacking fault is bounded by two partial dislocations (Sect. 4.3.1) formed by the dissociation of a perfect dislocation. A full (or *perfect*) dislocation with Burger's vector $a/2[110]$ in a III–V compound is dissociated into two Shockley partials according to (4.22) [310]. Since the dislocation energy is proportional to $|\mathbf{b}|^2$, the dissociation is energetically favored (see Sect. 4.3.1).

The stacking-fault energy in pure silicon is $\gamma = 47\,\mathrm{mJ\,m^{-2}}$ [311]. A similar value is found for Ge, $\gamma = 60\,\mathrm{mJ\,m^{-2}}$ [312] and undoped GaAs, $\gamma = 45\,\mathrm{mJ\,m^{-2}}$ [313]. In diamond a much larger value is found, $\gamma = 285\,\mathrm{mJ\,m^{-2}}$ [314]. Impurity incorporation typically reduces the stacking fault energy. The systematics of stacking fault energy for various III–V and II–VI compounds has been discussed [164, 315, 316]. It can be correlated with the s-parameter (2.11) as depicted in Fig. 4.29.

4.4.3 Grain Boundaries

The boundaries of crystal grains are called *grain boundaries*. They are defined by five parameters, three rotation angles (e.g. Euler angles) to describe how the orientation of grain II results from grain I and two parameters to define the boundary plane of the two grains in the coordinate system of reference grain I.

Such defects can have a large impact on the electric properties. They can collect point defects and impurities, act as barriers for transport (Sect. 8.3.8) or as carrier sinks due to (nonradiative) recombination. Details of their structure and properties can be found in [317, 318]. The two crystal grains meet each other with a relative tilt and/or twist. The situation is shown schematically in Fig. 4.30a for a small angle between the two crystals. A periodic pattern of dislocations forms at the interface that is called a small-angle grain boundary (SAGB) (Fig. 4.30b). In

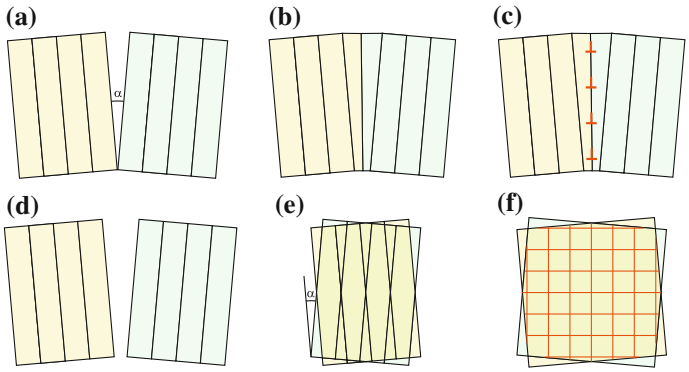


Fig. 4.30 Schemes of (**a**, **b**, **c**) pure tilt and (**d**, **e**, **f**) pure twist boundary, dislocation formation in (**c**) pure tilt and (**f**) twist boundaries

Fig. 4.31, experimental results for pure tilt SAGB are shown. The dislocation spacing is inversely proportional to the tilt angle θ. An image of a twist SAGB is shown in Fig. 4.32.

Special large angle boundaries possess (for a certain angle) a *coincident site lattice* (CSL). Some of these grain boundaries have a low energy and are thus commonly observed. The ratio of lattice points of the CSL and the lattice unit cell is an odd integer number n; the corresponding grain boundary is then labeled Σn. SAGB are also termed $\Sigma 1$. $\Sigma 3$ grain boundaries are always twin boundaries. An example with (111) grain boundary is schematically shown in an example in Fig. 4.33a. A $\Sigma 3$ (twin) boundary in silicon with {112} grain boundary [319, 320] is depicted in Fig. 4.34 together with the atomic arrangement of the grain boundary itself. A $\Sigma 5$ (001) grain boundary is schematically shown in Fig. 4.33b; the special angle is $\theta = \arctan 3/4 \approx 36.87°$.

Real grain boundaries may not be flat, contain impurities or precipitates and even consist of a thin amorphous layer.

4.4.4 Antiphase and Inversion Domains

Antiphase domains occur when one part of the crystal is shifted with respect to another by an antiphase vector **p**. This does not form a twin. If the polar direction changes between two domains they are called inversion domains.

In the zincblende structure the [110] and [$\bar{1}$10] directions are not equivalent. In one case there is a Zn-S lattice and in the other a S-Zn lattice. Both lattices vary by a 90° rotation or an inversion operation (which is not a symmetry operation of the zincblende crystal). If, e.g., a zincblende crystal is grown on a Si surface with monoatomic steps (Fig. 4.35, cmp. Fig. 11.6), adjoint regions have a different phase;

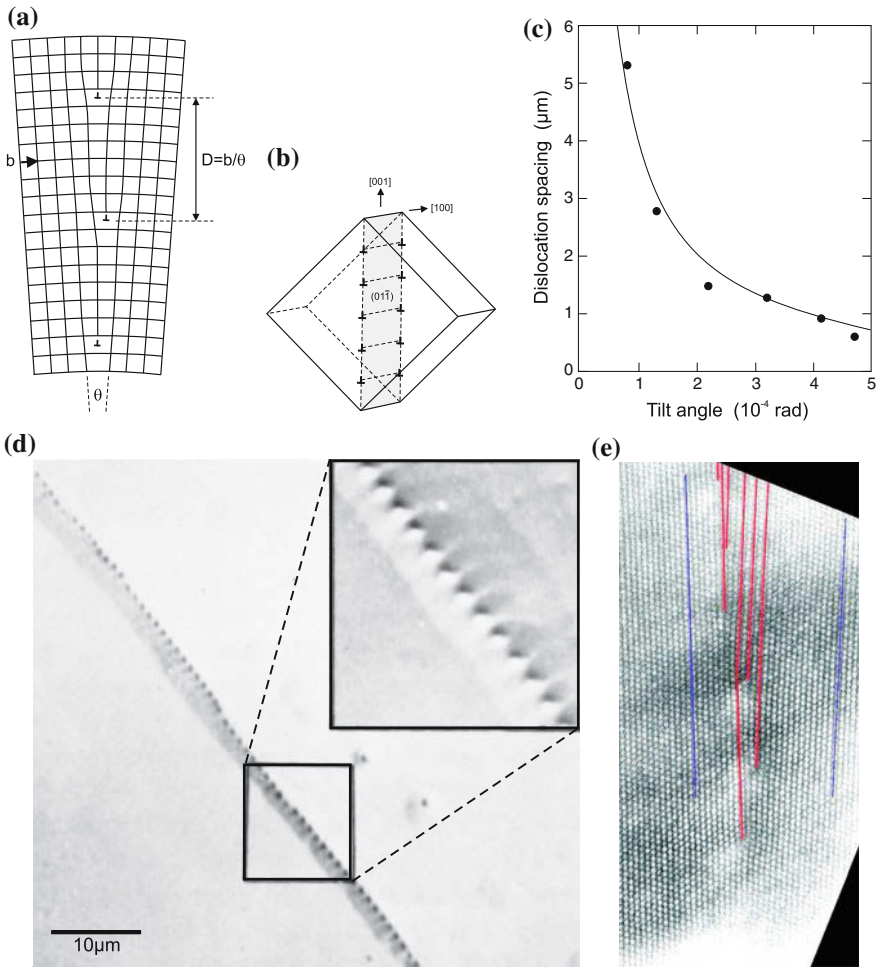


Fig. 4.31 (**a**) Scheme of a small-angle (pure tilt) grain boundary. (**b**) Model of edge dislocations in a {110} plane in Ge. (**c**) Relation of dislocation distance d and tilt angle θ for various small-angle grain boundaries in Ge. *Solid line* is relation $d = 4.0 \times 10^{-8}/\theta$. (**d**) Optical image of an etched (CP–4 etch) Ge sample with a small-angle grain boundary. Adapted from [321]. (**e**) HRTEM image of a small-angle grain boundary in Si with dislocations highlighted. From [322]

they are called antiphase domains (APD). The antiphase vector is $(0, 0, 1)\, a_0/4$. At the boundaries a two-dimensional defect, an antiphase domain boundary, develops. The APD boundary contains bonds between identical atom species. In Fig. 4.36, intertwining APD boundaries are shown on the surface of InP layers on Si. The antiphase domains can be visualized with an anisotropic etch.

In Fig. 4.37a, inversion domains in iron-doped ZnO are shown. Between domains the direction of the c-axis is reversed. The iron is found preferentially in the inversion domain boundary (IDB) (Fig. 4.37b) and plays an important role in its formation [326, 327].

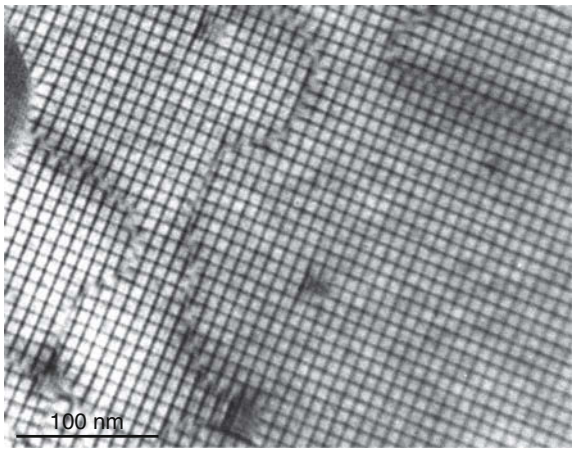


Fig. 4.32 Bright-field TEM image of pure twist boundary with network of pure twist dislocations fabricated by wafer bonding of two Si (001) surfaces with a relative twist. Adapted from [323]

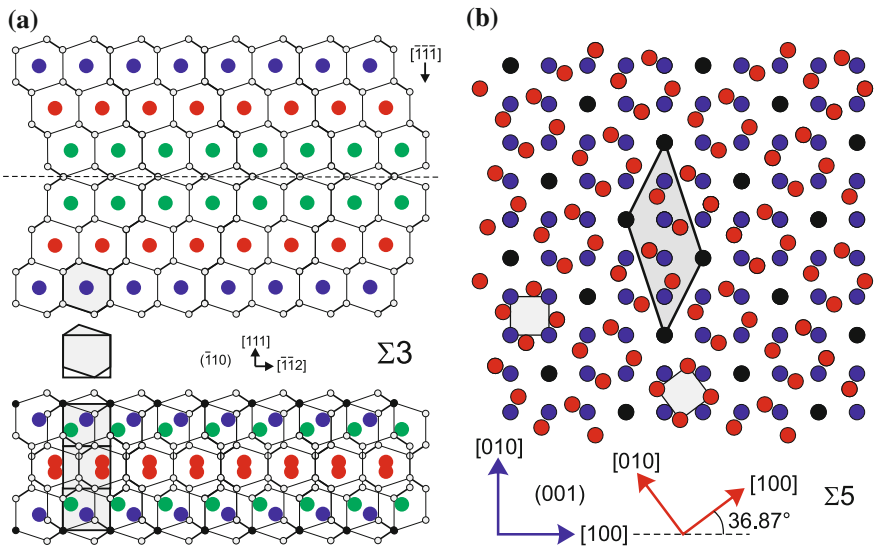


Fig. 4.33 (**a**) Schematic of $\Sigma 3$ (111) twin boundary in a diamond or zincblende structure (cmp. Fig. 4.28). The grain boundary is marked by a *dashed line* shown in side-view. The hexagonal and rectangular *grey boxes* have the same area. The lattice points of the coincident site lattice (CSL) are shown with *black circles* in the lower part of the figure. The unit cell of the CSL has three times the volume of the unit cell of the fcc lattice. (**b**) Schematic of a $\Sigma 5$ (001) grain boundary in a (simple) cubic crystal shown in plane-view. The *blue* and the *red* lattice are rotated by 36.86°, the lattice points of the CSL are shown in *black*. The unit cell of the CSL lattice (*dark grey*) has five times the volume of the cubic unit cell (*light grey*)

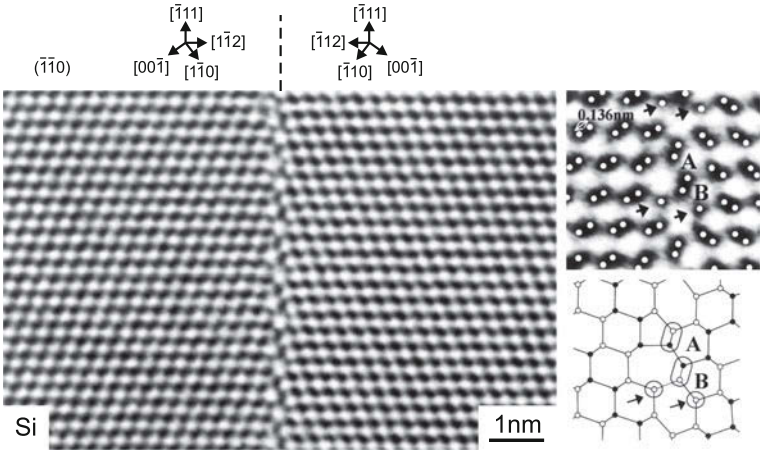


Fig. 4.34 TEM images in two magnifications of a $\Sigma 3\{112\}$ boundary in silicon together with a schematic of the atomic arrangement. Adapted from [320]

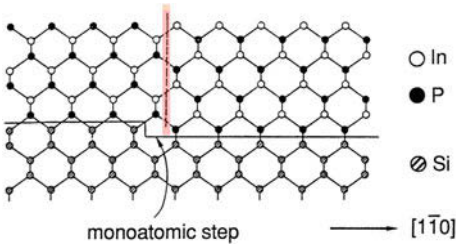


Fig. 4.35 Monoatomic step on the Si (001) surface and subsequent formation of an antiphase boundary in InP (zincblende)

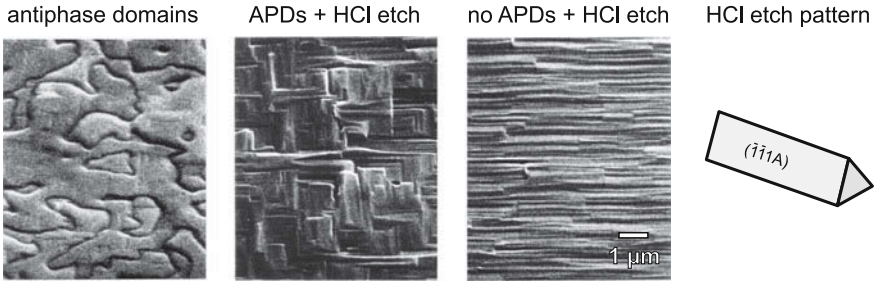


Fig. 4.36 Antiphase domains in InP on Si. HCl etchs InP anisotropically and prepares (111)A planes. The etch patterns of layers with (without) APDs are cross-hatched (linear). Adapted from [324]

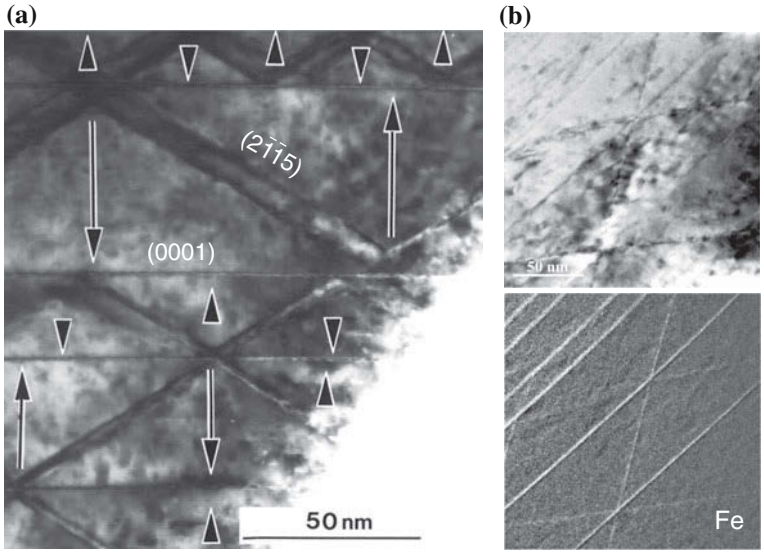


Fig. 4.37 Transmission electron microscopy of inversion domains in ZnO:Fe. (**a**) Inversion domains in iron-doped ZnO ($ZnO:Fe_2O_3 = 100:1$). *Arrows* denote the orientation of the *c*-axis in the respective domains. (**b**) *Top*: bright field TEM, *bottom* Fe distribution from energy-filtered image. Adapted from [325]

4.5 Disorder

Disorder is a general term for deviations from the ideal structure on a microscopic scale. Apart from the various structural defects discussed in the previous chapters, further examples of disorder are

- The presence of various isotopes of an element. This introduces disorder with regard to the mass of the atoms and impacts mostly phonon properties (see Fig. 8.24).
- The occupation of lattice sites in alloys (Sect. 3.7) ranging from a random alloy, clustering to (partially) ordered phases.
- The thermal and zero-point motion of atoms around their equilibrium position.

Chapter 5
Mechanical Properties

Abstract Lattice vibrations and phonons are treated with one-dimensional models and examples for real phonon dispersions for several semiconductors including phonons in alloys and disordered materials are given. Then the theory of linear elasticity and its application to semiconductors with regard to epitaxial strain, substrate bending and sheet-scrolling is given. Finally plastic relaxation effects such as critical thickness and wafer breakage are discussed.

5.1 Introduction

The atoms making up the solid have an average position from which they can deviate since they are elastically bonded. The typical atomic interaction potential looks like the one shown in Fig. 2.1. The atoms thus perform a vibrational motion (including zero point fluctuations) and the solid is elastic. The potential is essentially asymmetric, being steeper for small distances due to quantum-mechanical overlap of orbitals. However, for small amplitudes around the minimum a harmonic oscillator can be assumed (harmonic approximation). Beyond the elastic regime, plastic deformation occurs such as generation of defects, e.g. dislocations. Eventually also the crystal can break.

5.2 Lattice Vibrations

In the following we will discuss the dispersion relations for lattice vibrations, i.e. the connection between the frequency ν (or energy $h\nu = \hbar\omega$) of the wave and its wavelength λ (or k-vector $k = 2\pi/\lambda$). Acoustic and optical vibrations are introduced in one-dimensional models. A detailed treatment of the physics of lattice vibrations is given in [328].

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_5

5.2.1 Monoatomic Linear Chain

The essential physics of lattice vibrations can best be seen from a one-dimensional model that is called the linear chain. The mechanical vibrations will also be called *phonons*, although technically this term is reserved for the quantized lattice vibrations resulting from the quantum-mechanical treatment.

In the monoatomic linear chain the atoms of mass M are positioned along a line (x-axis) with a period (lattice constant) a at the positions $x_{n_0} = na$. This represents a one-dimensional Bravais lattice. The Brillouin zone of this system is $[-\pi/a, \pi/a]$.

The atoms will interact with a harmonic potential, i.e. the energy is proportional to the displacement $u_n = x_n - x_{n_0}$ to the second power. The total (mechanical) energy of the system is then:

$$U = \frac{1}{2} C \sum_n (u_n - u_{n+1})^2. \tag{5.1}$$

The model assumes that the mass points are connected via massless, ideal springs with a spring constant C. If $\phi(x)$ is the interaction energy between two atoms, C is given by $C = \phi''(a)$. Again, the harmonic approximation is only valid for small displacements, i.e. $u_n \ll a$. The displacement of the atoms can be along the chain (longitudinal wave) or perpendicular to the chain (transverse wave), see Fig. 5.1. We note that for these two types of waves the elastic constant C must not be the same.

When the sum in (5.1) has a finite number of terms ($n = 0, \ldots, N-1$), the boundary conditions have to be considered. There are typically two possibilities: The boundary atoms are fixed, i.e. $u_0 = u_{N-1} = 0$, the boundary conditions are periodic (Born–von Karman), i.e. $u_i = u_{N+i}$. If $N \gg 1$, the boundary conditions play no significant role anyway, thus those with the greatest ease for subsequent math are chosen. In solid-state physics typically periodic boundary conditions are

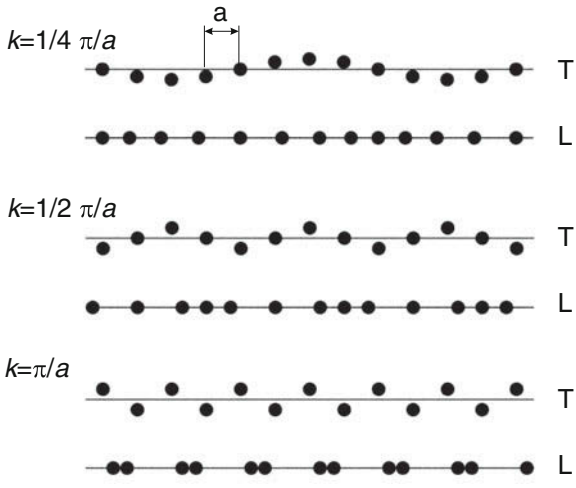


Fig. 5.1 Visualization of transverse ('T') and longitudinal ('L') waves in a linear monoatomic chain at different wavevectors

used. Boundary phenomena, such as at surfaces, are then treated separately (see Sect. 11.6.1).

The equations of motion derived from (5.1) are

$$M\ddot{u}_n = F_n = -\frac{\partial U}{\partial u_n} = -C\,(2u_n - u_{n-1} - u_{n+1}). \tag{5.2}$$

We solve for solutions that are periodic in time (harmonic waves), i.e. $u_n(x,t) = u_n \exp(-\mathrm{i}\omega t)$. Then the time derivative can be executed immediately as $\ddot{u}_n = -\omega^2 u_n$ and we obtain:

$$M\omega^2\, u_n = C\,(2u_n - u_{n-1} - u_{n+1}). \tag{5.3}$$

If, also, the solution is periodic in space, i.e. is a (one-dimensional) plane wave, i.e. $u_n(x,t) = v_0 \exp(\mathrm{i}(kx - \omega t))$ with $x = na$, we find from the periodic boundary condition $\exp(\mathrm{i}kNa) = 1$ and thus

$$k = \frac{2\pi}{a}\,\frac{n}{N},\ n \in \mathbf{N}. \tag{5.4}$$

It is important that, when k is altered by a reciprocal space vector, i.e. $k' = k + 2\pi n/a$, the displacements u_n are unaffected. This property means that there are only N values for k that generate independent solutions. These can be chosen as $k = -\pi/a, \ldots, \pi/a$, so that k lies in the Brillouin zone of the lattice. In the Brillouin zone there is a total number of N k-values, i.e. one for each lattice point. The distance between adjacent k-values is

$$\frac{2\pi}{Na} = \frac{2\pi}{L}, \tag{5.5}$$

L being the lateral extension of the system.

The displacements at the lattice points n and $n+m$ are now related to each other via

$$\begin{aligned} u_{n+m} &= v_0\, \exp(\mathrm{i}k(n+m)a) \\ &= v_0\, \exp(\mathrm{i}kna)\, \exp(\mathrm{i}kma) = \exp(\mathrm{i}kma)\, u_n. \end{aligned} \tag{5.6}$$

Thus, the equation of motion (5.3) reads

$$M\omega^2 u_n = C\,\left[2 - \exp(-\mathrm{i}ka) - \exp(\mathrm{i}ka)\right] u_n. \tag{5.7}$$

Using the identity $\exp(\mathrm{i}ka) + \exp(-\mathrm{i}ka) = 2\cos(ka)$, we find the dispersion relation of the monoatomic linear chain (Fig. 5.2):

$$\omega^2(k) = \frac{4C}{M}\,\frac{1-\cos(ka)}{2} = \frac{4C}{M}\,\sin^2\left(\frac{ka}{2}\right). \tag{5.8}$$

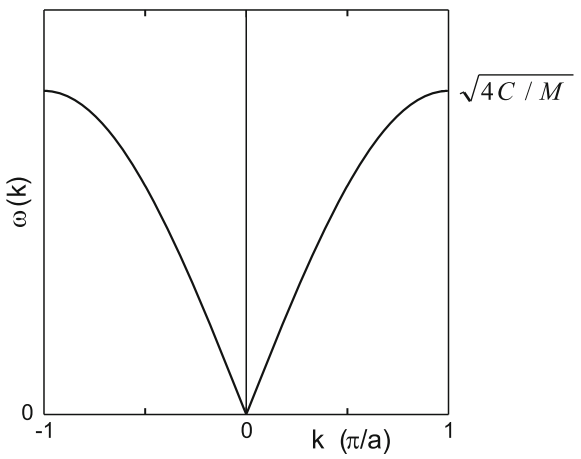


Fig. 5.2 Dispersion relation for a monoatomic linear chain

The solutions describe plane waves that propagate in the crystal with a phase velocity $c = \omega/k$ and a group velocity $v_g = \mathrm{d}\omega/\mathrm{d}k$

$$v_g = \sqrt{\frac{4C}{M}}\,\frac{a}{2}\,\cos\left(\frac{|k|a}{2}\right). \tag{5.9}$$

In the vicinity of the Γ point, i.e. $k \ll \pi/a$ the dispersion relation is linear in k

$$\omega(k) = a\sqrt{\frac{C}{M}}\,|k|. \tag{5.10}$$

We are used to such linear relations for sound (and also light) waves. The phase and group velocity are the same and do not depend on k. Thus, such solutions are called *acoustic*. The sound velocity of the medium is given by $v_\mathrm{s} = a\sqrt{C/M}$.

It is characteristic of the nonhomogeneous medium that, when k approaches the boundary of the Brillouin zone, the behavior of the wave is altered. For $k = \pi/a$ the wavelength is just $\lambda = 2\pi/k = 2a$, and thus samples the granularity of the medium. The maximum phonon frequency ω_m is

$$\omega_\mathrm{m} = \sqrt{\frac{4C}{M}}. \tag{5.11}$$

The group velocity is zero at the zone boundary, thus a standing wave is present.

Since the force constants of the longitudinal and transverse waves can be different, the dispersion relations are different. The transverse branch of the dispersion relation is 2-fold degenerate, unless the two directions that are perpendicular to x are not equivalent.

5.2.2 Diatomic Linear Chain

Now we consider the case that the system is made up from two different kinds of atoms (Fig. 5.3). This will be a model for semiconductors with a diatomic base, such as zincblende. We note that the diamond structure also needs to be modeled in this way, although both atoms in the base are the same.

The lattice will be the same and the lattice constant will be a. Alternating atoms of sort 1 and 2 with a relative distance of $a/2$ are on the chain. The displacements of the two atoms are labeled u_n^1 and u_n^2, both belonging to the lattice point n. The atoms have the masses M_1 and M_2. The force constants are C_1 (for the 1–2 bond within the base) and C_2 (for the 2–1 bond between different bases).

The total energy of the system is then given as

$$U = \frac{1}{2} C_1 \sum_n \left(u_n^1 - u_n^2\right)^2 + \frac{1}{2} C_2 \sum_n \left(u_n^2 - u_{n+1}^1\right)^2 . \tag{5.12}$$

The equations of motion are

$$M_1 \ddot{u}_n^1 = -C_1 \left(u_n^1 - u_n^2\right) - C_2 \left(u_n^1 - u_{n-1}^2\right) \tag{5.13a}$$

$$M_2 \ddot{u}_n^2 = -C_1 \left(u_n^2 - u_n^1\right) - C_2 \left(u_n^2 - u_{n+1}^1\right) . \tag{5.13b}$$

With the plane-wave ansatz $u_n^1(x,t) = v_1 \exp\left(\mathrm{i}(kna - \omega t)\right)$ and $u_n^2(x,t) = v_2 \exp\left(\mathrm{i}(kna - \omega t)\right)$ and periodic boundary conditions we find

$$0 = -M_1 \omega^2 v_1 + C_1 (v_1 - v_2) + C_2 \left[v_1 - v_2 \exp(-\mathrm{i}ka)\right] \tag{5.14a}$$

$$0 = -M_2 \omega^2 v_2 + C_1 (v_2 - v_1) + C_2 \left[v_2 - v_1 \exp(\mathrm{i}ka)\right]. \tag{5.14b}$$

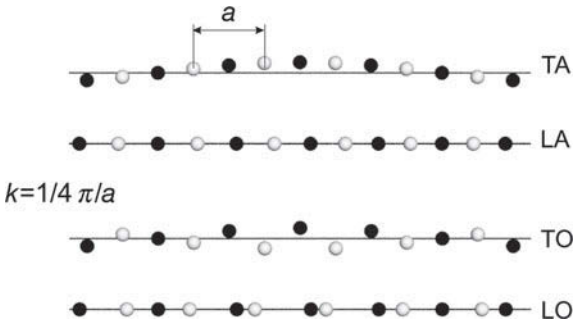


Fig. 5.3 Visualization of acoustic and optical waves in a diatomic linear chain

These equations for v_1 and v_2 can only be solved nontrivially if the determinant vanishes, i.e.

$$\begin{aligned} 0 &= \begin{vmatrix} M_1\,\omega^2 - (C_1 + C_2) & C_1 + \mathrm{e}^{-\mathrm{i}ka}\,C_2 \\ C_1 + \mathrm{e}^{\mathrm{i}ka}\,C_2 & M_2\,\omega^2 - (C_1 + C_2) \end{vmatrix} \\ &= M_1 M_2\,\omega^4 - (M_1 + M_2)(C_1 + C_2)\,\omega^2 + 2C_1C_2\,[1 - \cos(ka)]. \end{aligned} \tag{5.15}$$

Using the substitutions $C_+ = (C_1 + C_2)/2$, $C_\times = \sqrt{C_1C_2}$, the arithmetic and geometrical averages, and accordingly for M_+ and $M_\times$, the solution is

$$\omega^2(k) = \frac{2C_\times}{\gamma\,M_\times}\left[1 \pm \sqrt{1 - \gamma^2\,\frac{1 - \cos(ka)}{2}}\right], \tag{5.16}$$

with

$$\gamma = \frac{C_\times\,M_\times}{C_+M_+} \le 1. \tag{5.17}$$

The dispersion relation, as shown in Fig. 5.4, now has (for each longitudinal and transverse mode) two branches. The lower branch ('−' sign in (5.16)) is related to the acoustic mode; neighboring atoms have similar phase (Fig. 5.3). For the acoustic mode $\omega = 0$ at the Γ point and the frequency increases towards the zone boundary. The maximum phonon frequency ω_m is in the upper branch ('+' sign in (5.16)) at the zone center

$$\omega_\mathrm{m} = \sqrt{\frac{4C_\times}{\gamma\,M_\times}} = 2\sqrt{\frac{C_+\,M_+}{M_\times^2}}. \tag{5.18}$$

The upper branch is called the optical mode (since it can interact strongly with light, see Sect. 9.9) and neighboring atoms have opposite phase. In the vicinity of the Γ point the dispersion of optical phonons is parabolic with negative curvature:

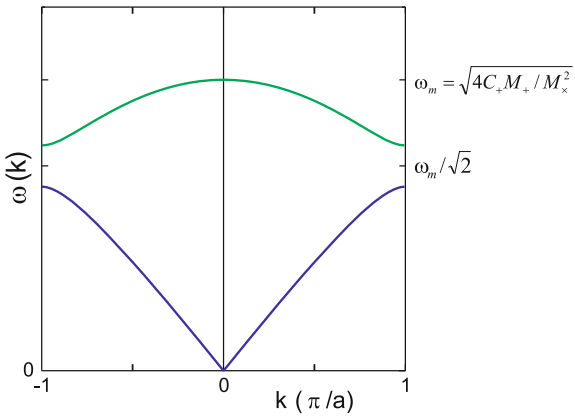


Fig. 5.4 Dispersion relation for a diatomic linear chain with acoustic (*blue*) and optical (*green*) branch

$$\omega(k) \cong \omega_{\mathrm{m}} \left[1 - \frac{1}{2} \left(\frac{\gamma a}{4} \right)^2 k^2 \right]. \tag{5.19}$$

Thus, four different vibrations exist that are labeled TA, LA, TO, and LO. Both the TA and TO branches are degenerate.

At the zone boundary (X point) a frequency gap exists. The gap center is at

$$\overline{\omega_{\mathrm{X}}} = \frac{\omega_{\mathrm{m}}}{\sqrt{2}}, \tag{5.20}$$

and the total width of the gap is

$$\Delta\omega_X = \omega_{\mathrm{m}} \sqrt{1-\gamma} = 2\sqrt{\frac{C_+ M_+ - C_\times M_\times}{M_\times^2}}. \tag{5.21}$$

The group velocity is zero for optical and acoustic phonons at $k = \pi/a$ and for optical phonons at the Γ point.

Usually two cases are treated explicitly: (i) atoms with equal mass ($M = M_1 = M_2$) and different force constants or (ii) atoms with unequal mass and identical force constants $C = C_1 = C_2$. For the case $C_1 = C_2$ and $M_1 = M_2$, $\gamma = 1$ and thus $\Delta\omega_X = 0$. Then the dispersion relation is the same as for the monoatomic chain, except that the k space has been folded since the actual lattice constant is now $a/2$.

$M_1 = M_2$

In this case, $M_+ = M_\times = M$ and the dispersion relation is

$$\omega^2 = \frac{2C_+}{M} \left[1 \pm \sqrt{1 - \frac{C_\times^2}{C_+^2} \frac{1-\cos(ka)}{2}} \right]. \tag{5.22}$$

At the zone boundary the frequencies for the acoustic and the optical branch are $\omega_{X,1} = \sqrt{2C_1/M}$ with $v_1 = v_2$ and $\omega_{X,2} = \sqrt{2C_2/M}$ with $v_1 = -v_2$, respectively (assuming $C_2 > C_1$). The motion for $k = \pi/a$ is phase shifted by 180° for adjacent bases. Additionally, for the acoustic branch the atoms of the base are in phase, while for the optical branch the atoms of the base are 180° out of phase. The vibration looks as if only one of the springs is strained.

$C_1 = C_2$

In this case, $C_+ = C_\times = C$ and the dispersion relation is

$$\omega^2 = \frac{2CM_+}{M_\times^2} \left[1 \pm \sqrt{1 - \frac{M_\times^2}{M_+^2} \frac{1-\cos(ka)}{2}} \right]. \tag{5.23}$$

At the zone boundary the frequencies for the acoustic and the optical branch are $\omega_{X,1} = \sqrt{2C/M_1}$ with $v_2 = 0$ and $\omega_{X,2} = \sqrt{2C/M_2}$ with $v_1 = 0$, respectively

(assuming $M_2 < M_1$). In the vibration for $k = \pi/a$ thus only one atom species oscillates, the other does not move. Close to the Γ point the atoms are in phase in the acoustic branch, i.e. $v_1 = v_2$. For the optical branch, the frequency at the Γ point is given by $\omega = \sqrt{2C/M_r}$ (with the reduced mass $M_r^{-1} = M_1^{-1} + M_2^{-1} = 2M_+/M_\times^2$) and the amplitude ratio is given by the mass ratio: $v_2 = -(M_1/M_2)\, v_1$, i.e. the heavier atom has the smaller amplitude.

5.2.3 *Lattice Vibrations of a Three-Dimensional Crystal*

When calculations are executed for a three-dimensional crystal with a monoatomic base, there are $3N$ equations of motion. These are transformed to normal coordinates and represent 3 acoustic branches (1 LA phonon mode and 2 TA phonon modes) of the dispersion relation. In a crystal with a base with p atoms, there are also 3 acoustic branches and $3(p-1)$ optical branches. For a diatomic base (as in the zincblende structure) there are 3 optical phonon branches (1 LO phonon mode and 2 TO phonon modes). The total number of modes is $3p$. The dispersion $\omega(\mathbf{k})$ now has to be calculated for all *directions* of $\mathbf{k}$.

In Figs. 5.5 and 5.6, the phonon dispersion in silicon, GaAs and GaP is shown along particular lines in the Brillouin zone (cf. Fig. 3.34b). The degeneracy of the LO and TO energies at the Γ point for the covalent group-IV semiconductor is lifted for the III–V semiconductors due to the ionic character of the bond and the macroscopic electric field connected with the long-wavelength LO phonon (see Sect. 5.2.9). Comparing GaAs [329] and GaP [330], the quite different mass of Ga- and P-atoms ($M_\times/M_+ \approx 0.92$) leads to the formation of a clear gap between the acoustic and optical branches, while for GaAs $M_\times/M_+ \approx 0.9994$ is close to 1 and no gap forms.

We note that the degeneracy of the TA phonon is lifted for propagation along the $\langle 110\rangle$ directions (Σ) because the two transverse directions $\langle 001\rangle$ and $\langle 1\bar{1}0\rangle$ are not equivalent.

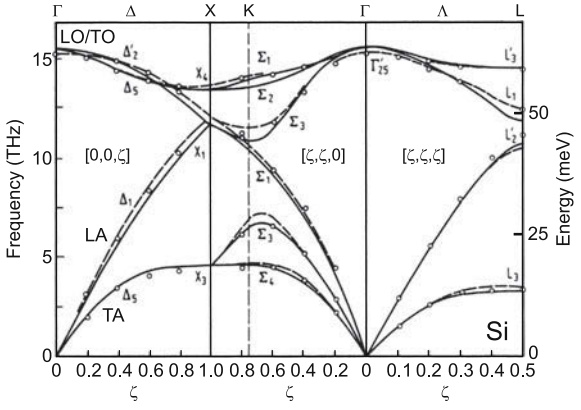


Fig. 5.5 Phonon dispersion in Si, experimental data and theory (*solid lines*: bond charge model, *dashed lines*: valence force field model). Adapted from [146]

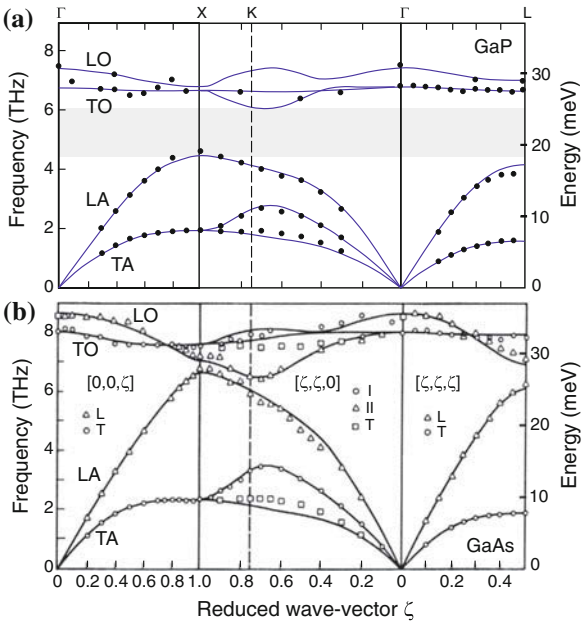


Fig. 5.6 Phonon dispersion in (**a**) GaP and (**b**) GaAs. Experimental data (*symbols*) and theory (*solid lines*, 14-parameter shell model). 'L' and 'T' refer to longitudinal and transverse modes, respectively. 'I' and 'II' (along $[\zeta, \zeta, 0]$) are modes whose polarization is in the $(1, \bar{1}, 0)$ plane. The *grey area* in (**a**) denotes the gap between acoustical and optical states. (**a**) Adapted from [331], (**b**) adapted from [329]

In boron nitride the masses of the two constituents are so similar that no gap exists between acoustical and optical branches (Fig. 5.7). Also the density of states (averaged over the entire Brillouin zone) is depicted (see next Sect.).

The displacement of atoms is shown in Fig. 5.8 for the different phonon modes present in zincblende crystals and in Fig. 5.9 for wurtzite crystals. The modes are labeled with their symmetry (in molecular notation) according to group theory (see remark in Sect. 6.2.5).

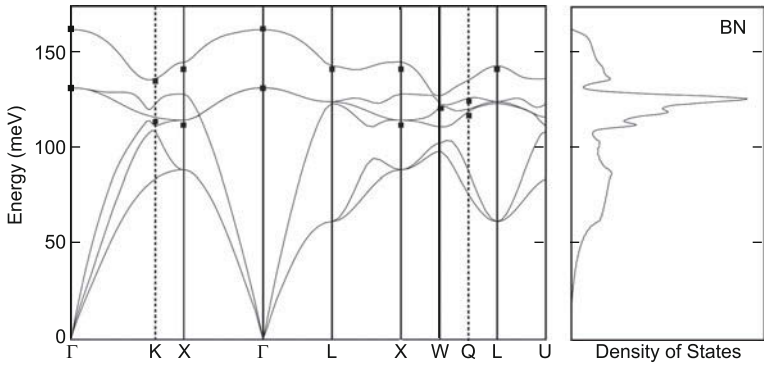


Fig. 5.7 Phonon dispersion in BN (*left panel*), experimental data (*symbols*) and theory (*solid lines*, first principles pseudopotential model). In the *right panel* the density of states is depicted. Adapted from [332]

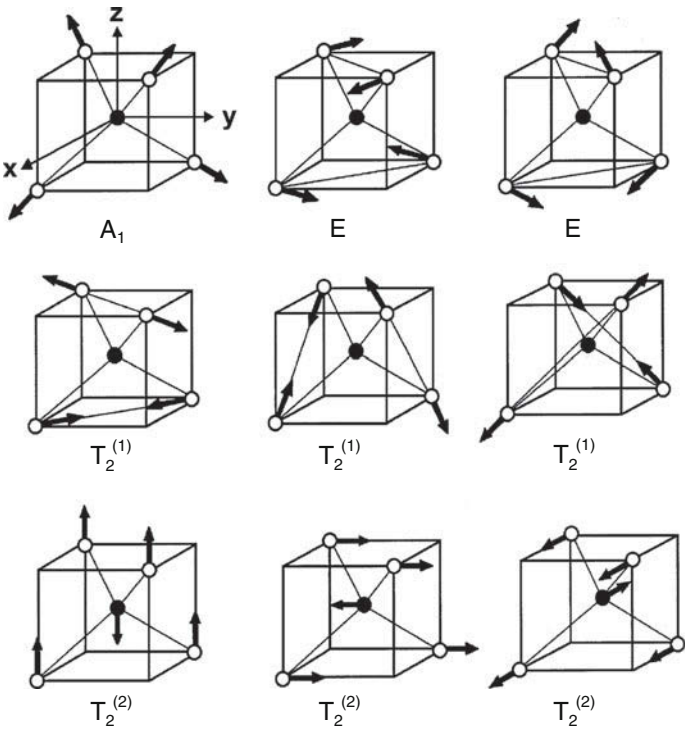


Fig. 5.8 Displacement of atoms for various phonon modes in zincblende crystals. Adapted from [333]

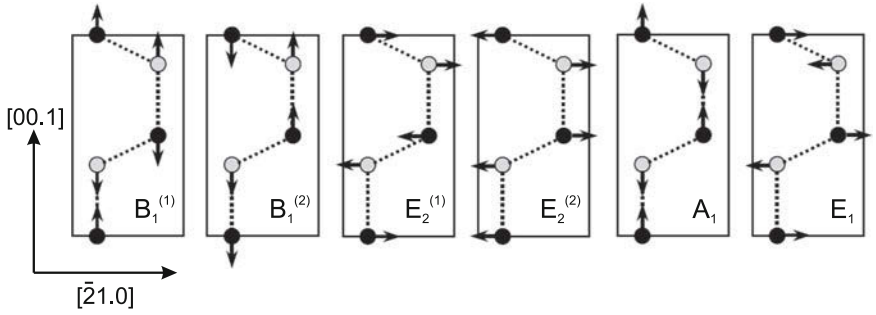


Fig. 5.9 Displacement of atoms for various phonon modes in wurtzite crystals. Adapted from [334]

The dependence of the phonon frequency on the mass of the atoms ($\propto 1/\sqrt{M}$) can be demonstrated with the isotope effect, visualized for GaAs in Fig. 5.10. The dependence of the phonon frequencies on the stiffness of the spring can be seen from Fig. 5.11; the smaller lattice constant provides the stiffer spring.

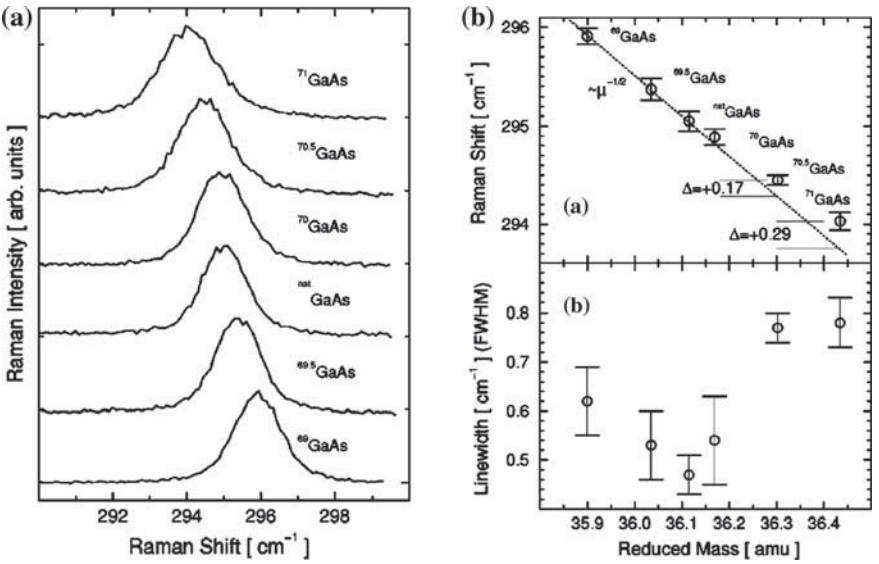


Fig. 5.10 (**a**) Raman spectra of GaAs with different isotope content as labeled. (**b**) Energy of optical phonons in GaAs with different isotope content [using the Raman spectra shown in (**a**)]. Reprinted with permission from [335], © 1999 APS

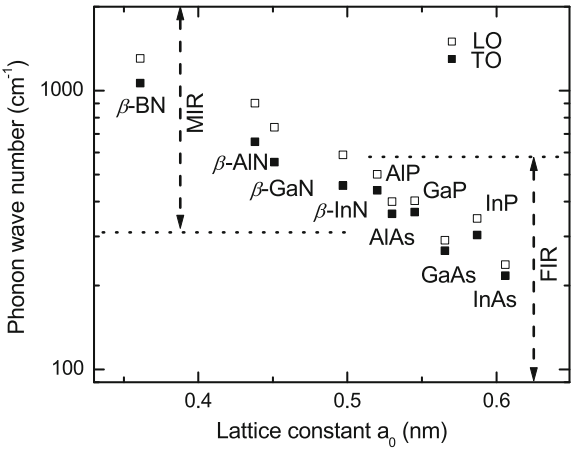


Fig. 5.11 Optical phonon frequencies (TO: *filled squares*, LO: *empty squares*) for a number of III–V compounds with different lattice constant a_0. 1 meV corresponds to 8.065 wave numbers (or cm^{-1}). Adapted from [336]

5.2.4 *Density of States*

The *density of states* (DOS) tells how many of the total $3pN$ modes are in a given energy interval. The states are spaced equally in **k**-space but not on the energy scale (see also Sect. 6.11).

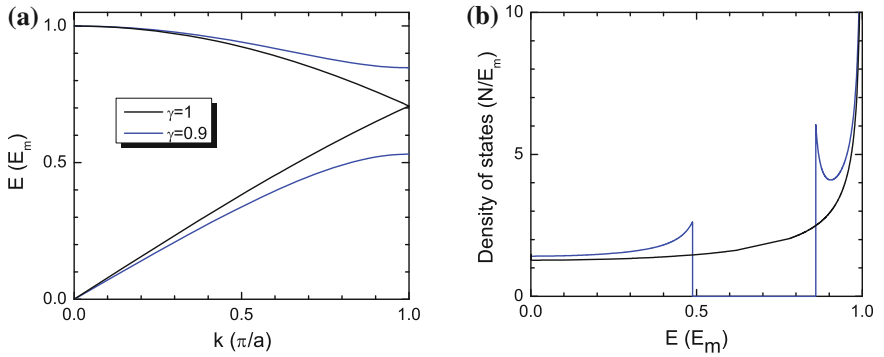


Fig. 5.12 (**a**) Phonon dispersion for the diatomic linear chain model for $\gamma = 1$ (*black line*) and $\gamma = 0.9$ (*blue lines*). (**b**) Corresponding density of states (in units of N/E_m)

For the monoatomic linear chain model, the number of states $N(E')$ from $E = 0$ up to $E = \hbar\omega = E'(k')$ for the dispersion of the acoustic phonons (5.8) is given as

$$N(E') = k' \frac{N}{\pi/a} = \frac{L}{\pi} k'. \tag{5.24}$$

Using (5.8), we find for one polarization ($E_\mathrm{m} = \hbar\omega_\mathrm{m}$)

$$N(E) = \frac{2N}{\pi} \arcsin\left(\frac{E}{E_\mathrm{m}}\right). \tag{5.25}$$

The DOS $D(E)$ is given by

$$D(E) = \frac{\mathrm{d}N(E)}{\mathrm{d}E} = \frac{2\,N}{\pi\,E_\mathrm{m}} \frac{1}{\sqrt{1-(E/E_\mathrm{m})^2}}. \tag{5.26}$$

Often the density of states is scaled by the (irrelevant) system size and given per atom (D/N) or per volume (D/L^3), per area (D/L^2) or per length (D/L) for three-, two- or one-dimensional systems, respectively.

In the diatomic linear chain model, additionally the optical phonons contribute to the density of states. In Fig. 5.12 the phonon density of states is shown for $\gamma = 0.9$ and for comparison for $\gamma = 1$ (gapless phonon dispersion). For small wavevector, the density of states is $4N/(\pi\,E_\mathrm{m})$.[1] Within the gap the density of states vanishes. At the edges of the band gap the density of states is enhanced. The total number of states for both dispersions is the same.

[1] The factor 2 compared to (5.26) stems from the folded Brillouin zone compared to the monoatomic chain model.

In a three-dimensional solid, (5.24) is modified to (for three degenerate polarizations)

$$N(E') = \frac{4\pi}{3} \frac{3}{(2\pi/L)^3} k'^3, \tag{5.27}$$

taking into account all states within a sphere in **k**-space of radius k'. Assuming a linear dispersion $\omega = v_s\, k$, we obtain

$$N(E) = \frac{V}{2\pi^2} \frac{E^3}{\hbar^3\, v_s^3}. \tag{5.28}$$

Thus the density of states is proportional to E^2,

$$D(E) = \frac{3\,V}{2\pi^2} \frac{E^2}{\hbar^3\, v_s^3}. \tag{5.29}$$

This dependence is the base for Debye's law for the T^3 temperature dependence of the heat capacity. As realistic example the phonon density of states for BN is depicted next to the dispersion in Fig. 5.7.

5.2.5 *Phonons*

Phonons are the quantized quasi-particles of the lattice vibrations (normal modes). The energy of a phonon can take the discrete values of a harmonic oscillator

$$E_{ph} = \left(n + \frac{1}{2}\right) \hbar\omega, \tag{5.30}$$

where n denotes the quantum number of the state, which corresponds to the number of energy quanta $\hbar\omega$ in the vibration. The amplitude of the vibration can be connected to n via the following discussion. For the classical oscillation $u = u_0\, \exp \mathrm{i}(kx - \omega t)$ the space and time average for the kinetic energy yields

$$E_{\mathrm{kin}} = \frac{1}{2}\, \rho\, V\, \overline{\left(\frac{\partial u}{\partial t}\right)^2} = \frac{1}{8}\, \rho\, V\, \omega^2\, u_0^2, \tag{5.31}$$

where ρ is the density and V the volume of the (homogeneous) solid. The energy of the oscillation is split in half between kinetic and potential energy. From $2E_{\mathrm{kin}} = E_{\mathrm{ph}}$ we find

$$u_0^2 = \left(n + \frac{1}{2}\right) \frac{4\hbar}{\rho\, V\, \omega}. \tag{5.32}$$

The number of phonons with which a vibrational mode is populated is thus directly related to the classical amplitude square.

Phonons act with a momentum $\hbar\mathbf{k}$, the so-called crystal momentum. When phonons are created, destroyed or scattered the crystal momentum is conserved, except for an arbitrary reciprocal-space vector $\mathbf{G}$. Scattering with $\mathbf{G} = 0$ is called a *normal* process, otherwise (for $\mathbf{G} \neq 0$) it is called an *umklapp* process.

5.2.6 Localized Vibrational Modes

A defect in the crystal can induce localized vibrational modes (LVM). The defect can be a mass defect, i.e. one of the masses M is replaced by M_d, or the force constants in the neighborhood are modified to C_d. A detailed treatment can be found in [337]. LVM are discussed, e.g., in [338–340].

First we consider the LVM for the one-dimensional, monoatomic chain. If the mass at lattice point $i = 0$ is replaced by $M_\mathrm{d} = M + \Delta M$ ($\epsilon_\mathrm{M} = \Delta M/M$), the displacements are given by $u_i = AK^{|i|}$, A being an amplitude, with

$$K = -\frac{1+\epsilon_\mathrm{M}}{1-\epsilon_\mathrm{M}}, \tag{5.33}$$

and the defect phonon frequency ω_d is

$$\omega_\mathrm{d} = \omega_\mathrm{m}\sqrt{\frac{1}{1-\epsilon_\mathrm{M}^2}}. \tag{5.34}$$

A real frequency is obtained for $|\epsilon_\mathrm{M}| < 1$. ω_d is then higher than the highest frequency of the bulk modes $\omega_\mathrm{m} = \sqrt{4C/M}$ (5.11). For $\epsilon_\mathrm{M} < 0$, i.e. the mass of the defect is smaller than the mass of the host atoms, K is negative and $|K| < 1$. Thus, the displacement can be written as

$$u_i \propto (-|K|)^{|i|} = (-1)^{|i|}\,\exp\left(+|i|\log|K|\right). \tag{5.35}$$

The numerical value of the exponent is negative, thus the amplitude decreases exponentially from the defect and indeed makes a localized vibrational mode. For small mass $M_\mathrm{d} \ll M$ (5.34) yields approximately $\omega_\mathrm{d} = \sqrt{2C/M_\mathrm{d}}$. This approximation is the so-called one-oscillator model. Since typically the extension of the localized mode is only a few lattice constants, the picture of LVM remains correct for impurity concentrations up to $\sim 10^{18} - 10^{20}\,\mathrm{cm}^{-3}$. For higher concentrations the concept of alloy modes has to be invoked (cf. Sect. 5.2.7).

For the case of group-III or -V substitutional impurities in group-IV semiconductors the change in force constants (treated below) can be neglected to some extent. For silicon ($M = 28$) and germanium ($M = 73$) the effect of various substitutions is shown in Fig. 5.13.

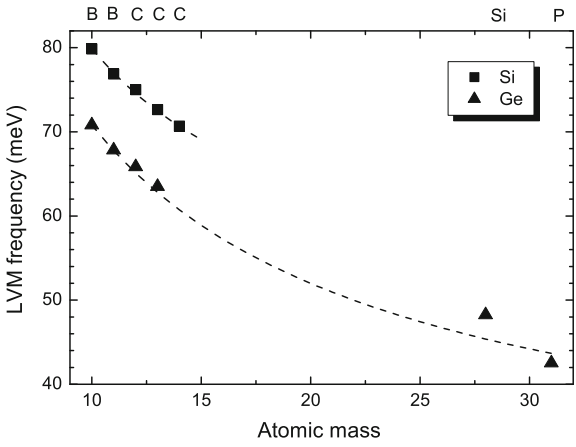


Fig. 5.13 Energy of local vibrational modes in Si and Ge. Experimental values at $T = 300\,\mathrm{K}$ (B in Ge: $T = 80\,\mathrm{K}$) taken from [337] and references therein and from [341] (C in Ge). The *dashed lines* are the mass dependence according to (5.34) scaled to the experimental frequency of the ^{10}B LVM

Now, additionally the force constants left and right of the defect are replaced by $C_\mathrm{d} = C + \Delta C$ ($\epsilon_\mathrm{C} = \Delta C/C$). The displacements are still given by $u_i = AK^{|i|}$, now with

$$K = -\frac{(1+\epsilon_\mathrm{M})\,(1+\epsilon_\mathrm{C})}{1-\epsilon_\mathrm{M}-2\epsilon_\mathrm{C}}. \tag{5.36}$$

An exponential decrease of the LVM amplitude occurs for negative K that is ensured for $\epsilon_\mathrm{M} + 2\epsilon_\mathrm{C} < 0$ (and $\epsilon_\mathrm{M} > -1$ and $\epsilon_\mathrm{C} > -1$). The defect frequency is given by

$$\omega_\mathrm{d} = \omega_\mathrm{m}\sqrt{\frac{(1+\epsilon_\mathrm{C})\,(2+\epsilon_\mathrm{C}\,(3+\epsilon_\mathrm{M}))}{2(1+\epsilon_\mathrm{M})\,(2\epsilon_\mathrm{C}+1-\epsilon_\mathrm{M})}}. \tag{5.37}$$

We note that for $\epsilon_\mathrm{C} = 0$ (5.33) and (5.34) are recovered.

For a given mass defect, the change of frequency with ΔC is (in linear order, i.e. for $\epsilon_\mathrm{C} \ll 1$)

$$\frac{\partial\omega_\mathrm{d}(\epsilon_\mathrm{M},\epsilon_\mathrm{C})}{\partial\epsilon_\mathrm{C}} = \frac{1-4\,\epsilon_\mathrm{M}-\epsilon_\mathrm{M}^2}{4\,(1-\epsilon_\mathrm{M})\sqrt{1-\epsilon_\mathrm{M}^2}}\,\epsilon_\mathrm{C}. \tag{5.38}$$

The linear coefficient diverges for $\epsilon_\mathrm{M} \to -1$. For ϵ_M between -0.968 and 0 the linear coefficient varies between 2 and 1/4. Therefore, a larger force constant ($\epsilon_\mathrm{C} > 0$) increases the LVM frequency of the defect, as expected for a stiffer spring.

In a binary compound the situation is more complicated. We assume here that the force constants remain the same and only the mass of the substitution atom M_d is different from the host. The host has the atom masses M_1 and M_2 with $M_1 < M_2$. Substitution of the heavy atom with a lighter one creates a LVM above the optical branch for $M_\mathrm{d} < M_2$. Additionally, a level in the gap between the optical and acoustic

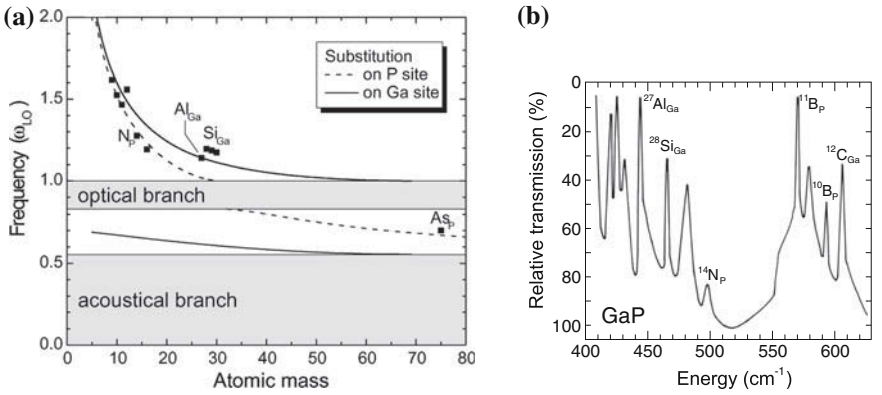


Fig. 5.14 (**a**) Numerical simulation of a linear chain model for GaP ($M_1 = 31$, $M_2 = 70$). Energy of local vibrational modes (*dashed* (*solid*) *line*): substitution on P (Ga) site in units of the optical phonon frequency at Γ ($\omega_m = 45.4$ meV). The *grey areas* indicate the acoustic and optical phonon bands. *Solid squares* are experimental data (from [337]), scaled to the theoretical curve for the $^{27}Al_{Ga}$ LVM frequency. (**b**) Differential transmission spectrum of GaP structure (nitrogen-doped layer on zinc-doped compensated substrate) against pure crystal ($T = 77$ K). Data from [342]

branch is induced. Such LVM is called a *gap mode*. Substitution of the lighter atom of the binary compound induces a LVM above the optical branch for $M_d < M_1$. A gap mode is induced for $M_d > M_1$. The situation for GaP is depicted in Fig. 5.14. LVM in GaAs have been reviewed in [338].

The energy position of a local vibrational mode is sensitive to the isotope mass of the surrounding atoms. In Fig. 5.15, a high-resolution ($0.03\,cm^{-1}$) spectrum of the $^{12}C_{As}$ LVM in GaAs is shown together with a theoretical simulation. The various theoretical peak positions are given as vertical bars, their height indicating the oscillator

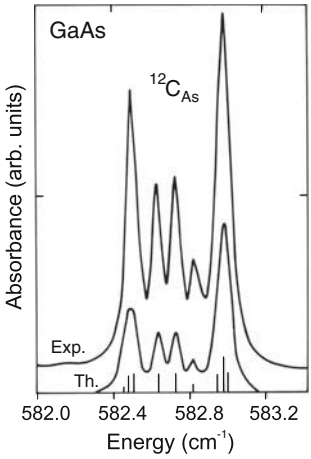


Fig. 5.15 Experimental (Exp., $T = 4.2$ K, resolution $0.03\,cm^{-1}$) and theoretical (Th., artificial Lorentzian broadening) infrared spectra of LVM of $^{12}C_{As}$ in GaAs. The positions and oscillator strengths of the theoretical transitions involving different configurations with ^{69}Ga and ^{71}Ga isotopes are shown as *vertical bars*. Data from [338]

strength. Five experimental peaks are obvious that are due to a total of nine different transitions. The C atom can experience five different surroundings (see Table 3.8) with the four neighbors being ^{69}Ga or ^{71}Ga. The natural isotope mix is an 'alloy' $^{69}Ga_x{}^{71}Ga_{1-x}As$ with $x = 0.605$. The configurations with T_d symmetry contribute one peak each, the lowest (^{71}Ga surrounding) and highest (^{69}Ga surrounding) energy transitions. The configurations with C_{3v} and C_{2v} symmetry contribute each with 2 and 3 nondegenerate modes, respectively.

The vibrations of impurity complexes have been discussed in [343].

5.2.7 *Phonons in Alloys*

In an alloy of the type $AB_{1-x}C_x$ the phonon frequencies will depend on the ternary composition. For the binary end materials AB and AC clearly TO and LO frequencies exist. The simplest behavior of the alloy is the one-mode behavior (Fig. 5.16d) where the mode frequencies vary continuously (and approximately linearly) with the composition. The oscillator strength (LO–TO splitting, (9.74)) remains approximately constant. In many cases, the two-mode behavior is observed where the LO–TO gap closes (accompanied by decreasing oscillator strength) and a localized vibrational mode and a gap mode occur for the binary end materials (Fig. 5.16a). Also, a mixed-mode behavior (Fig. 5.16b, c) can occur.

The masses of the three constituent atoms will be M_A, M_B, and M_C. Without limiting the generality of our treatment, we assume $M_B < M_C$. From the considerations in Sect. 5.2.6 on LVM and gap modes, the condition

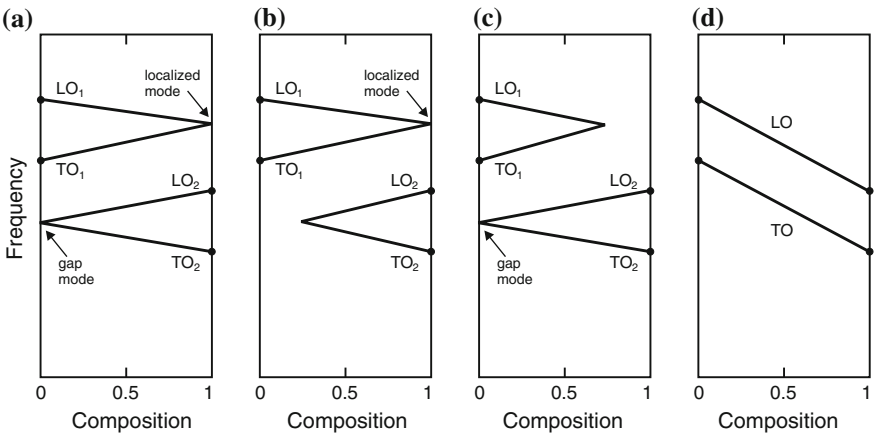


Fig. 5.16 Schematic behavior of phonon modes in an alloy. (**a**) Two-mode behavior with gap mode and localized mode, (**b**, **c**) mixed-mode behavior, (**b**) only localized mode allowed, (**c**) only gap mode allowed, (**d**) one-mode behavior with neither localized mode nor gap mode allowed

Table 5.1 Atomic masses of the constituents of various ternary compounds, reduced mass μ_{AC} (5.40), fulfillment of the relation from (5.40) ('+': fulfilled, '−': not fulfilled) and experimental mode behavior ('2': two-mode, '1': one-mode)

Alloy	A	B	C	M_A	M_B	M_C	μ_{AC}	Rel.	Modes
$GaP_{1-x}As_x$	Ga	P	As	69.7	31.0	74.9	36.1	+	2
$GaAs_{1-x}Sb_x$	Ga	As	Sb	69.7	74.9	121.8	44.3	−	1
$CdS_{1-x}Se_x$	Cd	S	Se	112.4	32.1	79.0	46.4	+	2
$Cd_xZn_{1-x}S$	S	Zn	Cd	32.1	65.4	112.4	25.0	−	1
$Mg_xZn_{1-x}O$	O	Mg	Zn	16.0	24.3	65.4	12.9	−	1

$$M_B < M_A, M_C \tag{5.39}$$

for two-mode behavior can be deduced. This ensures a LVM of atom B in the compound AC and a gap mode of atom C in the compound AB. However, it turns out that this condition is not sufficient, e.g. $Na_{1-x}K_xCl$ fulfills (5.39) but exhibits one-mode behavior. From a modified REI[2] model (for $\mathbf{k} \sim 0$ modes) it has been deduced that

$$M_B < \mu_{AC} = \frac{M_A\, M_C}{M_A + M_C} < M_A, M_C \tag{5.40}$$

is a necessary and sufficient condition (unless the force constants between A–B and A–C are significantly different) for two-mode behavior [344]. A detailed discussion is given in [345]. Equation (5.40) is a stronger condition than the previous one (5.39). If (5.40) is not fulfilled the compound exhibits one-mode behavior. As an example, we show the mass relations for $CdS_{1-x}Se_x$ and $Cd_xZn_{1-x}S$ in Table 5.1 and the experimental phonon energies in Fig. 5.17. Also in Table 5.1 the masses for $GaP_{1-x}As_x$ ($GaAs_{1-x}Sb_x$) exhibiting two- (one-) mode behavior are shown.

If the binary end components of a ternary alloy have different crystal structure, a transition between the two occurs which is reflected in the phonon structure (energies and mode symmetries). As an example, the optical phonon energies of $Mg_xZn_{1-x}O$ are depicted in Fig. 5.18 (cmp. Fig. 3.39).

5.2.8 *Disorder*

An example of 'small' disorder are the localized vibrational modes due to a single defect. Here we consider in our one-dimensional model random fluctuations of the model parameters. To that avail we set up a numerical implementation of an one-dimensional chain with masses $M_1 = M_2$ and spring constants $C_1 \neq C_2$, here $C_2 = 2\,C_1$. Now each spring constant varies randomly by $\pm\xi$ percent. The density

[2] Random element isodisplacement.

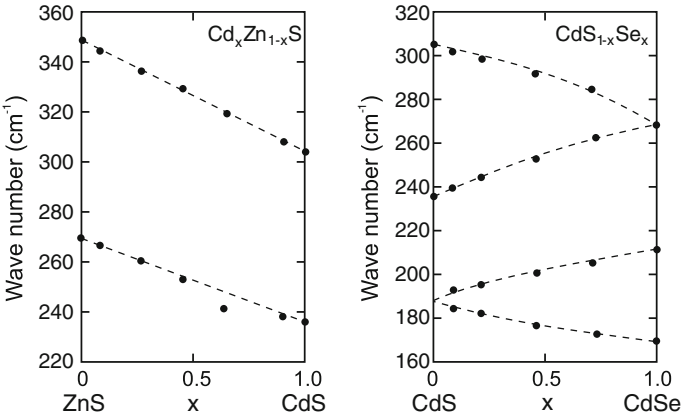


Fig. 5.17 Phonon energies of $Cd_xZn_{1-x}S$ and $CdS_{1-x}Se_x$ as a function of the ternary composition. Experimental data (*solid circles*) are from [344], *dashed lines* are guides to the eye

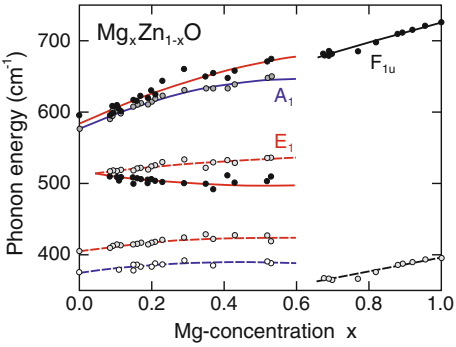


Fig. 5.18 LO (*solid lines*) and TO (*dashed lines*) phonon energies of $Mg_xZn_{1-x}O$ in the wurtzite structure (A_1 symmetry: *blue lines*, E_1 symmetry: *red lines*) and in the rocksalt phase (F_{1u} symmetry: *black lines*). Experimental data are shown as *symbols*. Adapted from data of [346]

of states is displayed for $\xi = 10, 20, 40$ and 60 %. The effects as shown in Fig. 5.19 are broadening of peaks in the DOS, broadening of the band edges, the development of band tails into the gap and eventually a closing of the gap. This is a typical behavior that also exists for electronic states (cmp. Fig. 6.46).

5.2.9 *Electric Field Created by Optical Phonons*

Adjacent atoms oscillate with opposite phase in an optical phonon. If the bond has (partial) ionic character, this leads to a time-dependent polarization and subsequently to a macroscopic electric field. This additional field will influence the phonon frequencies obtained from a purely mechanical approach. We consider in the following the case $\mathbf{k} \approx 0$. The phonon frequency for TO and LO vibrations is given by

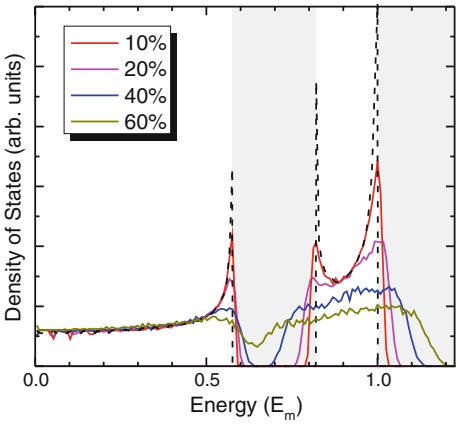


Fig. 5.19 Density of states versus energy (in units of maximum phonon energy $E_\mathrm{m} = \hbar\omega_\mathrm{m}$) for diatomic linear chain model ($M_1 = M_2$, $C_2 = 2\,C_1$, 2^9 atoms, average over 2^7 configurations) for various levels of random relative variations of the spring constants (*solid lines*). As *dashed lines* the density of states of the perfect chain is depicted (cmp. Fig. 5.12b) with the forbidden energy ranges shown in *light grey*

$$\omega_0 = \sqrt{\frac{2\,C}{M_\mathrm{r}}}, \tag{5.41}$$

where M_r is the reduced mass of the two different atoms (cf. Sect. 5.2.2). $\mathbf{u}$ is the relative displacement $\mathbf{u}_1 - \mathbf{u}_2$ of the two atoms in a diatomic base. When the interaction with the electric field $\mathbf{E}$ (which will be calculated self-consistently in the following) is considered, the Hamiltonian for the long-wavelength limit is given by [347]:

$$\hat{H}(\mathbf{p}, \mathbf{u}) = \frac{1}{2}\left(\frac{1}{M_\mathrm{r}}\,\mathbf{p}^2 + b_{11}\,\mathbf{u}^2 + 2b_{12}\,\mathbf{u}\cdot\mathbf{E} + b_{22}\,\mathbf{E}^2\right). \tag{5.42}$$

The first term is the kinetic energy ($\mathbf{p}$ stands for the momentum of the relative motion of the atoms 1 and 2 in the base, $\mathbf{p} = M_\mathrm{r}\dot{\mathbf{u}}$), the second the potential energy, the third the dipole interaction and the fourth the electric-field energy. The equation of motion for a plane wave $\mathbf{u} = \mathbf{u}_0 \exp\left[-\mathrm{i}(\omega t - \mathbf{k}\cdot\mathbf{r})\right]$ ($\ddot{\mathbf{u}} = -\omega^2\mathbf{u}$) yields

$$M_\mathrm{r}\,\omega^2\,\mathbf{u} = b_{11}\,\mathbf{u} + b_{12}\,\mathbf{E}. \tag{5.43}$$

Thus, the electric field is

$$\mathbf{E} = (\omega^2 - \omega_\mathrm{TO}^2)\,\frac{M_\mathrm{r}}{b_{12}}\,\mathbf{u}. \tag{5.44}$$

Here, the substitution $\omega_\mathrm{TO}^2 = b_{11}/M_\mathrm{r}$ was introduced that is consistent with (5.41) and $b_{11} = 2C$. ω_TO represents the mechanical oscillation frequency of the atoms

undisturbed by any electromagnetic effects. Already now the important point is visible. If ω approaches ω_{TO}, the system plus electric field oscillates with the frequency it has without an electric field. Therefore the electric field must be zero. Since the polarization $\mathbf{P} = (\epsilon - 1)\epsilon_0 \mathbf{E}$ is finite, the dielectric constant ϵ thus diverges.

The polarization is

$$\mathbf{P} = -\nabla_{\mathbf{E}} \hat{H} = -\left(b_{12}\,\mathbf{u} + b_{22}\,\mathbf{E}\right). \tag{5.45}$$

The displacement field is

$$\mathbf{D} = \epsilon_0\,\mathbf{E} + \mathbf{P} = \epsilon_0\,\mathbf{E} - \left(b_{22} - \frac{b_{12}^2/M_{\mathrm{r}}}{\omega_{\mathrm{TO}}^2 - \omega^2}\right)\mathbf{E} = \epsilon_0\,\epsilon(\omega)\,\mathbf{E}. \tag{5.46}$$

Therefore, the dielectric function is

$$\epsilon(\omega) = \epsilon(\infty) + \frac{\epsilon(0) - \epsilon(\infty)}{1 - (\omega/\omega_{\mathrm{TO}})^2}. \tag{5.47}$$

Here, $\epsilon(\infty) = 1 - b_{22}/\epsilon_0$ is the high-frequency dielectric constant and $\epsilon(0) = \epsilon(\infty) + b_{12}^2/(b_{11}\epsilon_0)$ the static dielectric constant. The relation (5.47) is shown in Fig. 5.20.

From the Maxwell equation $\nabla \cdot \mathbf{D} = 0$ for zero free charge we obtain the relation

$$\epsilon_0\,\epsilon(\omega)\,\nabla \cdot \mathbf{E} = 0. \tag{5.48}$$

Thus, either $\epsilon(\omega) = 0$ or $\nabla \cdot \mathbf{E} = 0$, i.e. $\mathbf{u}$ is perpendicular to $\mathbf{k}$. In the latter case we have a TO phonon and, neglecting retardation effects, using $\nabla \times \mathbf{E} = 0$ we find

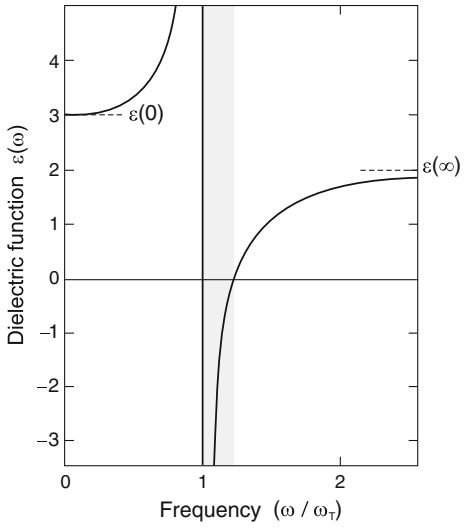


Fig. 5.20 Dielectric function according to (5.47) with $\epsilon(0) = 3$ and $\epsilon(\infty) = 2$ (without damping). *Grey area* denotes the region of negative ϵ

$\mathbf{E} = 0$ and therefore $\omega = \omega_{\mathrm{TO}}$, justifying our notation. In the case of $\epsilon(\omega) = 0$, we call the related frequency ω_{LO} and find the so-called Lyddane–Sachs–Teller (LST) relation

$$\frac{\omega_{\mathrm{LO}}^2}{\omega_{\mathrm{TO}}^2} = \frac{\epsilon(0)}{\epsilon(\infty)}. \tag{5.49}$$

This relation holds reasonably well for optically isotropic, heteropolar materials with two atoms in the basis, such as NaI and also GaAs. Since at high frequencies, i.e. $\omega \gg \omega_{\mathrm{TO}}$, only the individual atoms can be polarized, while for low frequencies the atoms can also be polarized against each other, $\epsilon(0) > \epsilon(\infty)$ and therefore also $\omega_{\mathrm{LO}} > \omega_{\mathrm{TO}}$. For GaAs, the quotient of the two phonon energies is 1.07. Using the LST relation (5.49), we can write for the dielectric function

$$\epsilon(\omega) = \epsilon(\infty) \left(\frac{\omega_{\mathrm{LO}}^2 - \omega^2}{\omega_{\mathrm{TO}}^2 - \omega^2} \right). \tag{5.50}$$

The (long-wavelength) TO-phonon does not create a long-range electric field. Using $\nabla \cdot \mathbf{D} = 0$ and (5.46) and looking at the longitudinal fields, we have

$$\epsilon_0 \, \mathbf{E} = b_{12} \, \mathbf{u} + b_{22} \, \mathbf{E}. \tag{5.51}$$

This can be rewritten as

$$\mathbf{E} = -\omega_{\mathrm{LO}} \sqrt{\frac{M_{\mathrm{r}}}{\epsilon_0}} \sqrt{\frac{1}{\epsilon(\infty)} - \frac{1}{\epsilon(0)}} \, \mathbf{u} \propto -\mathbf{u}. \tag{5.52}$$

The (long-wavelength) LO-phonon thus creates a long-range electric field acting *against* the ion displacement and represents an additional restoring force; this is consistent with the fact that $\omega_{\mathrm{LO}} > \omega_{\mathrm{TO}}$.

5.3 Elasticity

The elastic properties of the semiconductor are important if the semiconductor is subjected to external forces (pressure, temperature) or to lattice mismatch during heteroepitaxy.

5.3.1 Thermal Expansion

The lattice constant depends on temperature. The (linear) thermal expansion coefficient is defined as

$$\alpha(T_0) = \left. \frac{\partial a_0(T)}{\partial T} \right|_{T=T_0} \tag{5.53}$$

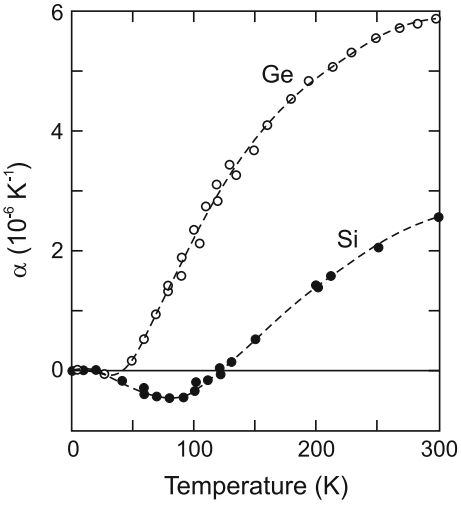


Fig. 5.21 Linear thermal expansion coefficient of silicon (*solid circles*) and germanium (*open circles*). Adapted from [348] based on experimental data from various sources. *Dashed lines* are guide to the eyes

and is temperature dependent. The temperature dependence of α for silicon and germanium is shown in Fig. 5.21. α is approximately proportional to the heat capacity (C_V) except at low temperatures. The negative values are due to negative Grüneisen parameters [348]. These anharmonicity effects are discussed in detail in [328].

5.3.2 Stress–Strain Relation

In this section, we recall the classical theory of elasticity [349]. The solid is treated as a continuous medium (piecewise homogeneous) and the displacement vector is thus a continuous function $\mathbf{u}(\mathbf{r})$ of the spatial coordinates. When the spatial variation $\nabla\mathbf{u}$ of $\mathbf{u}$ is small, the elastic energy can be written as

$$U = \frac{1}{2}\int \frac{\partial u_l}{\partial x_k}\, C_{klmn}\, \frac{\partial u_n}{\partial x_m}\, \mathrm{d}^3\mathbf{r}, \tag{5.54}$$

where $\mathbf{C}$ is the (macroscopic) tensor of the elastic coefficients. 21 components of this tensor can be independent. For crystals with cubic symmetry the number of independent constants is reduced to 3. An exchange $k \leftrightarrow l$ and $m \leftrightarrow n$ does not matter, only six indices have to be considered (xx, yy, zz, yz, xz, and xy). The strain components ϵ_{ij} are symmetrized according to

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j}\right). \tag{5.55}$$

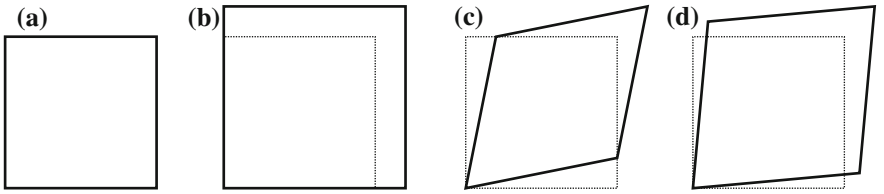


Fig. 5.22 Deformation of a square (**a**). (**b**) Pure hydrostatic deformation ($\epsilon_{xx} = \epsilon_{yy} = 0.2$, $\epsilon_{xy} = 0$), (**c**) pure shear deformation ($\epsilon_{xx} = \epsilon_{yy} = 0$, $\epsilon_{xy} = 0.2$), and (**d**) mixed deformation ($\epsilon_{xx} = \epsilon_{yy} = 0.1$, $\epsilon_{xy} = 0.1$)

The strains ϵ_{xx} are along the main axes of the crystal as visualized in Fig. 5.22. The stresses[3] σ_{kl} are then given by

$$\sigma_{kl} = C_{klmn}\,\epsilon_{mn}. \tag{5.56}$$

The inverse relation is mediated by the stiffness tensor **S**.

$$\epsilon_{kl} = S_{klmn}\,\sigma_{mn}. \tag{5.57}$$

Typically, the strain components e_{ij} or e_i are used with the convention $xx \to 1$, $yy \to 2$, $zz \to 3$, $yz \to 4$, $xz \to 5$, and $xy \to 6$ (Voigt notation):

$$e_{ij} = \epsilon_{ij}\,(2 - \delta_{ij}). \tag{5.58}$$

Then, $\sigma_m = C_{mn}e_n$ with the C_{ij} being the elastic constants. The x, y, and z directions are the main axes of the cubic solid, i.e. the $\langle 100 \rangle$ directions.

For *zincblende* material, the stress–strain relation reads[4]

$$\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \sigma_4 \\ \sigma_5 \\ \sigma_6 \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{pmatrix} \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{pmatrix}. \tag{5.59}$$

Values of the compliances for several semiconductors are given in Table. 5.2. The inverse relation is given by the matrix

[3] The stress is a force per unit area and has the dimensions of a pressure.

[4] $C_{11} = C_{1111}$, $C_{12} = C_{1122}$ and $C_{44} = C_{1212} = C_{1221} = C_{2121} = C_{2112}$.

Table 5.2 Elastic constants (in GPa) of some cubic semiconductors at room temperature. I_K refers to the Keating criterion (5.63)

Material	C_{11}	C_{12}	C_{44}	I_K
C	1076.4	125.2	577.4	1.005
Si	165.8	63.9	79.6	1.004
Ge	128.5	48.3	66.8	1.08
BN	820	190	480	1.11
GaAs	119	53.4	59.6	1.12
InAs	83.3	45.3	39.6	1.22
AlAs	120.5	46.86	59.4	1.03
ZnS	104.6	65.3	46.3	1.33
MgO	297	156	95.3	0.80

$$\begin{pmatrix} S_{11} & S_{12} & S_{12} & 0 & 0 & 0 \\ S_{12} & S_{11} & S_{12} & 0 & 0 & 0 \\ S_{12} & S_{12} & S_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & S_{44} \end{pmatrix}, \tag{5.60}$$

with the stiffness coefficients in this notation given by

$$S_{11} = \frac{C_{11} + C_{12}}{(C_{11} - C_{12})\,(C_{11} + 2C_{12})} \tag{5.61a}$$

$$S_{12} = \frac{C_{12}}{-C_{11}^2 - C_{11}C_{12} + 2C_{12}^2} \tag{5.61b}$$

$$S_{44} = \frac{1}{C_{44}}. \tag{5.61c}$$

We emphasize that in this convention (also called the *engineering* convention), e.g. $e_1 = \epsilon_{xx}$ and $e_4 = 2\,\epsilon_{yz}$. There is also another convention (the *physical* convention) without this factor of two; in this case the matrix in (5.59) contains the elements $2C_{44}$. We introduce

$$C_0 = 2\,C_{44} + C_{12} - C_{11}, \tag{5.62}$$

and note that $C_0 = 0$ for an isotropic material. The relation

$$I_K = \frac{2\,C_{44}\,(C_{11} + C_{12})}{(C_{11} - C_{12})\,(C_{11} + 3C_{12})} = 1 \tag{5.63}$$

known as the Keating criterion [350, 351], stems from the consideration of bending and stretching of the tetrahedral bonds in the valence force field (VFF) model. It is closely fulfilled (Table 5.2) for many tetrahedrally bonded semiconductors, in particular for the covalent ones. For MgO, the Keating criterion is not fulfilled because it has (six-fold coordinated) rocksalt structure and is thus not tetrahedrally bonded.

The Young's modulus Y,

$$\sigma_{\mathbf{nn}} = Y(\mathbf{n})\,\epsilon_{\mathbf{nn}}, \tag{5.64}$$

generally depends on the normal direction $\mathbf{n}$ of a strain. It is equivalent to $1/S_{11}$ of (5.61a).

For *isotropic* material Y and the Poisson ratio ν are related to the elastic constants of cubic material by

$$Y = C_{11} - \frac{2\,C_{12}^2}{C_{11} + C_{12}} \tag{5.65a}$$

$$\nu = \frac{C_{12}}{C_{11} + C_{12}}. \tag{5.65b}$$

For isotropic materials also Lamé's constants λ and μ are used. They are given by[5] $C_{11} = \lambda + 2\mu$, $C_{12} = \lambda$ and $C_{44} = \mu$ (note that C_0 according to (5.62) is zero).

The bulk modulus B (inverse of the compressibility),

$$\frac{1}{B} = -\frac{1}{V}\frac{\partial V}{\partial p}, \tag{5.66}$$

for the zincblende crystal is given as

$$B = \frac{C_{11} + 2\,C_{12}}{3}. \tag{5.67}$$

We note that Y, ν and C_{ij} of typical materials are both positive. Materials with negative Poisson ratio are called *auxetic* [352–354]. Also materials with negative compressibility are possible [355].

Beyond the dependence of the elastic constants on the bond length (as materialized in the phonon frequencies in Fig. 5.11), they depend on the ionicity. In Fig. 5.23, the elastic constants of various zincblende semiconductors are shown as a function of the ionicity f_i. The values for the elastic constants are normalized by e^2/a^4, a being the average nearest-neighbor distance.

For *wurtzite* crystals, five elastic constant are necessary for the stress–strain relation that reads[6]

$$C_{ij} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2}(C_{11} - C_{12}) \end{pmatrix}. \tag{5.68}$$

[5] For an isotropic material, $C_{ijkl} = \lambda\,\delta_{ij}\,\delta_{kl} + \mu\,(\delta_{ik}\,\delta_{jl} + \delta_{il}\,\delta_{jk})$.

[6] $(C_{11} - C_{12})/2 = C_{1212}$, $C_{44} = C_{1313} = C_{2323}$.

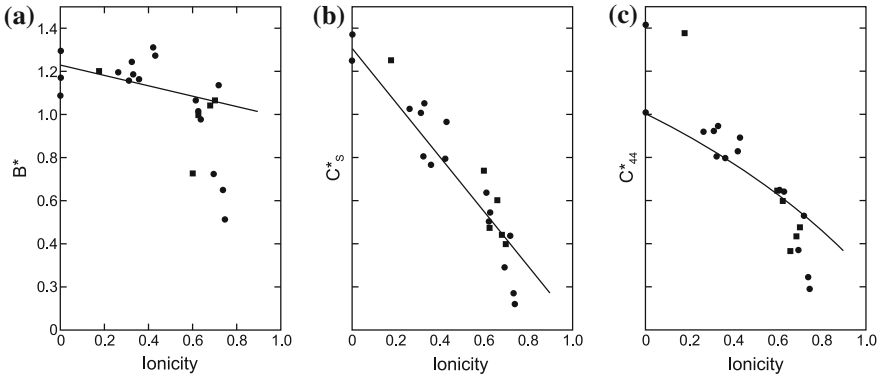


Fig. 5.23 Elastic constants as a function of ionicity for various semiconductors with diamond or zincblende (*circles*) and wurtzite (*squares*) structure. Constants are normalized by the modulus $C_0 = e^2/d^4$, d being the average nearest-neighbor distance. (**a**) Bulk modulus, $B^* = (C_{11} + 2C_{12})/(3C_0)$, (**b**, **c**) shear moduli, (**b**) $C_S^* = (C_{11} - C_{12})/C_0$, (**c**) $C_{44}^* = C_{44}/C_0$. *Solid lines* are a simple model as discussed in [356]. Adapted from [357]

Table 5.3 Elastic constants (in GPa) of some wurtzite semiconductors

Material	C_{11}	C_{12}	C_{13}	C_{33}	C_{44}	Reference
GaN	391	143	108	399	103	[359]
AlN	410	149	99	389	125	[360]
ZnS	124	60.2	45.5	140	28.6	[361]
ZnO	206	118	118	211	44	[362]

Experimental values for wurtzite materials are given in Table 5.3. The relation of the elastic tensor of wurtzite and zincblende materials, in particular viewed along the ⟨111⟩-direction has been discussed in [357, 358].

The bulk modulus of the wurtzite crystal is given by

$$B = \frac{(C_{11} + C_{12})\, C_{33} - 2\, C_{13}^2}{C_{11} + C_{12} + 2\, C_{33} - 4\, C_{13}}. \tag{5.69}$$

5.3.3 *Biaxial Strain*

In heteroepitaxy (cf. Sect. 12.2.6), a biaxial strain situation arises, i.e. layered material is compressed (or expanded in the case of tensile strain) in the interface plane and is expanded (compressed) in the perpendicular direction. Here, we assume that the substrate is infinitely thick, i.e. that the interface remains planar. Substrate bending is discussed in Sect. 5.3.5.

The simplest case is epitaxy on the (001) surface, i.e. $e_1 = e_2 = \epsilon_\parallel$. The component e_3 is found from the condition $\sigma_3 = 0$ (no forces in the z direction). All shear strains are zero. For zincblende material it follows

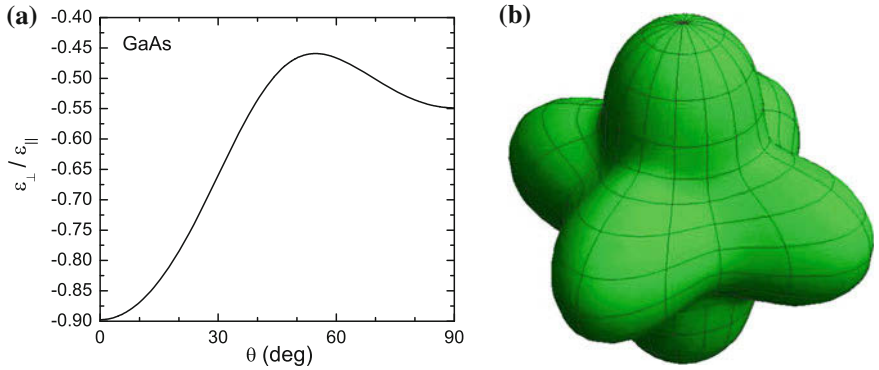


Fig. 5.24 Ratio $-\epsilon_\perp/\epsilon_\parallel$ for GaAs under symmetric biaxial strain. The angle θ denotes the surface normal in the $\langle 110\rangle$-azimuth ($\theta = 0$: [001], $\theta = 90°$: [110], the maximum of $\epsilon_\perp/\epsilon_\parallel$ is for [111]). (**b**) Is a three-dimensional visualization

$$\epsilon_\perp^{100} = e_3 = -\frac{C_{12}}{C_{11}}\,(e_1 + e_2) = -\frac{2\,C_{12}}{C_{11}}\,\epsilon_\parallel. \tag{5.70}$$

In Fig. 5.24 the ratio $\epsilon_\perp/\epsilon_\parallel$ is depicted for GaAs and various crystal orientations; the formulas for other orientations are more involved [363]:

$$\epsilon_\perp^{110} = -\frac{2C_{12} - C_0/2}{C_{11} + C_0/2}\,\epsilon_\parallel \tag{5.71}$$

$$\epsilon_\perp^{111} = -\frac{2C_{12} - 2C_0/3}{C_{11} + 2C_0/3}\,\epsilon_\parallel. \tag{5.72}$$

For wurtzite crystals and pseudomorphic growth along [00.1] the strain along the epitaxial direction (c-axis) is given by

$$\epsilon_\perp = -\frac{C_{13}}{C_{33}}\,(e_1 + e_2) = -\frac{2C_{13}}{C_{33}}\,\epsilon_\mathrm{a}, \tag{5.73}$$

where $\epsilon_\perp = \epsilon_\mathrm{c} = (c - c_0)/c_0$ and $\epsilon_\mathrm{a} = (a - a_0)/a_0$. For symmetric biaxial in-plane strain, the ratio $\epsilon_\perp/\epsilon_\parallel$ is shown in Fig. 5.25 for GaN and varying angle θ of the c-axis against the epitaxy direction. For the growth of wurtzite on wurtzite for $\theta \neq 0$, the epitaxial strain is actually *asymmetric* in the interface plane. For $\theta = 90°$, e.g. the epitaxy on m-plane substrate (cmp. Fig. 3.33) (c-axis lies in-plane), the in-plane strains are $e_1 = \epsilon_\mathrm{a}$ and $e_2 = \epsilon_\mathrm{c}$. For $\theta = 90°$, we find

$$\epsilon_\perp = -\frac{C_{12}\,\epsilon_\mathrm{a} + C_{13}\,\epsilon_\mathrm{c}}{C_{11}}. \tag{5.74}$$

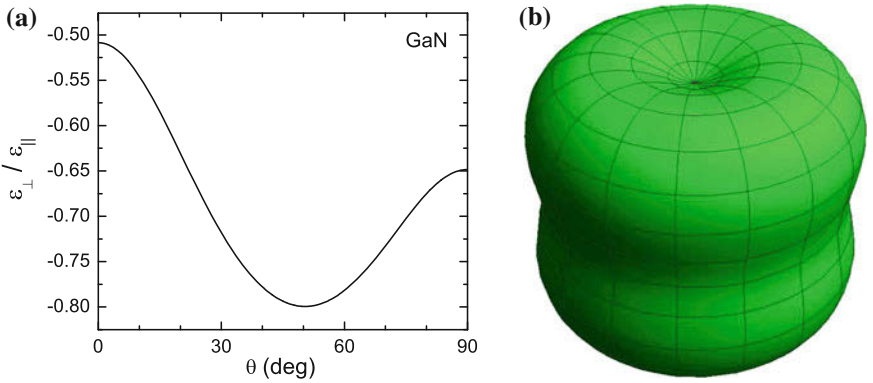


Fig. 5.25 Ratio $-\epsilon_\perp/\epsilon_\parallel$ for GaN under symmetric biaxial strain. In (**a**) θ denotes the angle of the c-axis with respect to the surface normal, (**b**) is a three-dimensional visualization, showing the in-plane isotropy

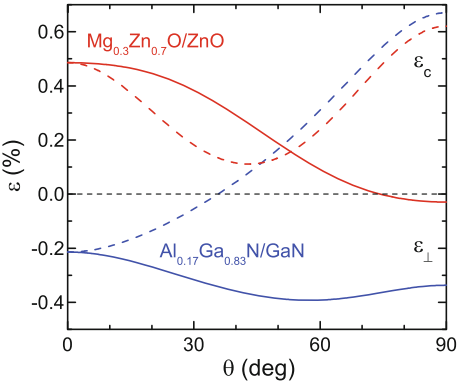


Fig. 5.26 Strains ϵ_c (*dashed lines*) and $\epsilon_\perp$ (*solid lines*) for $Al_{0.17}Ga_{0.83}N$/GaN (*blue*) and $Mg_{0.3}Ga_{0.7}O$/ZnO (*red*) as a function of the interface tilt angle θ with respect to [00.1]

The situation for pseudomorphic growth in the (Al,Ga,In)N system has been discussed for various interface orientations in [364] (cmp. also Fig. 15.14). The strains $\epsilon_\perp$ along the epitaxy direction and ϵ_c along the *c*-direction are depicted for $Al_{0.17}Ga_{0.83}N$/GaN and $Mg_{0.3}Ga_{0.7}O$/ZnO in Fig. 5.26. The different behavior of the nitride and the oxide system, e.g. regarding the sign change of ϵ_c, is due to the fact that ϵ_a is negative (positive) for $Al_xGa_{1-x}N$/GaN ($Mg_xGa_{1-x}O$/ZnO) ($\epsilon_c < 0$ for both cases) [365].

5.3.4 Three-Dimensional Strain

The strain distribution in two-dimensional or three-dimensional objects such as quantum wires and dots (see also Sect. 14) is more complicated.

A simple analytical solution for the problem of a strained inclusion is only possible for isotropic material parameters [366].

The solution for a sphere can be extended to yield the strain distribution of an inclusion of arbitrary shape. This scheme applies only for isotropic materials and identical elastic properties of the inclusion and the surrounding matrix. The solution will be given in terms of a surface integral of the boundary of the inclusion, which is fairly easy to handle. Several disconnected inclusions can be treated by a sequence of surface integrals.

The strain distribution for the inner and outer parts of a sphere with radius ρ_0 is given (in spherical coordinates) by

$$\epsilon_{\rho\rho}^{\text{in}} = \frac{2}{3}\,\epsilon_0\,\frac{1-2\nu}{1-\nu} = \epsilon_{\theta\theta}^{\text{in}} = \epsilon_{\phi\phi}^{\text{in}} \tag{5.75}$$

$$\epsilon_{\rho\rho}^{\text{out}} = \frac{2}{3}\,\epsilon_0\,\frac{1+\nu}{1-\nu}\left(\frac{\rho_0}{\rho}\right)^3 = -2\epsilon_{\theta\theta}^{\text{out}} = -2\epsilon_{\phi\phi}^{\text{out}}, \tag{5.76}$$

where ρ denotes the radius, ν the Poisson ratio, and ϵ_0 the relative lattice mismatch of the inclusion and the matrix. The radial displacements are

$$u_\rho^{\text{in}} = \frac{2}{3}\,\epsilon_0\,\frac{1-2\nu}{1-\nu}\,\rho \tag{5.77}$$

$$u_\rho^{\text{out}} = \frac{2}{3}\,\epsilon_0\,\frac{1-2\nu}{1-\nu}\,\rho_0^3\,\frac{1}{\rho^2}. \tag{5.78}$$

Dividing the displacement by the sphere's volume, we obtain the displacement per unit volume of the inclusion. From the displacement we can derive the stress σ_{ij}^0 per unit volume.

$$\sigma_{ii}^0 = \frac{1}{4\pi}\,\frac{Y\,\epsilon_0}{1-\nu}\,\frac{2x_i^2 - x_j - x_k}{\rho^5} \tag{5.79}$$

$$\sigma_{ij}^0 = \frac{3}{2}\,\frac{1}{4\pi}\,\frac{Y\,\epsilon_0}{1-\nu}\,\frac{x_i\,x_j}{\rho^5}, \tag{5.80}$$

where i, j and k are pairwise unequal indices. Due to the linear superposition of stresses, the stress distribution σ_{ij}^V for the arbitrary inclusion of volume V can be obtained by integrating over V

$$\sigma_{ij}^V = \int_V \sigma_{ij}^0\,(\mathbf{r} - \mathbf{r_0})\,\mathrm{d}^3\mathbf{r}. \tag{5.81}$$

The strains can be calculated from the stresses.

When ϵ_0 is constant within V, the volume integral can be readily transformed into an integral over the surface ∂V of V using Gauss' theorem. With the 'vector potentials' $\mathbf{A}_{ij}$ we fulfill $\mathrm{div}\mathbf{A}_{ij} = \sigma_{ij}$.

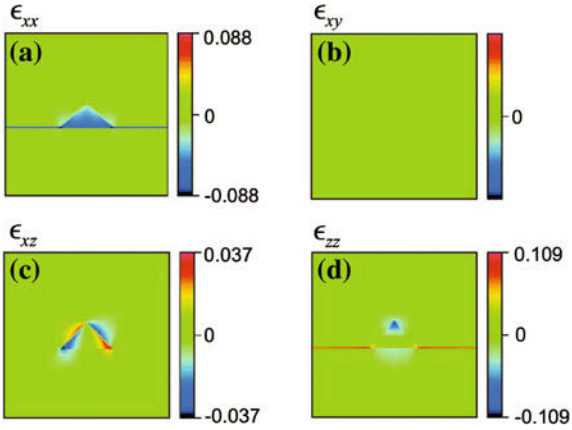


Fig. 5.27 Strain components in an InAs pyramid (quantum dot with {101} faces), embedded in GaAs. The cross section is through the center of the pyramid. The lattice mismatch between InAs and GaAs amounts to $\approx -7\,\%$. Reprinted with permission from [367], © 1995 APS

$$\mathbf{A}_{ii} = -\frac{1}{4\pi}\,\frac{Y\,\epsilon_0}{1-\nu}\,\frac{x_i \mathbf{e}_i}{\rho^3} \tag{5.82}$$

$$\mathbf{A}_{ij} = -\frac{1}{2}\,\frac{1}{4\pi}\,\frac{Y\,\epsilon_0}{1-\nu}\,\frac{x_i\,\mathbf{e}_j + x_j\,\mathbf{e}_i}{\rho^3}. \tag{5.83}$$

Equation (5.83) is valid for the case $i \neq j$. $\mathbf{e}_i$ is the unit vector in the ith direction. However, special care must be taken at the singularity $\mathbf{r} = \mathbf{r}_0$ if $\mathbf{r}_0$ lies within V because the stress within the 'δ-inclusion' is not singular (in contrast to the electrostatic analog of a δ-charge). Thus, we find

$$\sigma_{ij}^V(\mathbf{r}_0) = \oint_{\partial V} \mathbf{A}_{ij}\,\mathrm{d}\mathbf{S} + \delta_{ij}\,\frac{Y\,\epsilon_0}{1-\nu}\int_V \delta(\mathbf{r}-\mathbf{r}_0)\,\mathrm{d}^3\mathbf{r}. \tag{5.84}$$

As an example, we show in Fig. 5.27 the numerically calculated strain components [367] (taking into account the different elastic properties of the dot and matrix materials) in the cross section of a pyramidal InAs quantum dot in a GaAs matrix on top of a two-dimensional InAs layer. The strain component ϵ_{zz} is positive in the 2D layer, as expected from (5.70). However, in the pyramid ϵ_{zz} exhibits a complicated dependence and even takes negative values at the apex.

5.3.5 *Substrate Bending*

If a lattice-mismatched layer is pseudomorphically grown on top of a substrate it suffers biaxial strain. For finite substrate thickness part of the strain will relax via substrate bending. If the substrate is circular, a spherical cap is formed. If the lattice constant of the film is larger (smaller) than that of the substrate, the film is under compressive (tensile) strain and the curvature is convex (concave) with respect to

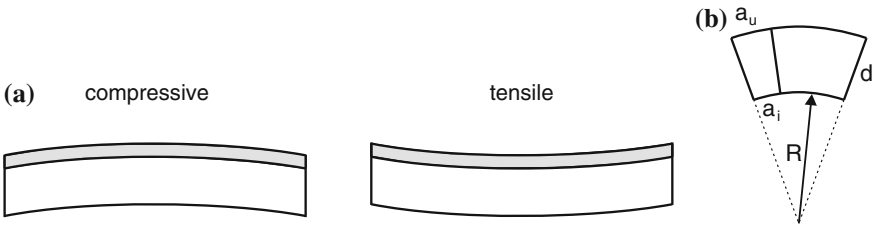


Fig. 5.28 (**a**) Schematic bending of a film/substrate system for compressive (*left*) and tensile (*right*) film strain. (**b**) Schematic deformation of curved film of thickness d. The lattice constants at the inner and outer surface are a_i and a_u, respectively

the outward normal given by the growth direction (Fig. 5.28a). Substrate bending can also be induced by a mismatch of the thermal expansion coefficients $\alpha_\mathrm{th}^\mathrm{f}$ and $\alpha_\mathrm{th}^\mathrm{s}$ of the film and substrate, respectively. If a film/substrate system is flat at a given temperature, e.g. growth temperature, a decrease of temperature, e.g. during cooling, will lead to compressive (tensile) strain if $\alpha_\mathrm{th}^\mathrm{f}$ is smaller (larger) than $\alpha_\mathrm{th}^\mathrm{s}$.

In a curved structure, the lattice constant in the tangential direction increases from a_i^t at the inner surface ($r = R = \kappa^{-1}$) to a_u^t at the outer surface ($r = R + d$). Thus, the tangential lattice constant varies with the radial position

$$a^\mathrm{t}(r) = a_\mathrm{i}^\mathrm{t}\,(1 + r\,\kappa)\,, \tag{5.85}$$

where d is the layer thickness (Fig. 5.28b). Therefore $a_\mathrm{u} = a_\mathrm{i}(1 + d/R)$. We note that (5.85) holds in all layers of a heterostructure, i.e. the film *and* the substrate.

The lattice constant in the radial direction a^r, however, depends on the lattice constant a_0 of the local material and is calculated from the biaxial strain condition, such as (5.70). The in-plane strain is $\epsilon_\parallel = (a^\mathrm{t} - a_0)/a_0$ (we assume a spherical cap with $\epsilon_\parallel = \epsilon_{\theta\theta} = \epsilon_{\phi\phi}$). For an isotropic material we find $a^\mathrm{r} = a_0\,(1 + \epsilon_\perp)$ with $\epsilon_\perp = -2\nu\epsilon_\parallel/(1 - \nu)$. The local strain energy density U is given by

$$U = \frac{Y}{1 - \nu}\,\epsilon_\parallel^2. \tag{5.86}$$

The total strain energy per unit area U' of a system of two layers with lattice constants a_1, a_2, Young's moduli Y_1, Y_2 and thickness d_1, d_2 (we assume the same Poisson constant ν in both layers) is

$$U' = \int_0^{d_1} U_1\,\mathrm{d}r + \int_{d_1}^{d_2} U_2\,\mathrm{d}r. \tag{5.87}$$

The total strain energy needs to be minimized with respect to a_i and R in order to find the equilibrium curvature κ. We find

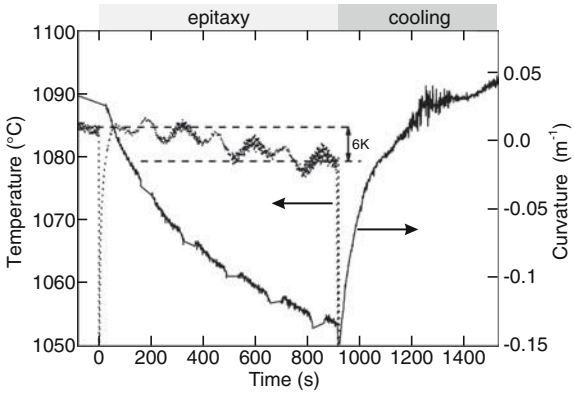


Fig. 5.29 Curvature of the middle of a Si wafer during GaN growth on an AlN interlayer grown at low temperatures on GaN and subsequent cooling. During growth the decrease in curvature indicates convex bowing due to compressive stress; during cooling the wafer flattens and becomes concave due to thermally induced tensile stress. Adapted from [372]

$$\kappa = \frac{6a_1a_2\,(a_2-a_1)\,d_1d_2\,(d_1+d_2)\,Y_1Y_2}{a_2^3\,d_1^4\,Y_1^2+\alpha\,Y_1Y_2+a_1^3\,d_2^4\,Y_2^4} \tag{5.88}$$
$$\alpha = a_1a_2\,d_1d_2\,\left[-a_2d_1(2d_1+3d_2)+a_1(6d_1^2+9d_1d_2+4d_2^2)\right].$$

For $a_2 = a_1\,(1+\epsilon)$ we develop κ to first order of ϵ and find ($\chi = Y_2/Y_1$) [368, 369]

$$\kappa = \frac{6\chi\,d_1\,d_2\,(d_1+d_2)}{d_1^4+4\chi\,d_1^3\,d_2+6\chi\,d_1^2\,d_2^2+4\chi\,d_1\,d_2^3+\chi^2\,d_2^4}\,\epsilon. \tag{5.89}$$

In the case of a substrate (d_s) with a thin epitaxial layer ($d_f \ll d_s$), the radius of curvature is approximately (Stoney's formula [370])

$$\kappa = 6\,\epsilon\,\frac{d_f}{d_s^2}\,\frac{Y_f}{Y_s}. \tag{5.90}$$

Conversely, if the radius of curvature is measured [371], e.g. optically, the film curvature (and through models also the film strain) can be determined during epitaxy as depicted in Fig. 5.29.

5.3.6 Scrolling

In some cases cylindrically scrolled structures are important, e.g. for thin-film flexible electronics, nanotubes, nanoscrolls or nanohelixes. The scrolling of thin layers must be avoided by suitable strain management for thin layers that are lifted off from their substrate for transfer to another flat substrate. If the film remains attached to its substrate, a scroll can be fabricated as schematically shown in Fig. 5.30. Such structures were first reported in [373], a review can be found in [374]. The shape of such scroll is investigated in [375] without a priori assumptions on its shape.

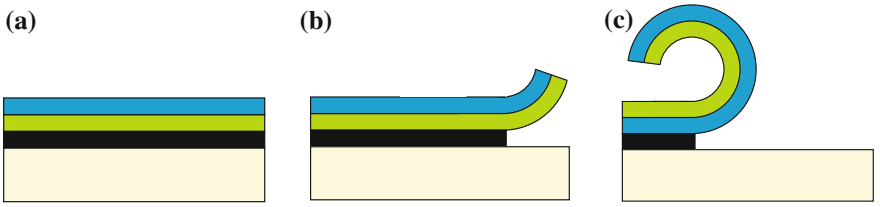


Fig. 5.30 Schematic representation of nanoscroll formation. (**a**) Strained heterostructure (*blue/green*) that is planar due to large substrate thickness, (**b**) starting removal of sacrificial layer (*black*), (**c**) release of thin film into nanoscroll geometry

If bending strain occurs only in *one* of the tangential directions, the energy density is given by

$$U = \frac{Y}{2\,(1-\nu^2)}\,(\epsilon_\mathrm{t}^2 + \epsilon_y^2 + 2\nu\,\epsilon_\mathrm{t}\,\epsilon_y), \tag{5.91}$$

where ϵ_y is the strain in the unbent direction (cylinder axis) as shown in Fig. 5.31a. For a strained heterostructure made up from two layers the curvature is given by (calculated analogous to (5.89), $\chi = Y_2/Y_1$ [369])

$$\kappa = \frac{6(1+\nu)\,\chi\,d_1\,d_2\,(d_1+d_2)}{d_1^4 + 4\chi\,d_1^3\,d_2 + 6\chi\,d_1^2\,d_2^2 + 4\chi\,d_1\,d_2^3 + \chi^2\,d_2^4}\,\epsilon, \tag{5.92}$$

which differs from (5.89) only by the factor $1+\nu$ in the nominator.

For cubic material and a (001) surface the energy is given as

$$U_{100} = \frac{C_{11}-C_{12}}{2C_{11}}\left[C_{11}\,(\epsilon_\mathrm{t}^2 + \epsilon_y^2 + C_{12}\,(\epsilon_\mathrm{t}+\epsilon_y)^2)\right] \tag{5.93}$$

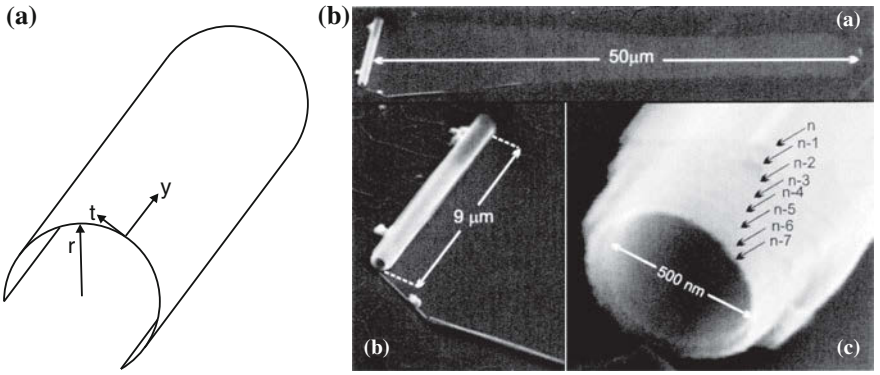


Fig. 5.31 (**a**) Schematic representation of a cylindrically rolled sheet with radial direction r, tangential direction t and direction along the cylinder axis y. (**b**) SEM images of multiwall InGaAs/GaAs nanoscroll rolled up over about 50 µm. Part (**b**) from [376]

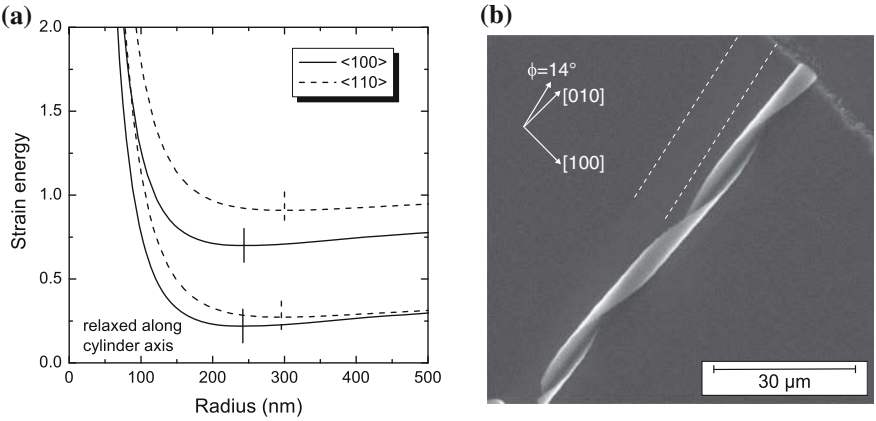


Fig. 5.32 (**a**) Strain energy (in units of the strain energy of the flat pseudomorphic layers) of a scroll of a 4-layer SiGe structure ($Si_{0.3}Ge_{0.7}$, $Si_{0.6}Ge_{0.4}$ and $Si_{0.8}Ge_{0.2}$, each 3 nm thick and a 1-nm Si cap) as a function of radius for winding directions along ⟨100⟩ and ⟨110⟩. *Top* (*bottom*) curves without (with complete) strain relaxation along the cylinder axis. *Vertical lines* indicate the positions of the respective energy minima [369]. (**b**) SEM image of curled InGaAs/GaAs nanoscroll rolled $\phi = 14°$ off ⟨100⟩. The stripe from which the film was rolled off is indicated by *white dashed lines*. Part (**b**) from [379]

for a scrolling direction along ⟨100⟩. When the (001)-oriented film winds up along a direction ⟨*hk*0⟩ having an angle ϕ with the [100] direction ($\phi = 45°$ for ⟨110⟩), the strain energy is given by (C_0 is given by (5.62))

$$U_\phi = U_{100} + C_0 \left(\frac{\epsilon_t - \epsilon_y}{2} \right)^2 \sin^2(2\phi). \tag{5.94}$$

The strain energy versus bending radius ($= \kappa^{-1}$) is shown for a SiGe nanoscroll in Fig. 5.32. First, the relaxation along the cylinder axis plays a minor role. The smallest strain energy is reached for scrolling along ⟨100⟩, also yielding the smaller bending radius (larger curvature). Therefore, the film preferentially scrolls along ⟨100⟩. This explains the observed 'curl' behavior of scrolls winding up for $\phi \neq 0$ [373, 377] (Fig. 5.32b). The effect of surface strain needs to be included to yield improved quantitative agreement with experimental values of $\kappa(\epsilon, d)$ [378].

5.4 Plasticity

5.4.1 Critical Thickness

Strained epitaxial films are called pseudomorphic when they do not contain defects and the strain relaxes elastically, e.g. by tetragonal distortion. When the layer thick-

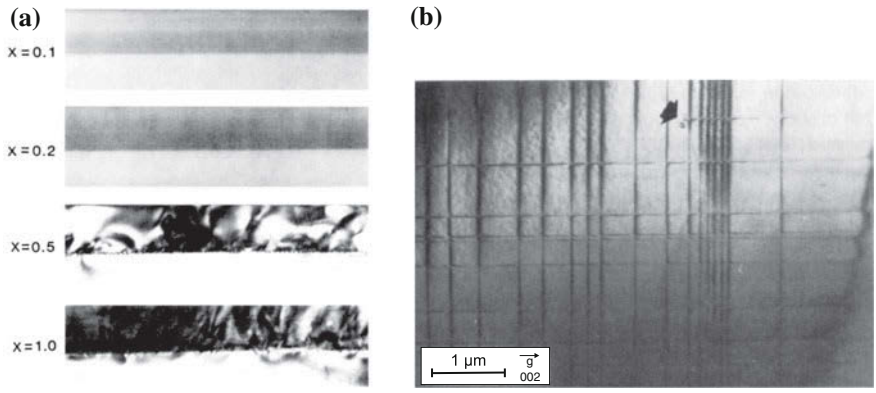


Fig. 5.33 (**a**) Series of cross-sectional TEM images of 100-nm thick Ge_xSi_{1-x} layers on Si(001) with different ternary compositions $x = 0.1, 0.2, 0.5$, and 1.0. The growth temperature was 550°C. The transition from commensurate to incommensurate growth is obvious. Adapted from [380]. (**b**) Plan view ⟨022⟩ TEM bright field image of a 250-nm $Ge_{0.15}Si_{0.85}$ layer on Si (001), annealed at about 700°C. The arrow denotes the position of a dislocation loop. Reprinted with permission from [381], © 1989 AVS

ness increases, however, strain energy is accumulated that will lead at some point to plastic relaxation via the formation of defects. In many cases, a grid of misfit dislocations forms at the interface (Figs. 4.18 and 5.33).

In Fig. 5.34 the strain around misfit dislocations at a GaAs/CdTe heterointerface, as calculated from a TEM image (Fig. 4.14), is shown.

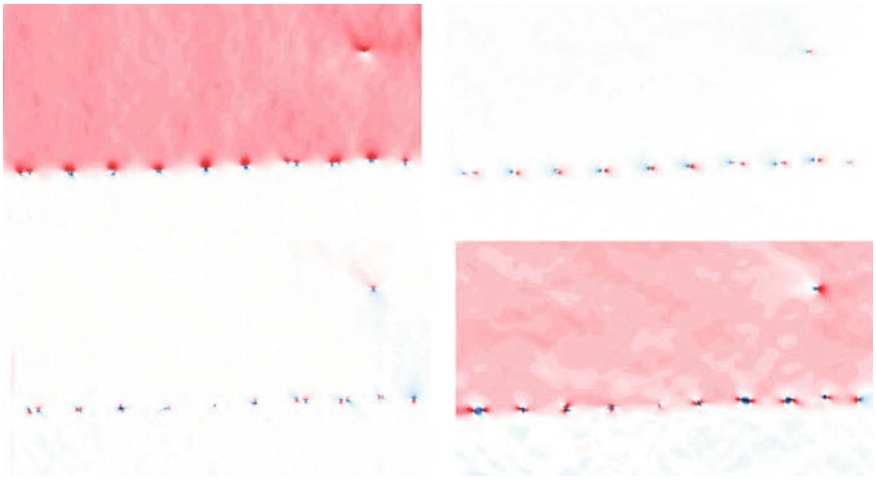


Fig. 5.34 Components $\begin{pmatrix} \epsilon_{xx} & \epsilon_{xz} \\ \epsilon_{zx} & \epsilon_{zz} \end{pmatrix}$ of the strain tensor (with respect to the GaAs lattice constant) of the dislocation array shown in Fig. 4.14, *red*/*blue*: positive/negative value, *white*: zero. From [286]

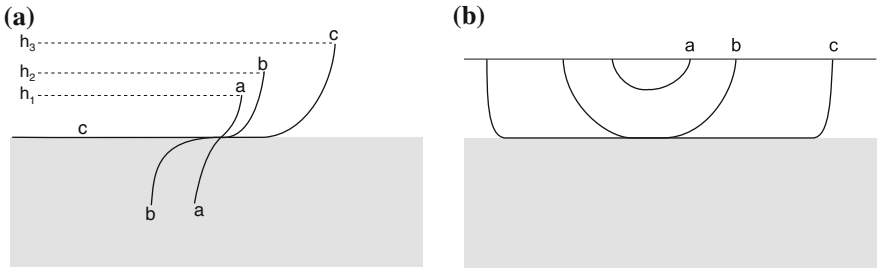


Fig. 5.35 Schematic formation of misfit dislocations by (**a**) elongation of a grown-in threading dislocation and (**b**) by the nucleation and growth of dislocation half-loops. (**a**) depicts a threading dislocation. Initially, for thickness h_1 the interface is coherent 'a', for larger thickness h_2 the interface is critical and the force of the interface on the dislocation is equal to the tension in the dislocation line, 'b'. For larger thickness, e.g. h_3, the dislocation line is elongated in the plane of the interface, 'c'. In (**b**) 'a' denotes a subcritical dislocation half-loop, 'b' depicts a half-loop being stable under the misfit stress and for 'c' the loop has grown under the misfit stress into a misfit dislocation line along the interface

The average distance p of the dislocations is related to the misfit $f = (a_1 - a_2)/a_2$ and the edge component $b_\perp$ of the Burger's vector (for a 60° dislocation $b_\perp = a_0/\sqrt{8}$)

$$p = \frac{b_\perp}{f}. \tag{5.95}$$

Two mechanisms have been proposed for the formation of misfit dislocations (Fig. 5.35), the elongation of a grown-in threading dislocation [382] and the nucleation and growth of dislocation half-loops [383]. For the modeling of such systems a mechanical approach based on the forces on dislocations [382] or an energy consideration based on the minimum strain energy necessary for defect formation [383–386] can be followed. Both approaches have been shown to be equivalent [387] (if a periodic array of dislocations is considered). In [388] it was pointed out that the finite speed of plastic flow also has to be considered to explain experimental data. Temperature affects the observed critical thickness and a kinetic model is needed. Another way of introducing dislocations is the plastic relaxation at the edge of coherent strained islands (cmp. Fig. 14.26).

In the following, isotropic materials and identical elastic constants of substrate and thin film are assumed, following [387]. The interface plane is the (x,y)-plane, the growth direction is z. The energy E_d of a periodic dislocation array with period p and Burgers vector $\mathbf{b} = (b_1, b_2, b_3)$ is

$$E_\mathrm{d} = \frac{Y}{8\pi\,(1-\nu^2)}\,\beta^2$$

$$\beta^2 = \left[b_1^2 + (1-\nu)\,b_2^2 + b_3^2\right] \ln\left(\frac{p\,\left[1-\exp(-4\pi h/p)\right]}{2\pi\,q}\right)$$

$$+\left(b_1^2-b_3^2\right)\frac{4\pi h}{p}\,\frac{\exp(-4\pi h/p)}{1-\exp(-4\pi h/p)}$$
$$-\frac{1}{2}\left(b_1^2+b_3^2\right)\left(\frac{4\pi h}{p}\right)^2\frac{\exp(-4\pi h/p)}{\left[1-\exp(-4\pi h/p)\right]^2}$$
$$+b_3^2\,\frac{2\pi h}{p}\,\frac{\exp(-2\pi h/p)}{1-\exp(-2\pi h/p)}, \tag{5.96}$$

where h is the film thickness and q denotes the cutoff length for the dislocation core, taken as $q = b$. The misfit strain including the relaxation due to dislocations with Burger's vectors $\mathbf{b}$ and $\hat{\mathbf{b}}$ in the two orthogonal interface $\langle 110\rangle$ directions $\mathbf{n}$ and $\hat{\mathbf{n}}$. We chose the coordinate system such that $\mathbf{n} = (1, 0, 0)$ and $\hat{\mathbf{n}} = (0, 1, 0)$ (the z direction remains). With respect to these axes the Burger's vectors are $\left(\pm\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right) a_0/\sqrt{2}$. The misfit strain $\epsilon_{ij}^{\mathrm{m}}$ is reduced due to the dislocation formation to the 'relaxed' misfit strain $\epsilon_{ij}^{\mathrm{r}}$ with

$$\epsilon_{ij}^{\mathrm{r}} = \epsilon_{ij}^{\mathrm{m}} + \frac{b_i\, n_j + b_j\, n_i}{2p} + \frac{\hat{b}_i\, \hat{n}_j + \hat{b}_j\, \hat{n}_i}{2p}, \tag{5.97}$$

with an associated stress σ_{ij}. The strain energy E_{s} of the layer due to the relaxed misfit is then

$$E_{\mathrm{s}} = \frac{1}{2}\, h\, \sigma_{ij}\, \epsilon_{ij}^{\mathrm{r}} \tag{5.98}$$
$$\lim_{p\to\infty} E_{\mathrm{s}} = 2h\,\frac{Y(1+\nu)}{1-\nu}\, f^2. \tag{5.99}$$

The total strain energy E is given by

$$pE = 2E_{\mathrm{d}} + 2E_{\mathrm{c}} + p\, E_{\mathrm{s}} \tag{5.100}$$
$$E_\infty = \lim_{p\to\infty} E, \tag{5.101}$$

with the core energy E_{c} of the dislocation that needs to be calculated with an atomistic model (not considered further here). This energy is shown in Fig. 5.36a for the material parameters of $Ge_{0.1}Si_{0.9}$/Si(001) (misfit -0.4%) for various layer thicknesses as a function of $1/p$. This plot looks similar to that for a first-order phase transition (with $1/p$ as the order parameter). For a certain critical thickness h_{c1} the energy of the layer without any dislocation and the layer with a particular dislocation density p_1 are identical ($E - E_\infty = 0$) and additionally $\partial E/\partial p|_{p=p_1} = 0$. However, between $p \to \infty$ and $p = p_0$ there is an energy barrier. The critical thickness h_{c2} is reached when

$$\partial E/\partial p\,|_{p\to\infty} = 0, \tag{5.102}$$

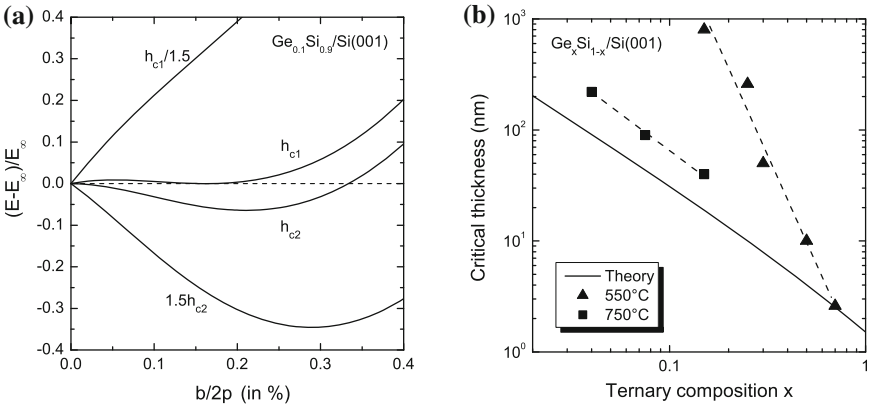


Fig. 5.36 (**a**) Theoretical calculation for the strain energy versus inverse dislocation density for various thicknesses of $Ge_{0.1}Si_{0.9}$ layers on Si (001). The ordinate is $b/2p$, $b/2$ being the edge component of the Burgers vector and p being the dislocation spacing. The abscissa is the strain energy E scaled with E_∞ (5.101). (**b**) Critical thickness for Ge_xSi_{1-x} layers on Si (001). The *solid line* is theory (h_{c2}) according to (5.103). Data points are from [389] (*squares*, growth temperature of 750 °C) and from [380] (*triangles* for growth temperature of 550 °C)

i.e. the energy decreases monotonically for increasing dislocation density up to the global energy minimum at a certain equilibrium dislocation spacing p_2. Equation (5.102) leads to the following implicit equation for the determination of h_{c2}:

$$h_{c2} = \frac{b\left[-16 + 3b^2 + 8\,(-4+\nu)\,\ln\,(2h_{c2}/q)\right]}{128\,f\,\pi\,(1+\nu)}, \tag{5.103}$$

with the length of the Burgers vector $b = a_0/\sqrt{2}$.

The theoretical dependence of h_{c2} for Ge_xSi_{1-x}/Si(001) with varying composition is shown in Fig. 5.36b together with experimental data. The critical thickness for a fairly high growth temperature is much closer to the energetic equilibrium than that deposited at lower temperature. This shows that there are kinetic limitations for the system to reach the mechanical equilibrium state. Also, the experimental determination of the critical thickness is affected by finite resolution for large dislocation spacing, leading generally to an overestimate of h_c.

In zincblende materials two types of dislocations are possible, α and β, with Ga- and As-based cores, respectively. They have $[\bar{1}10]$ and $[110]$ line directions for a compressively strained interface. The α dislocation has the larger glide velocity. Therefore, strain relaxation can be anisotropic with regard to the $\langle 110\rangle$ directions for zincblende material, e.g. InGaAs/GaAs [390, 391].

5.4.2 *Cleaving*

The cleavage planes of the diamond structure are {111} planes (Fig. 5.37a). It is easiest to break the bonds connecting the double layers in the ⟨111⟩ directions.

The cleavage planes of the zincblende structure are {110} planes (Fig. 5.37b). Due to the ionic character, breaking the bonds connecting the double layers in the ⟨111⟩ directions would leave charged surfaces, which is energetically unfavorable. The {100} planes contain only one sort of atom and would also leave highly charged surfaces. The {110} planes contain equal amounts of A and B atoms and are neutral. Ideally, the cleaving plane is atomically flat (Fig. 5.38a) or exhibits large mono-atomically flat terraces. However, certain dopants in high concentrations, e.g. GaAs:Te, can induce a rough surface due to lattice distortion [392].

The natural cleavage planes of wurtzite (GaN) are $\{1\bar{1}.0\}$ (m-type) planes [393].

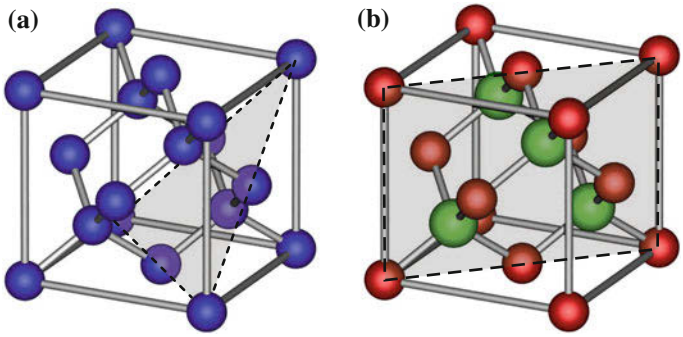


Fig. 5.37 Cleaving planes of the (**a**) diamond and (**b**) zincblende lattice

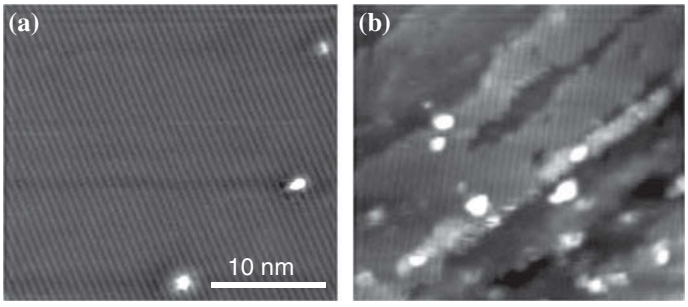


Fig. 5.38 Scanning tunneling microscopy images of a cleaved GaAs (110) surface with (**a**) good cleave and (**b**) bad cleave with defects dominating. Adapted from [394]

5.4.3 Wafer Breakage

The thickness and thus strength of wafers for semiconductor production (cmp. Sect. 12.2.2) is an important issue. The wafer should be as thin as possible for saving expensive materials but thick enough to avoid loss due to stress during handling, in particular during the later steps in a process since the value of a wafer increases with number of process steps it has undergone.

Reasons for wafer breakage is the mechanical handling (pick, place, transport) [395], stress loads due to processing (dielectrics, metals, asymmetric structures) and stress during processing for example due to thermal loads in annealing or deposition steps and cutting/dicing. The problem is less important in microelectronics but especially severe in photovoltaics (PV) industry, handling large total areas; on the other hand the profit loss per broken wafer is much higher in microelectronics industry. An additional problem poses the grain structure of multicrystalline silicon wafers used for PV [396] and the effect of surface cracks and irregularities at wafer edges and corners. Just going from a wafer thickness of 270 to 250 μm can more than double the breakage rate at certain process steps [397]. The minimum strength of a wafer with surface cracks is about 100 MPa, while the strength of wafers with cracks at the edge can reach rather low values around 20 MPa. Also the careful shaping of the wafer edge is important to avoid breakage [398].

Chapter 6
Band Structure

Silicon is a metal.

A.H. Wilson, 1931 [67]

Abstract A treatment of electron states in one-dimensional potentials introduces into the concepts of band gap and effective mass. The band structures of various semiconductors are reviewed. The systematics of band gaps, symmetry considerations, band gaps in alloys, amorphous semiconductors and the effect of strain and temperature are discussed. Electron and hole dispersions are treated and the density of states in various dimensions is derived.

6.1 Introduction

Valence electrons that move in the crystals feel a periodic potential

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R}) \tag{6.1}$$

for all vectors $\mathbf{R}$ of the direct lattice. The potential[1] is due to the effect of the ion cores and all other electrons. Thus a serious many-body problem is present. In principle, the band structure can be calculated from the periodic arrangements of the atoms and their atomic order number. We note that for some problems, e.g. the design of optimal solar cells, a certain band structure is known to be ideal and a periodic atomic arrangement, i.e. a material, needs to be found that generates the optimal band structure. This problem is called the *inverse band structure problem*.

[1] In this book the form of the potential will never be explicitly given.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_6

6.2 Electrons in a Periodic Potential

6.2.1 Bloch's Theorem

First, we will deduce some general conclusions about the structure of the solution as a consequence of the periodicity of the potential. We first investigate the solution of a Schrödinger equation of the type

$$H\,\Psi(\mathbf{r}) = \left[-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r})\right]\Psi(\mathbf{r}) = E\,\Psi(\mathbf{r}) \tag{6.2}$$

for an electron. U will be periodic with the lattice, i.e. it will obey (6.1).

Bloch's theorem says that the eigenstates Ψ of a one-particle Hamiltonian as in (6.2) can be written as the product of plane waves and a lattice-periodic function, i.e.

$$\Psi_{n\mathbf{k}}(\mathbf{r}) = A\,\exp(\mathrm{i}\,\mathbf{k}\,\mathbf{r})\,u_{n\mathbf{k}}(\mathbf{r}). \tag{6.3}$$

The normalization constant A is often omitted. If $u_{n\mathbf{k}}(\mathbf{r})$ is normalized, $A = 1/\sqrt{V}$, where V is the integration volume. The wavefunction is indexed with a quantum number n and the wavevector $\mathbf{k}$. The key is that the function $u_{n\mathbf{k}}(\mathbf{r})$, the so-called Bloch function, is periodic with the lattice, i.e.

$$u_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}\,(\mathbf{r} + \mathbf{R}) \tag{6.4}$$

for all vectors $\mathbf{R}$ of the direct lattice. The proof is simple in one dimension and more involved in three dimensions with possibly degenerate wavefunctions, see [399].

If $E_{n\mathbf{k}}$ is an energy eigenvalue, then $E_{n\mathbf{k}+\mathbf{G}}$ is also an eigenvalue for all vectors $\mathbf{G}$ of the reciprocal lattice, i.e.

$$E_n(\mathbf{k}) = E_n\,(\mathbf{k} + \mathbf{G}). \tag{6.5}$$

Thus the energy values are periodic in reciprocal space. The proof is simple, since the wavefunction (for $\mathbf{k} + \mathbf{G}$) $\exp(\mathrm{i}(\mathbf{k} + \mathbf{G})\mathbf{r})\,u_{n(\mathbf{k}+\mathbf{G})}(\mathbf{r})$ is for $u_{n(\mathbf{k}+\mathbf{G})}(\mathbf{r}) = \exp(-\mathrm{i}\mathbf{G}\mathbf{r})\,u_{n\mathbf{k}}(\mathbf{r})$ obviously an eigenfunction to $\mathbf{k}$.

A band structure along one $\mathbf{k}$-direction can be displayed in various zone schemes as depicted in Fig. 6.1. The most frequently used scheme is the *reduced* zone scheme. In three dimensions, the band structure is typically shown along particular paths in the Brillouin zone, as depicted, e.g., in Fig. 6.2c.

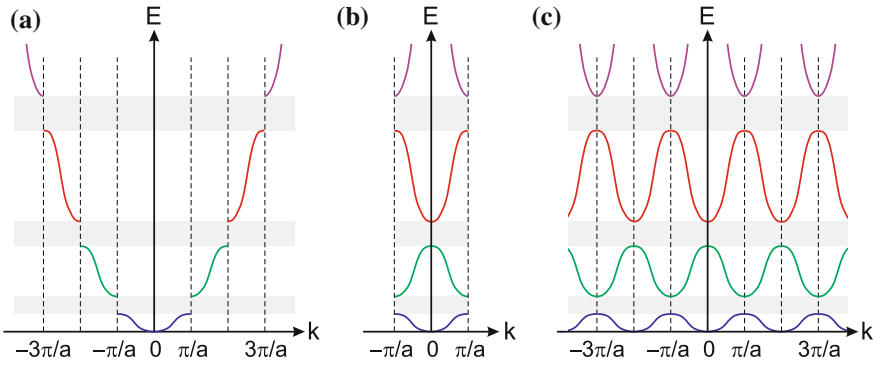


Fig. 6.1 Zone schemes for a band structure: (**a**) extended, (**b**) reduced and (**c**) repetitive zone scheme

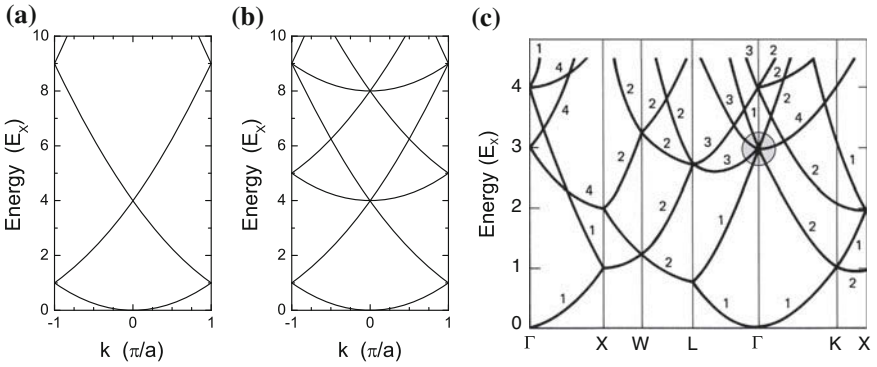


Fig. 6.2 Dispersion of free electrons (empty lattice calculation, $U = 0$, shown in the first Brillouin zone) in (**a**) a one-dimensional lattice ($\mathbf{G} = n\,2\pi/a$), (**b**) a simple cubic lattice ($\mathbf{G} = (h, k, l)\,2\pi/a$) and (**c**) in a fcc lattice. The energy is measured in units of the energy at the X-point, $E_\mathrm{X} = (\hbar^2/2m)(2\pi/a)^2$. The *shaded circle* in (**c**) represents the region where the band gap develops for finite periodic potential $U \neq 0$

6.2.2 *Free-Electron Dispersion*

If the entire wavefunction (from (6.3)) obeys the Schrödinger equation (6.2), the Bloch function $u_{n\mathbf{k}}$ fulfills the equation

$$\left[\frac{1}{2m}\,(\mathbf{p} + \hbar\,\mathbf{k})^2 + U(\mathbf{r})\right] u_{n\mathbf{k}}(\mathbf{r}) = E_{n\mathbf{k}}\, u_{n\mathbf{k}}(\mathbf{r}), \tag{6.6}$$

which is easy to see from $\mathbf{p} = -i\hbar\nabla$.

First, we discuss the simplest case of a periodic potential, $U \equiv 0$. This calculation is also called the *empty lattice* calculation. The solution of (6.6) is then just constant,

i.e. $u_{\mathbf{k}} = c$ and $\Psi_{\mathbf{k}}(\mathbf{r}) = c \exp(\mathrm{i}\mathbf{k}\mathbf{r})$. The dispersion of the free electron is then given by

$$E(\mathbf{k}) = \frac{\hbar^2}{2m}\mathbf{k}^2, \tag{6.7}$$

where $\mathbf{k}$ is an arbitrary vector in the reciprocal space. $\mathbf{k}'$ is a vector from the Brillouin zone such that $\mathbf{k} = \mathbf{k}' + \mathbf{G}$ with a suitable reciprocal lattice vector $\mathbf{G}$. Because of (6.5) the dispersion relation can be written also as

$$E(\mathbf{k}) = \frac{\hbar^2}{2m}(\mathbf{k}' + \mathbf{G})^2, \tag{6.8}$$

where $\mathbf{k}'$ denotes a vector from the Brillouin zone. Thus, many branches of the dispersion relation arise from using various reciprocal lattice vectors in (6.8).

The resulting dispersion relation for the free electron is shown in Fig. 6.2a for a one-dimensional system ($\mathbf{k}'$ and $\mathbf{G}$ are parallel) and in Fig. 6.2b for the simple cubic lattice (in the so-called reduced zone scheme). In Fig. 6.2c, the (same) dispersion of the free electron is shown for the fcc lattice.

6.2.3 Non-Vanishing Potential

Now the effect of a non-vanishing periodic potential on electron motion will be discussed. A simple, analytically solvable model that visualizes the effect of a periodic potential on the dispersion relation of the electrons and the formation of a (one-dimensional) band structure with gaps is the Kronig–Penney model [64] which is discussed in the Appendix F.

General Wave Equation

In this section, we will discuss the solution of a general wave equation for electrons in a periodic potential. The solution is investigated particularly at the zone boundary. The potential U is periodic with the lattice (6.1). It can be represented as a Fourier series with the reciprocal lattice vectors (lattice vector expansion, cf. (3.18)):

$$U(\mathbf{r}) = \sum_{\mathbf{G}} U_{\mathbf{G}} \exp(\mathrm{i}\,\mathbf{G}\,\mathbf{r})\,. \tag{6.9}$$

Since U is a real function, $U_{-\mathbf{G}} = U_{\mathbf{G}}^*$. The deeper reason for the success of such an approach is that for typical crystal potentials, the Fourier coefficients decrease rapidly with increasing $\mathbf{G}$, e.g. for the unscreened Coulomb potential $U_{\mathbf{G}} \propto 1/G^2$. The wavefunction is expressed as a Fourier series (or integral) over all allowed (Bloch) wavevectors $\mathbf{K}$,

$$\Psi(\mathbf{r}) = \sum_{\mathbf{K}} C_{\mathbf{K}} \exp(\mathrm{i}\,\mathbf{K}\,\mathbf{r})\,. \tag{6.10}$$

The kinetic and potential energy terms in the Schrödinger equation (6.6) are

$$\nabla^2\Psi = -\sum_{\mathbf{K}} \mathbf{K}^2\, C_{\mathbf{K}} \exp(\mathrm{i}\,\mathbf{K}\,\mathbf{r}) \tag{6.11a}$$

$$U\Psi = \sum_{\mathbf{G}}\sum_{\mathbf{K}} U_{\mathbf{G}}\, C_{\mathbf{K}} \exp(\mathrm{i}\,(\mathbf{G}+\mathbf{K})\,\mathbf{r})\,. \tag{6.11b}$$

With $\mathbf{K}' = \mathbf{K} + \mathbf{G}$, (6.11b) can be rewritten as

$$U\Psi = \sum_{\mathbf{G}}\sum_{\mathbf{K}'} U_{\mathbf{G}}\, C_{\mathbf{K}'-\mathbf{G}} \exp\left(\mathrm{i}\,\mathbf{K}'\,\mathbf{r}\right). \tag{6.12}$$

Now, the Schrödinger equation can be written as an (infinite) system of algebraic equations:

$$(\lambda_{\mathbf{K}} - E)\, C_{\mathbf{K}} + \sum_{\mathbf{G}} U_{\mathbf{G}}\, C_{\mathbf{K}-\mathbf{G}} = 0, \tag{6.13}$$

with $\lambda_{\mathbf{K}} = \hbar^2\,\mathbf{K}^2/(2m)$.

Solution for One Fourier Coefficient

The simplest (non-trivial) potential energy has only one important Fourier coefficient U for the shortest reciprocal lattice vector $\mathbf{G}$. Also, we have $U_{-\mathbf{G}} = U_{\mathbf{G}}$. Thus, the (one-dimensional) potential has the form $U(x) = 2\,U\,\cos(Gx)$. Then the equation system (6.13) has only two equations for $C_{\mathbf{K}}$ and $C_{\mathbf{K}-\mathbf{G}}$, leading to the condition

$$\begin{vmatrix} \lambda_{\mathbf{K}} - E & U \\ U & \lambda_{\mathbf{K}-\mathbf{G}} - E \end{vmatrix} = 0. \tag{6.14}$$

We find two solutions

$$E_{\pm} = \frac{\lambda_{\mathbf{K}} + \lambda_{\mathbf{K}-\mathbf{G}}}{2} \pm \sqrt{\left(\frac{\lambda_{\mathbf{K}} - \lambda_{\mathbf{K}-\mathbf{G}}}{2}\right)^2 + U^2}. \tag{6.15}$$

Solution at the Zone Boundary

We consider the solution at the zone boundary, i.e. at $\mathbf{K} = \mathbf{G}/2$. The kinetic energy is then the same for $\mathbf{K} = \pm\mathbf{G}/2$, i.e. $\lambda_{\mathbf{K}} = \lambda_{\mathbf{K}-\mathbf{G}} = (\hbar^2/2m)\,(G^2/4) = \lambda$. The determinant (6.14) reads then

$$(\lambda - E)^2 - U^2 = 0. \tag{6.16}$$

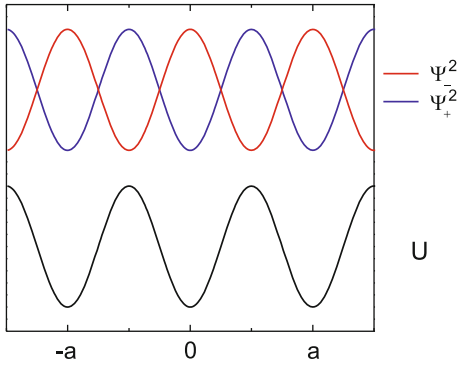


Fig. 6.3 Periodic potential U (one-dimensional cosine, *black*) and the squares of the wavefunctions Ψ_- (*red*) and Ψ_+ (*blue*) for the wavevector at the zone boundary, $K = G/2 = \pi/a$

Thus the energy values at the zone boundary are

$$E_\pm = \lambda \pm U = \frac{\hbar^2}{2m}\frac{G^2}{4} \pm U. \tag{6.17}$$

At the zone boundary, a splitting of the size $E_+ - E_- = 2U$ occurs. The center of the energy gap is given by the energy $\lambda_\mathbf{K}$ of the free-electron dispersion. The ratio of the coefficients is $C_{\mathbf{G}/2}/C_{-\mathbf{G}/2} = \mp 1$. The '$-$' solution of (6.17) (lower energy) is a standing cosine wave (Ψ_-), the '$+$' solution (Ψ_+) is a standing sine wave as visualized in Fig. 6.3. For the lower-energy (binding) state the electrons are localized at the potential minima, i.e. at the atoms, for the upper state (antibinding) the electrons are localized between the atoms. Both wavefunctions have the same periodicity since they belong to the same wavevector $\mathbf{K} = \mathbf{G}/2$. We note that the periodicity of Ψ is $2a$, while the periodicity of Ψ^2 is equal to the lattice constant a.

Gap States

For energies within the gap, solutions with a *complex* wavevector $K = G/2 + \mathrm{i}\,q$ exist. Solving (6.16) results (in terms of $q'^2 = (\hbar^2/2m)\,q^2$) to

$$E_\pm = \lambda - q'^2 \pm \sqrt{-4\,\lambda\,q'^2 + U^2}. \tag{6.18}$$

For energies $E = \lambda + \epsilon$ with $-U \le \epsilon \le U$, the complex part of the wavevector is given by

$$q'^2 = -(\epsilon + 2\lambda) + \sqrt{4\lambda\,(\epsilon + \lambda) + U^2}. \tag{6.19}$$

The maximum value of q is in the center of the band gap ($\epsilon = 0$); for $|U| \ll 2\,\lambda$, it is $q'^2_{\max} \approx U^2/(4\lambda)$. At the band edges ($\epsilon = \pm U$), $q = 0$. q is the characteristic length of an exponentially decaying wave function. Such solutions occur at surfaces or interfaces. For larger band gaps, the localization length is smaller (larger q) (Fig. 6.4).

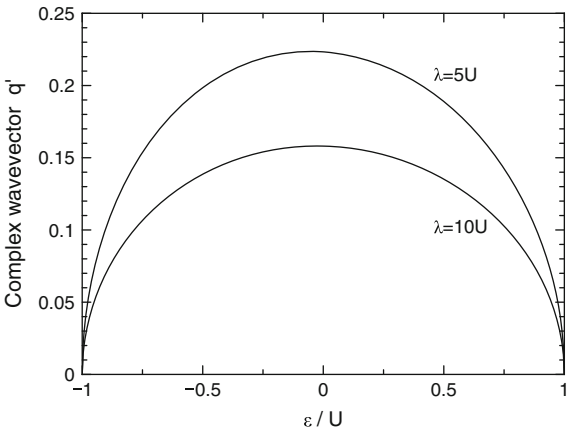


Fig. 6.4 Complex band structure $q'(\epsilon)$ according to (6.19) for two different values of λ/U

Solution in the Vicinity of the Zone Boundary

For $\mathbf{K}$ in the vicinity of the zone boundary the solutions (6.15) can be developed. Therefore, we use the (small) distance from the zone boundary $\widetilde{\mathbf{K}} = \mathbf{K} - \mathbf{G}/2$. With $\lambda = (\hbar^2/2m)\,(G^2/4)$ we rewrite still exactly (6.15):

$$E_{\pm}\left(\widetilde{\mathbf{K}}\right) = \frac{\hbar^2}{2m}\left(\frac{1}{4}\mathbf{G}^2 + \widetilde{\mathbf{K}}^2\right) \pm \left(4\lambda\,\frac{\hbar^2\widetilde{\mathbf{K}}^2}{2m} + U^2\right)^{1/2}. \tag{6.20}$$

For small $\widetilde{\mathbf{K}}$ with $\frac{\hbar^2\mathbf{G}\widetilde{\mathbf{K}}}{2m} \ll |U|$, the energy is then approximately given by

$$E_{\pm}\left(\widetilde{\mathbf{K}}\right) \cong \lambda \pm U + \frac{\hbar^2\widetilde{\mathbf{K}}^2}{2m}\left(1 \pm \frac{2\,\lambda}{U}\right). \tag{6.21}$$

Thus the energy dispersion in the vicinity of the zone boundary is parabolic. The lower state has a negative curvature, the upper state a positive curvature. The curvature is

$$m^* = m\,\frac{1}{1 \pm 2\lambda/U} \approx \pm m\,\frac{U}{2\lambda}, \tag{6.22}$$

and will be later related to the *effective mass*. The approximation in (6.22) is valid for $|U| \ll 2\lambda$. We note that in our simple model m^* increases linearly with increasing band gap $2U$ (see Fig. 6.28 for experimental data).

6.2.4 Kramer's Degeneracy

$E_{\mathrm{n}}(\mathbf{k})$ is the dispersion in a band. The time-reversal symmetry (Kramer's degeneracy) implies

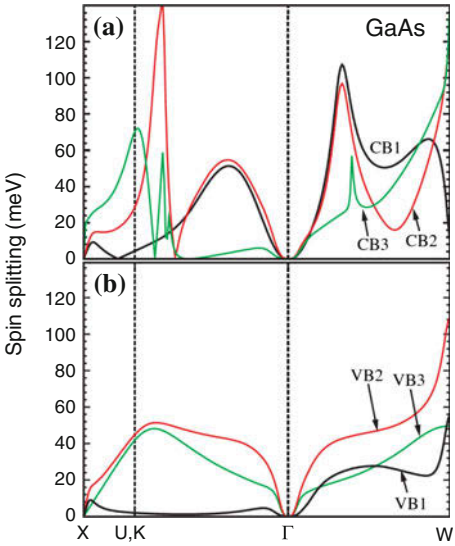


Fig. 6.5 Theoretical calculation of the spin splitting of (**a**) the three *lowest* conduction bands (CB1, CB2, and CB3) and (**b**) the *top* three valence bands (VB1, VB2, and VB3) of GaAs. Adapted from [401]

$$E_{n\uparrow}(\mathbf{k}) = E_{n\downarrow}(-\mathbf{k}), \tag{6.23}$$

where the arrow refers to the direction of the electron spin. If the crystal is symmetric under inversion, we have additionally

$$E_{n\uparrow}(\mathbf{k}) = E_{n\uparrow}(-\mathbf{k}). \tag{6.24}$$

With both time reversal and inversion symmetry the band structure fulfills

$$E_{n\uparrow}(\mathbf{k}) = E_{n\downarrow}(\mathbf{k}). \tag{6.25}$$

The inversion symmetry is particularly important for the spin-orbit interaction. In the absence of inversion symmetry, e.g. in (non-centrosymmetric) zincblende crystals (Fig. 3.16b) or in heterostructures (Fig. 12.34b), a spin splitting, e.g. $E_{n\uparrow}(\mathbf{k}) \neq E_{n\downarrow}(\mathbf{k})$, is present. It can be thought of as provoked by an effective magnetic field. Bulk inversion asymmetry (BIA) leads to the Dresselhaus spin splitting [400, 401] that is shown in Fig. 6.5 for GaAs (cmp. Fig. 6.7a). The spin splitting due to structural inversion asymmetry (SIA) is described by the Bychkov-Rashba Hamiltonian [402, 403]. A review on these topics can be found in [404].

6.2.5 *Symmetry Considerations*

In general the symmetry of the lattice is a symmetry of the system's Hamiltonian and thus transfers into the electronic (and other) properties of the semiconductor.

The means to formulate this mathematically is group theory. At a given reciprocal lattice point, the wave function must fulfill the given spatial symmetry. Additional symmetry due to spin and spin-orbit interaction enters via the double-group scheme. This problem has been treated for the 32 point groups (cmp. Table B.2) in [405] and in [406] particularly for the pc, fcc, bcc and hcp lattices. A detailed treatment for the zincblende structure has been given in [407]. The symmetry at particular lattice points is denoted by the irreducible representations of the symmetry group, e.g. by the Γ_i-symbols used in Figs. 6.6, 6.7 or also Fig. 6.38. As an example base functions with the symmetry of the irreducible representations of tetraeder group

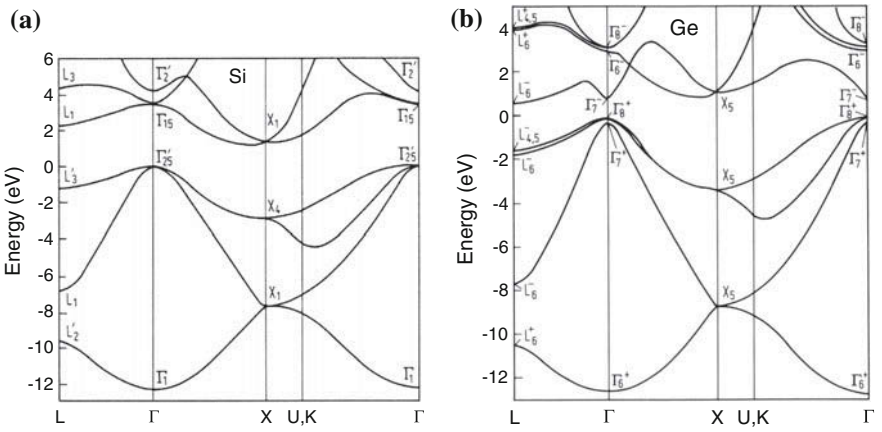


Fig. 6.6 Band structure of (**a**) silicon (indirect) and (**b**) germanium (indirect). In Si, the minima of the conduction band are in the ⟨100⟩ direction, for germanium in the ⟨111⟩ direction. Adapted from [146], based on [409]

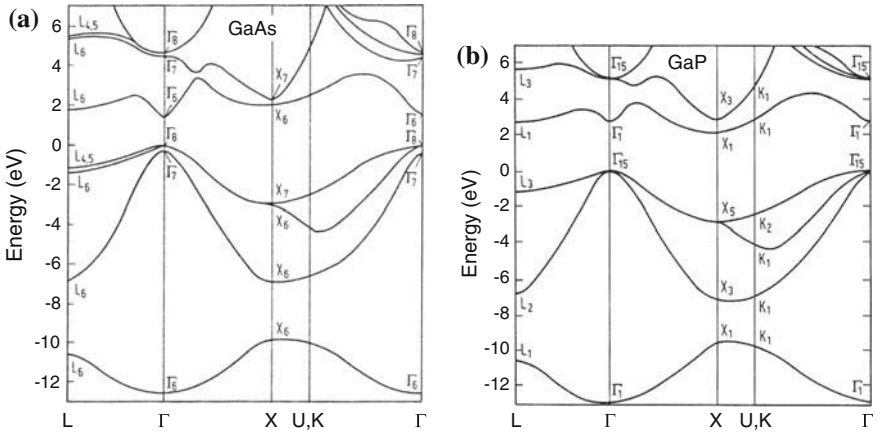


Fig. 6.7 Band structure of (**a**) GaAs (direct) and (**b**) GaP (indirect). For GaAs the minimum of the conduction band is at Γ, for GaP in the ⟨100⟩ direction. Adapted from [146], based on [409]

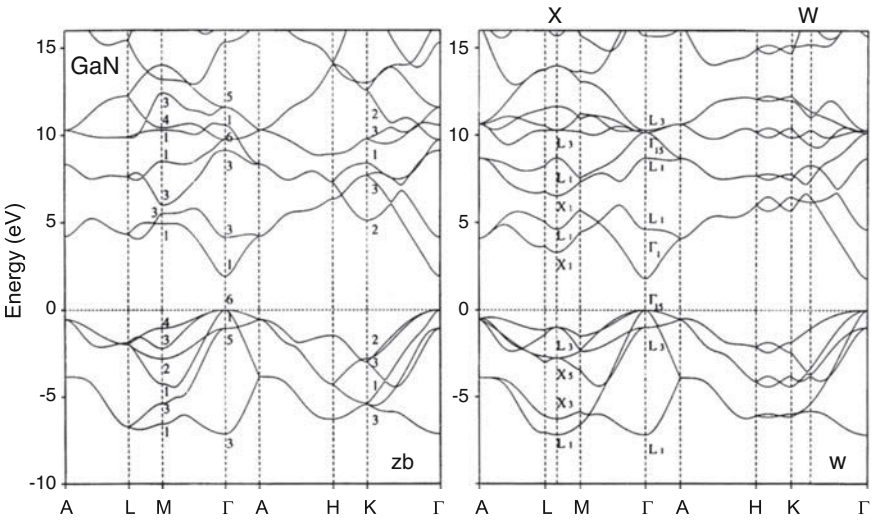


Fig. 6.8 Band structure of GaN (direct) in zincblende (zb) modification (*left*) and wurtzite (w) modification (*right*), both displayed in the wurtzite Brillouin zone to facilitate comparison

Table 6.1 Representations of the tetraeder group (zincblende structure) in molecular, BSW [408] and Koster [405] notation and corresponding base functions (c.p.: cyclic permutations)

Molecular	BSW	Koster	Base functions
A_1	Γ_1	Γ_1	$x\,y\,z$
A_2	Γ_2	Γ_2	$x^4\,(y^2 - z^2)$ + c.p.
E	Γ_{12}	Γ_3	$2\,z^2 - (x^2 + y^2)$, $(x^2 - y^2)$
T_2	Γ_{15}	Γ_4	x, y, z
T_1	Γ_{25}	Γ_5	$z\,(x^2 - y^2)$ and c.p.

are listed in Table 6.1. With the knowledge of the wave functions at the points of high symmetry, it is possible to deduce the general nature of the energy bands in the vicinity of such symmetry points.

6.3 Band Structures of Selected Semiconductors

In the following, the band structures of various important and prototype semiconductors are discussed. The band below the energy gap is called the valence band; the band above the gap is the conduction band. The band gap ΔE_{cv}, mostly denoted as E_{g}, is the energy separation between the highest valence-band state and the lowest conduction-band state. The maximum of the valence band is for most semiconductors at the Γ point.

6.3.1 Silicon

For silicon, an elemental semiconductor, (Fig. 6.6a) the minimum of the conduction band is located close to the X point at $0.85\pi/a$ in the $\langle 100 \rangle$ direction. Thus, it is not at the same point in **k** space as the top of the valence band. Such a band structure is called *indirect*. Since there are six equivalent $\langle 100 \rangle$ directions, there are six equivalent minima of the conduction band.

6.3.2 Germanium

Germanium, another elemental semiconductor, (Fig. 6.6b) also has an indirect band structure. The conduction minima are at the L point in the $\langle 111 \rangle$ direction. Due to symmetry there are eight equivalent conduction-band minima.

6.3.3 GaAs

GaAs (Fig. 6.7a) is a compound semiconductor with a *direct* band gap since the top of the valence band and the bottom of the conduction band are at the same position in **k** space (at the Γ point). The next highest (local) minimum in the conduction band is close to the L point.

6.3.4 GaP

GaP (Fig. 6.7b) is an indirect compound semiconductor. The conduction-band minima are along the $\langle 100 \rangle$ directions.

6.3.5 GaN

GaN (Fig. 6.8) is a direct semiconductor that has wurtzite structure but can also occur in the metastable cubic (zincblende) phase.

6.3.6 Lead Salts

The band gap of PbS (Fig. 6.9), PbSe and PbTe is direct and located at the L point. The lead chalcogenide system shows the anomaly that with increasing atomic weight

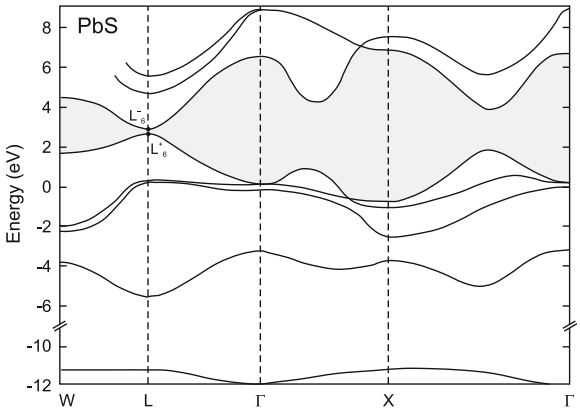


Fig. 6.9 Calculated band structure of PbS (direct). The energy gap is at the L point. The forbidden band is shown in *grey*. Adapted from [410]

the band gap does not decrease monotonically. At 300 K, the band gaps are 0.41, 0.27 and 0.31 eV for PbS, PbSe and PbTe, respectively.

6.3.7 MgO, ZnO, CdO

Cadmium oxide is a cubic semiconductor in the rocksalt structure. Due to symmetry considerations, coupling (repulsion) of oxygen 2p- and cadmium 3d-orbitals does not occur at the zone center in the rocksalt structure. Repulsion occurs though away from the Γ-point and therefore the valence band maximum is not located at the zone center (Fig. 6.10). Thus CdO is an indirect semiconductor. A similar effect would occur in rs-ZnO due to zinc 3d orbitals; however, ZnO has wurtzite structure for

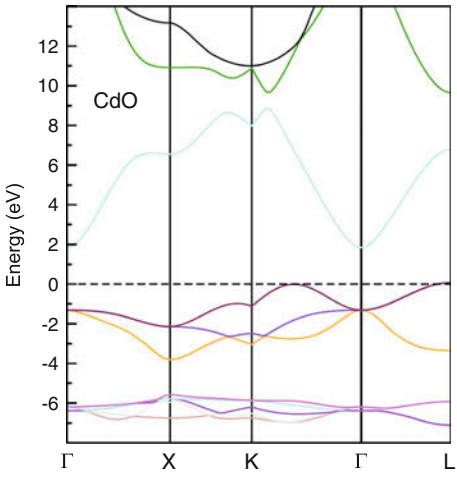


Fig. 6.10 Calculated indirect band structure of CdO. The *top* of the valence band is at $E = 0$. Adapted from [412]

which p–d coupling at the Γ-point is allowed; thus ZnO is direct. In MgO, Mg of course only possesses populated s- and p-orbitals and no such repulsion is present; thus MgO even with its rocksalt structure is also direct [411].

6.3.8 Chalcopyrites

The experimental band gaps of a number of chalcopyrite semiconductors are listed in Table 6.2. The band structures of $CuAlS_2$, $CuAlSe_2$, and $CuGaSe_2$ are compared in Fig. 6.11.

Table 6.2 Band gaps of various chalcopyrite semiconductors

Material	E_g (eV)		E_g (eV)		E_g (eV)
$CuAlS_2$	3.5	$CuGaS_2$	2.5	$CuInS_2$	1.53
$CuAlSe_2$	2.71	$CuGaSe_2$	1.7	$CuInSe_2$	1.0
$CuAlTe_2$	2.06	$CuGaTe_2$	1.23	$CuInTe_2$	1.0–1.15
$AgAlS_2$	3.13	$AgGaS_2$	2.55	$AgInS_2$	1.87
$AgAlSe_2$	2.55	$AgGaSe_2$	1.83	$AgInSe_2$	1.24
$AgAlTe_2$	2.2	$AgGaTe_2$	1.1–1.3	$AgInTe_2$	1.0
$ZnSiP_2$	2.96	$ZnGeP_2$	2.34	$ZnSnP_2$	1.66
$ZnSiAs_2$	2.12	$ZnGeAs_2$	1.15	$ZnSnAs_2$	0.73
$CdSiP_2$	2.45	$CdGeP_2$	1.72	$CdSnP_2$	1.17
$CdSiAs_2$	1.55	$CdGeAs_2$	0.57	$CdSnAs_2$	0.26

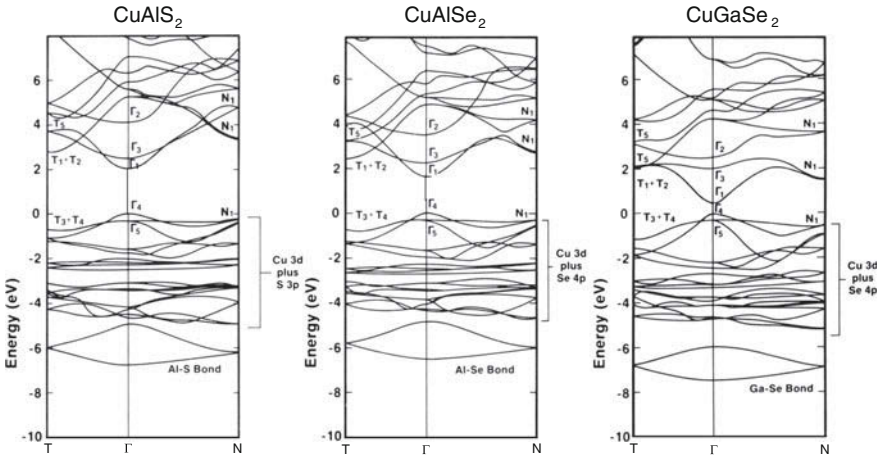


Fig. 6.11 Calculated band structures of $CuAlS_2$, $CuAlSe_2$, and $CuGaSe_2$. The absolute values of the gap energies are incorrect due to LDA calculation. Adapted from [413]

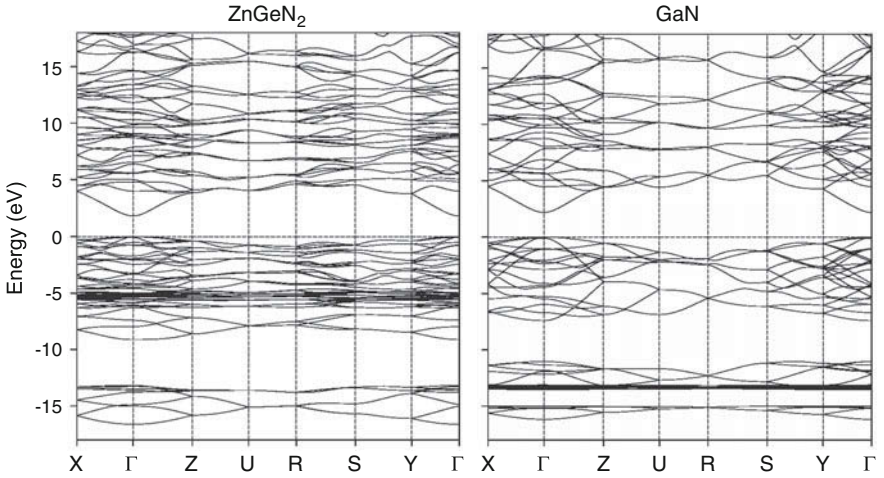


Fig. 6.12 Calculated (within LDA) band structures of $ZnGeN_2$ and its related III–V compound GaN, both displayed in the chalcopyrite (orthorhombic) Brillouin zone to facilitate comparison. Adapted from [414]

In Fig. 6.12, the theoretical band structure of GaN and its closest related chalcopyrite $ZnGeN_2$ are compared, both shown in the chalcopyrite (orthorhombic) Brillouin zone. The band gap of $ZnGeN_2$ is smaller than that of GaN and the difference of 0.4 eV is fairly well reproduced by the calculation[2] (giving 0.5 eV).

6.3.9 Spinels

The band structure of spinels (in particular $CdIn_2S_4$) has been discussed in [415], the band structure of ZnM_2O_4 has been calculated in [412] for (M = Co, Rh, Ir) (Fig. 6.13) and in [416] for (M = Al, Ga, In).

6.3.10 Delafossites

In Fig. 6.14, the theoretical band structures of the delafossites $CuAlO_2$, $CuGaO_2$, and $CuInO_2$ are shown. The maximum of the valence band is not at Γ but near the F point. The direct band gap at Γ decreases for the sequence Al → Ga → In, similar to the trend for AlAs, GaAs and InAs. The direct band gap at F and L, causing the optical absorption edge, increases, however (experimental values are 3.5, 3.6, and 3.9 eV).

[2] Due to the local density approximation (LDA) the absolute values of the band gaps are too small by about 1 eV.

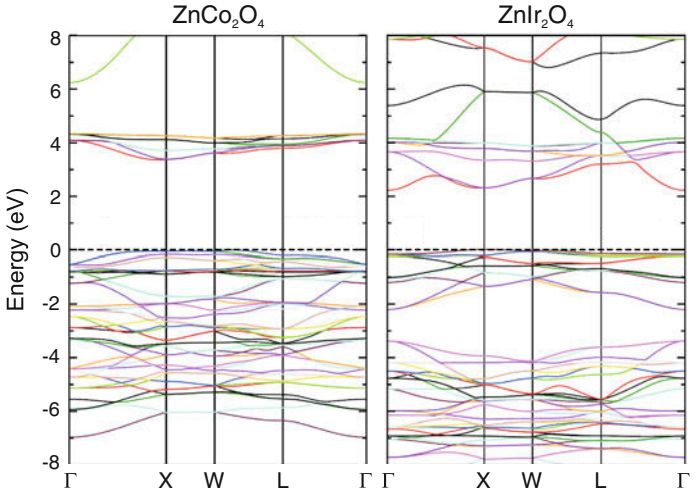


Fig. 6.13 Calculated band structures of $ZnCo_2O_4$ and $ZnIr_2O_4$. Adapted from [412]

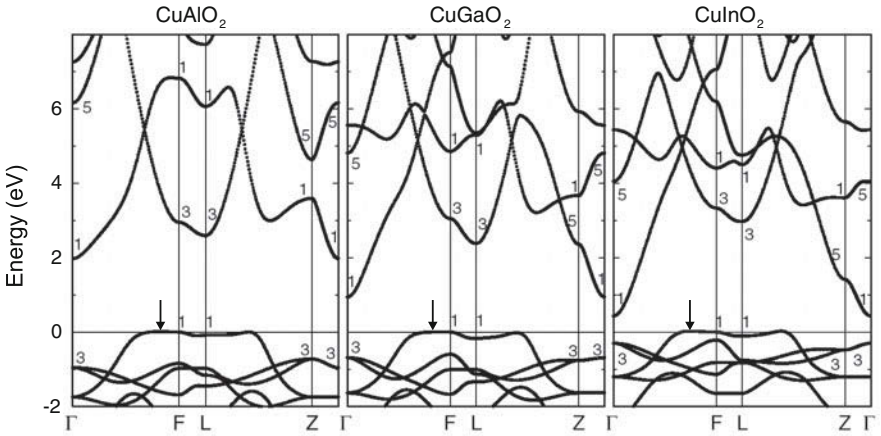


Fig. 6.14 Band structures of $CuAlO_2$, $CuGaO_2$, and $CuInO_2$, calculated with LDA (underestimating the absolute value of the band gaps). The *arrows* denote the maximum of the valence band that has been set to zero energy for each material. Adapted from [200]

6.3.11 Perovskites

The calculated band structure of $BaTiO_3$ in the tetragonal phase is shown in Fig. 6.15. The minimum of the conduction band is at the Γ point. The maximum of the valence band is not at the Γ point but at the M point. The band gap of the LDA[3] calculation is too small (2.2 eV) compared to the experimental value ∼3.2 eV.

[3] local density approximation.

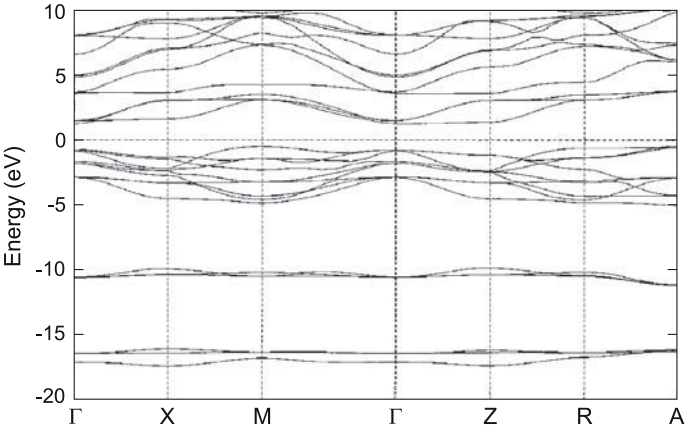


Fig. 6.15 Calculated energy band structure of $BaTiO_3$ along the major symmetry directions. The Fermi level (E_F) is set at zero energy. Adapted from [417]

6.4 Systematics of Semiconductor Band Gaps

The trends with regard to the size of the band gap for elemental, III–V and II–VI semiconductors can essentially be understood in terms of the bond strength and ionicity. In Fig. 6.16, the band gaps of many important semiconductors are shown as a function of the lattice constant. For elemental semiconductors, the band gap decreases with reduced bond strength, i.e. lattice constant (C→Si→Ge). A similar trend exists both for the III–V and the II–VI semiconductors.

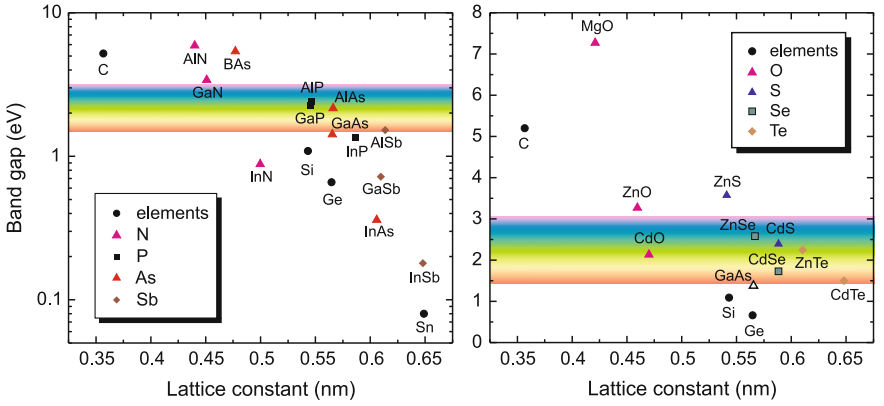


Fig. 6.16 Band gaps as a function of the lattice constant for various elemental, III–V and II–VI semiconductors. The lattice constant of wurtzite semiconductors has been recalculated for a cubic cell

For the same lattice constant, the band gap increases with increasing ionicity, i.e. IV–IV→III–V→II–VI→I–VII. A typical example is the sequence Ge→ GaAs→ ZnSe→CuBr, for which all materials have almost the same lattice constant, and the band gaps increase 0.66 eV→1.42 eV→2.7 eV→2.91 eV.

This behavior can be understood within the framework of a modified Kronig–Penney model [418] (Appendix F). Double potential wells ($b/a = 3$) are chosen to mimic the diatomic planes along the ⟨111⟩ direction in the zincblende structure (Fig. 6.17a). Symmetric wells (depth P_0) are chosen to model covalent semiconductors and asymmetric wells with depths $P_0 \pm \Delta P$ to model partially ionic semiconductors. Results are shown in Fig. 6.17a for $P_0 = -3$. With increasing asymmetry, i.e. ionicity, the band gap increases, mostly due to a downward shift of the valence band. The case of III–V (II–VI) semiconductors is reached for $\Delta P \approx 2$ (4).

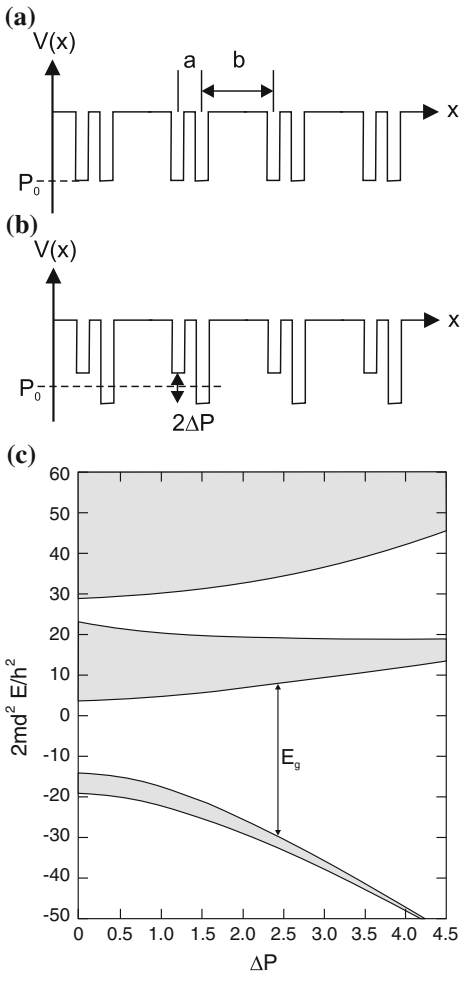


Fig. 6.17 Kronig–Penney model (along ⟨111⟩, $b/a = 3$) for a (**a**) IV–IV semiconductor and (**b**) for a III–V (or II–VI) semiconductor, (**c**) resulting band structure ($P_0 = -3$). d is the lattice constant ($b + a$). Adapted from [418]

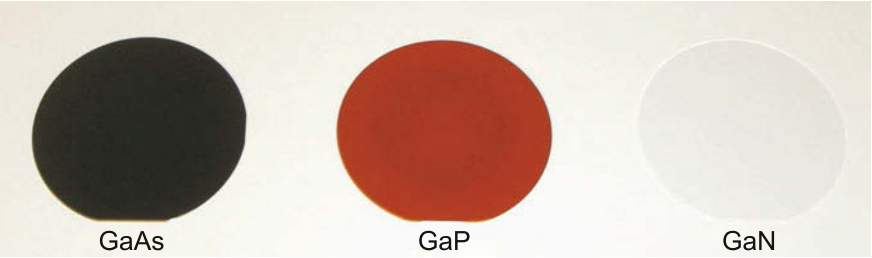


Fig. 6.18 Optical image of two inch wafers of GaAs (*left*), GaP (*center*) and ZnO (*right*). A GaN wafer would look like the ZnO wafer

Table 6.3 Comparison of band gap, lattice constant and ionicity of gallium–group V semiconductors for various anions

Anion	E_g (eV)	a_0 (nm)	f_i
N	3.4	0.45	0.50
P	2.26	0.545	0.33
As	1.42	0.565	0.31
Sb	0.72	0.61	0.26

Lattice constant for GaN has been recalculated for a cubic cell

In Fig. 6.18, the visual impression of 2" wafers of GaAs, GaP and GaN on white paper is shown. GaAs (and GaSb) is opaque since the band gap is below the visible spectral range. GaP has a band gap in the green and appears red, GaN has a band gap in the ultra-violet and thus appears transparent. As can be seen from Table 6.3, the anion sequence Sb, As, P, and N leads to smaller lattice constant and higher ionicity. A notable deviation from this rule is InN whose band gap (0.7 eV) is much smaller than that of InP [419].

6.5 Alloy Semiconductors

In alloy semiconductors, the size of the band gap and the character of the band gap will depend on the composition. The dependence of the band gap on the ternary composition is mostly nonlinear and can usually be expressed with a bowing parameter b that is mostly positive. For a compound $A_xB_{1-x}C$ the band gap is written as

$$E_g(A_xB_{1-x}C) = E_g(BC) + x\left[E_g(AC) - E_g(BC)\right] - b\,x\,(1-x). \qquad (6.26)$$

Even on the virtual crystal approximation (VCA) level (Sect. 3.7.3) a nonzero bowing parameter b is predicted. However, a more thorough analysis shows that the bowing cannot be treated adequately within VCA and is due to the combined effects of volume deformation of the band structure with the alloy lattice constant, charge exchange in

the alloy with respect to the binary end components, a structural contribution due to the relaxation of the cation–anion bond lengths in the alloy and a small contribution due to disorder [420]. The discussion of Sect. 6.11.3 is related.

The Si_xGe_{1-x} alloy has diamond structure for all concentrations and the position of the conduction-band minimum in **k**-space switches from L to X at about $x = 0.15$ (Fig. 6.19a). However, for all concentrations the band structure is indirect. The

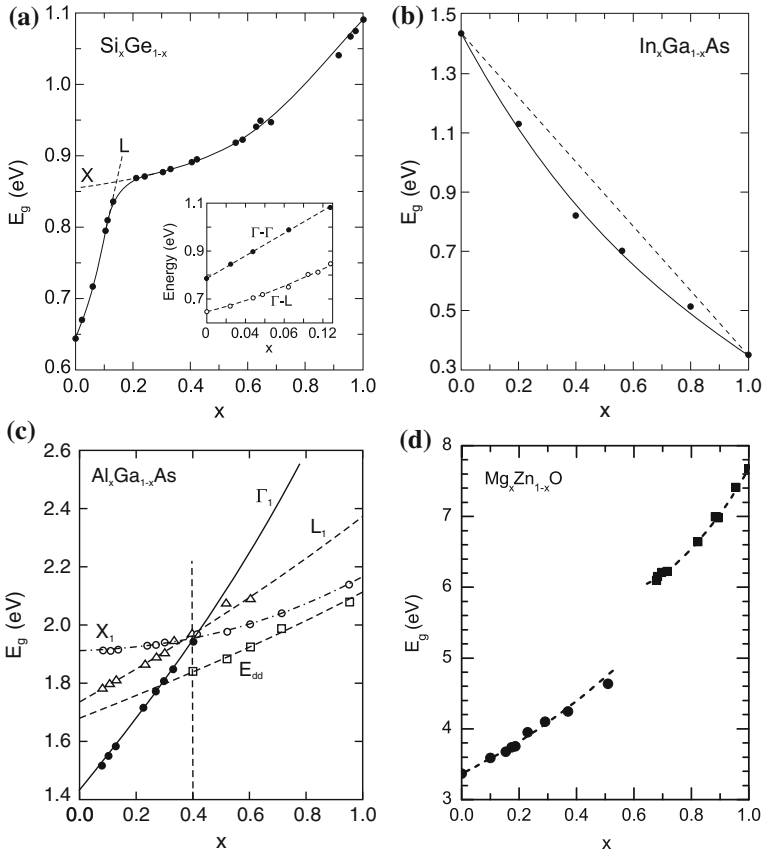


Fig. 6.19 (**a**) Band gap of Si_xGe_{1-x} alloy ($T = 296\,K$) with a change from the conduction-band minimum at L (Ge-rich) to X. The *inset* depicts the transition energy of the indirect (Γ–L) and direct (Γ–Γ) absorption edge for low Si content. Adapted from [424]. (**b**) Band gap (at room temperature) of $In_xGa_{1-x}As$. The *solid line* is an interpolation with bowing ($b = 0.6\,eV$) and the *dashed line* is the linear interpolation. Data from [421]. (**c**) Band gap (at room temperature) in the ternary system $Al_xGa_{1-x}As$. For $x < 0.4$ the alloy is a direct, for $x > 0.4$ an indirect, semiconductor. E_{dd} denotes the energy position of a deep donor (cf. Sect. 7.7.6). Adapted from [425]. (**d**) Band gap (at room temperature) in the ternary system $Mg_xZn_{1-x}O$. Data (from spectroscopic ellipsometry [426, 427]) are for hexagonal wurtzite phase (*circles*), and Mg-rich cubic rocksalt phase (*squares*). *Dashed lines* are fits to data with a different bowing parameter for each phase

$In_xGa_{1-x}As$ alloy has zincblende structure for all compositions. The band gap is direct and decreases with a bowing parameter of $b = 0.6\,eV$ [421] (Fig. 6.19b). This means that for $x = 0.5$ the band gap is 0.15 eV smaller than expected from a linear interpolation between GaAs and InAs, as reported by various authors [422].

If one binary end component has a direct band structure and the other is indirect, a transition occurs from direct to indirect at a certain composition. An example is $Al_xGa_{1-x}As$ where GaAs is direct and AlAs is indirect. For all concentrations the crystal has zincblende structure. In Fig. 6.19c, the Γ, L and X conduction-band minima for ternary $Al_xGa_{1-x}As$ are shown. Up to an aluminum concentration of $x = 0.4$ the band structure is direct. Above this value the band structure is indirect with the conduction-band minimum being at the X point. The particularity of $Al_xGa_{1-x}As$ is that the lattice constant is almost independent of x. For other alloys lattice match to GaAs or InP substrates is only obtained for specific compositions, as shown in Fig. 6.20.

Bowing in the group-III–nitride system has been discussed in [423].

If the two binary end components have different crystal structure, a phase transition occurs at a certain composition (range). An example is $Mg_xZn_{1-x}O$, where ZnO has wurtzite structure and MgO has rocksalt structure. The band gap is shown in Fig. 6.19d. In this case, each phase has its own bowing parameter.

All alloys of Fig. 6.19b–d have mixed cations. The band gap also varies upon anion substitution in a similar way as shown in Fig. 6.21 for ternary alloys with the cation Zn and the chalcogenides S, Se, Te and O.

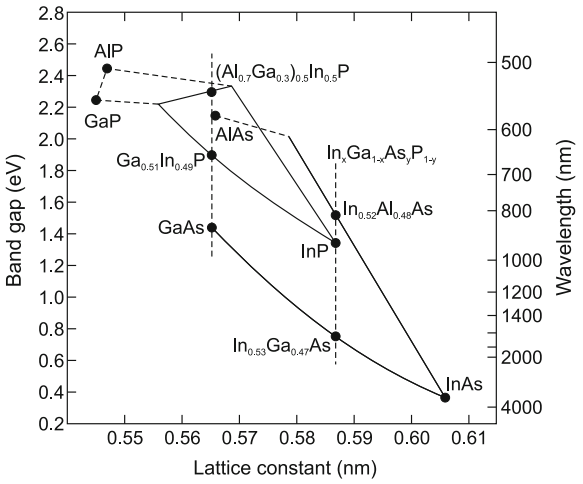


Fig. 6.20 Band gap versus lattice constant for $Ga_xIn_{1-x}P$ and $Al_xIn_{1-x}P$ (lattice matched to GaAs) as well as for $In_xAl_{1-x}As$ and $In_xGa_{1-x}As$ alloys (lattice matched to InP)

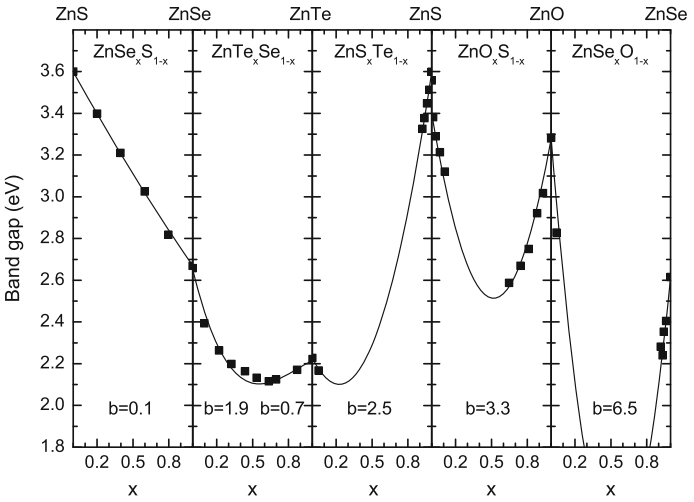


Fig. 6.21 Band gap of various Zn-based alloys. The *lines* are fits with (6.26), the bowing parameter b is labeled. Data for Zn(S,Se,Te) from [428], for Zn(O,Se/Te) from [429]

6.6 Amorphous Semiconductors

Since the crystal lattice in an amorphous semiconductor is not periodic, the concept of **k**-space and the related concepts such as band structure $E(\mathbf{k})$ break down at least partially. The density of states, however, remains a meaningful and useful quantity (Sect. 6.12.2).

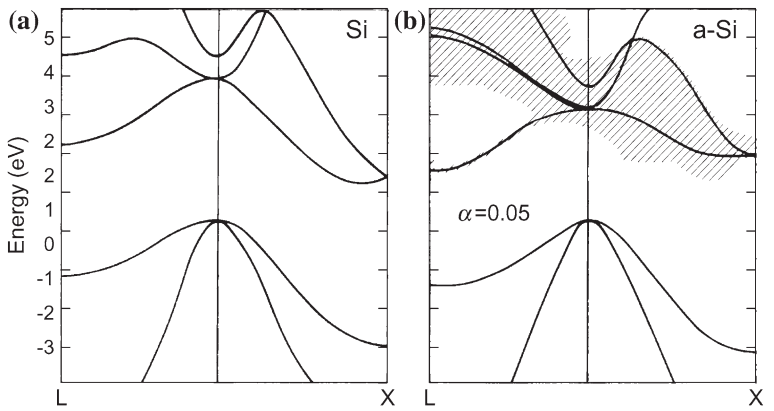


Fig. 6.22 (**a**) Calculated band structure of crystalline silicon. (**b**) Calculated band structure of amorphous silicon with $\alpha = 0.05$ (cf. (3.7)). The *solid lines* denote the real part of the energy, the *shaded areas* denote the regions with a width of twice the imaginary part of the energies centered around the real part. Adapted from [431]

In a perfectly crystalline semiconductor the eigenenergies of the states in the bands are real. An amorphous semiconductor can be modeled using a spectrum of complex energies [430]. In Fig. 6.22 the band structure of crystalline silicon is shown next to that calculated for amorphous silicon with $\alpha = 0.05$.

6.7 Temperature Dependence of the Band Gap

The band gap typically decreases with increasing temperature (see Fig. 6.23 for Si and GaAs). The reasons for this are the change of electron–phonon interaction and the expansion of the lattice. The temperature coefficient may be written as

$$\left(\frac{\partial E_g}{\partial T}\right)_p = \left(\frac{\partial E_g}{\partial T}\right)_V - \frac{\alpha}{\beta}\left(\frac{\partial E_g}{\partial p}\right)_T , \tag{6.27}$$

where α is the volume coefficient of thermal expansion and β is the volume compressibility. A recommendable discussion of the thermodynamic role of the band gap as chemical potential for the mass action law (7.12), entropy contributions and its temperature dependence can be found in [432].

An anomaly is present for the lead salts (PbS, PbSe, PbTe) for which the temperature coefficient is positive (Fig. 6.24a). Theoretical calculations [434] show that both terms in (6.27) are positive for the lead salts. The L_6^+ and L_6^- levels (see Fig. 6.9) shift as a function of temperature in such a way that their separation increases (Fig. 6.24b).

Also in copper and silver halides [435, 436] (Fig. 6.25a) and chalcopyrites [437] (Fig. 6.25b) the increase of band gap with increasing temperature has been found, sometimes only for a certain temperature range. This effect is attributed to the p-d

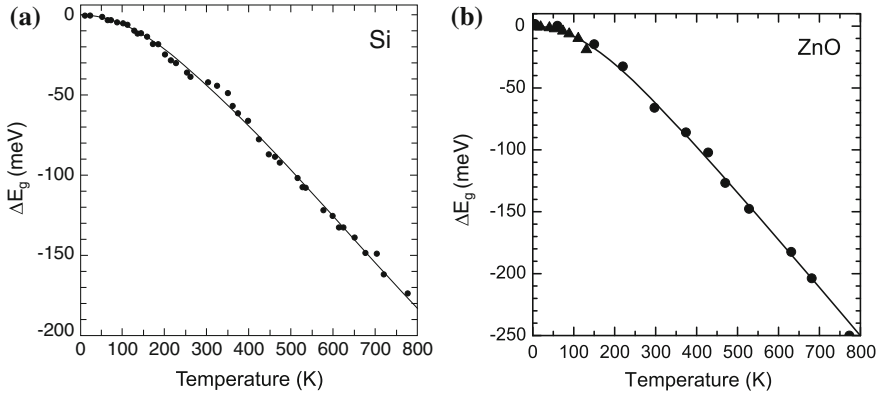


Fig. 6.23 Temperature dependence of the band gap of (**a**) Si (data from [433]) and (**b**) ZnO (experimental data from photoluminescence (*triangles*) and ellipsometry (*circles*)). The *solid lines* are fits with (6.30) and the parameters given in Table 6.4

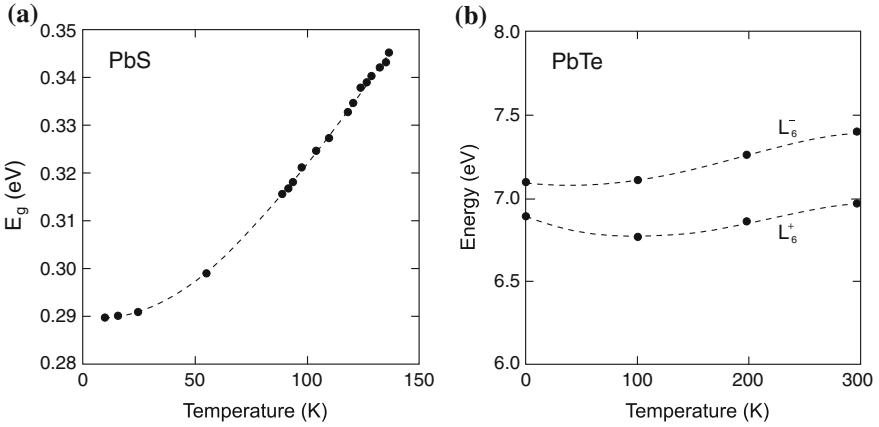


Fig. 6.24 (**a**) Band gap versus temperature for PbS. (**b**) Theoretical position of L_6^+ and L_6^- as a function of temperature for PbTe. Adapted from [410]

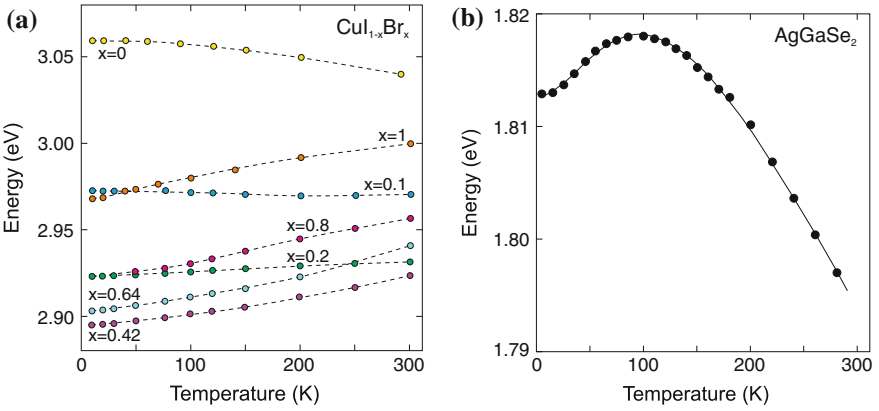


Fig. 6.25 (**a**) Band gap versus temperature for zincblende $CuI_{1-x}Br_x$ alloys with various compositions x (including binary CuI and CuBr) as labeled. *Dashed lines* are guide to the eyes. Adapted from [435]. (**b**) Band gap versus temperature for chalcopyrite $AgGaSe_2$. *Solid line* is fit with two-oscillator Bose-Einstein model. Adapted from [437]

electron hybridization in the valence band with Cu 3d electrons and to even stronger effect with Ag 4d electrons.

For many semiconductors the temperature dependence can be described with the empirical, three-parameter Varshni formula [438],

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{T + \beta}, \tag{6.28}$$

Table 6.4 Parameters for the temperature dependence of the band gap (6.29) and (6.30) for various semiconductors

	α (10^{-4} eV/K)	Θ (K)	Δ	α_B (10^{-4} eV/K)	Θ_B (K)
Si	3.23	446	0.51	2.82	351
Ge	4.13	253	0.49		
GaAs	4.77	252	0.43	5.12	313
InP	3.96	274	0.48		
InAs	2.82	147	0.68		
ZnSe	5.00	218	0.36		
ZnO	3.8	659	0.54		

where $E_g(0)$ is the band gap at zero temperature. A more precise and physically motivated formula (based on a Bose–Einstein phonon model) has been given in [439]

$$E_g(T) = E_g(0) - 2\,\alpha_B\,\Theta_B \left[\coth\left(\frac{\Theta_B}{2T}\right) - 1\right], \tag{6.29}$$

where α_B is a coupling constant and $k\Theta_B$ is a typical phonon energy; typical values are given in Table 6.4. This model reaches a better description of the fairly flat dependence at low temperatures.

The more elaborate model of [440] takes into account a more variable phonon dispersion, including optical phonons, and proposes the four-parameter formula

$$E_g(T) = E_g(0) - \alpha\,\Theta \left[\frac{1 - 3\,\Delta^2}{\exp(2/\gamma) - 1} + \frac{3\,\Delta^2}{2}\left(\sqrt[6]{1+\beta} - 1\right)\right] \tag{6.30}$$

$$\beta = \frac{\pi^2}{3\,(1+\Delta^2)}\,\gamma^2 + \frac{3\Delta^2 - 1}{4}\,\gamma^3 + \frac{8}{3}\,\gamma^4 + \gamma^6$$

$$\gamma = 2\,T/\Theta,$$

where α is the high-temperature limiting magnitude of the slope (of the order of several 10^{-4} eV/K), Θ is an effective average phonon temperature and Δ is related to the phonon dispersion. Δ takes typically values between zero (Bose–Einstein model) and 3/4 [440].

6.8 Isotope Dependence of the Band Gap

The band edge slightly depends on the isotope composition of semiconductor, as shown for GaAs in Fig. 6.26. The effect is discussed in detail in [441].

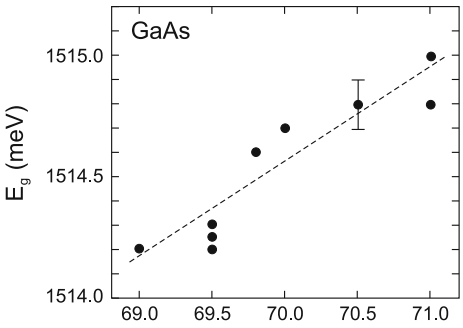


Fig. 6.26 Band gap of GaAs (at $T = 10\,\mathrm{K}$) as a function of the Ga isotope content. *Dashed line* is linear fit. Adapted from [441]

6.9 Electron Dispersion

6.9.1 Equation of Electron Motion

The equation of motion for the electron in the band structure is no longer given by Netwon's law $\mathbf{F} = \mathrm{d}(m\mathbf{v})/\mathrm{d}t$ as in vacuum. Instead, the propagation of quantum-mechanical electron wave packets has to be considered. Their group velocity is given by ($v_\mathrm{g} = \partial\omega/\partial k$)

$$\mathbf{v} = \frac{1}{\hbar}\,\nabla_\mathbf{k} E(\mathbf{k}), \tag{6.31}$$

where $\nabla_\mathbf{k}$ is the gradient with respect to $\mathbf{k}$. Through the dispersion relation the influence of the crystal and its periodic potential on the motion enters the equation.

An electric field $\mathcal{E}$ acts on an electron during the time δt the work $\delta E = -e\mathcal{E} v_g\,\delta t$. This change in energy is related to a change in k via $\delta E = \mathrm{d}E/\mathrm{d}k\,\delta k = \hbar v_g\,\delta k$. Thus, we arrive at $\hbar\,\mathrm{d}k/\mathrm{d}t = -e\mathcal{E}$. For an external force we thus have

$$\hbar\,\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = -e\,\mathbf{E} = \mathbf{F}. \tag{6.32}$$

Thus, the crystal momentum $\mathbf{p} = \hbar\mathbf{k}$ takes the role of the momentum. A more rigorous derivation can be found in [399].

In the presence of a magnetic field $\mathbf{B}$ the equation of motion is:

$$\hbar\,\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = -e\,\mathbf{v}\times\mathbf{B} = -\frac{e}{\hbar}\,(\nabla_\mathbf{k} E)\times\mathbf{B}. \tag{6.33}$$

The motion in a magnetic field is thus perpendicular to the gradient of the energy, i.e. the energy of the electron does not change. It oscillates therefore on a surface of constant energy perpendicular to $\mathbf{B}$.

6.9.2 *Effective Mass of Electrons*

From the free-electron dispersion $E = \hbar^2 k^2/(2m)$ the mass of the particle is inversely proportional to the curvature of the dispersion relation, i.e. $m = \hbar^2/(\mathrm{d}^2 E/\mathrm{d}k^2)$. This relation will now be generalized for arbitrary dispersion relations. The tensor of the effective mass is defined as

$$(m^{*-1})_{ij} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k_i \, \partial k_j}. \tag{6.34}$$

The equation $\mathbf{F} = m^* \dot{\mathbf{v}}$ must be understood as a tensor equation, i.e. for the components of the force $F_i = m^*_{ij} a_j$. Force and acceleration must no longer be collinear. In order to find the acceleration from the force, the inverse of the effective-mass tensor must be used, $\mathbf{a} = (\mathbf{m}^*)^{-1} \mathbf{F}$.

In Fig. 6.27 the energy dispersion of the (lowest) conduction band in a typical semiconductor, the related electron velocity and the effective mass are shown schematically.

In (6.22) the ratio of the effective mass and the free-electron mass is of the order of $m^*/m \approx U/\lambda$, the ratio of the free particle energy and the band gap. For typical semiconductors, the width of the (valence) band is of the order of 20 eV, and the gap is about 0.2–2 eV. Thus, the effective mass is expected to be 10–100 times smaller than the free-electron mass. Additionally, the relation $m^* \propto E_\mathrm{g}$ is roughly fulfilled (Fig. 6.28).

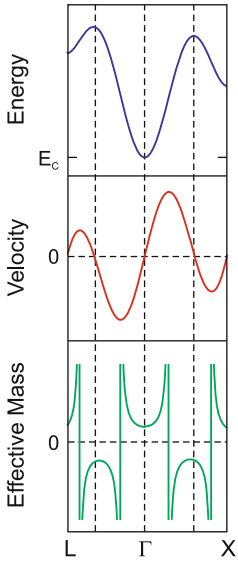


Fig. 6.27 Schematic diagram of the electron dispersion $E(\mathbf{k})$ in a typical semiconductor (*blue*) and corresponding carrier velocity ($\propto \partial E/\partial k$) (*red*) and effective mass ($\propto 1/(\partial^2 E/\partial k^2)$) (*green*)

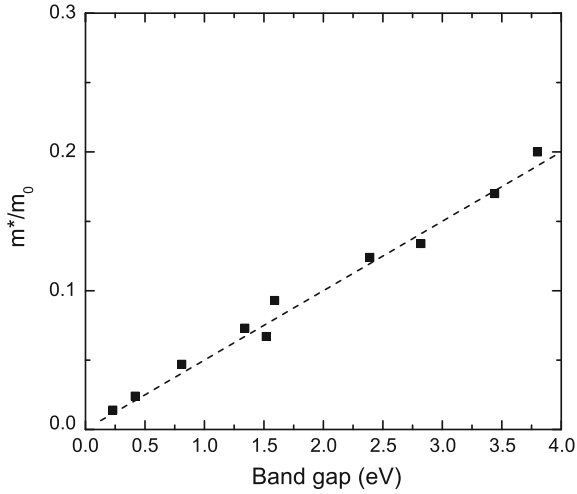


Fig. 6.28 Effective electron mass (in units of the free-electron mass m_0) as a function of the (low-temperature) band gap for several (direct band gap) semiconductors. The *dashed line* fulfills $m^*/m_0 = E_g/20\,\mathrm{eV}$

From so-called $\mathbf{k}\cdot\mathbf{p}$ theory [442] (see Appendix G) the effective electron mass is predicted to be related to the momentum matrix element $\mathbf{p}_{cv}$

$$\mathbf{p}_{cv} = \langle c|\mathbf{p}|v\rangle = \int_{\Omega_0} u^*_{c,\mathbf{k}}(\mathbf{r})\,\mathbf{p}\,u_{c,\,\mathbf{k}}(\mathbf{r})\,\mathrm{d}^3\mathbf{r}, \tag{6.35}$$

with Ω_0 being the unit cell volume and the Bloch functions $|c\rangle$ and $|v\rangle$ of the conduction and valence band, respectively, given as

$$|c\rangle = u_{c,\mathbf{k}_c}(\mathbf{r})\,\exp(\mathrm{i}\,\mathbf{k}_c\mathbf{r}) \tag{6.36a}$$

$$|v\rangle = u_{v,\mathbf{k}_v}(\mathbf{r})\,\exp(\mathrm{i}\,\mathbf{k}_v\mathbf{r})\,. \tag{6.36b}$$

Typically, the $\mathbf{k}$-dependence of the matrix element is small and neglected. The momentum matrix element will also be important for optical transitions between the valence and conduction bands (Sect. 9.5). Other related quantities that are often used are the energy parameter E_P

$$E_P = \frac{2\,|\mathbf{p}_{cv}|^2}{m_0}, \tag{6.37}$$

and the bulk momentum matrix element M_b^2 that is given by

$$M_b^2 = \frac{1}{3}\,|\mathbf{p}_{cv}|^2 = \frac{m_0}{6}\,E_P. \tag{6.38}$$

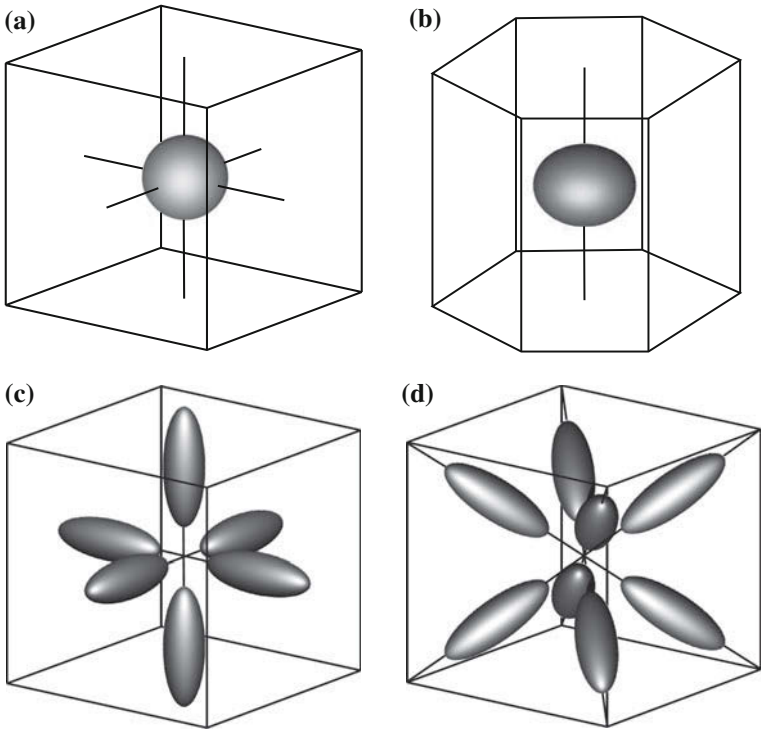


Fig. 6.29 Energy isosurfaces in **k**-space in the vicinity of the conduction-band minima for (**a**) GaAs with isotropic (spehrical) minimum at Γ point, (**b**) ZnO with anisotropic (ellipsoidal) minimum at Γ point (anisotropy exaggerated), (**c**) silicon with six equivalent anisotropic minima ($m_\mathrm{l}/m_\mathrm{t} = 5$ not to scale) along $\langle 100\rangle$ and (**d**) germanium with eight equivalent anisotropic minima along $\langle 111\rangle$. The *cube* indicates the $\langle 100\rangle$ directions for the cubic materials. For the wurtzite material (part **b**) the vertical direction is along [00.1]

The electron mass is given by[4]

$$\begin{aligned}\frac{m_0}{m_\mathrm{e}^*} &= 1 + \frac{E_\mathrm{P}}{3}\left(\frac{2}{E_\mathrm{g}} + \frac{1}{E_\mathrm{g}+\Delta_0}\right) \\ &= 1 + E_\mathrm{P}\,\frac{E_\mathrm{g}+2\Delta_0/3}{E_\mathrm{g}\left(E_\mathrm{g}+\Delta_0\right)} \approx 1 + \frac{E_\mathrm{P}}{E_\mathrm{g}+\Delta_0/3} \approx \frac{E_\mathrm{P}}{E_\mathrm{g}}.\end{aligned} \tag{6.39}$$

Comparison with the fit from Fig. 6.28 yields that E_P is similar for all semiconductors [443] and of the order of 20 eV (InAs: 22.2 eV, GaAs: 25.7 eV, InP: 20.4 eV, ZnSe: 23 eV, CdS: 21 eV).

In silicon there are six conduction-band minima. The surfaces of equal energy are schematically shown in Fig. 6.29c. The ellipsoids are extended along the $\langle 100\rangle$

[4] Δ_0 is the spin-orbit splitting discussed in Sect. 6.10.2.

Table 6.5 Longitudinal direction of effective mass ellipsoid, longitudinal and transverse effective electron mass in several semiconductors

	Long. dir.	m_l	m_t	$m_\mathrm{l}/m_\mathrm{t}$	$m_\mathrm{d,e}$	References
C	⟨100⟩	1.4	0.36	3.9	1.9	[446]
Si	⟨100⟩	0.98	0.19	5.16	1.08	[447]
Ge	⟨111⟩	1.59	0.082	19.4	0.88	[447]
ZnO	[00.1]	0.21	0.25	0.88		[448]
CdS	[00.1]	0.15	0.17	0.9		[449]

Mass values in units of the free electron mass m_0. For the density of states mass $m_\mathrm{d,e}$ see (6.68)

direction because the longitudinal mass (along the Δ path) is larger than the transverse mass in the two perpendicular directions (Table 6.5). For example, the dispersion relation around the [100] minimum is then given as (k_x^0 denotes the position of the conduction-band minimum)

$$E(\mathbf{k}) = \hbar^2 \left(\frac{(k_x - k_x^0)^2}{2\, m_\mathrm{l}} + \frac{k_y^2 + k_z^2}{2\, m_\mathrm{t}} \right). \tag{6.40}$$

For germanium surfaces of constant energy around the eight conduction-band minima in the ⟨111⟩ directions are depicted in Fig. 6.29d. The longitudinal and the transverse masses are again different. For GaAs, the conduction-band dispersion around the Γ point is isotropic, thus the surface of constant energy is simply a sphere (Fig. 6.29a). In wurtzite semiconductors the conduction-band minimum is at the Γ point. The mass along the c-axis is typically smaller than the mass within the (00.1) plane [444] ($m_\mathrm{l}/m_\mathrm{t} \approx 0.8$ for ZnO [445]), see Fig. 6.29b. In [444] also an anisotropy within the (00.1) plane is predicted.

The directional dependence of the mass can be measured with cyclotron resonance experiments with varying direction of the magnetic field. In Fig. 6.30, the field **B** is in the (110) plane with different azimuthal directions. When the (static) magnetic field makes an angle θ with the longitudinal axis of the energy surface, the effective mass is given as [450]

$$\frac{1}{m^*} = \sqrt{\frac{\cos^2\theta}{m_\mathrm{t}^2} + \frac{\sin^2\theta}{m_\mathrm{t}\, m_\mathrm{l}}}. \tag{6.41}$$

6.9.3 *Nonparabolicity of Electron Mass*

The dispersion around the conduction-band minimum is only parabolic for small **k**. The further away the wavevector is from the extremum, the more the actual dispersion deviates from the ideal parabola (see, e.g., Fig. 6.7). This effect is termed

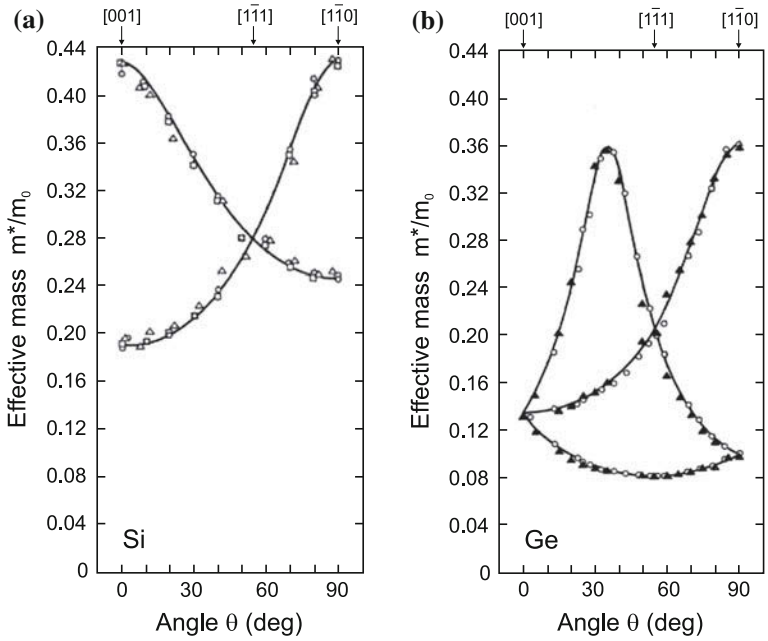


Fig. 6.30 Effective electron mass from cyclotron resonance experiments (at $T = 4\,\mathrm{K}$) on (**a**) Si and (**b**) Ge for the magnetic field in the (110) plane and various azimuthal directions θ. Experimental data and fits (*solid lines*) using (6.41) with (**a**) $m_\mathrm{l} = 0.98$, $m_\mathrm{t} = 0.19$ and (**b**) $m_\mathrm{l} = 1.58$, $m_\mathrm{t} = 0.082$. Data from [447]

nonparabolicity. Typically, the energy increases less quickly with k than in the parabolic model. This can be described in a so-called two-level model with the dispersion relation

$$\frac{\hbar^2 k^2}{2m_0^*} = E\left(1 + \frac{E}{E_0^*}\right), \tag{6.42}$$

where $E_0^* > 0$ parameterizes the amount of nonparabolicity (a parabolic band corresponds to $E_0^* = \infty$). The nonparabolic dispersion for GaAs is shown in Fig. 6.31a. The curvature is reduced for larger $\mathbf{k}$ and thus the effective mass is energy dependent and increases with the energy. Equation (6.42) leads to the energy-dependent effective mass

$$m^*(E) = m_0^*\left(1 + \frac{2E}{E_0^*}\right), \tag{6.43}$$

where m_0^* denotes here the effective mass at $\mathbf{k} = 0$. Theory and experimental data for the effective electron mass of GaAs are shown in Fig. 6.31b.

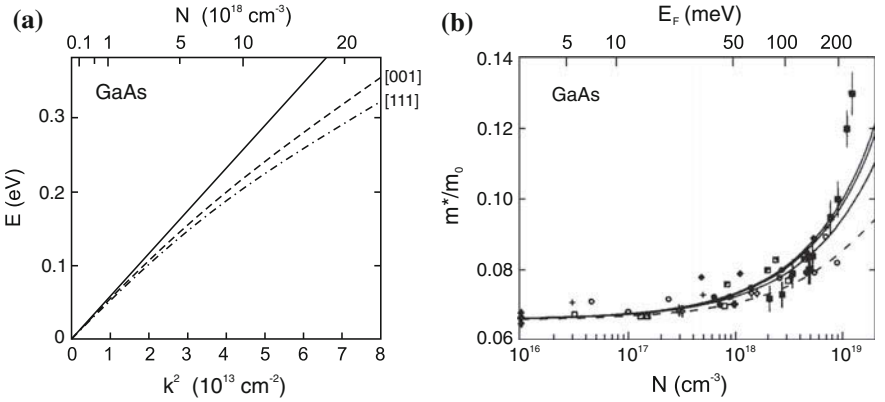


Fig. 6.31 (**a**) Dispersion relations for the conduction band of GaAs. The *solid line* is parabolic dispersion (constant effective mass). The *dashed* (*dash-dotted*) *line* denotes the dispersion for **k** along [001] ([111]) from a five-level $\mathbf{k} \cdot \mathbf{p}$ model (5LM). (**b**) Cyclotron resonance effective mass of electrons in GaAs as a function of the Fermi level (*upper* abscissa) and the corresponding electron concentration (*lower* abscissa). The *dashed line* is from a 2LM according to (6.43) with $E_0^* = 1.52\,\text{eV}$. The *solid lines* are for a 5LM for the three principal directions of the magnetic field. The *symbols* represent experimental data from different sources. Data from [451]

6.10 Holes

6.10.1 Hole Concept

Holes are missing electrons in an otherwise filled band. A Schrödinger-type wave-equation for holes (unoccupied electron states) was derived by Heisenberg [63] to interpret Hall effect data. The hole concept is useful to describe the properties of charge carriers at the top of the valence band. The hole is a new quasi-particle whose dispersion relation is schematically shown in Fig. 6.32 in relation to the dispersion of electrons in the valence band.

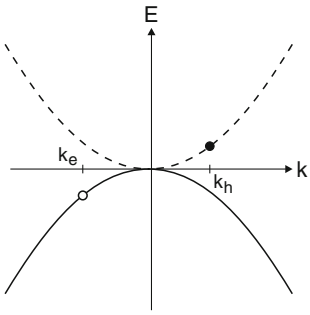


Fig. 6.32 Hole dispersion (*dashed line*) in relation to the electron dispersion in the valence band (*solid line*)

The wavevector of the hole (filled circle in Fig. 6.32) is related to that of the 'missing' electron (empty circle in Fig. 6.32) by $k_\mathrm{h} = -k_\mathrm{e}$. The energy is $E_\mathrm{h}(k_\mathrm{h}) = -E_\mathrm{e}(k_\mathrm{e})$, assuming that $E_\mathrm{V} = 0$, otherwise $E_\mathrm{h}(k_\mathrm{h}) = -E_\mathrm{e}(k_\mathrm{e}) + 2E_\mathrm{V}$. The hole energy is larger for holes that are further away from the top of the valence band, i.e. the lower the energy state of the missing electron. The velocity of the hole, $v_\mathrm{h} = \hbar^{-1}\,\mathrm{d}E_\mathrm{h}/\mathrm{d}k_\mathrm{h}$, is the same, $v_\mathrm{h} = v_\mathrm{e}$, and the charge is positive, $+e$. The effective mass of the hole is positive at the top of the valence band, $m_\mathrm{h}^* = -m_\mathrm{e}^*$. Therefore, the drift velocities of an electron and hole are opposite to each other. The resulting current, however, is the same.

6.10.2 Hole Dispersion Relation

The valence band at the Γ point is 3-fold degenerate. The band developed from the atomic (bonding) p states; the coupling of the spin $s = 1/2$ electrons with the orbital angular momentum $l = 1$ leads to a total angular momentum $j = 1/2$ and $j = 3/2$. The latter states are degenerate at Γ in zincblende bulk material and are called *heavy* holes (hh) for $m_j = \pm 3/2$ and *light* holes (lh) for $m_j = \pm 1/2$ due to their different dispersion (Fig. 6.33a). The two ($m_j = \pm/2$) states of the $j = 1/2$ state are split-off from these states by an energy Δ_0 due to spin-orbit interaction and are called *split-off* (s-o) holes. The spin-orbit interaction increases with increasing atomic order number Z of the anion since the electrons are located preferentially there (Fig. 6.34).

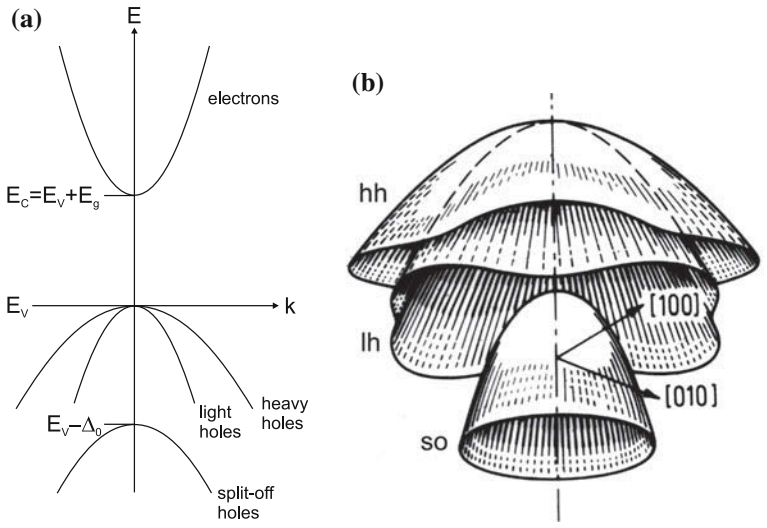


Fig. 6.33 (**a**) Simplified band structure with conduction band and three valence bands and (**b**) three-dimensional visualization (E vs. (k_x, k_y)) of the valence bands of Ge (including warping). Part (**b**) from [453]

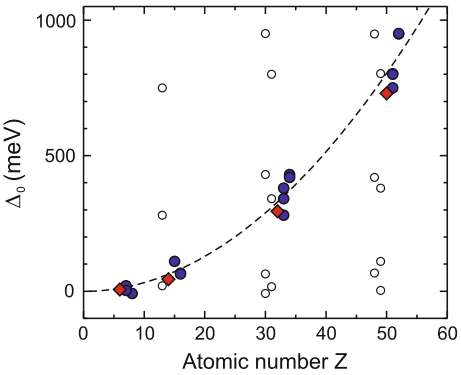


Fig. 6.34 Spin-orbit splitting Δ_0 for elemental (*diamonds*) and various III–V and II–VI (*circles*) semiconductors. The data are plotted as *empty* (*filled*) *circles* as a function of the cation (anion) order number. Obviously, Δ_0 correlates with the anion Z. The *dashed line* is proportional to Z^2

A detailed discussion of the spin-orbit splitting in zincblende semiconductors is given in [452].

All three holes have different mass. In the vicinity of the Γ point the dispersion for heavy and light holes can be described with (+:hh, −:lh)

$$E(\mathbf{k}) = A\,k^2 \pm \sqrt{B^2\,k^4 + C^2\,\left(k_x^2\,k_y^2 + k_y^2\,k_z^2 + k_x^2\,k_z^2\right)}. \tag{6.44}$$

For heavy and light holes there is a dependence of the dispersion, i.e. the mass, in the (001) plane. This effect, sketched in Fig. 6.33b, is called *warping*. The warping at the GaAs valence-band edge is shown in Fig. 6.35. Equation (6.44) can also be expressed in terms of angular coordinates [454].

The s-o holes have the dispersion

$$E(\mathbf{k}) = -\Delta_0 + A\,k^2. \tag{6.45}$$

Values for A, B, C^2 and Δ_0 for a number of semiconductors are given in Table 6.6. The valence-band structure is often described with the Luttinger parameters γ_1, γ_2, and γ_3 that can be represented through A, B, and C via

$$\frac{\hbar^2}{2m_0}\,\gamma_1 = -A \tag{6.46a}$$

$$\frac{\hbar^2}{2m_0}\,\gamma_2 = -\frac{B}{2} \tag{6.46b}$$

$$\frac{\hbar^2}{2m_0}\,\gamma_3 = \frac{\sqrt{B^2 + C^2/3}}{2}. \tag{6.46c}$$

The mass of holes in various directions can be derived from (6.44). The mass along the [001] direction, i.e. $\hbar^2/\left(\partial^2 E(\mathbf{k})/\partial k_x^2\right)$ for $k_y = 0$ and $k_z = 0$, is

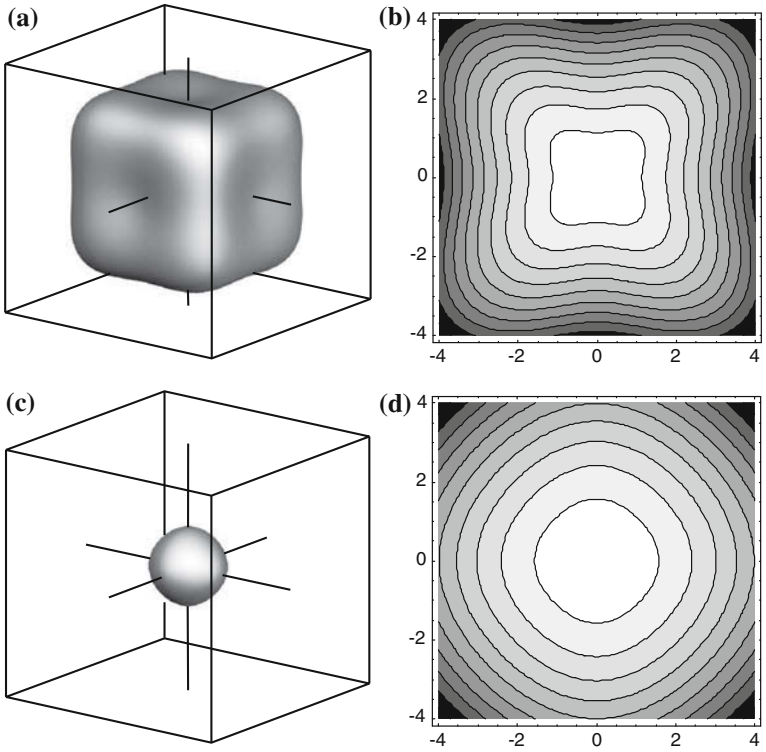


Fig. 6.35 Dispersion at the valence band edge of GaAs for (**a**, **b**) heavy holes and (**c**, **d**) light holes. (**a**, **c**) Constant energy surfaces and (**b**, **d**) isolines in the (k_x, k_y)-plane ((**b**) and (**d**) have different energy scales)

Table 6.6 Valence-band parameters (for (6.44)) A and B in units of $(\hbar^2/2m_0)$, C^2 in units of $(\hbar^2/2m_0)^2$, and Δ_0 in eV

Material	A	B	C^2	Δ_0
C	−4.24	−1.64	9.5	0.006
Si	−4.28	−0.68	24	0.044
Ge	−13.38	−8.5	173	0.295
GaAs	−6.9	−4.4	43	0.341
InP	−5.15	−1.9	21	0.11
InAs	−20.4	−16.6	167	0.38
ZnSe	−2.75	−1.0	7.5	0.43

From [146, 455, 456]

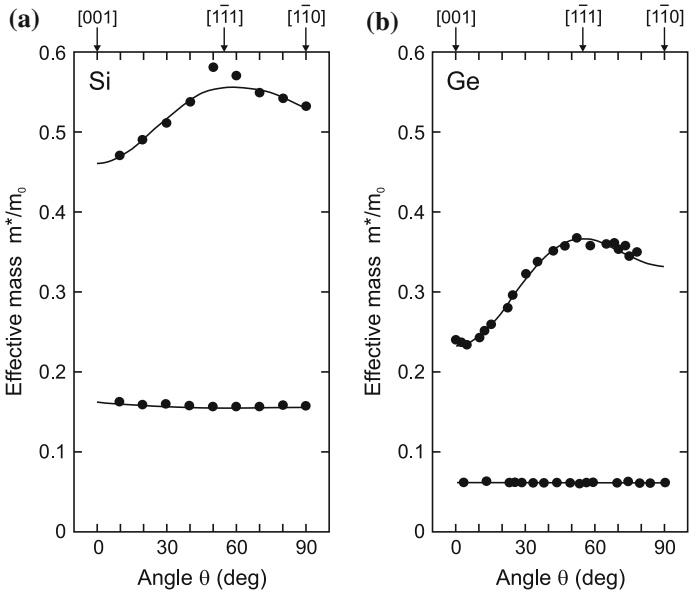


Fig. 6.36 Effective hole masses from cyclotron resonance experiments ($T = 4\,\mathrm{K}$) for heavy and light holes in (**a**) Si and (**b**) Ge for the magnetic field in the (110) plane and various azimuthal directions θ. Experimental data (*symbols*) and fits (*solid lines*) using (6.48). Adapted from [447]

$$\frac{1}{m_{\mathrm{hh}}^{100}} = \frac{2}{\hbar^2}(A+B) \tag{6.47a}$$

$$\frac{1}{m_{\mathrm{lh}}^{100}} = \frac{2}{\hbar^2}(A-B). \tag{6.47b}$$

The anisotropy of hole masses has been investigated with cyclotron resonance experiments (Fig. 6.36). For θ being the angle between the magnetic field and the [001] direction, the effective heavy hole (upper sign) and light hole (lower sign) mass in cubic semiconductors is given by [447]

$$m^* = \frac{\hbar^2}{2}\frac{1}{A \pm \sqrt{B^2 + C^2/4}} \times \left\{ \frac{C^2\,(1 - 3\cos^2\theta)^2}{64\sqrt{B^2 + C^2/4}\left[A \pm \sqrt{B^2 + C^2/4}\right]} + \ldots \right\}. \tag{6.48}$$

For $C^2 = 0$ the hole bands are isotropic, as is obvious from (6.44). In this case $\gamma_2 = \gamma_3$, the so-called spherical approximation. The average of the hole masses over all directions is

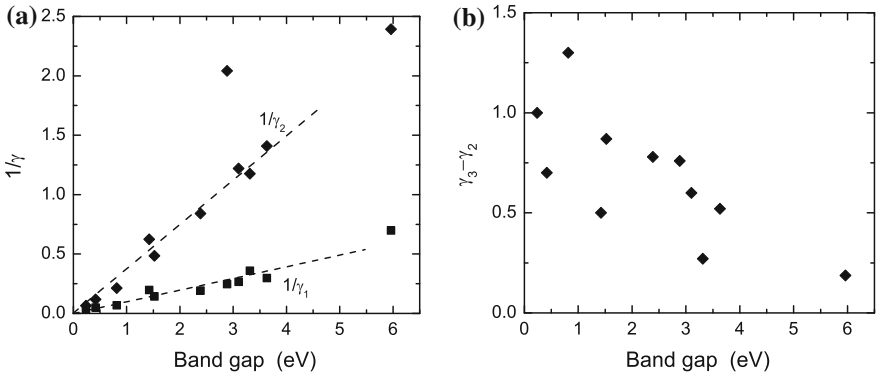


Fig. 6.37 Luttinger parameters for various III-V semiconductors versus their band gap. (**a**) Inverse values of γ_1 (*squares*) and γ_2 (*diamonds*). *Dashed lines* are guides to the eyes. (**b**) $\gamma_3 - \gamma_2$ versus band gap

$$\frac{1}{m_{\mathrm{hh}}^{\mathrm{av}}} = \frac{2}{\hbar^2}\left(A + B\left[1 + \frac{2\,C^2}{15\,B^2}\right]\right) \tag{6.49a}$$

$$\frac{1}{m_{\mathrm{lh}}^{\mathrm{av}}} = \frac{2}{\hbar^2}\left(A - B\left[1 + \frac{2\,C^2}{15\,B^2}\right]\right). \tag{6.49b}$$

Similar to the correlation of the electron mass with the band gap (Fig. 6.28), the Luttinger parameters are correlated with the band gap as shown in Fig. 6.37. The parameters $1/\gamma_1$ and $1/\gamma_2$ increase about linearly with E_{g}. The parameter $\gamma_3 - \gamma_2$, which is responsible for the valence band warping, decreases with increasing band gap.

6.10.3 Valence-Band Fine Structure

In Fig. 6.38, the schematic structure of the band edges for zincblende structure semiconductors is shown. The s-o holes in the zincblende structure are split-off due to the spin-orbit interaction Δ_{so}, the Γ_8 band is degenerate (heavy and light holes). Degeneracies for the holes are removed in the wurtzite and chalcopyrite structures by the additional crystal field splitting Δ_{cf} due to the anisotropy between the a- and c-axes. Typically, e.g. for CdS, the topmost valence band in the wurtzite structure has Γ_9 symmetry (allowed optical transitions only for $\mathbf{E} \perp \mathbf{c}$); an exception is ZnO for which the two upper bands are believed to be reversed. In the chalcopyrite structure optical transitions involving the Γ_6 band are only allowed for $\mathbf{E} \perp \mathbf{c}$. The three hole bands are usually labeled A, B, and C from the top of the valence band.

The energy positions of the three bands (with respect to the position of the Γ_{15} band) in the presence of spin-orbit interaction and crystal field splitting are given

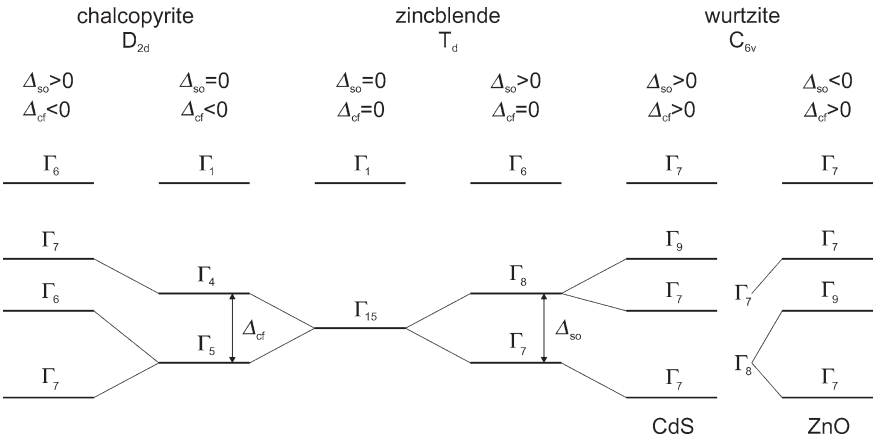


Fig. 6.38 Schematic band structure of zincblende and the valence-band splitting due to spin-orbit interaction Δ_{so} and crystal field splitting Δ_{cf} for chalcopyrites (typically $\Delta_{\mathrm{cf}} < 0$, see Fig. 6.39) and wurtzites. For the wurtzites the situation is schematically shown for CdS ($\Delta_{\mathrm{so}} = 67\,\mathrm{meV}$, $\Delta_{\mathrm{cf}} = 27\,\mathrm{meV}$) (or GaN) and ZnO ($\Delta_{\mathrm{so}} = -8.7\,\mathrm{meV}$, $\Delta_{\mathrm{cf}} = 41\,\mathrm{meV}$)

within the quasi-cubic approximation [457] by

$$E_1 = \frac{\Delta_{\mathrm{so}} + \Delta_{\mathrm{cf}}}{2} \tag{6.50a}$$

$$E_{2,3} = \pm\sqrt{\left(\frac{\Delta_{\mathrm{so}} + \Delta_{\mathrm{cf}}}{2}\right)^2 - \frac{2}{3}\,\Delta_{\mathrm{so}}\,\Delta_{\mathrm{cf}}}\,. \tag{6.50b}$$

In chalcopyrites the crystal field splitting is typically negative (Fig. 6.39). It is approximately linearly related to $1 - \eta$ (for $\eta = c/2a$ see Sect. 3.4.6).

6.10.4 Band Inversion

In certain compounds the band gap can shrink to zero (zero-gap semiconductor, cf. Sect. 18.1.2) and even become negative in the sense that the s-orbital type Γ_6 symmetry (conduction) band is inverted below the valence-band edge. HgTe is a classical example for such material as shown in Fig. 6.40, but similar effects are also present in other semiconductors, for example various chalcopyrites [459].

For the zero-gap case, the dispersion of some bands is linear (cmp. Sect. 18.1.2); this corresponds to a very strong non-parabolicity. The dielectric function of zero-gap semiconductors is discussed in [460].

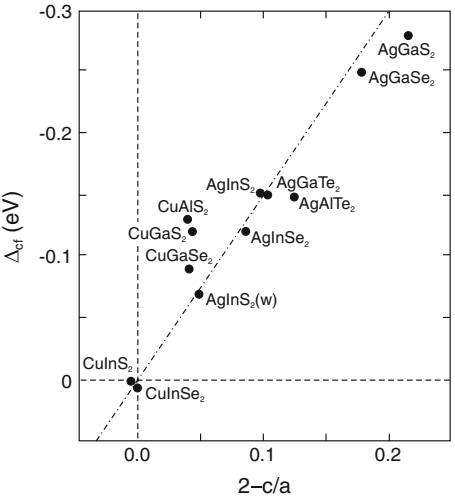


Fig. 6.39 Crystal field splitting Δ_{cf} for various chalcopyrite compounds versus the tetragonal distortion $2 - c/a = 2\,(1 - \eta)$. *Dash-dotted line* represents $\Delta_{cf} = 1.5\,b\,(2 - c/a)$ for $b = 1\,\text{eV}$. Data from [458]

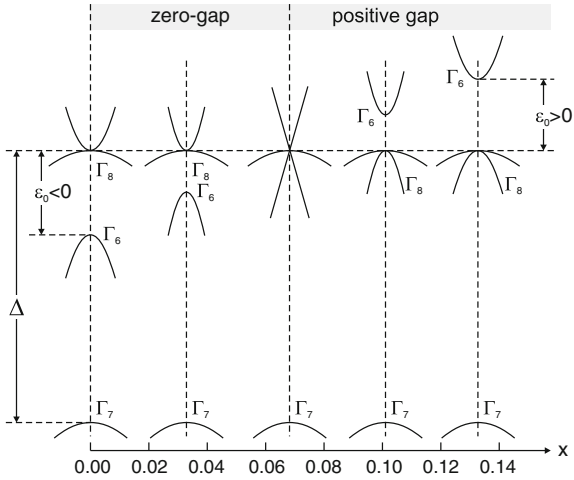


Fig. 6.40 Schematic band structure of zincblende with vanishing energy gap for the ternary compounds $Mn_xHg_{1-x}Te$. Note the linear dispersion for the zero-gap case at $x \approx 0.07$

6.11 Strain Effects on the Band Structure

A mechanical strain (or equivalently stress) causes changes in the bond lengths. Accordingly, the band structure is affected. These effects have been exhaustively treated in [461, 462]. For small strain, typically $\epsilon \lesssim 0.01$ the shift of the band edges is linear with the strain, for large strain it becomes nonlinear [463]. Often homogeneous strain is assumed, the effect of inhomogeneous strain is discussed in [464].

6.11.1 Strain Effect on Band Edges

In a direct-gap zincblende material the position of the conduction-band edge is only affected by the hydrostatic component of the strain

$$E_{\mathrm{C}} = E_{\mathrm{C}}^{0} + a_{\mathrm{c}} \left(\epsilon_{\mathrm{xx}} + \epsilon_{\mathrm{yy}} + \epsilon_{\mathrm{zz}}\right) = E_{\mathrm{C}}^{0} + a_{\mathrm{c}}\,\mathrm{Tr}(\epsilon), \tag{6.51}$$

where $a_{\mathrm{c}} < 0$ is the conduction-band hydrostatic deformation potential and E_{C}^{0} is the conduction-band edge of the unstrained material. Similarly, the valence-band edge is

$$E_{\mathrm{V}} = E_{\mathrm{V}}^{0} + a_{\mathrm{v}}\,\mathrm{Tr}(\epsilon), \tag{6.52}$$

where $a_{\mathrm{v}} > 0$ is the valence-band hydrostatic deformation potential. Therefore the band gap increases by

$$\Delta E_{\mathrm{g}} = a\,\mathrm{Tr}(\epsilon) = a\,\left(\epsilon_{\mathrm{xx}} + \epsilon_{\mathrm{yy}} + \epsilon_{\mathrm{zz}}\right), \tag{6.53}$$

with $a = a_{\mathrm{C}} - a_{\mathrm{V}}$. Such linear behavior upon hydrostatic pressure has been found for many semiconductors and is shown in Fig. 6.41a for $Ga_{0.92}In_{0.08}As$. The anomaly for N-doping is discussed below in Sect. 6.11.3. In Fig. 6.42 the dependence of the direct and indirect gaps of GaAs is shown. The dependence of the direct gap on pressure is non-linear, that on the density is linear [465].

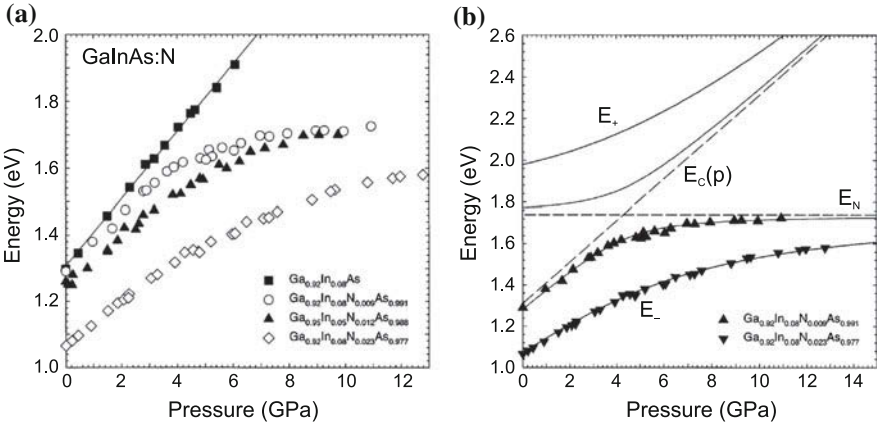


Fig. 6.41 (**a**) Dependence of the band gap of $Ga_{0.92}In_{0.08}As$ alloy (*squares*) and nitrogen-doped (Ga,In)As on (compressive) hydrostatic pressure, determined by photomodulated transmission at $T = 295\,\mathrm{K}$. (**b**) Pressure dependence of band gap for two (Ga,In)(As,N) samples together with model calculation (6.58). The coupling parameter is $V = 0.12\,\mathrm{eV}$ (0.4 eV) for a nitrogen content of 0.9 % (2.3 %). Adapted from [466]

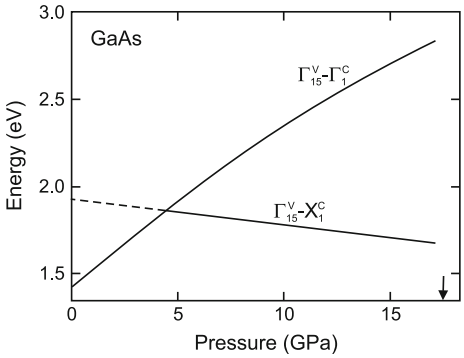


Fig. 6.42 Dependence of the direct Γ_{15}^{V}–Γ_{1}^{C} and indirect Γ_{15}^{V}–X_{1}^{C} band gap of GaAs ($T = 300\,\mathrm{K}$) on pressure. *Solid lines* are interpolations of experimental data, *dashed line* is extrapolation to $p = 0$. The crossing of the direct and indirect band gap occurs at 4.2 GPa. The *arrow* denotes the pressure of the phase transition from zincblende to an orthorhomic structure around 17 GPa. Adapted from [465]

Biaxial and shear strains affect the valence bands and lead to shifts and splitting of the heavy and light holes at the Γ point:

$$E_{\mathrm{v,hh/lh}} = E_{\mathrm{v}}^{0} \pm E_{\epsilon\epsilon} \tag{6.54a}$$

$$E_{\epsilon\epsilon}^{2} = b^{2}/2\left[\left(\epsilon_{\mathrm{xx}} - \epsilon_{\mathrm{yy}}\right)^{2} + \left(\epsilon_{\mathrm{yy}} - \epsilon_{\mathrm{zz}}\right)^{2} + \left(\epsilon_{\mathrm{xx}} - \epsilon_{\mathrm{zz}}\right)^{2}\right] + d^{2}\left[\epsilon_{\mathrm{xy}}^{2} + \epsilon_{\mathrm{yz}}^{2} + \epsilon_{\mathrm{xz}}^{2}\right],$$

where E_{v}^{0} denotes the bulk valence-band edge. b and d are the optical deformation potentials. For compressive strain the heavy-hole band is above the light-hole band. For tensile strain there is strong mixing of the bands (Fig. 6.43). In Table 6.7 the deformation potentials for some III–V semiconductors are listed. Typical values are in the eV regime.

In a wurtzite crystal, seven (or eight) deformation potentials are needed that are termed a (for the change of band gap with hydrostatic strain, again $a = a_{\mathrm{C}} - a_{\mathrm{V}}$) and D_1–D_6 (for the valence band structure) [467, 468].

In Si and Ge, three deformation potentials, termed a, b, d, are needed for the valence band and two for each conduction band minimum, Ξ_{u} and Ξ_{d} [469]. The energy position of the ith conduction-band edge (with unit vector $\mathbf{a}_i$ pointing to the valley) is

$$E_{\mathrm{C},i} = E_{\mathrm{C},i}^{0} + \Xi_{\mathrm{d}}\,\mathrm{Tr}(\boldsymbol{\epsilon}) + \Xi_{\mathrm{u}}\,\mathbf{a}_i\,\boldsymbol{\epsilon}\,\mathbf{a}_i, \tag{6.55}$$

where $E_{\mathrm{C},i}^{0}$ denotes the energy of the unstrained conduction-band edge. The deformation potentials for Si and Ge are given in Table 6.8.

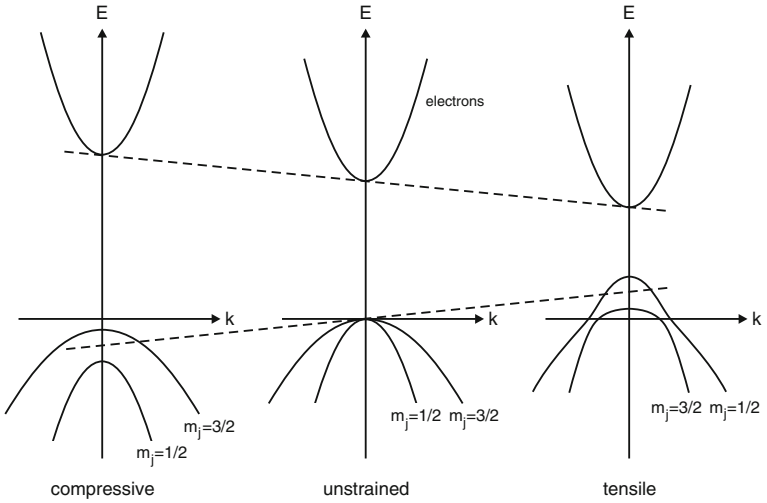


Fig. 6.43 Schematic band structure of GaAs in unstrained state (*center*) and under compressive and tensile biaxial strain as labeled. *Dashed lines* indicate shift of band edges due to hydrostatic part of strain

Table 6.7 Deformation potentials for some III–V semiconductors

Material	a	b	d
GaAs	−9.8	−1.7	−4.6
InAs	−6.0	−1.8	−3.6

All values in eV

Table 6.8 Deformation potentials for silicon and germanium

Material	$\Xi_d^{(\Delta)}$	$\Xi_u^{(\Delta)}$	$\Xi_d^{(L)}$	$\Xi_u^{(L)}$	a	b	d
Si	1.1	10.5	−7.0	18.0	2.1	−2.33	−4.75
Ge	4.5	9.75	−4.43	16.8	2.0	−2.16	−6.06

All values in eV from [470]

6.11.2 Strain Effect on Effective Masses

In the presence of strain the band edges are shifted (cf. Sect. 6.11). Since the electron mass is related to the band gap, it is expected that the mass will also be effected. In the presence of hydrostatic strain ϵ_H the electron mass is [471] (cf. to (6.39) for $\epsilon_H \to 0$)

$$\frac{m_0}{m_e^*} = 1 + \frac{E_P}{E_g + \Delta_0/3}\left[1 - \epsilon_H\left(2 + \frac{3a}{E_g + \Delta_0/3}\right)\right], \tag{6.56}$$

with a being the hydrostatic deformation potential and $\epsilon_H = \mathrm{Tr}(\epsilon)$. In [471], formulas are also given for biaxial and shear strain and also for hole masses. Since the effective

mass enters the mobility, the electrical conductivity depends on the stress state of the semiconductor (piezoresistivity, see Sect. 8.3.11).

6.11.3 Interaction with a Localized Level

The normal dependence of the band gap on hydrostatic pressure is linear and given by (6.53). (Ga,In)As containing nitrogen exhibits a remarkable deviation from this behavior as shown in Fig. 6.41a. This is due to the interaction of the continuum states of the conduction band with the electron level of the isoelectronic nitrogen impurity (Sect. 7.7.9) E_N, being within the conduction band. For GaAs it is 0.2 eV above the conduction band edge E_C. This phenomenon has been investigated theoretically within microscopic detail [472]. Within a simple 'band anticrossing' two-level model, the coupling of the pressure-dependent conduction band edge E_C and the nitrogen level can be obtained from the Eigenwert equation

$$\begin{vmatrix} E - E_C & V \\ V & E - E_N \end{vmatrix} = 0, \tag{6.57}$$

V being the coupling constant. The determinant vanishes for

$$E_\pm = \frac{1}{2}\left(E_C + E_N \pm \sqrt{(E_C - E_N)^2 + 4V^2}\right). \tag{6.58}$$

Here the weak pressure dependence of E_N is neglected for simplicity. This model can explain the pressure dependence of the band gap of (Ga,In)As:N fairly well [466] (Fig. 6.41b). The coupling parameter V is in the order of a few 0.1 eV for small

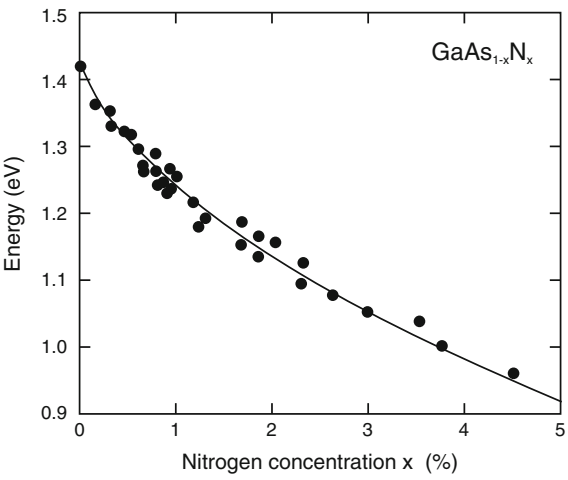


Fig. 6.44 Bandgap of $GaAs_{1-x}N_x$, experimental data from various sources (*symbols*) and model (*curve*) according to (6.58) with $V = V_0\sqrt{x}$ for $V_0 = 2.7$ eV. Adapted from [473]

nitrogen content. In photomodulated reflection also the E_+ levels can be observed [473]. The anticrossing model can also model the dependence of the $GaAs_{1-x}N_x$ bandgap on the nitrogen concentration [473] (Fig. 6.44).

6.12 Density of States

6.12.1 General Band Structure

The dispersion relation yields how the energy of a (quasi-) particle depends on the **k** vector. Now we want to know how many states are at a given energy. This quantity is called the *density of states* (DOS) and is written as $D(E)$. It is defined in an infinitesimal sense such that the number of states between E and $E + \delta E$ is $D(E)\delta E$. In the vicinity of the extrema of the band structure many states are at the same energy such that the density of states is high.

The dispersion relation of a band will be given as $E = E(\mathbf{k})$. If several bands overlap, the densities of state of all bands need to be summed up. The density of states at the energy $\tilde{E}$ for the given band is

$$D(\tilde{E})\,\mathrm{d}E = 2\int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi/L)^3}\,\delta(\tilde{E} - E(\mathbf{k})), \tag{6.59}$$

where, according to (5.5), $(2\pi/L)^3$ is the **k**-space volume for one state. The factor 2 is for spin degeneracy. The integral runs over the entire **k**-space and selects only those states that are at $\tilde{E}$. The volume integral can be converted to a surface integral over the isoenergy surface $S(\tilde{E})$ with $E(\mathbf{k}) = \tilde{E}$. The volume element $\mathrm{d}^3\mathbf{k}$ is written as $\mathrm{d}^2S\,\mathrm{d}\mathbf{k}_\perp$. The vector $\mathrm{d}\mathbf{k}_\perp$ is perpendicular to $S(\tilde{E})$ and proportional to $\nabla_\mathbf{k}E(\mathbf{k})$, i.e. $\mathrm{d}E = |\nabla_\mathbf{k}E(\mathbf{k})|\,\mathrm{d}\mathbf{k}_\perp$.

$$D(\tilde{E}) = 2\int_{S(\tilde{E})} \frac{\mathrm{d}^2S}{(2\pi/L)^3}\,\frac{1}{|\nabla_\mathbf{k}E(\mathbf{k})|}. \tag{6.60}$$

In this equation, the dispersion relation is explicitly contained. At band extrema the gradient diverges, however, in three dimensions the singularities are integrable and the density of states takes a finite value. The corresponding peak is named a van-Hove singularity. The concept of the density of states is valid for all possible dispersion relations, e.g. for electrons, phonons or photons.

The density of states for the silicon band structure (see Fig. 6.6a) is shown in Fig. 6.45.

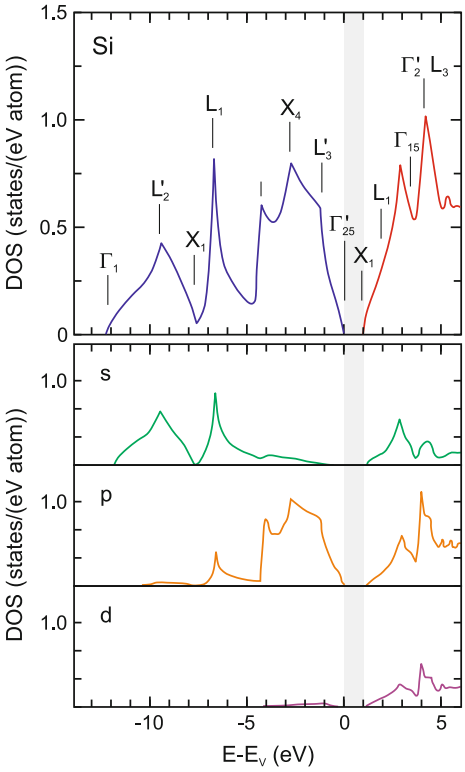


Fig. 6.45 Density of states in the silicon valence- (*blue*) and conduction-band (*red*) as obtained from theoretical calculation using empirical pseudopotentials. *Grey* regions denotes the band gap. Critical points (cf. Fig. 6.6a) are labeled. In the *lower* three graphs, the DOS is decomposed into contributions from different angular momentum states (s (*green*), p (*orange*) and d (*purple*)). *Top part* adapted from [474], *bottom part* adapted from [155]

6.12.2 Amorphous Semiconductors

If disorder is introduced, the density of states is modified as shown in Fig. 6.46 for amorphous germanium using a calculation with complex eigenenergies. The defects, as compared to the perfect lattice, introduced states in the band gap and generally wash out the sharp features from the crystalline DOS.

Several models exist for the defect level distributions within the band gap. The first model was the Mott model which has band tails at the valence and conduction band edges [475]. In the Cohen-Fritzsche-Ovshinsky (CFO) model [476], the band tails are more severe and overlap; the Fermi energy lies at the minimum of the density of states. In the Davis–Mott model [477] deep states were added in the gap and eventually the Marshall-Owen model [478] assumes band tails and donor- and acceptor-like deep states. The four models are schematically shown in Fig. 6.47. These model densities of states allow also the interpretation of carrier transport in amorphous semiconductors, taking into account localized and delocalized states (see Sect. 8.8).

The density of states for an amorphous semiconductor is best calculated from atomistic models, possibly averaging over many configurations. The typical features,

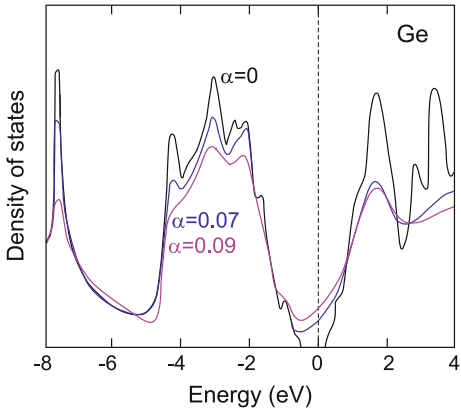


Fig. 6.46 Theoretical calculation for the density of states of amorphous Ge models as obtained for various degrees of disorder α (3.7). $\alpha = 0.09$ corresponds to a mean short-range order distance of about 2.4 lattice constants (cmp. Fig. 3.14b). Adapted from [177]

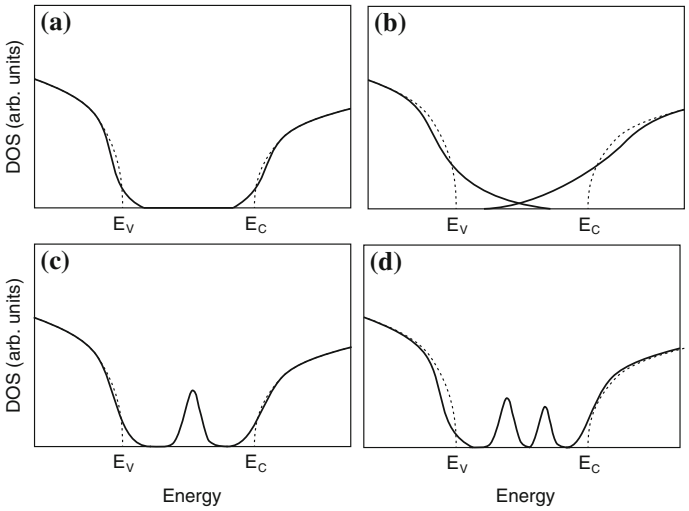


Fig. 6.47 Model density of states in amorphous semiconductors (*solid lines*) according to Mott [475], Cohen-Fritzsche-Ovshinsky [476], Davis–Mott [477] and Marshall–Owen [478]. *Dashed lines* represent the DOS of the same material without disorder

compared to the clear band gap of a similar ordered material, are band tails due to disorder (cmp. Sect. 5.2.8) and deep levels within the gap due to specific atomic arrangements not present in ordered bulk. The most investigated system is amorphous silicon; in Fig. 6.48 a numerical calculation of the density of states is shown together with charge distribution of four states at selected energies [479]. The further the states are in the band tail, the stronger their localization is. The two most right states shown in Fig. 6.48 are not conducting.

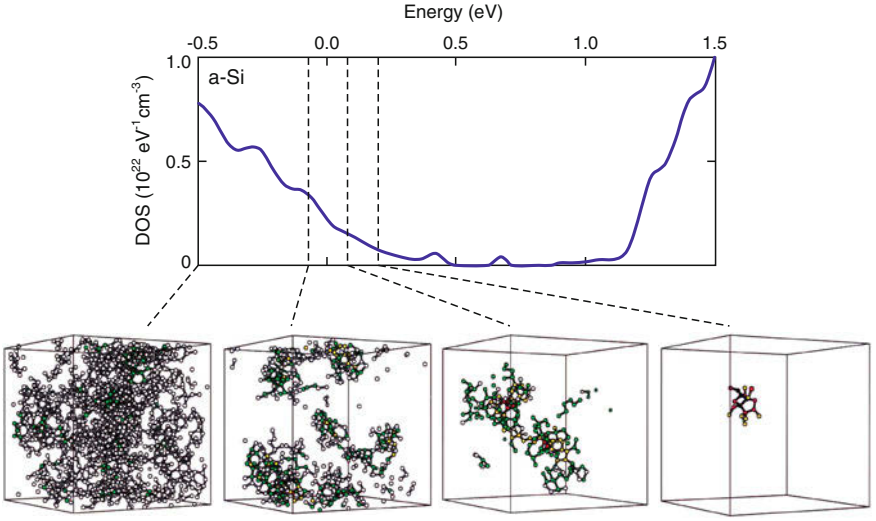


Fig. 6.48 Theoretical calculation of the density of electronic states of amorphous silicon. The charge distribution in four selected states at the indicated energies is shown, from *right* to *left* with decreasing localization. Adapted from [479]

As another example, simulations of $ZnSnO_3$ are shown in Fig. 6.49. The band tail between 0 and 0.5 eV is due to disorder of oxygen 2p orbitals [480]. At 0.9 eV a level due to under-coordinated oxygen appears. Deep levels are due to metal-metal bonds. Band tails due to chemically disordered oxygen have been experimentally observed for amorphous GIZO [481].

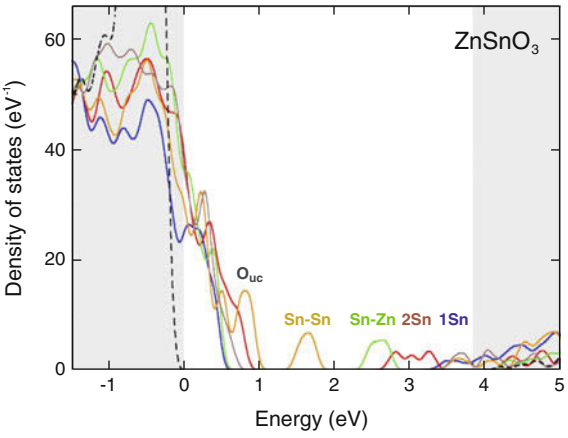


Fig. 6.49 Theoretical calculation for the density of states of crystalline (*dashed lines*, conduction and valence bands indicated by *greay areas*) and amorphous $ZnSnO_3$ with different configurations (*solid lines*). States due to under-coordinated oxygen (O_{uc}) and metal-metal binds are labelled. Adapted from [480]

6.12.3 Free-Electron Gas

In M dimensions, the energy states of a free-electron gas are given as

$$E(\mathbf{k}) = \frac{\hbar^2}{2\,m^*} \sum_{i=1}^{M} k_i^2. \tag{6.61}$$

The k_i can take the values $\pm\pi n/L$ (in the first Brillouin zone) with $n \leq N$, N being the number of unit cells in one dimension. These values are equidistant in $\mathbf{k}$-space. Each M-dimensional $\mathbf{k}$-point takes a volume of $(2\pi/L)^M$. The number of states $N(E_\mathrm{F})$ up to the energy $E_\mathrm{F} = \frac{\hbar^2}{2m} k_\mathrm{F}^2$ (later used as Fermi energy E_F and Fermi vector k_F) is

$$N(E_\mathrm{F}) = \frac{2}{(2\pi/L)^M} \int_{\mathbf{k}=0}^{|\mathbf{k}|=k_\mathrm{F}} \mathrm{d}^M k. \tag{6.62}$$

The factor 2 is for spin degeneracy, the integration runs over M dimensions. The density of states is the derivative

$$D(E) = \frac{\mathrm{d}N}{\mathrm{d}E}. \tag{6.63}$$

In the following, the density of states for $M = 3, 2, 1$ and zero dimensions is derived. A visualization is given in Fig. 14.1.

M = 3

This case relates to bulk material in which electrons are free to move in all three dimensions. Performing the integral (6.62) for $M = 3$ yields for an isotropic mass,

$$N^{3\mathrm{D}} = \frac{V}{3\pi^2} k_\mathrm{F}^3 = \frac{V}{3\pi^2} \left(\frac{2\,m\,E_\mathrm{F}}{\hbar^2} \right)^{3/2}. \tag{6.64}$$

Therefore, k_F and E_F are given by

$$k_\mathrm{F} = \left(\frac{3\pi^2\,N}{V} \right)^{1/3} \tag{6.65}$$

$$E_\mathrm{F} = \frac{\hbar^2}{2\,m^*} \left(\frac{3\pi^2\,N}{V} \right)^{2/3}, \tag{6.66}$$

and the density of states in three dimensions is

$$D^{3\mathrm{D}}(E) = \frac{V}{2\pi^2} \left(\frac{2\,m^*}{\hbar^2} \right)^{3/2} \sqrt{E}. \tag{6.67}$$

Mostly the density of states is used as density of states per volume, then the factor V in (6.67) is omitted.

If a conduction-band minimum is degenerate, a factor g_v (valley degeneracy) must be included in the density of states, i.e. $g_\mathrm{v} = 6$ for Si and $g_\mathrm{v} = 8$ for Ge ($g_\mathrm{v} = 1$ for GaAs). This factor is typically included in the mass used in (6.67) that then becomes the density of states mass $m_\mathrm{d,e}$. If the conduction-band minimum has cylindrical symmetry in $\mathbf{k}$-space, such as for Si and Ge, the mass that has to be used is

$$m_\mathrm{d,e} = g_\mathrm{v}^{2/3} \left(m_\mathrm{t}^2 m_\mathrm{l}\right)^{1/3}. \tag{6.68}$$

In the case of a degeneracy of the valence band, the states of several bands need to be summed. In bulk material, typically the heavy and light hole bands are degenerate at the Γ point. If the split-off band is not populated because of insufficient temperature, the valence-band edge density of states is expressed by the density of states hole mass

$$m_\mathrm{d,h} = \left(m_\mathrm{hh}^{3/2} + m_\mathrm{lh}^{3/2}\right)^{2/3}. \tag{6.69}$$

The density of states (per volume) at the conduction and valence band edges are thus given by

$$D_\mathrm{e}^\mathrm{3D}(E) = \frac{1}{2\pi^2}\left(\frac{2\,m_\mathrm{d,e}}{\hbar^2}\right)^{3/2}\sqrt{E - E_\mathrm{C}},\ E > E_\mathrm{C} \tag{6.70}$$

$$D_\mathrm{h}^\mathrm{3D}(E) = \frac{1}{2\pi^2}\left(\frac{2\,m_\mathrm{d,h}}{\hbar^2}\right)^{3/2}\sqrt{E_\mathrm{V} - E},\ E < E_\mathrm{V}. \tag{6.71}$$

M = 2

This case is important for thin layers in which the electron motion is confined in one direction and free in a plane. Such structures are called quantum wells (see Sect. 12.3.2). We find for the 2D density of states (for each subband over which it is not summed here, including spin degeneracy)

$$N^\mathrm{2D} = \frac{A}{2\pi}k_\mathrm{F}^2 = \frac{A}{\pi}\frac{m^*}{\hbar^2}E, \tag{6.72}$$

where A is the area of the layer. The density of states is thus constant and given by

$$D^\mathrm{2D}(E) = \frac{A}{\pi}\frac{m^*}{\hbar^2}. \tag{6.73}$$

M = 1

The case $M = 1$ describes a quantum wire in which the electron motion is confined in two dimensions and free in only one dimension. For this case, we find for a wire

of length L

$$N^{1D} = \frac{2L}{\pi} k_F = \frac{2L}{\pi} \left(\frac{2m^* E}{\hbar^2} \right)^{1/2}. \tag{6.74}$$

The density of states becomes singular at $E = 0$ and is given by (for one subband)

$$D^{1D}(E) = \frac{L}{\pi} \left(\frac{2m^*}{\hbar^2} \right)^{1/2} \frac{1}{\sqrt{E}}. \tag{6.75}$$

M = 0

In this case electrons have no degrees of freedom, as, e.g., in a quantum dot (Sect. 14.3), and each state has a δ-like density of states at each of the quantized levels.

Chapter 7
Electronic Defect States

Über Halbleiter sollte man nicht arbeiten, das ist eine Schweinerei, wer weiß ob es überhaupt Halbleiter gibt.
One should not work on semiconductors. They are a mess. Who knows whether semiconductors exist at all.
W. Pauli, 1931 [482]

Abstract After the carrier statistics for intrinsic conduction and general doping principles, donors and acceptors, compensation and high doping effects are treated in detail. The concept of quasi-Fermi levels is introduced. Finally for deep levels and their thermodynamics general remarks and several examples are given.

7.1 Introduction

One centimeter cube of a semiconductor contains about 5×10^{22} atoms. It is practically impossible to achieve perfect purity. Typical low concentrations of impurity atoms are in the 10^{12}–$10^{13}\,\mathrm{cm}^{-3}$ regime. Such a concentration corresponds to a purity of 10^{-10}, corresponding to about one alien in the world's human population. In the beginning of semiconductor research the semiconductors were so impure that the actual semiconducting properties could only be used inefficiently. Nowadays, thanks to large improvements in high-purity chemistry, the most common semiconductors, in particular silicon, can be made so pure that the residual impurity concentration plays no role in the physical properties. However, the most important technological step for semiconductors is *doping*, the controlled incorporation of impurities, in order to manage the semiconductor's conductivity. Typical impurity concentrations used in doping are 10^{15}–$10^{20}\,\mathrm{cm}^{-3}$. A milestone in the understanding of doping and the spreading of semiconductor technology was the 1950 textbook by Shockley [483].

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_7

7.2 Carrier Concentration

Generally, the density of electrons in the conduction band is given by

$$n = \int_{E_C}^{\infty} D_e(E)\, f_e(E)\, \mathrm{d}E, \tag{7.1}$$

and accordingly the density of holes in the valence band is

$$p = \int_{-\infty}^{E_V} D_h(E)\, f_h(E)\, \mathrm{d}E. \tag{7.2}$$

The energy of the top of the valence band is denoted by E_V, the bottom of the conduction band as E_C. The distribution function for holes is $f_h = 1 - f_e$. Thus,

$$f_h(E) = 1 - \frac{1}{\exp\left(\frac{E-E_F}{kT}\right)+1} = \frac{1}{\exp\left(-\frac{E-E_F}{kT}\right)+1}. \tag{7.3}$$

If several hole bands (hh, lh, so) are considered, the same distribution is valid for all hole bands in thermal equilibrium.

We assume parabolic band edges, i.e. effective masses m_e and m_h for electrons and holes, respectively. The density of states (per volume) in the conduction and valence bands is given by (6.70) and (6.71).

In thermodynamic equilibrium, the distribution function $f_e(E)$ for electrons is given by the Fermi–Dirac distribution (Fermi function) $f_e(E)$ (E.22)

$$f_e(E) = \frac{1}{\exp\left(\frac{E-E_F}{kT}\right)+1}, \tag{7.4}$$

If the Boltzmann distribution (E.23) is a good approximation, the carrier distribution is called *nondegenerate*. If the Fermi distribution needs to be invoked, the carrier ensemble is called *degenerate*. If the Fermi level is within the band, the ensemble is highly degenerate.

If the Boltzmann approximation (E.23) cannot be applied, i.e. at high temperatures or for very small band gaps, the integral over Df cannot be analytically evaluated. In this case the Fermi integral is needed that is defined[1] as

$$F_n(x) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{y^n}{1+\exp(y-x)}\, \mathrm{d}y. \tag{7.5}$$

[1] Equation (7.5) is restricted to $n > -1$. A form without restriction is $\mathcal{F}_n(x) = \frac{1}{\Gamma(n+1)} \int_0^{\infty} \frac{y^n}{1+\exp(y-x)}\, \mathrm{d}y$. The factor $2/\sqrt{\pi}$ is often omitted but must be then added explicitly in, e.g. (7.6).

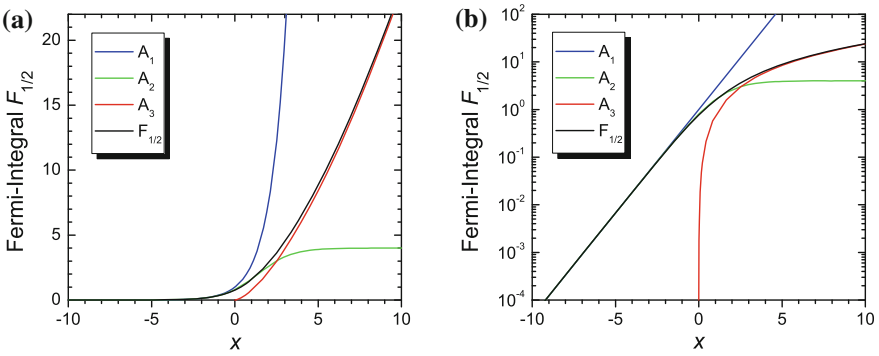


Fig. 7.1 Fermi integral $\hat{F}_{1/2} = (\sqrt{\pi}/2)F_{1/2}$ with approximations in three regions of the argument: $A_1(x) = (\sqrt{\pi}/2)\exp(x)$ for $x < 2$, $A_2(x) = (\sqrt{\pi}/2)(1/4 + \exp(-x))^{-1}$ for $-2 < x < 2$, $A_3(x) = 2/3x^{3/2}$ for $x > 2$. (**a**) linear, (**b**) semilogarithmic plot

In the present case of bulk materials $n = 1/2$. For large negative argument, i.e. $x < 0$ and $|x| \gg 1$, $F_{1/2}(x) \approx \exp(x)$, which is the Boltzmann approximation. $F_{1/2}(0) = 0.765\ldots \approx 3/4$. For large argument, i.e. $x \gg 1$, $F_{1/2}(x) \approx (2/\sqrt{\pi})(2/3)$ $x^{3/2}$. Such fairly simple approximations are plotted in Fig. 7.1 in comparison with the Fermi integral. For computations, analytical [484–486] or numerical approximations [487, 488] are used.

The derivative of the Fermi integral is given by $F_n'(x) = nF_{n-1}(x)$, $n > 0$. For $n = 0$, i.e. a two-dimensional system, the integral can be executed explicitly, $F_0(x) = (2/\sqrt{\pi})\ln\left[1 + \exp(x)\right]$.

With the Fermi integral $F_{1/2}$ (7.10) and (7.11) the free-carrier densities can be written as

$$n = N_{\mathrm{C}}\, F_{1/2}\left(\frac{E_{\mathrm{F}} - E_{\mathrm{C}}}{kT}\right) \tag{7.6}$$

$$p = N_{\mathrm{V}}\, F_{1/2}\left(-\frac{E_{\mathrm{F}} - E_{\mathrm{V}}}{kT}\right), \tag{7.7}$$

with

$$N_{\mathrm{C}} = 2\left(\frac{m_{\mathrm{e}}\, kT}{2\pi\,\hbar^2}\right)^{3/2} \tag{7.8}$$

$$N_{\mathrm{V}} = 2\left(\frac{m_{\mathrm{h}}\, kT}{2\pi\,\hbar^2}\right)^{3/2}, \tag{7.9}$$

where N_{C} (N_{V}) is called the conduction-band (valence-band) edge density of states. The masses in (7.8) and (7.9) are the density of states masses given in (6.68) and (6.69). Values of $N_{\mathrm{C,V}}$ for Si, Ge and GaAs are given in Table 7.1.

Now, we assume that the Boltzmann approximation (Appendix E) can be used, i.e. the probability that a band state is populated is $\ll 1$. Then, the integral (7.1) can

Table 7.1 Band gap, intrinsic carrier concentration, conduction band and valence-band edge density of states at $T = 300\,\mathrm{K}$ for various semiconductors

	E_g (eV)	n_i (cm^{-3})	N_C (cm^{-3})	N_V (cm^{-3})
InSb	0.18	1.6×10^{16}		
InAs	0.36	8.6×10^{14}		
Ge	0.67	2.4×10^{13}	1.04×10^{19}	6.0×10^{18}
Si	1.124	1.0×10^{10}	7.28×10^{19}	1.05×10^{19}
GaAs	1.43	1.8×10^{6}	4.35×10^{17}	5.33×10^{18}
GaP	2.26	2.7×10^{0}		
GaN	3.3	$\ll 1$		

be executed analytically and the concentration n of electrons in the conduction band is given as

$$n = 2\left(\frac{m_e kT}{2\pi\hbar^2}\right)^{3/2} \exp\left(\frac{E_F - E_C}{kT}\right) = N_C \exp\left(\frac{E_F - E_C}{kT}\right). \tag{7.10}$$

For the Boltzmann approximation and a parabolic valence band, the density of holes is given by

$$p = 2\left(\frac{m_h kT}{2\pi\hbar^2}\right)^{3/2} \exp\left(-\frac{E_F - E_V}{kT}\right) = N_V \exp\left(-\frac{E_F - E_V}{kT}\right). \tag{7.11}$$

The product of the electron and hole density is

$$\begin{aligned} n\,p &= N_V\,N_C \exp\left(-\frac{E_C - E_V}{kT}\right) = N_V\,N_C \exp\left(-\frac{E_g}{kT}\right) \\ &= 4\left(\frac{kT}{2\pi\hbar^2}\right)^3 (m_{d,e}\,m_{d,h})^{3/2} \exp\left(-\frac{E_g}{kT}\right). \end{aligned} \tag{7.12}$$

Thus, the product np is *independent* of the position of the Fermi level, as long as the Boltzmann approximation is fulfilled, i.e. the Fermi level is not in the vicinity of one of the band edges within several kT:

$$E_V + 4kT < E_F < E_C - 4kT. \tag{7.13}$$

The relation (7.12) is called the mass-action law.

In Fig. 7.2, the product np is shown for silicon over a wide range of Fermi energies. If E_F is within the band gap, np is essentially constant. If the Fermi level is in the valence or conduction band, np decreases exponentially.

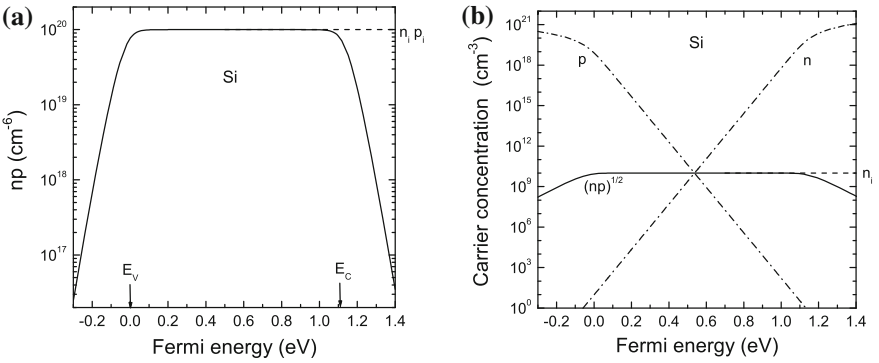


Fig. 7.2 (**a**) np for silicon at $T = 300\,\mathrm{K}$ as a function of the position of the Fermi level. The valence-band edge E_V is chosen as $E = 0$. np is constant for the range of Fermi energies given by (7.13) ($4kT \approx 0.1\,\mathrm{eV}$). (**b**) n, p and $\sqrt{np}$ as a function of the Fermi level

7.3 Intrinsic Conduction

First, we consider the conductivity of the intrinsic, i.e. an ideally pure, semiconductor. At $T = 0$ all electrons are in the valence band, the conduction band is empty and thus the conductivity is zero (a completely filled band cannot conduct current). Only at finite temperatures the electrons have a finite probability to be in a conduction-band state and to contribute to the conductivity. Due to neutrality, the electron and hole concentrations in the intrinsic semiconductors are the same, i.e. each electron in the conduction band comes from the valence band,

$$-n + p = 0, \tag{7.14}$$

or $n_\mathrm{i} = p_\mathrm{i}$. Therefore

$$\begin{aligned} n_\mathrm{i} = p_\mathrm{i} &= \sqrt{N_\mathrm{V} N_\mathrm{C}}\, \exp\left(-\frac{E_\mathrm{g}}{2kT}\right) \\ &= 2\left(\frac{kT}{2\pi\hbar^2}\right)^{3/2} (m_\mathrm{e} m_\mathrm{h})^{3/4}\, \exp\left(-\frac{E_\mathrm{g}}{2kT}\right). \end{aligned} \tag{7.15}$$

The mass-action law

$$np = n_\mathrm{i}\, p_\mathrm{i} = n_\mathrm{i}^2 = p_\mathrm{i}^2 \tag{7.16}$$

will be essential also for light and moderately doped semiconductors. The intrinsic carrier concentration is exponentially dependent on the band gap. Thus, in thermodynamic equilibrium intrinsic wide-gap semiconductors have much smaller electron concentrations than intrinsic small-gap semiconductors (see Table 7.1). The intrinsic carrier concentration of Si (in cm^{-3}) has been determined to be (within 1 %, T in K)

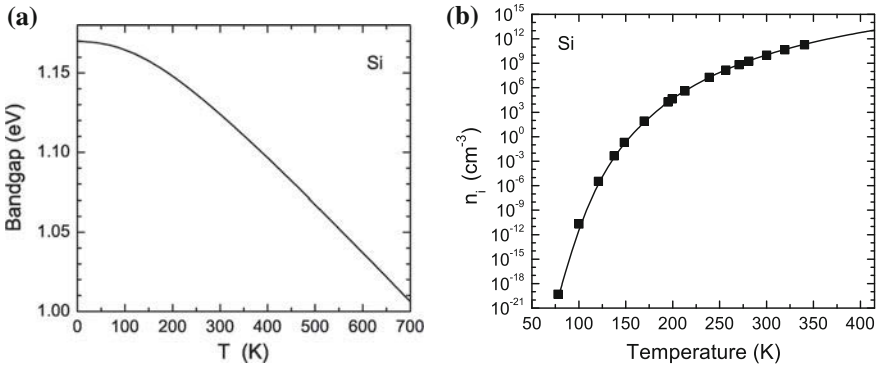


Fig. 7.3 (**a**) Band gap of silicon versus temperature. (**b**) Intrinsic carrier concentration of silicon versus temperature. *Solid line* is (7.17) using $E_g = 1.204\,\text{eV} - (2.73 \times 10^{-4}\,\text{eV/K})\,T$ [491], *symbols* are experimental data from [492]

$$n_i^{Si} = 1.640 \times 10^{15}\ T^{1.706}\ \exp\left(-\frac{E_g(T)}{2kT}\right) \tag{7.17}$$

for temperatures between 77 and 400 K [489, 490] (Fig. 7.3).

As we will see later in Part II, many semiconductor devices rely on regions of low conductivity (depletion layers) in which the carrier concentration is small. Since the carrier concentration cannot be smaller than the intrinsic concentration ($n + p \geq 2n_i$), an increase of temperature leads to increasing ohmic conduction in the depletion layers and thus to a reduction or failure of device performance. The small band gap of Ge leads to degradation of bipolar device performance already shortly above room temperature. For silicon, intrinsic conduction limits operation typically to temperatures below about 300 °C. For higher temperatures, as required for devices in harsh environments, such as close to motors or turbines, other semiconductors with wider band gaps need to be used, such as GaN, SiC or even diamond.

From the neutrality condition for the intrinsic semiconductor (7.14), (7.10) and (7.11), the Fermi level of the intrinsic semiconductor can be determined as

$$E_F = E_i = \frac{E_V + E_C}{2} + \frac{kT}{2}\ln\left(\frac{N_V}{N_C}\right) = \frac{E_V + E_C}{2} + \frac{3}{4}kT\ln\left(\frac{m_h}{m_e}\right). \tag{7.18}$$

Since the hole mass is perhaps a factor of ten larger than the electron mass, the second term has the order of kT. Thus, for typical semiconductors where $E_g \gg kT$, the intrinsic Fermi level, denoted by E_i, is close to the middle of the band gap, i.e. $E_i \approx (E_C + E_V)/2$.

The situation for an intrinsic semiconductor is schematically shown in Fig. 7.4b.

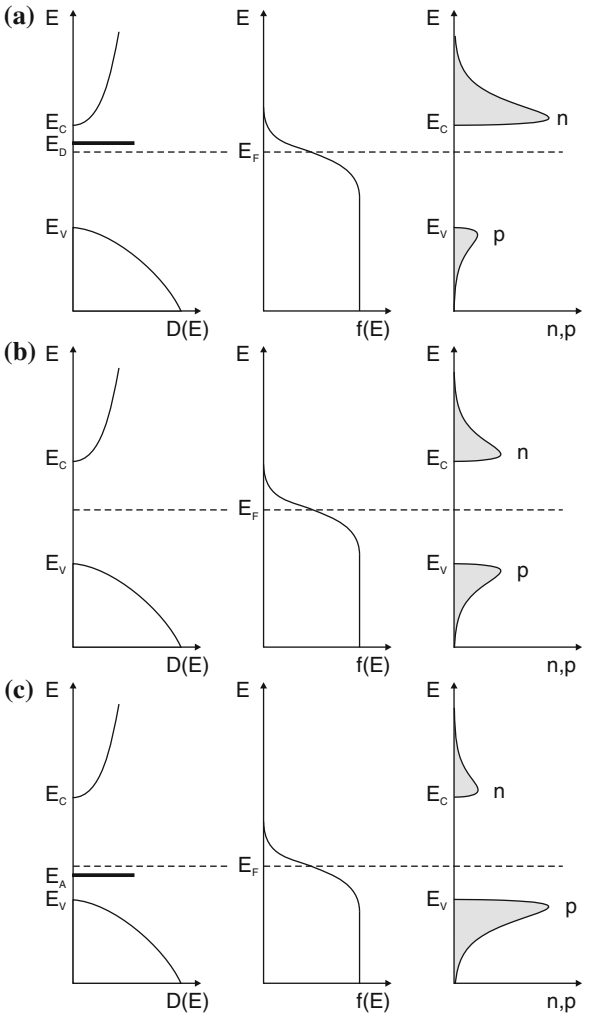


Fig. 7.4 Density of states (*left column*), Fermi distribution (*center column*) and carrier concentration (*right column*) for (**a**) n-type, (**b**) intrinsic and (**c**) p-type semiconductors in thermal equilibrium

7.4 Doping

7.4.1 Concept

The modification of the conductivity of a semiconductor using point defects is termed *doping*. In 1930 electrical conduction of semiconductors was attributed solely to impurities [493]. However 'chemically pure' substances become conductive upon deviation from stoichiometry, e.g. historically found for changes in the anion concentration and conductivity in CuI [31] (p-type) and ZnO [70] (n-type). The modification of CuI by exposure to different partial pressure of iodine in organic solutions

with different iodine concentration [34] and subsequently various concentrations of copper vacancies [494] can be considered the first doping of a semiconductor (1909).

The electronic levels of a defect or an impurity can exist within the forbidden gap of the bulk host material. These levels can lie close to the band edges or in the vicinity of the middle of the band gap. In a simplified approach, the first stem from *shallow* defects (Sect. 7.5), the latter from *deep* defects (Sect. 7.7).

7.4.2 Doping Principles

In [495] various doping principles are formulated. Essentially, the amount of impurities that lead to electrically active dopants is limited by the increasingly probable formation of compensating defects. In the case of donors, these are electron killers, e.g. n-type doping of Si:As is limited by the formation of V_{Si} [496]. In the case of acceptors, the compensating defects are hole killers. The so-called n-type *pinning energy* $E_F^{n,pin}$ is the Fermi level at which such killer defect (e.g. a cation vacancy) forms. When the Fermi level reaches the pinning energy, no further progress in n-type doping can be made, since the spontaneously generated electron killers will negate the introduced (impurity) donors. As a tendency, materials with low lying conduction band, i.e. large electron affinity (difference between vacuum level and conduction band) can be doped n-type. Similarly, p-type doping by acceptors, shifting the Fermi level towards the valence band, will meet at some point $E_F^{p,pin}$, called p-type pinning energy, when native hole killers, such as anion vacancies or cation interstitials form spontaneously. At this point, further p-type doping is no longer possible. p-type doping is facilitated by materials whose valance band maximum is close to the vacuum level [495].

A comparison of the wide gap materials ZnO, NiO and MgO is depicted in Fig. 7.5. The position of the pinning levels is marked on a common energy scale. From the position of $E_F^{n,pin}$ it can be understood, that ZnO can be highly n-doped while NiO

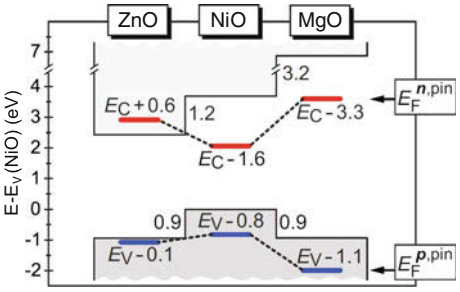


Fig. 7.5 Comparison of ZnO, NiO and MgO on a common energy scale, comparing conduction band and valence band edges and n-type (*red*) and p-type (*blue*) pinning energies (determined for metal-rich and oxygen-rich conditions, respectively). Adapted from [497]

and MgO cannot [497]. From $E_{\mathrm{F}}^{\mathrm{p,pin}}$, NiO can be doped p-type, while MgO cannot be doped at all.

For dopability, generally, it is also important that the ionized charges from impurities are free and thus contribute to the free charge carrier density and do not form localized states, e.g. due to polaronic effects (Sect. 8.6).

7.5 Shallow Defects

In Fig. 7.6, the positions of the energy levels of a variety of impurities are shown for Ge, Si and GaAs. An impurity for which the long-range Coulomb part of the ion-core potential determines the energetic level is termed a *shallow* impurity. The extension of the wavefunction is given by the Bohr radius. This situation is in contrast to a *deep* level where the short-range part of the potential determines the energy level. The extension of the wavefunction is then of the order of the lattice constant. A view on the history of the science of shallow impurity states is given in [498, 499].

We will consider first a group-IV semiconductor, Si, and (impurities) dopants from the groups III and V of the periodic system. When these are incorporated on a lattice site (with tetrahedral bonds), there is one electron too few (group III, e.g. B) or one electron too many (group V, e.g. As). The first case is called an *acceptor*, the latter a *donor.* The doping of III–V semiconductors is detailed in [501].

7.5.1 Donors

Silicon doped with arsenic is denoted as Si:As. The situation is schematically shown in Fig. 7.7. The arsenic atom has, after satisfying the tetrahedral bonds, an extra electron. This electron is bound to the arsenic atom via the Coulomb interaction since the ion core is positively charged compared to the silicon cores. If the electron is ionized, a fixed positive charge remains at the As site.

Without being in the silicon matrix, an arsenic atom has an ionization energy of 9.81 eV. However, in the solid the Coulomb interaction is screened by the dielectric constant of the material, typically ϵ_{r} is of the order of 10 for typical semiconductors. Additionally, the mass is renormalized (effective mass) by the periodic potential to a value that is smaller than the free electron mass. Within effective-mass theory (Appendix H) Bohr's theory of the hydrogen problem is scaled with the (isotropic) effective mass m_{e}^* and the dielectric constant ϵ_{r}, the binding energy (ionization energy) E_{D}^b of the electron to the shallow donor is (relative to the continuum given by the conduction-band edge E_{C})

$$E_{\mathrm{D}}^b = \frac{m_{\mathrm{e}}^*}{m_0}\,\frac{1}{\epsilon_{\mathrm{r}}^2}\,\frac{m_0\, e^4}{2\,(4\pi\epsilon_0\,\hbar)^2}. \tag{7.19}$$

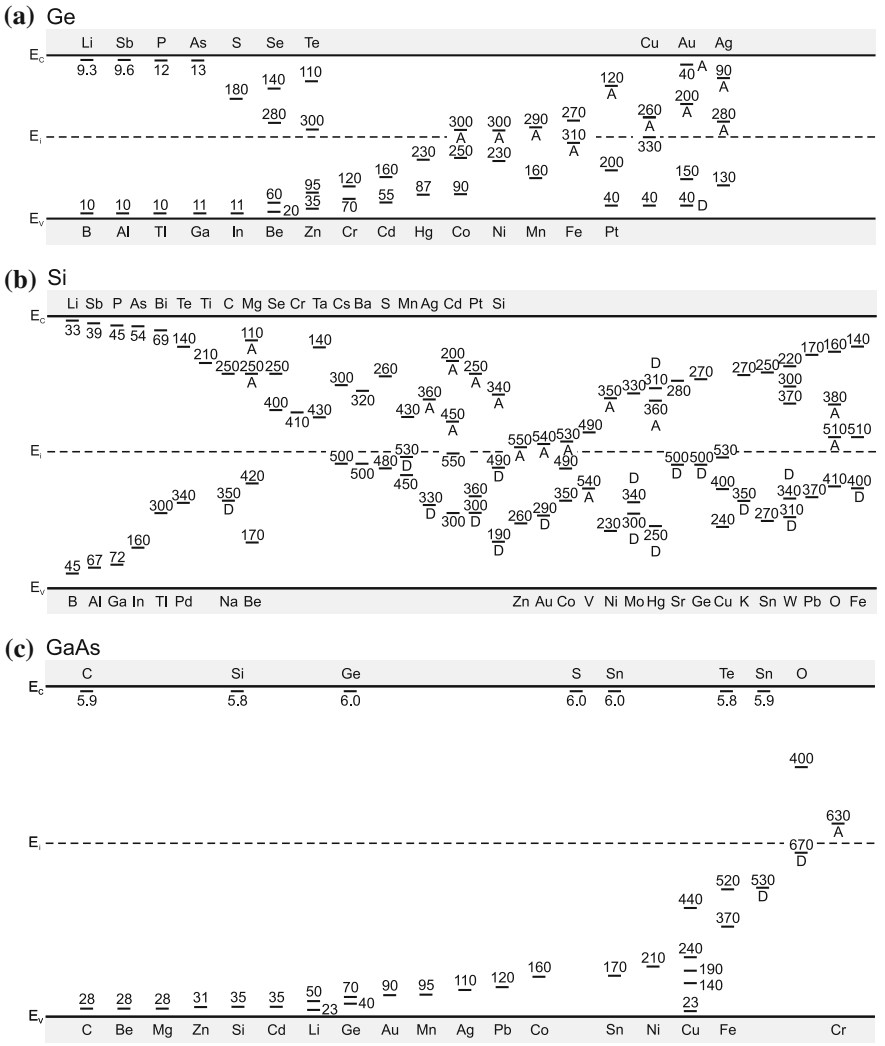


Fig. 7.6 Energetic position (ionization energy labeled in meV) of various impurities (A: acceptor, D: donor) in (**a**) Ge, (**b**) Si and (**c**) GaAs. Based on [500]

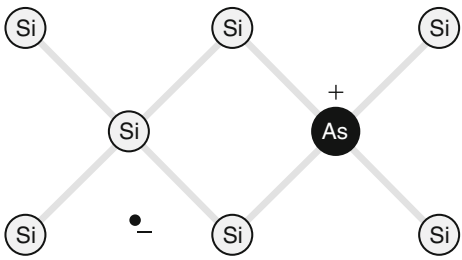


Fig. 7.7 Arsenic impurity in silicon. Arsenic donates one electron, and a fixed positive charge remains

The absolute energy position of the level is $E_D = E_C - E_D^b$. The first factor in the right side of (7.19) is the ratio of effective and free-electron mass, typically 1/10, the second factor is typically 1/100. The third factor is the ionization energy of the hydrogen atom, i.e. the Rydberg energy of 13.6 eV. Thus, the binding energy in the solid is drastically reduced by a factor of about 10^{-3} to the 10 meV regime. The excited states of the hydrogen-like spectrum can also be investigated experimentally (Sect. 9.7).

The extension of the wavefunction of the electron bound to the fixed ion is given by the Bohr radius

$$a_D = \frac{m_0}{m_e^*}\,\epsilon_r\,a_B, \tag{7.20}$$

where $a_B = 0.053$ nm denotes the hydrogen Bohr radius. For GaAs $a_D = 10.3$ nm. A similar value has been determined for InP [502]. For semiconductors with a non-isotropic band minimum, such as Si, Ge or GaP, an 'elliptically deformed' hydrogen problem with the masses m_l and m_t has to be treated [503].

An impurity that fulfills (7.19) is called an *effective-mass* impurity. For GaAs, the effective-mass donor has a binding energy of 5.715 meV, which is closely fulfilled for several chemical species (Table 7.3). In GaP, experimental values deviate considerably from the effective-mass donor (59 meV). For silicon, considering the anisotropic tensor of the effective masses, the result for the effective-mass donor binding energy is 29 meV [503]. Some experimentally observed values are summarized in Table 7.2.

Table 7.2 Binding energies E_D^b of Li and group-V donors in elemental semiconductors

	Li	N	P	As	Sb
C		1700	≈500		
Si	33		45	49	39
Ge	9.3		12.0	12.7	9.6

Data for carbon from [505]. All values in meV

Table 7.3 Binding energies E_D^b of donors in GaAs (data from [506]), GaP (data from [507]) and GaN (low concentration limits, data from [508, 509])

	V site		III site	
GaAs	S	5.854	C	5.913
	Se	5.816	Si	5.801
	Te	5.786	Ge	5.937
GaP	O	897	Si	85
	S	107	Ge	204
	Se	105	Sn	72
	Te	93		
GaN	O	39	Si	22
			Ge	19

All values in meV

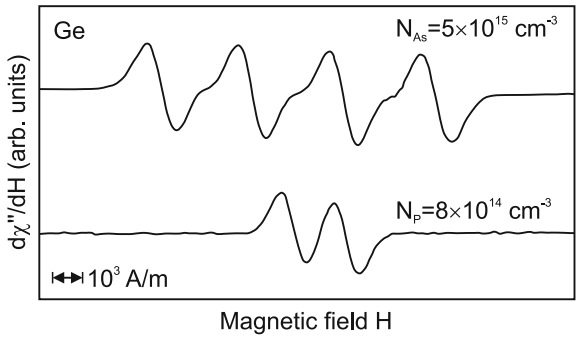


Fig. 7.8 Electron spin resonance signal from As and P in Ge with the magnetic field **H** parallel to [100], $T \approx 1.3$ K. Adapted from [504]

Deviations from the effective-mass theory are due to modification of the potential in the immediate vicinity of the impurity atom and breakdown of the effective-mass formalism.

Different impurities can have quite similar binding energies. They can be distinguished, e.g. by electron spin resonance (ESR). At low temperatures the electron is localized on the impurity and the hyperfine interaction with the nucleus can be resolved in ESR. In Fig. 7.8 data are shown for As and P in germanium. The multiplets distinguish the nuclear spins $I = 3/2$ for arsenic (^{75}As) and $I = 1/2$ for phosphorus (^{31}P) [504].

The donors are typically distributed statistically (randomly) in the solid. Otherwise their distribution is called clustered. The concentration of donors is labeled N_D and usually given in cm^{-3}.

The concentration of donors populated with an electron (neutral donors) is denoted by N_D^0, the concentration of ionized donors (positively charged) is N_D^+. Other conventions in the literature label the concentrations N_1 and N_0, respectively:

$$N_1 = N_\mathrm{D}^0 = N_\mathrm{D}\, f_\mathrm{e}(E_\mathrm{D}) \tag{7.21a}$$

$$N_0 = N_\mathrm{D}^+ = N_\mathrm{D}\,(1 - f_\mathrm{e}(E_\mathrm{D})), \tag{7.21b}$$

with $f_\mathrm{e}(E_\mathrm{D}) = \left[1 + \exp(E_\mathrm{D} - E_\mathrm{F})\right]^{-1}$. For the sum of these quantities the condition

$$N_\mathrm{D} = N_\mathrm{D}^+ + N_\mathrm{D}^0 \tag{7.22}$$

holds.

The ratio of the two concentrations is first given as (caveat: this formula will be modified below)

$$\frac{N_\mathrm{D}^0}{N_\mathrm{D}^+} = \frac{N_1}{N_0} = \frac{f}{1-f} = \exp\left(\frac{E_\mathrm{F} - E_\mathrm{D}}{kT}\right). \tag{7.23}$$

Now, the degeneracy of the states has to be considered. The donor charged with one electron has a 2-fold degeneracy $g_1 = 2$ since the electron can take the spin up and

down states. The degeneracy of the ionized (empty) donor is $g_0 = 1$. Additionally, we assume here that the donor cannot be charged with a second electron (cmp. Sect. 7.7.2). Due to Coulomb interaction, the energy level of the possible N_D^- state is in the conduction band. Otherwise, a multiply charged center would be present. We also do not consider excited states of N_D^0 that might be in the band gap as well. In the following, we will continue with $\hat{g}_D = g_1/g_0 = 2$ as suggested in [510].[2] We note that the definition of the degeneracy factor for donors (and acceptors, see (7.39)) is not consistent in the literature as summarized in [511]. Considering now the degeneracy, (7.23) is modified to

$$\frac{N_D^0}{N_D^+} = \frac{N_1}{N_0} = \hat{g}_D \exp\left(\frac{E_F - E_D}{kT}\right). \tag{7.24}$$

This can be understood from thermodynamics (cf. Sect. 4.2.2), a rate analysis or simply the limit $T \to \infty$.

The probabilities f^1 and f^0 for a populated or empty donor, respectively, are

$$f^1 = \frac{N_1}{N_D} = \frac{1}{\hat{g}_D^{-1} \exp\left(\frac{E_D - E_F}{kT}\right) + 1} \tag{7.25a}$$

$$f^0 = \frac{N_0}{N_D} = \frac{1}{\hat{g}_D \exp\left(-\frac{E_D - E_F}{kT}\right) + 1}. \tag{7.25b}$$

First, we assume that no carriers in the conduction band stem from the valence band (no intrinsic conduction). This will be the case at sufficiently low temperatures when $N_D \gg n_i$. Then the number of electrons in the conduction band is equal to the number of ionized donors, i.e.

$$n = f^0 N_D = N_0 = \frac{N_D}{1 + \hat{g}_D \exp\left(\frac{E_F - E_D}{kT}\right)} = \frac{1}{1 + n/n_1} N_D,$$

with $n_1 = (N_C/\hat{g}_D) \exp\left(-E_D^b/kT\right)$. The neutrality condition (its general from is given in (7.41)) is

$$-n + N_D^+ = -n + N_0 = 0, \tag{7.26}$$

leading to the equation (n is given by (7.10))

$$N_C \exp\left(\frac{E_F - E_C}{kT}\right) - \frac{N_D}{1 + \hat{g} \exp\left(\frac{E_F - E_D}{kT}\right)} = 0. \tag{7.27}$$

[2] We do not agree with the treatment of the conduction band valley degeneracy in [510] for the donor degeneracy factor for Ge and Si.

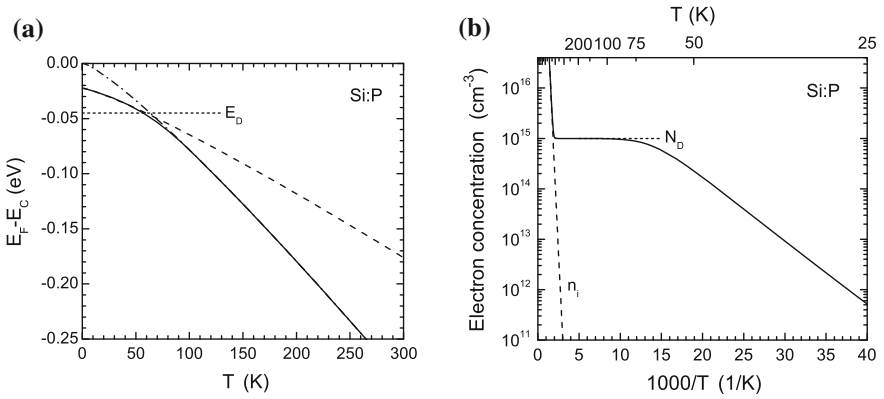


Fig. 7.9 (**a**) Position of the Fermi level in Si:P ($N_D = 10^{15}\,\mathrm{cm}^{-3}$, $E_D^b = 45\,\mathrm{meV}$, no acceptors) as a function of temperature *without* consideration of intrinsic carriers. Zero energy refers to the (temperature-dependent, Table 6.4) conduction-band edge E_C with approximative solutions for low (*dashed line*, (7.29)) and high (*dash-dotted line*, (7.30)) temperatures. (**b**) Corresponding density of conduction-band electrons as a function of temperature

Solving this equation will yield the Fermi level (as a function of temperature T, doping level E_D and doping concentration N_D).[3] The solution is

$$E_F = E_C - E_D^b + kT \ln\left(\frac{\left[1 + 4\hat{g}_D \frac{N_D}{N_C} \exp\left(\frac{E_D^b}{kT}\right)\right]^{1/2} - 1}{2\,\hat{g}_D}\right). \tag{7.28}$$

For $T \to 0$ the Fermi level is, as expected, in the center between the populated and unpopulated states, i.e. at $E_F = E_C - E_D^b/2$. In Fig. 7.9a the position of the Fermi is shown for a donor with 45 meV binding energy in Si. For low temperatures the solution can be approximated as (dashed curve in Fig. 7.9b)

$$E_F \cong E_C - \frac{1}{2} E_D^b + \frac{1}{2} kT \ln\left(\frac{N_D}{\hat{g}_D N_C}\right). \tag{7.29}$$

The freeze-out of carriers in n-type silicon has been discussed in detail in [512], taking into account the effects of the fine structure of the donor states. We note that the fairly high donor binding energy in silicon leads to freeze-out of carriers at about 40 K and is thus limiting for the low-temperature performance of devices. Ge has smaller donor ionization energies and subsequently a lower freeze-out temperature of 20 K. For n-type GaAs, conductivity is preserved down to even lower temperatures.

[3] As usual, the Fermi level is determined by the global charge neutrality, see also Sect. 4.2.2.

We note that the freeze-out of carriers involves the recombination of free electrons with the ionized donors. This aspect is considered in Sect. 10.9. Microscopically this process is equal to the emission of a (far infrared) photon [513, 514]. Similarly the release of an electron from the donor is due to the absorption of a photon.

For higher temperatures, when the electron density saturates towards N_D, the approximate solution is (dash-dotted curve in Fig. 7.9a)

$$E_F \cong E_C + kT \ln\left(\frac{N_D}{N_C}\right). \tag{7.30}$$

The electron density n is given (still in the Boltzmann approximation) by

$$\begin{aligned} n &= N_C \exp\left(-\frac{E_D^b}{kT}\right) \frac{\left[1 + 4\,\hat{g}_D \frac{N_D}{N_C} \exp\left(\frac{E_D^b}{kT}\right)\right]^{1/2} - 1}{2\,\hat{g}_D} \\ &= \frac{2\,N_D}{1 + \left[1 + 4\,\hat{g}_D \frac{N_D}{N_C} \exp\left(\frac{E_D^b}{kT}\right)\right]^{1/2}}. \end{aligned} \tag{7.31}$$

The theoretical electron density as a function of temperature is shown in Fig. 7.9b. It fits very well to experimental data for arsenic doped germanium [515] as shown in Fig. 7.10 (Arrhenius plot, $\ln n$ vs. $1/T$).

For low temperatures, the solution (7.31) is close to

$$n \cong \sqrt{\frac{N_D N_C}{\hat{g}_D}} \exp\left(-\frac{E_D^b}{2kT}\right) = \sqrt{n_1\, N_D}. \tag{7.32}$$

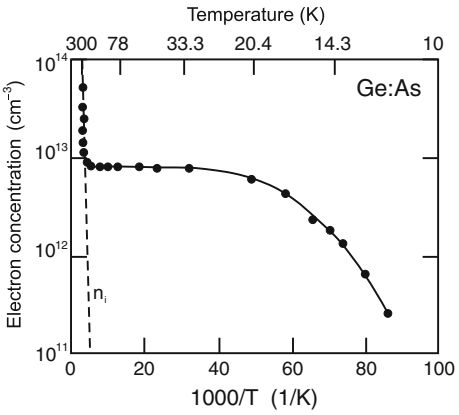


Fig. 7.10 Electron concentration as a function of temperature for a Ge:As sample with $N_D \approx 10^{13}\,\mathrm{cm}^{-3}$. *Solid line* is fit to the data with a donor binding energy of 12.7 meV. Adapted from [515]

For high temperatures, $n \cong N_\mathrm{D}$. This regime is called *exhaustion* or *saturation* since all possible electrons have been ionized from their donors. We note that even in this case $np = n_\mathrm{i} p_\mathrm{i}$ holds, however, $n \gg p$.

While the characteristic energy for the ionization of electrons from donors is E_D^b, at high enough temperatures electrons are transferred also from the valence band into the conduction band. Thus, in order to make the above consideration valid for all temperatures, the intrinsic conduction also has to be considered. The neutrality condition (still in the absence of any acceptors) is

$$-n + p + N_\mathrm{D}^+ = 0. \tag{7.33}$$

Using (7.10) and $p = n_\mathrm{i}^2/n$, the equation reads:

$$N_\mathrm{C} \exp\left(\frac{E_\mathrm{F} - E_\mathrm{C}}{kT}\right) - \frac{n_\mathrm{i}^2}{N_\mathrm{C} \exp\left(\frac{E_\mathrm{F}-E_\mathrm{C}}{kT}\right)} - \frac{N_\mathrm{D}}{1 + \hat{g}_\mathrm{D} \exp\left(\frac{E_\mathrm{F}-E_\mathrm{D}}{kT}\right)} = 0. \tag{7.34}$$

The solution is

$$E_\mathrm{F} = E_\mathrm{C} - E_\mathrm{D}^b + kT \ln\left[\frac{\beta/\gamma + \gamma/N_\mathrm{C}^2 - 1}{3\,\hat{g}_\mathrm{D}}\right], \tag{7.35}$$

with

$$\gamma = \left(\frac{-N_\mathrm{C}^4\,\alpha + \sqrt{(N_\mathrm{C}^4\,\alpha)^2 - 4(N_\mathrm{C}^2\,\beta)^3}}{2}\right)^{1/3} \tag{7.36a}$$

$$\beta = N_\mathrm{C}^2 + 3\hat{g}_\mathrm{D} N_\mathrm{C} N_\mathrm{D} \exp\left(\frac{E_\mathrm{D}}{kT}\right) + 3\hat{g}_\mathrm{D}^2 n_\mathrm{i}^2 \exp\left(\frac{2E_\mathrm{D}}{kT}\right) \tag{7.36b}$$

$$\alpha = 2N_\mathrm{C}^2 + 9\hat{g}_\mathrm{D} N_\mathrm{C} N_\mathrm{D} \exp\left(\frac{E_\mathrm{D}}{kT}\right) - 18\hat{g}_\mathrm{D}^2 n_\mathrm{i}^2 \exp\left(\frac{2E_\mathrm{D}}{kT}\right). \tag{7.36c}$$

The temperature-dependent position of the Fermi level is shown in Fig. 7.11. The carrier concentration is given by

$$n = N_\mathrm{C} \exp\left(-\frac{E_\mathrm{D}^b}{kT}\right) \frac{\beta/\gamma + \gamma/N_\mathrm{C}^2 - 1}{3\,\hat{g}_\mathrm{D}}. \tag{7.37}$$

The three important regimes are the intrinsic conduction at high temperatures when $n_\mathrm{i} \gg N_\mathrm{D}$, the exhaustion at intermediate temperatures when $n_\mathrm{i} \ll N_\mathrm{D}$ and $kT > E_\mathrm{D}$, and finally the freeze-out regime for $kT \ll E_\mathrm{D}$ at low temperatures when the electrons condense back into the donors. The three regimes can be seen in the experimental data of Fig. 7.10 for donors (n-Ge) and Fig. 7.15 for acceptors (p-Ge).

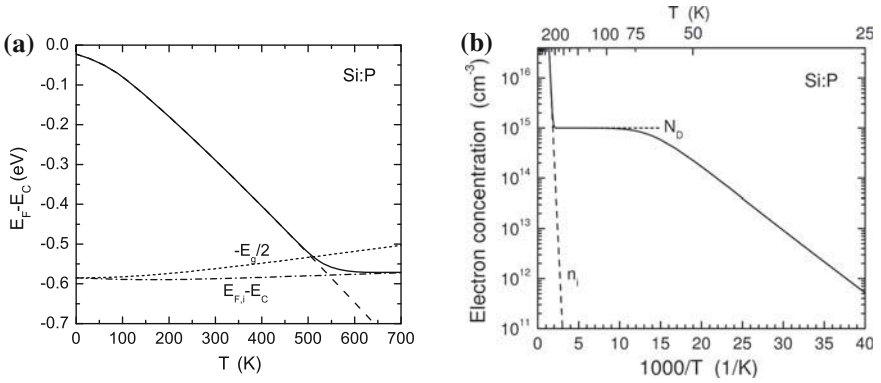


Fig. 7.11 (**a**) Position of the Fermi level in Si:P ($N_D = 10^{15}\,cm^{-3}$, $E_D^b = 45\,meV$, no acceptors) as a function of temperature. The temperature dependence of the band gap (as given in Table 6.4) has been taken into account. Zero energy refers to the conduction-band edge for all temperatures. The *dotted curve* shows $E_g/2$. The *dashed* (*dash-dotted*) *line* shows the low- (high-) temperature limit according to (7.29) and (7.18), respectively. (**b**) Corresponding electron concentration as a function of temperature. The *dashed line* shows the intrinsic carrier density

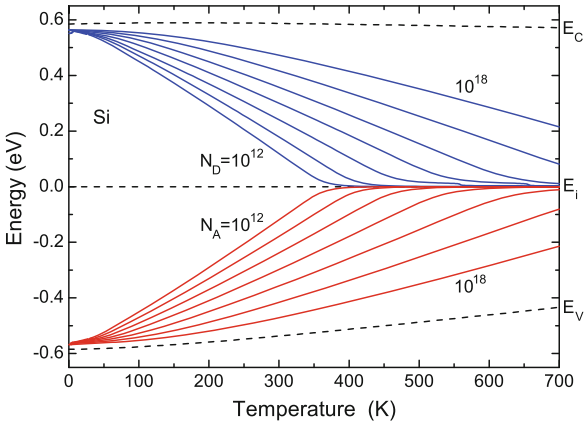


Fig. 7.12 Fermi level in silicon as a function of temperature for various doping levels (n-type (*blue lines*) and p-type (*red lines*)) of $10^{12}, 10^{13}, \ldots, 10^{18}\,cm^{-3}$. The intrinsic Fermi level is chosen as zero energy for all temperatures. The (temperature-dependent) conduction and valence band edges are shown as *dashed lines*

A similar plot as in Fig. 7.11a is shown in Fig. 7.12. With increasing temperature the Fermi level shifts from close to the band edge towards the band center. At higher doping the shift begins at higher temperatures.

The electronic states of individual donors can be directly visualized by scanning tunneling microscopy (STM) as shown in Fig. 7.13 for Si:P. For small negative bias, tunneling occurs through the charged dopant that is located within the first three monolayers. At high negative bias the large contribution from the filled valence band masks the effect of the donor. This image, however, shows that the contrast attributed to the dopant atom is not due to surface defects or absorbates.

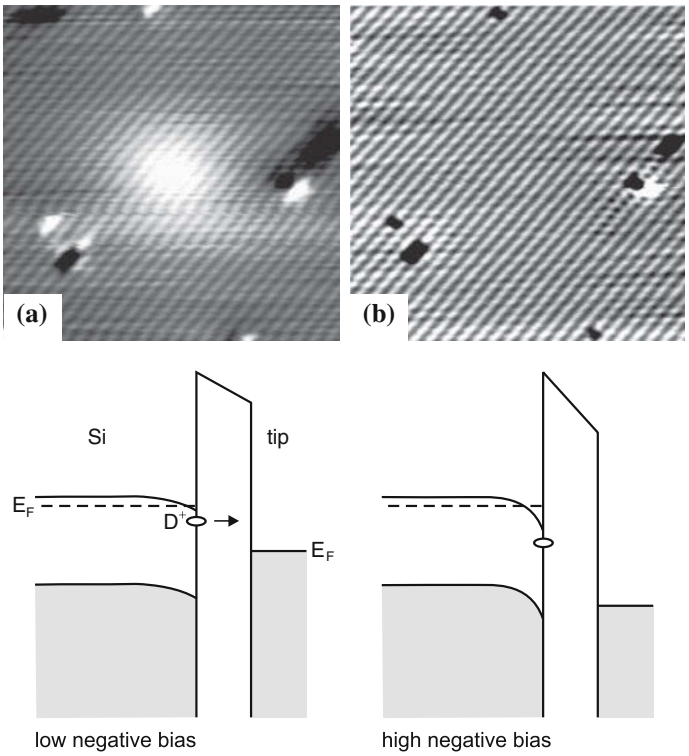


Fig. 7.13 Filled-state image of a phosphorus atom underneath a Si (001) surface at a tunneling current of 110 pA. The doping level is $5 \times 10^{17}\,\mathrm{cm}^{-3}$. (**a**) Sample bias -0.6 V, (**b**) sample bias -1.5 V between Si:P and tip. Image sizes are $22 \times 22\,\mathrm{nm}^2$. Reprinted with permission from [516], © 2004 APS. *Lower row* Schematic band diagrams for the two bias situations

7.5.2 Acceptors

A group-III atom in Si has one electron too few for the tetrahedral bond. Thus, it 'borrows' an electron from the electron gas (in the valence band) and thus leaves a missing electron (termed hole) in the valence band (Fig. 7.14). The energy level of the impurity is in the gap close to the valence-band edge. The latter consideration is made in the electron picture. In the hole picture, the acceptor ion has a hole and the hole ionizes (at sufficient temperature) into the valence band. After ionization the acceptor is charged negatively. Also, for this system a Bohr-like situation arises that is, however, more complicated than for donors because of the degeneracy of the valence bands and their warping.

In Table 7.4 the acceptor binding energies E_{A}^{b} for group-III atoms in C, Ge and Si are listed. The absolute acceptor energy is given as $E_{\mathrm{A}} = E_{\mathrm{V}} + E_{\mathrm{A}}^{b}$. In Table 7.5 acceptor binding energies are listed for GaAs, GaP and GaN. While in GaAs some

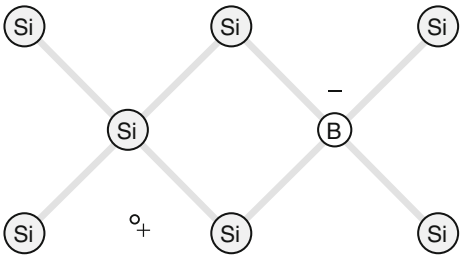


Fig. 7.14 Boron impurity in silicon. Boron accepts one electron and a fixed negative charge remains

Table 7.4 Binding energies E_A^b of group-III acceptors in elemental semiconductors

	B	Al	Ga	In
C	369			
Si	45	57	65	16
Ge	10.4	10.2	10.8	11.2

Data for diamond from [517, 518]. All values in meV

Table 7.5 Binding energies E_A^b of acceptors in GaAs, GaP and GaN (low concentration values, data from [519, 520])

	V site		III site	
GaAs	C	27	Be	28
	Si	34.8	Mg	28.8
	Ge	40.4	Zn	30.7
	Sn	167	Cd	34.7
GaP	C	54	Be	57
	Si	210	Mg	60
	Ge	265	Zn	70
			Cd	102
GaN	C	230	Mg	220
	Si	224	Zn	340
			Cd	550

All values in meV

acceptors are close to the effective mass value of 27 meV, in GaP the deviation from the effective-mass value ≈50 meV is large.

When the conductivity is determined by holes or electrons, the material is called p-type or n-type, respectively. We note that some metals also show hole conduction (e.g. Al). However, for metals the conductivity type is fixed, while the same semiconductor can be made n- or p-type with the appropriate doping.

The acceptor concentration is denoted by N_A. The concentration of neutral acceptors is N_A^0, the concentration of charged acceptors is N_A^-. Of course

$$N_A = N_A^0 + N_A^-. \tag{7.38}$$

The ratio of the degeneracy of the (singly) filled and empty acceptor level is $\hat{g}_A$. In Ge $\hat{g}_A = 4$ since the localized hole wave function may be formed in EMA with four Bloch wave functions (heavy and light holes) [521]. For Si with its small split-off energy (Table 6.6) $\hat{g}_A = 6$ according to [522]. For doubly ionized acceptors, e.g. Zn in Si and Ge (see Sect. 7.7.3), the more shallow level ($Zn^- \rightarrow Zn^0$) has $\hat{g}_A = 6/4 = 1.5$ in Ge [522]. A more general discussion of the degeneracy factor for multiply charged acceptors can be found in [510, 523]. Similar to the considerations for electrons and donors we have

$$\frac{N_A^0}{N_A^-} = \hat{g}_A \exp\left(-\frac{E_F - E_A}{kT}\right). \tag{7.39}$$

The population of the acceptor levels is given by

$$N_A^- = \frac{N_A}{1 + \hat{g}_A \exp\left(-\frac{E_F - E_A}{kT}\right)}. \tag{7.40}$$

The formulas for the position of the Fermi level and the hole density are analogous to those obtained for electrons and donors and will not be explicitly given here. The analogue to Fig. 7.11b is shown for data on p-doped Ge [524, 525] in Fig. 7.15. The acceptor activation energy is 11 meV which could be due to various impurities (cf. Table 7.4). The different impurities (B, Al, Ga) can be distinguished by photothermal ionization spectroscopy [525] (cmp. Sect. 9.7).

In Fig. 7.12, the temperature dependence of the Fermi level is included for p-type Si. With increasing temperature the Fermi level shifts from the valence-band edge (For $T = 0$, $E_F = E_V + E_A^b/2$) towards the middle of the band gap (intrinsic Fermi level).

Also, the wavefunction at acceptors can be imaged using scanning tunneling microscopy [526]. In [527] images of ionized and neutral Mn in GaAs have been

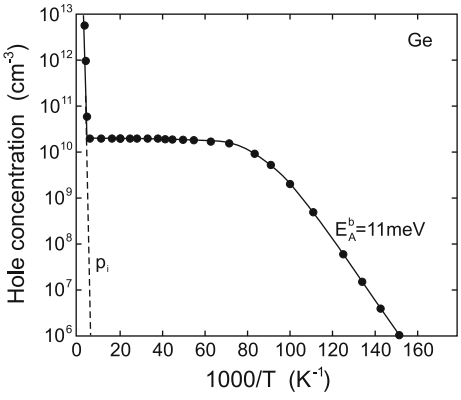


Fig. 7.15 Carrier concentration as a function of temperature for p-type Ge. The net shallow level concentration is $2 \times 10^{10}\,cm^{-3}$. *Solid line* is fit to the data, the *dashed line* indicates the intrinsic hole concentration p_i. Adapted from [525]

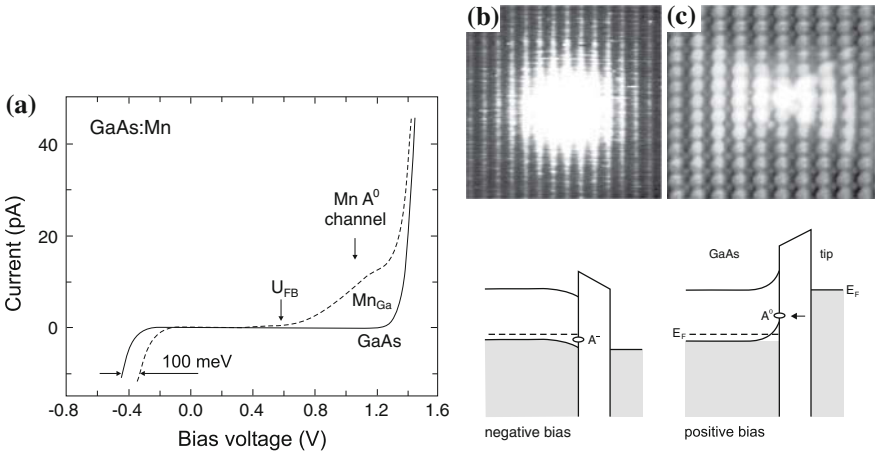


Fig. 7.16 (**a**) Tunneling I–V characteristic of GaAs:Mn sample. *Solid* (*dashed*) line is for pure GaAs (subsurface Mn on Ga site). U_{FB} denotes the simulated flat-band voltage. Adapted from [527]. (**b**, **c**) STM images of a Mn atom underneath a GaAs (110) surface. The doping level is $3 \times 10^{18}\,\mathrm{cm}^{-3}$. (**b**) Sample bias $-0.7\,\mathrm{V}$, (**c**) sample bias $+0.6\,\mathrm{V}$. Below the images are schematic band diagrams of GaAs:Mn and tip. Image sizes are (**b**) $8 \times 8\,\mathrm{nm}^2$ and (**c**) $5.6 \times 5\,\mathrm{nm}^2$. Reprinted with permission from [527], ©2004 APS. *Lower row* under parts (**a**, **b**): Schematic band diagrams for the two bias situations

reported (Fig. 7.16b). The tunneling I–V characteristics are shown in Fig. 7.16a. At negative bias, the acceptor is ionized and appears spherically symmetric due to the effect of the A^- ion Coulomb potential on the valence-band states. At intermediate positive voltages, tunneling is through the neutral state. The wavefunction of A^0 looks like a bow-tie due to the admixture of d-wavefunctions [528]. The Mn atom is presumably in the third subsurface atomic layer. At even higher positive bias the contrast due to the dopant is lost because the image is dominated by a large tunneling current from the tip to the empty conduction band.

7.5.3 Compensation

When donors *and* acceptors are simultaneously present, some of the impurities will compensate each other. Electrons from donors will recombine with holes on the acceptors. Depending on the quantitative situation the semiconductor can be n- or p-type. This situation can be invoked by intentional doping with donors or acceptors or by the unintentional background of donors (acceptors) in p-doped (n-doped) material. Also the formation of pairs, exhibiting a new defect level different from the single donor or single acceptor, has been described, e.g. for Se and B in silicon [273].

The charge-neutrality condition (now finally in its most general form) reads

$$-n + p - N_{\mathrm{A}}^{-} + N_{\mathrm{D}}^{+} = 0. \tag{7.41}$$

We will now discuss the case of the presence of donors and acceptors, but limit ourselves to sufficiently low temperatures (or wide band gaps) such that the intrinsic carrier density can be neglected. We assume Boltzmann statistics and assume here $N_{\mathrm{D}} > N_{\mathrm{A}}$. Then it is a very good approximation to use $N_{\mathrm{A}}^{-} = N_{\mathrm{A}}$ since there are enough electrons from the donors to recombine with (and thus compensate) all acceptors. Under the given assumptions regarding the temperature $p = 0$ and the material is n-type. Thus, in order to determine the position of the Fermi level, the charge-neutrality condition

$$n + N_{\mathrm{A}} - N_{\mathrm{D}}^{+} = 0 \tag{7.42}$$

must be solved (compare to (7.27))

$$N_{\mathrm{C}} \exp\left(\frac{E_{\mathrm{F}} - E_{\mathrm{C}}}{kT}\right) + N_{\mathrm{A}} - \frac{N_{\mathrm{D}}}{1 + \hat{g} \exp(\frac{E_{\mathrm{F}} - E_{\mathrm{D}}}{kT})} = 0. \tag{7.43}$$

We rewrite (7.42) and find $N_{\mathrm{D}} - N_{\mathrm{A}} - n = N_{\mathrm{D}}^{0} = N_{\mathrm{D}}^{+}\, \hat{g}_{\mathrm{D}} \exp\left(\frac{E_{\mathrm{F}} - E_{\mathrm{D}}}{kT}\right)$ using (7.24). Using again (7.42) and also (7.10), (7.43) can be written as

$$\frac{n\,(n + N_{\mathrm{A}})}{N_{\mathrm{D}} - N_{\mathrm{A}} - n} = \frac{N_{\mathrm{C}}}{\hat{g}_{\mathrm{D}}} \exp\left(-\frac{E_{\mathrm{D}}^{\mathrm{b}}}{kT}\right), \tag{7.44}$$

a form given in [529]. Analogously for compensated p-type material

$$\frac{p\,(p + N_{\mathrm{D}})}{N_{\mathrm{A}} - N_{\mathrm{D}} - p} = \frac{N_{\mathrm{V}}}{\hat{g}_{\mathrm{A}}} \exp\left(-\frac{E_{\mathrm{A}}^{\mathrm{b}}}{kT}\right) \tag{7.45}$$

holds.

The solution of (7.43) is

$$E_{\mathrm{F}} = E_{\mathrm{C}} - E_{\mathrm{D}}^{b} + kT \ln\left(\frac{\left[\alpha^2 + 4\hat{g}_{\mathrm{D}} \frac{N_{\mathrm{D}} - N_{\mathrm{A}}}{N_{\mathrm{C}}} \exp\left(\frac{E_{\mathrm{D}}^{b}}{kT}\right)\right]^{1/2} - \alpha}{2\,\hat{g}_{\mathrm{D}}}\right), \tag{7.46}$$

with

$$\alpha = 1 + \hat{g}_{\mathrm{D}} \frac{N_{\mathrm{A}}}{N_{\mathrm{C}}} \exp\left(\frac{E_{\mathrm{D}}^{b}}{kT}\right) = 1 + \frac{N_{\mathrm{A}}}{\beta} \tag{7.47a}$$

$$\beta = \frac{N_\mathrm{C}}{\hat{g}_\mathrm{D}} \exp\left(-\frac{E_\mathrm{D}^\mathrm{b}}{kT}\right). \tag{7.47b}$$

The carrier density is best obtained from (7.44),

$$n = \sqrt{(N_\mathrm{A} - \beta)^2 + 4\, N_\mathrm{D}\, \beta} - \frac{N_\mathrm{A} + \beta}{2}. \tag{7.48}$$

For $N_\mathrm{A} = 0$ we have $\alpha = 1$ and (7.28) is reproduced, as expected. For $T = 0$ (and $N_\mathrm{A} \neq 0$) the Fermi energy lies at $E_\mathrm{F} = E_\mathrm{D}$ since the donor level is *partially* filled ($N_\mathrm{D}^0 = N_\mathrm{D} - N_\mathrm{A}$). For low temperatures the Fermi level is approximated by

$$E_\mathrm{F} \cong E_\mathrm{C} - E_\mathrm{D}^b + kT\, \ln\left(\frac{N_\mathrm{D}/N_\mathrm{A} - 1}{\hat{g}_\mathrm{D}}\right). \tag{7.49}$$

The corresponding carrier density at low temperatures is

$$n = \frac{N_\mathrm{C}}{\hat{g}_\mathrm{D}} \exp\left(-\frac{E_\mathrm{D}^b}{kT}\right)\left(\frac{N_\mathrm{D}}{N_\mathrm{A}} - 1\right). \tag{7.50}$$

For higher temperatures (7.32) holds approximately for $n > N_\mathrm{A}$; the slope is now given by $E_\mathrm{D}^b/2$ as in the uncompensated case (Fig. 7.17b). For sufficiently high

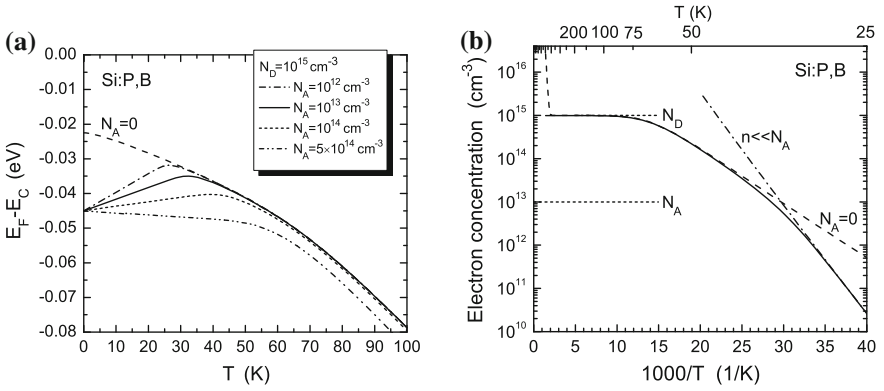


Fig. 7.17 (**a**) Position of Fermi level in partially compensated Si:P,B ($N_\mathrm{D} = 10^{15}\,\mathrm{cm}^{-3}$, $E_\mathrm{D}^b = 45\,\mathrm{meV}$, $E_\mathrm{A}^b = 45\,\mathrm{meV}$, *solid line* $N_\mathrm{A} = 10^{13}\,\mathrm{cm}^{-3}$, *dashed line* $N_\mathrm{A} = 0$, *dash-dotted line* $N_\mathrm{A} = 10^{12}\,\mathrm{cm}^{-3}$, *short-dashed line* $N_\mathrm{A} = 10^{14}\,\mathrm{cm}^{-3}$, *dash-double dotted line* $N_\mathrm{A} = 5 \times 10^{14}\,\mathrm{cm}^{-3}$) as a function of temperature. (**b**) Corresponding electron concentration for $N_\mathrm{A} = 10^{13}\,\mathrm{cm}^{-3}$ as a function of temperature (neglecting intrinsic carriers), *dashed line* for $N_\mathrm{A} = 0$ according to (7.31), *dash-dotted line* approximation for $n \ll N_\mathrm{A}$ as in (7.50)

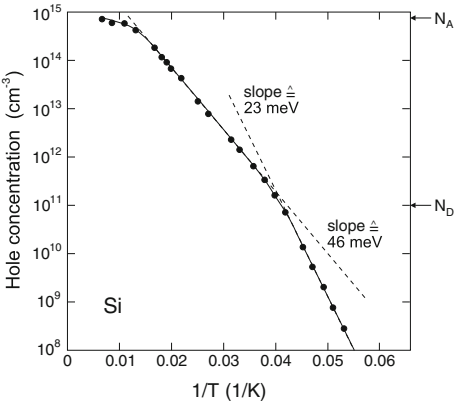


Fig. 7.18 Hole density in p-type silicon ($N_A = 7.4 \times 10^{14}\,\mathrm{cm}^{-3}$, $E_A^b = 46\,\mathrm{meV}$ (probably boron) and partial compensation with $N_D = 1.0 \times 10^{11}\,\mathrm{cm}^{-3}$). Adapted from [530]

temperatures in the exhaustion regime (but still $n_i < n$) the electron density is given by

$$n \cong N_D - N_A. \tag{7.51}$$

At even higher temperatures the electron density will be determined by the intrinsic carrier concentration. Only in this case $p \neq 0$!

An experimental example is shown in Fig. 7.18 for partially compensated p-Si (with $N_D \ll N_A$). The change of slope around $p \approx N_D$ is obvious.

If donors are added to a p-type semiconductor, first the semiconductor remains p-conducting as long as $N_D \ll N_A$. If the donor concentration becomes larger than the acceptor concentration, the conductivity type switches from p- to n-conduction. If the impurities are exhausted at room temperature, the lowest carrier concentration is reached for $N_D = N_A$. Such a scenario is shown for p-type $In_xGa_{1-x}As_{1-y}N_y$ doped with various concentrations of Si in Fig. 7.19. At high Si incorporation, the number of charge carriers saturates due to autocompensation (see Sect. 7.5.5) and the formation of Si precipitates. Since the ionization energies of donors and acceptors are typically different, the situation for $N_D \approx N_A$ needs, in general, to be investigated carefully and will depend on the temperature.

7.5.4 *Multiple Impurities*

If more than one donor species is present, (7.43) can be generalized, e.g. for the case of two donors D1 and D2 in the presence of compensating acceptors,

$$n + N_A - \frac{N_{D1}}{1 + \hat{g}_1 \exp(\frac{E_F - E_{D1}}{kT})} - \frac{N_{D2}}{1 + \hat{g}_2 \exp(\frac{E_F - E_{D2}}{kT})} = 0. \tag{7.52}$$

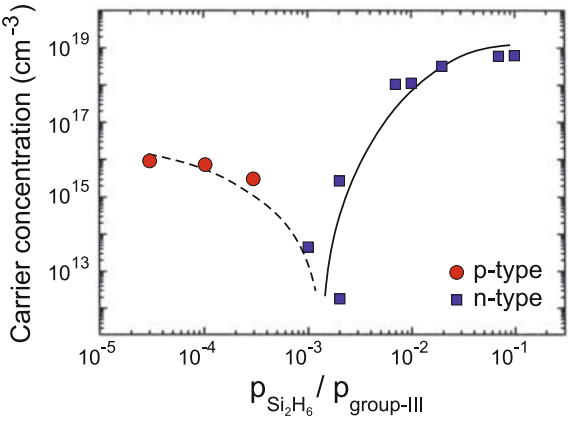


Fig. 7.19 Carrier concentration and conductivity type (*red circles* p, *blue squares* n) for MOVPE-grown $In_xGa_{1-x}As_{1-y}N_y$ layers on GaAs (001) (layer thickness $\approx 1\,\mu m$, $x \approx 5\%$, $y \approx 1.6\%$) doped with different amounts of silicon. The ordinate is the ratio of the partial pressures of disilane and the group-III precursors (TMIn and TMGa) in the gas phase entering the MOVPE reactor. *Lines* are guides to the eye. Experimental data from [531]

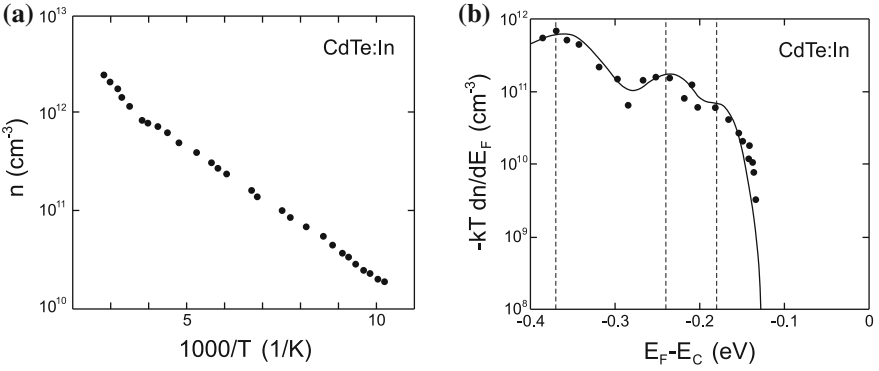


Fig. 7.20 (**a**) Electron concentration versus temperature as determined from Hall effect for a CdTe sample doped with indium. (**b**) $-kT\ \mathrm{d}n/\mathrm{d}E_F$, as determined from the experimental Hall data (*symbols*). The *solid line* is theory for three donor levels ($E_{D1} = E_C - 0.37\,eV$, $N_{D1} = 2.5 \times 10^{12}\,cm^{-3}$; $E_{D2} = E_C - 0.24\,eV$, $N_{D2} = 7.0 \times 10^{11}\,cm^{-3}$; $E_{D3} = E_C - 0.18\,eV$, $N_{D3} = 2.5 \times 10^{11}\,cm^{-3}$) whose energy positions are indicated by *dashed lines*. Adapted from [533]

This case is treated in [532]. Simple high and low temperature approximations can be found where the trap with the larger and smaller activation energy, respectively, dominates. The case for multiple acceptors (and compensating donors) is treated analogously. As detailed in [533], the function $\mathrm{d}n/\mathrm{d}E_F$ has a maximum at the donor level position; this can be used to visualize the contribution of several donors (with sufficiently different binding energies) from $n(T)$ as measured by Hall effect (Fig. 7.20).

7.5.5 *Amphoteric Impurities*

If an impurity atom can act as a donor and acceptor it is called amphoteric. This can occur if the impurity has several levels in the band gap (such as Au in Ge or Si). In this case, the nature of the impurity depends on the position of the Fermi level. Another possibility is the incorporation on different lattice sites. For example, carbon in GaAs is a donor if incorporated on the Ga-site. On the As-site carbon acts as an acceptor.

Thus, e.g. crystal growth kinetics can determine the conductivity type. In Fig. 7.21 the conductivity due to carbon background is shown for GaAs grown using MOVPE under various growth conditions. At high (low) arsine partial pressure incorporation of carbon on As-sites is less (more) probable, thus the conductivity is n-type (p-type). Also, growth on different surfaces can evoke different impurity incorporation, e.g. n-type on (001) GaAs and p-type on (311)A GaAs, since the latter is Ga-stabilized.

The charge density at an impurity nucleus can be investigated via the isomer shift as determined by Mössbauer spectroscopy [535, 536]. The incorporation of the isotope ^{119}Sn can be controlled in III-V compounds to be on cation or anion site as donor or acceptor, respectively. This is accomplished by introducing ^{119}In or ^{119}Sb on group-III and group-V site, respectively, both decaying into ^{119}Sn without leaving their lattice site. The isomer shifts of ^{119}Sn in various III-V compounds are shown in Fig. 7.22. In [536] it is concluded from these data that the tin donor is formed by a positive tin ion and the electron charge transfer to its neighboring (group-V) atoms is rather small. For tin as an acceptor, for the present conditions an ionized, i.e. negatively charged acceptor, the isomer shift follows closely the trend from substitution in group-IV semiconductors. Therefore four electrons form the tetrahedral bond, while the extra electron is located rather at the (positively charged) group-III next neighbors and not in the impurity cell. The difference to the point charge Coulomb distribution is called central-cell correction.

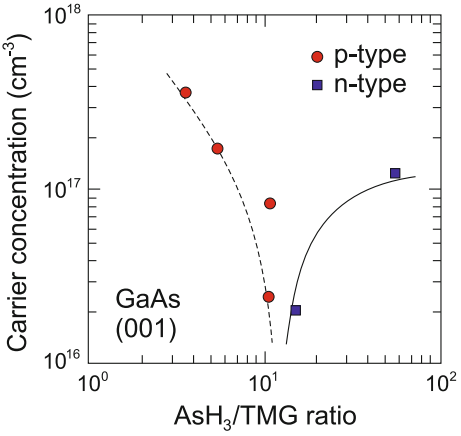


Fig. 7.21 Background doping of GaAs due to carbon in MOVPE for different ratios of the partial pressures of AsH_3 and TMG (trimethylgallium). The conductivity type (*blue squares* n-type, *red circles* p-type) depends on the incorporation of C from CH_3 radicals on Ga- or As-site. *Lines* are guides to the eye. Experimental data from [534]

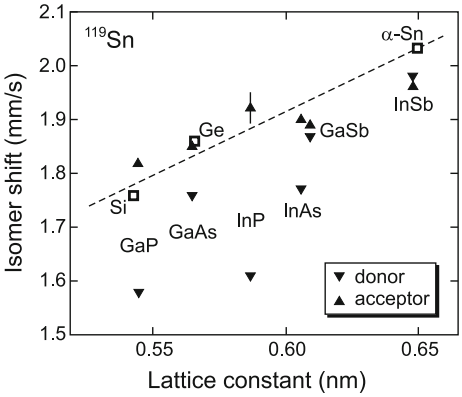


Fig. 7.22 Isomer shift (relative to $CaSnO_3$) of ^{119}Sn in various group-IV and III-V compound semiconductors as labeled. *Dashed line* is trend from isoelectronic substitution. Experimental data from [536]

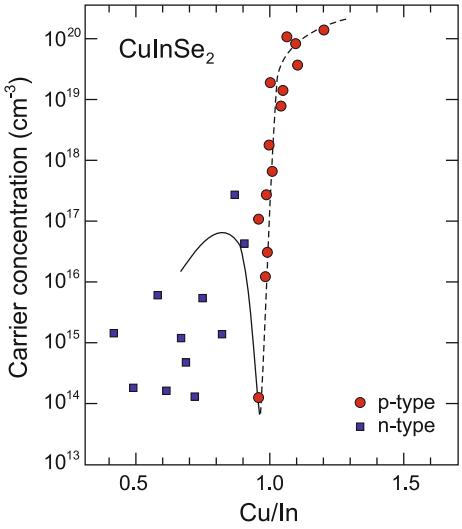


Fig. 7.23 Carrier concentration and conductivity type (*blue squares* n-type, *red circles* p-type) as a function of stoichiometry for $CuInSe_2$ thin films. *Lines* are guides to the eye. Experimental data from [537]

Deviation from the ideal stoichiometry introduces point defects that can be electrically active and change conductivity type and carrier concentration. In the case of $CuInSe_2$, excess Cu could go on interstitial positions or promote selenium vacancies, both leading to n-type behavior. This material is particularly sensitive to deviations from ideal stoichiometry for both Cu/In ratio (Fig. 7.23) and Se deficiency [537].

7.5.6 Autodoping

If intrinsic defects such as vacancies or interstitials, possibly as a result of non-stoichiometry, or anti-site defects cause electronic levels relevant for conductivity one speaks of autodoping. An example is the role of A-B antisites in AB_2O_4 spinels

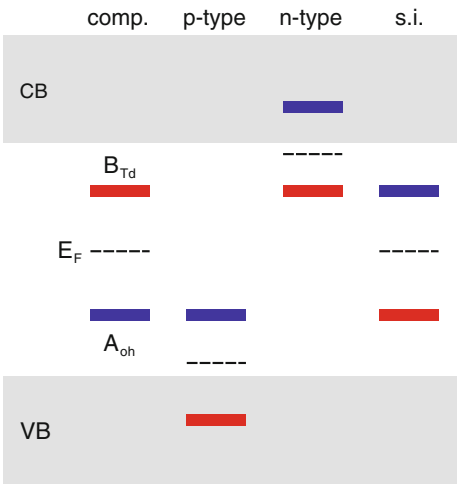


Fig. 7.24 Schematic position of electronic levels of A_B (*blue*, 0/− transition level) and B_A (*red*, +/0 transition level) defects in AB_2O_4 spinels and resulting material properties (compensated, n- or p-type or semi-insulating). After [538]

(Sect. 3.4.7). In the perfect crystal the A (B) atoms occupy tetraeder (octaeder) places. Typical charges are A^{2+} and B^{3+}. Thus (without charge transfer) the A atom on octaeder site (A_{Oh}) acts like a donor and the B atom on a tetraeder site (B_{Td}) as an acceptor. Such defects have been classified in [538] as being able to create compensated, semi-insulating, n-type or p-type material depending on the defect formation energies and the position of the electronic levels of the A_B and B_A defects in the band gap (Fig. 7.24). An example for a p-type spinel oxide is $ZnCo_2O_4$ [539].

7.5.7 *High Doping*

For low doping concentrations, the impurity atoms can be considered to be decoupled. At low temperature, only hopping from one impurity to the next is possible due to thermal emission or tunneling and the semiconductor becomes an insulator.

With increasing concentration, the distance between impurities decreases and their wavefunctions can overlap. Then, an impurity band develops (Fig. 7.25). A periodic arrangement of impurity atoms would result in well-defined band edges as found in the Kronig–Penney model. Since the impurity atoms are randomly distributed, the band edges exhibit tails. For high doping, the impurity band overlaps with the conduction band. In the case of compensation, the impurity band is not completely filled and contains (a new type of) holes. In this case, conduction can take place within the impurity band even at low temperature, making the semiconductor a metal. This metal–insulator transition has been discussed by Mott [540]. Examples for highly doped semiconductors are transparent conductive oxides (Chap. 20), the contact layer for an ohmic contact (Sect. 21.2.6) or the active

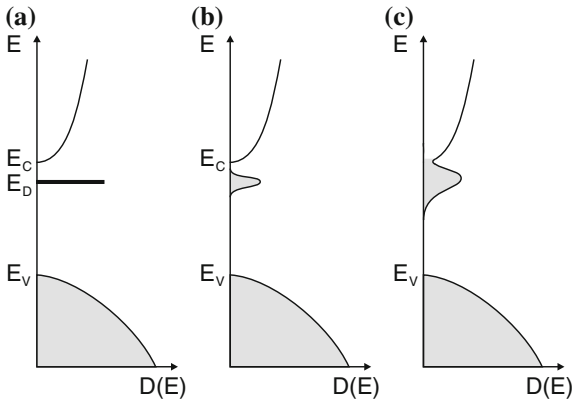


Fig. 7.25 Principle of the formation of a (donor) impurity band. (**a**) Small doping concentration and sharply defined impurity state at E_D, (**b**) increasing doping and development of an impurity band that (**c**) widens further and eventually overlaps with the conduction band for high impurity concentration. The *shaded areas* indicate populated states at $T = 0\,\mathrm{K}$

layers in a tunneling diode (Sect. 21.5.9). The physics, properties and preparation of highly doped semiconductors are treated in detail in [541].

The formation of the impurity band leads to a reduction of the impurity ionization energy as known from (7.19). Typical results are shown in Fig. 7.26a for n-type Ge [515] and Fig. 7.26b for p-type ZnTe [542]. At the critical doping concentration of $N_c = 1.5 \times 10^{17}$, the activation energy for the carrier concentration disappears. Similar effects have been observed for Si [543] and GaAs [544]. The freeze-out of the carrier concentration (see Fig. 7.9) disappears as shown in Fig. 7.27. Critical doping concentrations are listed in Table 7.6. The decrease of the ionization energy E^b (donor or acceptor) follows the dependence [515, 543]

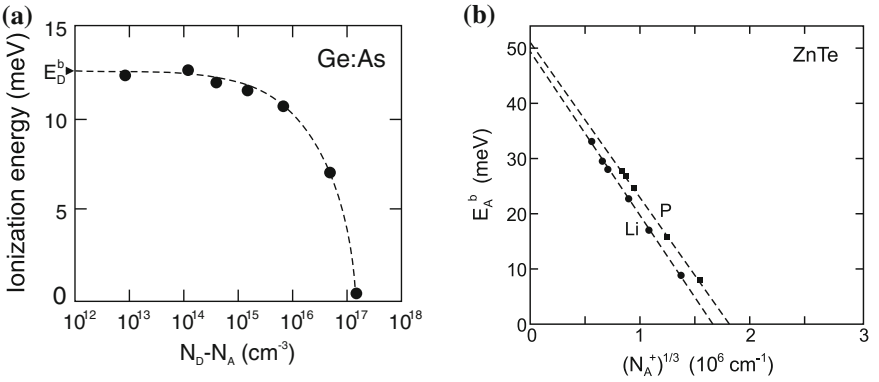


Fig. 7.26 (**a**) Donor ionization energy in n-type Ge for various doping concentrations. *Dashed line* is a guide to the eye. The arrow labeled E_D^b denotes the low-concentration limit (cf. Table 7.2). Experimental data from [515]. (**b**) Acceptor ionization energy for ZnTe:Li and ZnTe:P as a function of the third root of the ionized acceptor concentration. Data from [542]

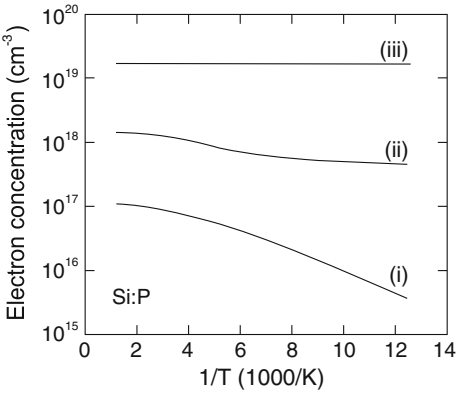


Fig. 7.27 Electron concentration versus inverse temperature for Si:P for three different doping concentrations ((i): $1.2 \times 10^{17}\,\mathrm{cm}^{-3}$, (ii): $1.25 \times 10^{18}\,\mathrm{cm}^{-3}$, (iii): $1.8 \times 10^{19}\,\mathrm{cm}^{-3}$). Experimental data from [543]

Table 7.6 Critical doping concentration for various semiconductors (at room temperature)

Material	Type	$N_c\,(\mathrm{cm}^{-3})$	Ref.
C:B	p	2×10^{20}	[518]
Ge:As	n	1.5×10^{17}	[515]
Si:P	n	1.3×10^{18}	[543]
Si:B	p	6.2×10^{18}	[543]
GaAs	n	1.0×10^{16}	[544]
GaP:Si	n	6×10^{19}	[546]
GaP:Zn	p	2×10^{19}	[547]
GaN:Si	n	2×10^{18}	[548]
GaN:Mg	p	4×10^{20}	[519]
$Al_{0.23}Ga_{0.77}N$:Si	n	3.5×10^{18}	[549]
ZnTe:Li	p	4×10^{18}	[542]
ZnTe:P	p	6×10^{18}	[542]
ZnO:Al	n	8×10^{18}	[550]

$$E^{\mathrm{b}} = E_0^{\mathrm{b}} - \alpha\, N_{\mathrm{i}}^{1/3} = E_0^{\mathrm{b}} \left[1 - \left(\frac{N_{\mathrm{i}}}{N_{\mathrm{c}}} \right)^{1/3} \right], \tag{7.53}$$

where N_{i} is the concentration of ionized dopants. A refined theory, considering screening, shift and tails of the conduction band and most importantly broadening of the donor level has been presented in [545].

The critical density can be estimated from the Mott criterion when the distance of the impurities becomes comparable to their Bohr radius (7.20)

$$2a_{\mathrm{D}} = \frac{3}{2\pi}\, N_{\mathrm{c}}^{1/3}. \tag{7.54}$$

Table 7.7 Maximum electrically active doping concentration for GaAs

Material	Type	N_c (cm^{-3})	Ref.
GaAs:Te	n	2.6×10^{19}	[554]
GaAs:Si	n	1.8×10^{19}	[555]
GaAs:C	p	1.5×10^{21}	[556]
GaAs:Be	p	2×10^{20}	[557]

The pre-factor $3/(2\pi)$ stems from the random distribution of impurities and disappears for a periodic arrangement. The Mott criterion is (rewriting (7.54))

$$a_D N_c^{1/3} \approx 0.24. \tag{7.55}$$

For GaAs with $a_D = 10.3\,nm$, the criterion yields $N_c = 1.2 \times 10^{16}\,cm^{-3}$, in agreement with experiment.

The achievable maximum concentration of electrically active dopants is limited by the concentration dependence of the diffusion coefficient, Coulomb repulsion, autocompensation and the solubility limit [501]. In Table 7.7 the maximum carrier concentrations for GaAs with various dopants are listed.

As an example we show the Ga-doping of epitaxial ZnO layers on sapphire in Fig. 7.28. Under slightly Zn-rich (O-polar) conditions the growth mode is two-dimensional and the carrier concentration increases linearly with the Ga concentration, $n \approx c_{Ga}$, up to high values in the $10^{20}\,cm^{-3}$ range [551]. For O-rich (Zn-polar) conditions the growth mode changes to three-dimensional growth and the activation ratio of Ga donors becomes low [552]. Above a gallium content of 2 %, the octahedral coordination of gallium and thus the partial segregation into a parasitic $ZnGa_2O_4$ spinel phase is observed for [Ga] = 4 % [553].

The random distribution of dopants essentially makes a doped semiconductor a disordered system. The physics of electronic states in disordered systems has been reviewed in [558]. A metal–insulator transition is observed at a certain value of

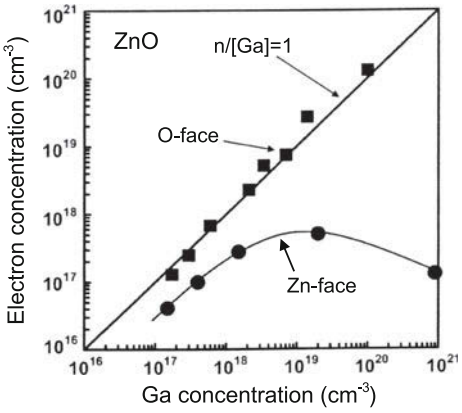


Fig. 7.28 Electron concentration as a function of gallium concentration in MBE grown ZnO:Ga on sapphire for the two different polarities. Adapted from [551, 552]

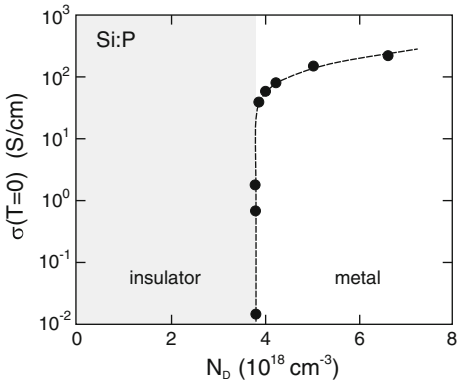


Fig. 7.29 Zero temperature conductivity of Si:P for various (donor) doping concentrations. Experimental data (*symbols*) and guide to the eye (*dashed line*). Adapted from [559]

doping ($N_P = 3.8 \times 10^{18}\,\mathrm{cm}^{-3}$), as shown in Fig. 7.29 for Si:P [559]. For a certain value of disorder all states become localized (Anderson localization [560], cmp. Sect. 8.8).

7.6 Quasi-Fermi Levels

The carrier concentrations were given by (7.6) and (7.7). So far, we have only considered semiconductors in thermodynamic equilibrium for which $np = n_i^2$. In a non-equilibrium situation, e.g. for external excitation or carrier injection in a diode, the electron and hole densities can each take arbitrary values, in principle. In particular, np will no longer be equal to n_i^2 and there is no Fermi level constant throughout the structure. In this case, however, quasi-Fermi levels F_n and F_p for electrons and holes, respectively, are defined via

$$n(\mathbf{r}) = N_C\, F_{1/2}\left(\frac{F_n(\mathbf{r}) - E_C}{kT}\right) \tag{7.56a}$$

$$p(\mathbf{r}) = N_V\, F_{1/2}\left(-\frac{F_p(\mathbf{r}) - E_V}{kT}\right). \tag{7.56b}$$

A quasi-Fermi level is sometimes called imref[4] and can also be denoted as E_{F_n} or E_{F_p}. We emphasize that the quasi-Fermi levels are *only* a means to describe the local carrier density in a logarithmical way. The quasi-Fermi levels can be obtained from the density via

[4] W. Shockley had asked E. Fermi for permission to use his name reversed. Fermi was not too enthusiastic but granted permission.

$$F_\mathrm{n} = E_\mathrm{C} + kT \ln\left(\frac{n}{N_\mathrm{C}}\right) \tag{7.57a}$$

$$F_\mathrm{p} = E_\mathrm{V} - kT \ln\left(\frac{p}{N_\mathrm{V}}\right). \tag{7.57b}$$

The quasi-Fermi levels do not imply that the carrier distribution is actually a Fermi distribution. This is generally no longer the case in thermodynamical nonequilibrium. However, in 'well-behaved' cases the carrier distribution in nonequilibrium can be approximated locally as a Fermi distribution using a local quasi-Fermi level and a local temperature, i.e.

$$f_\mathrm{e}(\mathbf{r}, E) \cong \frac{1}{\exp\left(\frac{E - F_\mathrm{n}(\mathbf{r})}{kT(\mathbf{r})}\right) + 1}. \tag{7.58}$$

Using the quasi-Fermi levels, np is given by

$$n(\mathbf{r})\, p(\mathbf{r}) = n_\mathrm{i}^2 \exp\left(\frac{F_\mathrm{n}(\mathbf{r}) - F_\mathrm{p}(\mathbf{r})}{kT}\right). \tag{7.59}$$

We note that for an inhomogeneous semiconductor or a heterostructure (cf. Chap. 12), n_i may also depend on the spatial position. In the case of thermodynamic equilibrium the difference of the quasi-Fermi levels is zero, i.e. $F_\mathrm{n} - F_\mathrm{p} = 0$ and $F_\mathrm{n} = F_\mathrm{p} = E_\mathrm{F}$.

7.7 Deep Levels

For deep levels the short-range part of the potential determines the energy level. The long-range Coulomb part will only lead to a correction. The term 'deep level' implies that the level is within the band gap and far from the band edges. However, some deep levels (in the sense of the potential being determined by the ion core) have energy levels close to the band edges or even within a band. Details can be found in [233, 561–564].

The wavefunction is strongly localized. Thus, it cannot be composed of Bloch functions, as has been done for the shallow levels for the effective-mass impurity. The localization in $\mathbf{r}$ space leads to a delocalization in $\mathbf{k}$ space. Examples are Si:S, Si:Cu or InP:Fe, GaP:N, ZnTe:O. Deep levels can also be due to intrinsic defects such as vacancies or antisite defects.

Due to the larger distance to the band edges, deep levels are not efficient at providing free electrons or holes. Quite the opposite, they rather capture free carriers and thus lead to a reduction of conductivity. Centers that can capture electrons *and* holes lead to nonradiative recombination of electrons through the deep level into the valence band (see also Chap. 10). This can be useful for the fabrication of semi-insulating

layers with low carrier concentration and fast time response of, e.g. switches and photodetectors.

While the electronic properties of deep levels can be readily characterized, the microscopic origin is not immediately apparent. Next to theoretical modeling of defects and correlation with experimental results, paramagnetic hyperfine interactions have proven useful to identify the microscopic nature of various defects [565].

7.7.1 Charge States

The deep level can have different charge states depending on the occupancy with electrons. The energy position within the gap varies with the charge state due to the Coulomb interaction. Also, the lattice relaxation around the defect depends on the charge state and modifies the energy level.

The localized charge q_d at the defect is the integral over the change $\Delta\rho$ of the charge density compared to the perfect crystal over a sufficiently large volume V_∞ around the defect

$$q_\mathrm{d} = \int_{V_\infty} \Delta\rho(\mathbf{r})\, \mathrm{d}^3\mathbf{r} = \frac{n\,e}{\epsilon_\mathrm{r}}. \tag{7.60}$$

In semiconductors, the charge $q_\mathrm{d}\epsilon_\mathrm{r}$ is an integer multiple of the elementary charge. The defect is said to be in the nth charge state. Each charge state has a certain stable atomic configuration $\mathbf{R}_n$. Each charge state has a ground state and excited states that can each have different stable atomic configurations.

Now, we discuss how the concentration of the various charge states depends on the position of the Fermi level. The overall constraint of global charge neutrality determines the chemical potential of the electron, i.e. the Fermi level in Fermi–Dirac statistics. We use the approximation that the concentration of defects is so small that the mutual interaction of defects becomes negligible.

As an example, we treat the possible reaction $V^0 \rightleftharpoons V^+ + e^-$, where V^0 denotes a neutral vacancy and V^+ is a positively charged vacancy, created by the ionization of an electron from the vacancy into the conduction band. The free energy G depends on the numbers n_0 of neutral and n_+ of positively charged vacancies. The minimum condition is met by

$$\mathrm{d}G = \frac{\partial G}{\partial n_0}\,\mathrm{d}n_0 + \frac{\partial G}{\partial n_+}\,\mathrm{d}n_+ = 0. \tag{7.61}$$

The neutrality constraint is $\mathrm{d}n_0 + \mathrm{d}n_+ = 0$ and therefore the minimum condition reads

$$\frac{\partial G}{\partial n_0} = \frac{\partial G}{\partial n_+}. \tag{7.62}$$

For noninteracting defects and using (4.9) we write

$$\frac{\partial G}{\partial n_0} = G^{\mathrm{f}}(V^0) + kT \ln\left(\frac{n_0}{N_0}\right) \tag{7.63a}$$

$$\frac{\partial G}{\partial n_+} = \frac{\partial G(V^+)}{\partial n_+} + \frac{\partial G(e^-)}{\partial n_+} = G^{\mathrm{f}}_{V^+} + kT \ln\left(\frac{n_+}{N_+}\right) + \mu_{\mathrm{e}^-}, \tag{7.63b}$$

where $N_0 = NZ_0$ and $N_+ = NZ_+$ are the number of available sites, given by the number N of atomic sites and including possible internal degeneracies Z_0 and Z_+, respectively. Degeneracy factors of deep levels are not a simple subject [522] and, e.g. the degeneracy factors of Au donor and acceptor levels in Si are under discussion [566–568]. G_{f} denotes the free enthalpy of formation of the respective defect, as in (4.3). We have written the free enthalpy of the separated pair V^+ and e^- as the sum $G(V^+) + G(e^-)$. $\mu_{\mathrm{e}^-} = \partial G(e^-)/\partial n_+$ is (by definition) the chemical potential of the electron, i.e. the Fermi energy E_{F} of Fermi–Dirac statistics.[5] From (7.63a, b) we find for the ratio of the concentrations of defects $c_0 = n_0/N$ and $c_+ = n_+/N$

$$\frac{c_0}{c_+} = \frac{Z_+}{Z_0} \exp\left(-\frac{G^{\mathrm{f}}_{V^+} - G^{\mathrm{f}}_{V^0} + E_{\mathrm{F}}}{kT}\right) = \frac{Z_+}{Z_0} \exp\left(\frac{E_{\mathrm{t}}(V^0) - E_{\mathrm{F}}}{kT}\right), \tag{7.64}$$

where the trap level energy (for the particular charge transition), $E_{\mathrm{t}}(V^0) = G^{\mathrm{f}}_{V^0} - G^{\mathrm{f}}_{V^+}$, is the free enthalpy of ionization of V^0. We note that c_0 can be obtained from (4.9) and E_{F} is determined by the charge-neutrality condition.

As example experimental data on the charge transition $\mathrm{Fe}^0 \rightleftharpoons \mathrm{Fe}^+ + \mathrm{e}^-$ of interstitial iron (in tetrahedral position, Fig. 7.30a, cmp. Fig. 3.18) in silicon is shown. The concentration of Fe^0 is tracked via the EPR signal from the neutral $S = 1$ state[6] with g-factor $g = 2.07$ [569]. For n-type samples the iron is in neutral state and the maximum EPR signal is found. For strongly p-type samples, the Fermi energy is below the trap level and all iron is in Fe^+ state, yielding no EPR signal at the given g-factor. From the investigation of various silicon samples with different doping levels and consequently different position of the Fermi level, the trap (deep donor) energy is found to be $E_{\mathrm{V}} + 0.375\,\mathrm{eV}$ as indicated in Fig. 7.30b.

[5] The chemical potential in a one-component system is $\mu = \partial G/\partial n = G/n$. In a multicomponent system it is, for the ith component, $\mu_i = \partial G/\partial n_i \neq G/n_i$.

[6] The electron configuration is $3\mathrm{d}^8$ with two paramagnetic electrons. Under uniaxial stress along [100] the EPR line splits into a doublet. [569] Further details can be found in [570].

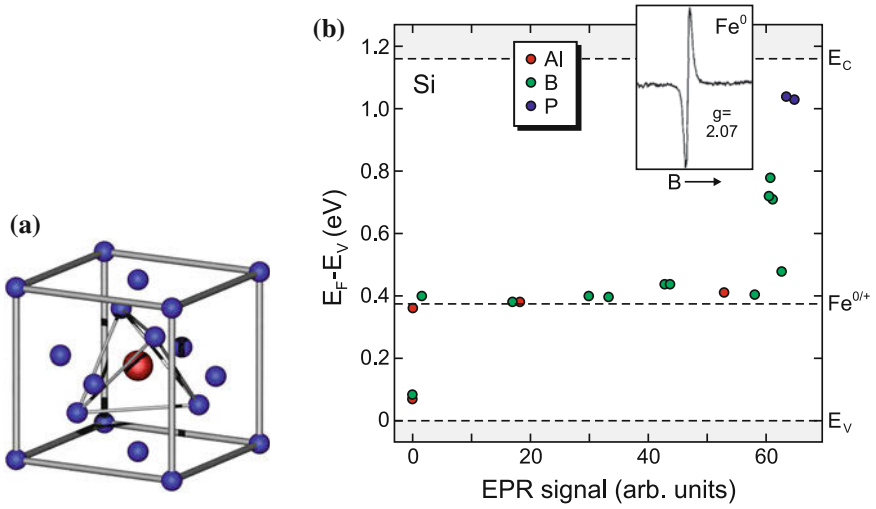


Fig. 7.30 (**a**) Silicon cubic unit cell with an interstitial iron atom (*red*) at tetrahedral site. (**b**) EPR intensity (at $T = 95\,\mathrm{K}$ from interstitial iron in neutral state, Fe^0 with $S = 1$) versus Fermi level position for iron-doped silicon with varying Fermi level due to different amounts of shallow impurity levels from to Al, B and P as labeled. The *shaded areas* indicate the valence and conduction band. The *dashed line* at $E_t = E_V + 0.375\,\mathrm{eV}$ indicates the trap level. The *inset* shows a typical EPR spectrum of Fe^0. Adapted from [571], *inset* adapted from [572]

7.7.2 Double Donors

An impurity that has two extra electrons available after bonding in the matrix may give rise to a double donor. Typical examples are substitutional chalcogenide atoms (S, Se or Te) in silicon [573] and germanium [574], interstitial impurities such as Mg_i in Si [575], or group-V atoms on a group-III site in a III–V compound (antisite defect), such as P_{Ga} in GaP [576] or As_{Ga} in GaAs [577].

The double donor is electronically similar to a helium atom. Due to the repulsive Coulomb interaction of the two electrons on the neutral double donor, the (single) ionization energy E_1 (also often labeled $E(0, 1)$ or $E(0, +)$) of D^0 is smaller than that of D^+ (E_2, also labeled $E(1, 2)$ or $E(+, ++)$). For He and He^+ the ratio of ionization energies is 0.45; for chalcogenides in Si and Ge similar ratios have mostly been found (Table 7.8).

The carrier statistics and the degeneracy factors for a double donor have been discussed in [510, 578]. Typically, the degeneracy factor for the ionization of the double donor $D^0 \rightarrow D^+$ is $\hat{g}_D = g_2/g_1 = 1/2$ and for the ionization $D^+ \rightarrow D^{++}$ is $\hat{g}_D = g_1/g_0 = 2/1 = 2$.

For the probabilities to find a neutral, single and double ionized donor we find following the treatment in [578]

Table 7.8 Binding energies (to conduction band) of double donor chalcogenide impurities in Si and Ge

Host	State	S	Se	Te
Si	D^0	318	307	199
	D^+	612	589	411
Ge	D^0	280	268	93
	D^+	590	512	330

All energies in meV, data from [573, 574]

$$d^0 = \frac{N_D^0}{N_D} = \frac{\exp\frac{2E_F}{kT}}{\exp\frac{E_1+E_2}{kT} + \exp\frac{2E_F}{kT} + 2\exp\frac{E_1+E_F}{kT}} \tag{7.65a}$$

$$d^+ = \frac{N_D^+}{N_D} = \frac{\exp\frac{E_1+E_2}{kT}}{\exp\frac{E_1+E_2}{kT} + \exp\frac{2E_F}{kT} + 2\exp\frac{E_1+E_F}{kT}} \tag{7.65b}$$

$$d^{++} = \frac{N_D^{++}}{N_D} = \frac{2\exp\frac{E_1+E_F}{kT}}{\exp\frac{E_1+E_2}{kT} + \exp\frac{2E_F}{kT} + 2\exp\frac{E_1+E_F}{kT}} \tag{7.65c}$$

The probabilities are depicted in Fig. 7.31a. The maximum of d^+ is at the energy $(E_1 + E_2)/2$. Its value is

$$d^+\left(\frac{E_1+E_2}{2}\right) = \frac{1}{1+\exp\left(-\frac{E_1-E_2}{2kT}\right)} \tag{7.66}$$

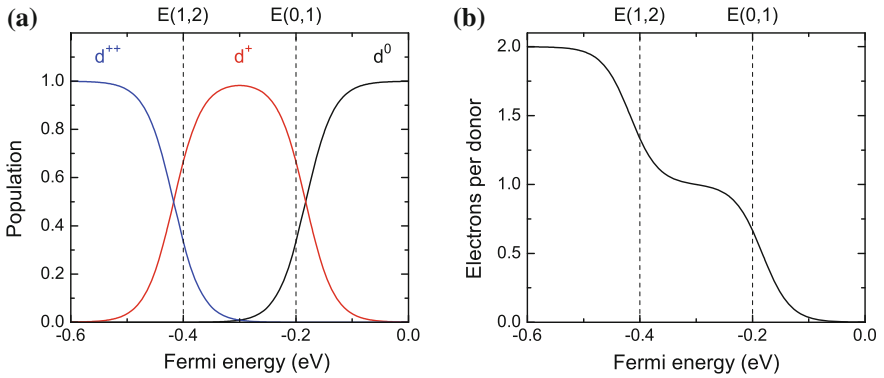


Fig. 7.31 (**a**) Population of states of a double donor (neutral: *black*, single ionized: *red*, double ionized: *blue*) according to (7.65a–c) as a function of the Fermi level. The ionization energies have been chosen as $E_1 = -0.2\,\text{eV}$ and $E_2 = -0.4\,\text{eV}$ and are indicated by *dashed lines* ($kT = 25\,\text{meV}$); these energies are similar to Si:Te (cmp. Table 7.8). The conduction band edge is taken as zero energy. (**b**) depicts the according number of electrons $\tilde{n}$ ionized from the donor.

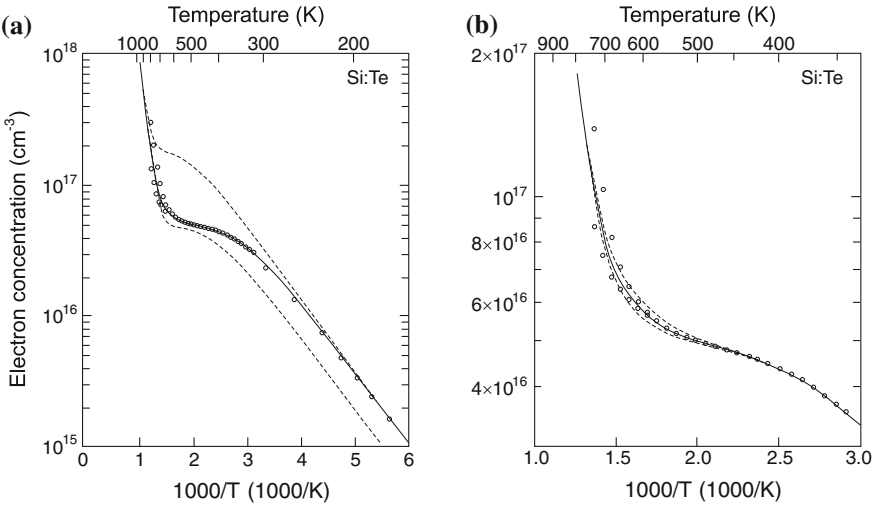


Fig. 7.32 Temperature dependent electron concentration (from Hall data) for Si:Te. (**a**) Experimental data and fit with double donor model using $N_{\mathrm{Te}} = 5 \times 10^{16}\,\mathrm{cm}^{-3}$, $E_1 = 200\,\mathrm{meV}$ and $E_2 = 440\,\mathrm{meV}$ (*solid line*). Single donor models would fail ($N_{\mathrm{Te}} = 5 \times 10^{16}\,\mathrm{cm}^{-3}$ and $N_{\mathrm{Te}} = 2 \times 10^{17}\,\mathrm{cm}^{-3}$, *dashed lines*). (**b**) Second ionization step in more detail with fits using different values for E_2; the *solid line* is for $E_2 = 440\,\mathrm{meV}$, the other *dashed lines* for $E_2 = 420$ and $460\,\mathrm{meV}$. Adapted from [579]

and reaches a value close to one for $(E_1 - E_2)/kT \gg 1$. In Fig. 7.31b the number of electrons per donor $\tilde{n} = (N_{\mathrm{D}}^{+} + 2\,N_{\mathrm{D}}^{++})/N_{\mathrm{D}}$ is shown as a function of the Fermi level; at $(E_1 + E_2)/2$, exactly $\tilde{n} = 1$. In Fig. 7.32 the temperature dependent electron concentration in Si:Te is depicted. Up to 570 K the single ionization is visible (other shallow impurities present in the sample in lower concentrations $< 10^{14}\,\mathrm{cm}^{-3}$ play no role). From the fit $E_1 = 200 \pm 2.7\,\mathrm{meV}$ is determined [579]. Single donor models would fail. The second ionization step is somewhat masked by the onset of intrinsic conduction. According to (7.15), the slope of n_i is $E_{\mathrm{g}}/2 \approx 500\,\mathrm{meV}$ which is similar to $E_2 \approx 440\,\mathrm{meV}$.

7.7.3 Double Acceptors

In analogy to double donor defects, double acceptors can introduce up to two holes into the valence band. A typical example is Zn in silicon [580], exhibiting its 'normal' acceptor level ($\mathrm{Zn}^0/\mathrm{Zn}^-$) at $E_{\mathrm{V}} + 0.31\,\mathrm{eV}$. In moderately n-doped silicon another level ($\mathrm{Zn}^-/\mathrm{Zn}^{2-}$) is observed at $E_{\mathrm{C}} - 0.55\,\mathrm{eV}$, when the n-doping is sufficient to partially compensate the Zn and supply one electron for each Zn atom but not two ($2N_{\mathrm{Zn}} > N_{\mathrm{D}} > N_{\mathrm{Zn}}$). A similar situation has been observed for Zn in germanium, exhibiting the levels $E_{\mathrm{V}} + 0.03\,\mathrm{eV}$ and $E_{\mathrm{V}} + 0.09\,\mathrm{eV}$ [581]. In Fig. 7.33 three

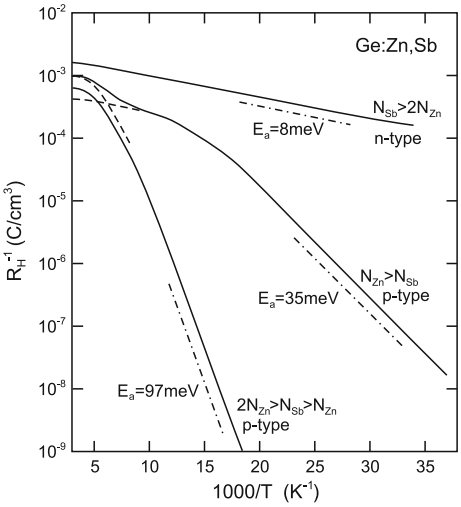


Fig. 7.33 Inverse (absolute) Hall coefficient (cmp. Sect. 13.2.1) R_H^{-1}, i.e. charge concentration, for three Ge:Zn samples with different degree of compensation with Sb donors as labeled. The *dash-dotted lines* indicate typical slopes. The *dashed lines* sketch the $Zn^0 \rightarrow Zn^-$ and the $Zn^- \rightarrow Zn^{--}$ processes. Adapted from [581]

different Ge:Zn samples are compared. If the additional Sb donor concentration ($N_D \approx 3.4 \times 10^{16}\,cm^{-3}$) is larger than $2N_{Zn}$ ($N_{Zn} \approx 1.2 \times 10^{16}\,cm^{-3}$), the sample is n-type (upper curve). The slope is similar to the Ge:Sb donor binding energy (Table 7.2). If compensation with donors is weak ($N_{Zn} > N_D$, middle curve) first the shallow donor level with 0.03 eV activation energy is activated and subsequently the deeper one with 0.09 eV activation energy, creating p-conduction with a saturated hole density of $p \approx 2N_A - N_D > N_{Zn}$ (negative Hall coefficient). The two individual activation processes are sketched as dashed lines in Fig. 7.33. If the Sb concentration is larger than N_{Zn} but smaller than $2N_{Zn}$, the shallow acceptor level is filled with electrons, leaving still the only partially filled deeper acceptor level available for ionization (lower curve). In this case the sample is still p-type, but the saturation hole density is $p \approx 2N_A - N_D < N_{Zn}$. The degeneracy factors for Zn in Si and Ge have been discussed in [522].

7.7.4 Jahn–Teller Effect

The lattice relaxation can reduce the symmetry of the defect. Many defects, such as a vacancy, a tetrahedral interstitial or an impurity, occupy initially tetrahedral sites in the zincblende structure. The lattice relaxation reduces the symmetry, e.g. to tetragonal or trigonal, and therefore causes initially degenerate levels to split. Such splitting is called the static Jahn–Teller effect [561, 582]. The energy change in terms of the atomic displacement Q can be denoted (using perturbation theory for the simplest, nondegenerate case) as $-I\,Q^2$ ($I > 0$). Including the elastic contribution

with a force constant C, the energy of a configuration Q is

$$E = -I\,Q + \frac{1}{2}\,C\,Q^2. \tag{7.67}$$

The stable configuration $Q_{\min}$, for which the energy is minimal ($E_{\min}$), is therefore given by

$$Q_{\min} = \frac{I}{C} \tag{7.68a}$$

$$E_{\min} = -\frac{I^2}{2C}. \tag{7.68b}$$

Several equivalent lattice relaxations may exist, e.g. a 3-fold minimum for remaining C_{3v} symmetry. The energy barrier between them has a finite height. Therefore, e.g. at sufficient temperature, the defect can switch between different configurations and eventually again becomes isotropic (dynamic Jahn–Teller effect). The experimental observation depends on the relation between the characteristic time of the experiment and the reorientation time constant of the defect.

7.7.5 *Negative-U Center*

We explain the principle of a so-called negative-U center [583] for the Si vacancy [584] (cf. Fig. 4.2). It was first proposed by Anderson to explain the properties of amorphous chalcogenide glasses [585]. Many defects in semiconductors exhibit negative-U behavior, e.g. also the boron interstitial in Si [584, 586]. Coulomb energy and the Jahn–Teller effect compete for the position of the occupancy level for different charge states. U refers to the additional energy upon charging of the defect with an additional electron. The Coulomb repulsion of electrons leads to an *increase* of the energy, i.e. positive U, which has been calculated to be 0.25 eV for the Si vacancy [587] for all charge states. The occupation level (cf. Sect. 4.2.2) $E_0(1, 2)$ (the index 0 indicates effects only due to many-electron Coulomb interaction), separating the domination of V^{++} and V^{+} (Fig. 7.34) is 0.32 eV above the valance-band edge. Therefore, the occupation level $E_0(0, 1)$ is expected to lie at about 0.57 eV about E_V.

The Jahn–Teller effect may lead to a splitting of the otherwise 4-fold degenerate states of the vacancy. A detailed experimental study using hyperfine interactions can be found in [588]. The schematic level diagram for the Jahn–Teller splitting is shown in Fig. 7.35. The V^{++} state (A_1 is always populated with two electrons) is resonant with the valence band. The T_2 state lies in the band gap. When the Jahn–Teller effect (now on the T_2 state) is included, the energies of the different charge states depend on the configuration coordinate (a mostly tetragonal distortion in the case of the Si vacancy).

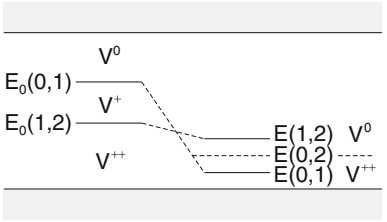


Fig. 7.34 Charge states of the vacancy in silicon. *Left* level scheme without lattice relaxation, *right* level scheme including the Jahn–Teller effect. For a Fermi level below (above) $E(0, 2)$ the charge state V^{++} (V^0) is dominant

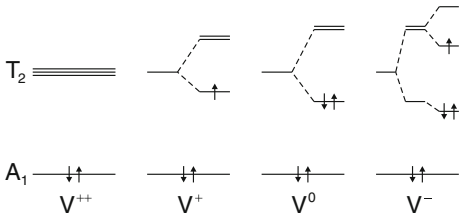


Fig. 7.35 Jahn–Teller splitting for different charge states of the vacancy. A_1 and T_2 refer to irreducible representations of the T_d point symmetry group. A_1 is nondegenerate and therefore does not exhibit a Jahn–Teller effect. T_2 is triply degenerate. The arrows represent electrons and their spin orientation

$$E_{V^0} = E(0, Q) = E(0, Q = 0) - 2IQ + \frac{1}{2}CQ^2 \tag{7.69a}$$

$$E_{V^+} = E(1, Q) = E(1, Q = 0) - IQ + \frac{1}{2}CQ^2 \tag{7.69b}$$

$$E_{V^{++}} = E(2, Q) = E(2, Q = 0) + \frac{1}{2}CQ^2. \tag{7.69c}$$

For the $n = 2$ state the T_2 gap state is empty and thus no degeneracy and Jahn–Teller term arises. For $n = 1$ there is a linear Jahn–Teller term. The occupation with two electrons (V^0) causes an approximately twice as large Jahn–Teller splitting for the $n = 0$ state. The force constant is assumed to be independent of the charge state. The energies for the minimum configurations $Q^n_{\min}$ are therefore

$$E(0, Q^0_{\min}) = E(0, Q = 0) - 4\,\frac{I^2}{2C} \tag{7.70a}$$

$$E(1, Q^1_{\min}) = E(1, Q = 0) - \frac{I^2}{2C} \tag{7.70b}$$

$$E(2, Q^2_{\min}) = E(2, Q = 0). \tag{7.70c}$$

The Jahn–Teller energy $E_{\mathrm{JT}} = I^2/2C$ lowers the position of the occupancy levels E_0 calculated with Coulomb terms only. The occupancy levels *including* the Jahn–Teller contribution are therefore given as

$$E(1,2) = E_0(1,2) - E_{\mathrm{JT}} \tag{7.71a}$$

$$E(0,1) = E_0(0,1) - 3\,E_{\mathrm{JT}}. \tag{7.71b}$$

For the vacancy in silicon the Jahn–Teller energy E_{JT} is about 0.19 eV. Thus the $E(1,2)$ level is lowered from 0.32 to 0.13 eV. The E(0,1) occupancy level, however, is reduced from 0.57 to 0.05 eV [584, 589] (see Fig. 7.34). The occupancy level $E(0,2)$ is in the middle between $E(0,1)$ and $E(1,2)$ ($E(0,2) = (E(0,1) + E(1,2))/2$) and indicated in Fig. 7.36a. At this energy, $c(V^0) = c(V^{++})$ and the value of $c(V^+)$ is small ($\approx \exp\frac{E_1 - E_2}{2kT}$) since $E(0,1) < E(1,2)$ (cmp. (7.66)).

The relative concentrations of the three charge states are determined by (7.64) (degeneracy and entropy terms have been neglected)

$$\frac{c(V^{++})}{c(V^{+})} = \exp\left(\frac{E(1,2) - E_{\mathrm{F}}}{kT}\right) \tag{7.72a}$$

$$\frac{c(V^{+})}{c(V^{0})} = \exp\left(\frac{E(0,1) - E_{\mathrm{F}}}{kT}\right). \tag{7.72b}$$

They are depicted in Fig. 7.36a in a plot related to Fig. 7.31a. Therefore, V^{++} dominates if $E_{\mathrm{F}} < E(0,1)$ and V^0 dominates for $E_{\mathrm{F}} > E(1,2)$. In the intermediate range $E(0,1) < E_{\mathrm{F}} < E(1,2)$ we know from (7.72a, b) that V^+ is dominated by V^0 and V^{++}. However, at this point it is not clear whether V^{++} or V^0 dominates

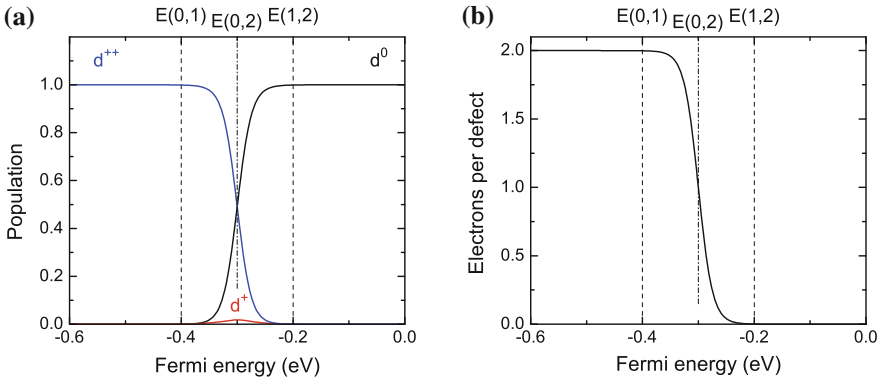


Fig. 7.36 (**a**) Population of states of a negative-U defect (neutral: *black*, single ionized: *red*, double ionized: *blue*) according to (7.65a–c) as a function of the Fermi level. The ionization energies have been chosen as $E_1 = -0.4$ eV and $E_2 = -0.2$ eV (cmp. Fig. 7.31) and are indicated by *dashed lines* ($kT = 25$ meV). The occupancy level $E(0,2)$ is indicated with a *dash-dotted line*. The conduction band edge is taken as zero energy. (**b**) depicts the according number of electrons ionized from the defect.

overall. The ratio of the concentrations of V^{++} and V^0 is given by

$$\frac{c(V^{++})}{c(V^0)} = \exp\left(\frac{E(1,2) + E(0,1) - 2E_F}{kT}\right) = e^2 \exp\left(\frac{E(0,2) - E_F}{kT}\right). \quad (7.73)$$

The occupancy level $E(0, 2)$ is thus again given as

$$E(0,2) = \frac{E(0,1) + E(1,2)}{2}, \quad (7.74)$$

and is shown in Fig. 7.34. V^{++} dominates if $E_F < E(0, 2)$ and V^0 dominates for $E_F > E(0, 2)$. V^+ is, for no position of the Fermi level, the dominating charge state of the Si vacancy. We note that for n-doped Si the V^- and V^{--} can also be populated. The population of the V^0 state with an extra electron introduces another Jahn–Teller splitting (Fig. 7.35) that has trigonal symmetry.

Generally, the Jahn–Teller effect can make the addition of an electron cause an effectively negative charging energy; in this case the center is termed a negative-U center. We note that the single vacancy in germanium is not a negative-U center due to smaller Jahn–Teller distortion and smaller electron-lattice coupling [590].

7.7.6 DX Center

The DX center is a deep level that was first investigated for n-doped (e.g. Si-doped) $Al_xGa_{1-x}As$. It dominates the transport properties of the alloy for $x > 0.22$. For smaller Al concentrations and GaAs the DX level lies in the conduction band. DX-type deep levels have also been found for other alloys and dopants, e.g. GaAsP:S.

It is experimentally found that the capture process of electrons into the DX center is thermally activated. The capture energy E_c depends on the AlAs mole fraction (Fig. 7.37). The (average) barrier for electron capture has a minimum of 0.21 eV for $x \approx 0.35$, near the crossover point between direct and indirect band gap

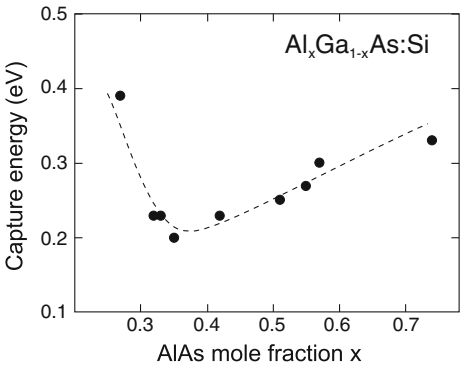


Fig. 7.37 Energy barrier for electron capture E_c at the Si-DX center in $Al_xGa_{1-x}As$ for various compositions. Experimental data from [591]

(cf. Fig. 6.19). For lower Al concentrations, the capture barrier increases to 0.4 eV for $x = 0.27$; for $x > 0.35$ the capture barrier increases to about 0.3 eV for x around 0.7 [591]. The barrier for thermally releasing carriers from the DX center has been determined to be about 0.43 eV, independent of the Al mole fraction [591].

Carriers can be removed from the DX center by optical absorption of photons with energy larger than about 1.2 eV. If carriers are removed by optical excitation at low temperatures the (re-)capture is so slow ($\sigma < 10^{-30}\,\mathrm{cm}^2$) that the carriers remain in the conduction band and cause persistent photoconductivity (PPC). The PPC is only reduced upon increasing the sample temperature. The concentration of the DX center is about the same as the net doping concentration.

The properties of the DX center are reviewed in [592, 593]. So far, no definite microscopic model of the DX center has been agreed on. Lang [594] proposed that the DX center involves a donor and an unknown defect (probably a vacancy). It probably involves large lattice relaxation as in the configuration coordinates model of Fig. 7.38 where the donor binding energy E_D^b with respect to the conduction-band minimum, the barrier for electron capture E_c, the barrier for electron emission E_e and the optical ionization energy E_o are labeled. The donor binding energy is measured with Hall effect (cf. Sect. 13.2.1) at temperatures sufficient to overcome the capture and emission barriers, the emission barrier is measured with deep level transient spectroscopy (DLTS). The capture barrier manifests itself in PPC experiments. We note that the DX center is related to the L-conduction band. For small Al mole fraction, the DX level is degenerate with the Γ-related conduction band (see Fig. 7.38b).

Theoretical models and experimental evidence hint at a vacancy-interstitial model for the Si-DX center [595]. The donor (Si) is displaced along the ⟨111⟩ direction from the Ga substitution site. The displacement is predicted to be 0.117 nm and the

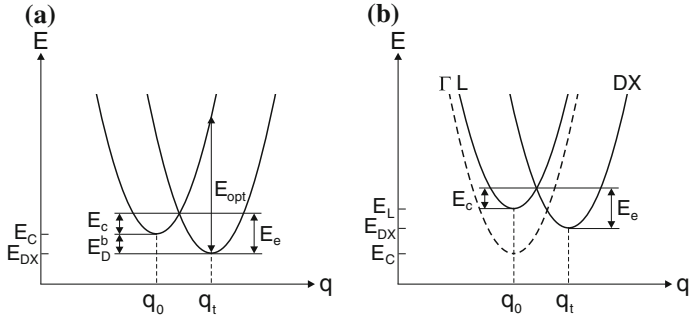


Fig. 7.38 (**a**) Schematic configuration coordinate diagram for the DX level with large lattice relaxation. q_0 is the configuration of the empty defect, q_t is the configuration of the filled defect. The donor binding energy E_D^b, the barrier for electron capture E_c, the barrier for electron emission E_e and the optical ionization energy E_o are labeled. E_C denotes the conduction-band edge. We note that in AlGaAs the DX level is associated with the L conduction band (see Fig. 6.19). (**b**) Schematic configuration coordinate diagram for the DX level in $Al_{0.14}Ga_{0.86}As$ with the DX level being degenerate with the (Γ-related) conduction band

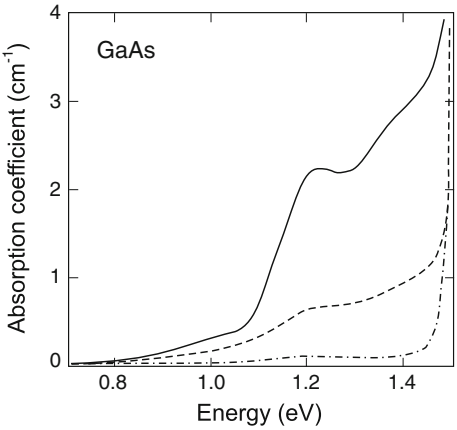


Fig. 7.39 Absorption spectrum of GaAs at low temperatures ($T = 10\,\mathrm{K}$) when cooled in the dark (*solid line*). The *dashed* (*dash-dotted*) *line* is the absorption after illuminating the sample for 1 min (10 min) with white light, leading to quenching of the EL2-related absorption. Adapted from [597]

distorted geometry can be viewed as a Ga vacancy and a Si interstitial. The charge state of the (filled) DX center is proposed to be a two-electron negative-U state.

7.7.7 EL2 Defect

The EL2 defect is a deep donor in GaAs. It is not related to impurities but occurs for intrinsic material, in particular grown under As-rich conditions. It has physical properties similar to the DX center. The bleaching of absorption due to EL2, i.e. the optical removal of electrons from the defect at low temperatures, is shown in Fig. 7.39. The microscopic model [596] describes the EL2 defect as an arsenic antisite defect, i.e. an arsenic atom on a Ga site, $\mathrm{As_{Ga}}$. In the charged state the arsenic atom is displaced from the lattice position and a complex of a Ga vacancy (symmetry T_{3d}) and an interstitial As (symmetry C_{3v}) with 0.14 nm displacement along $\langle 111 \rangle$ forms ($\mathrm{V_{Ga}}$-$\mathrm{As_i}$). The charged state is filled with two electrons.

7.7.8 Semi-insulating Semiconductors

Semiconductors with high resistivity (10^7–$10^9\,\Omega\,\mathrm{cm}$) are called semi-insulating ('s.i.' or 'si'). Semi-insulating substrates are needed for high-speed devices. The high resistivity should stem from a small free-carrier density at finite temperature and not from a small mobility due to poor crystal quality. For sufficiently wide band gap, the intrinsic carrier concentration is small and such pure material is semi-insulating, e.g. GaAs with $n_\mathrm{i} = 1.47 \times 10^6\,\mathrm{cm}^{-3}$ and $5.05 \times 10^8\,\Omega\,\mathrm{cm}$ [598]. Since shallow impurities are hard to avoid, another route is used technologically. Impurities that form deep levels are incorporated in the semiconductor in order to compensate free carriers. For

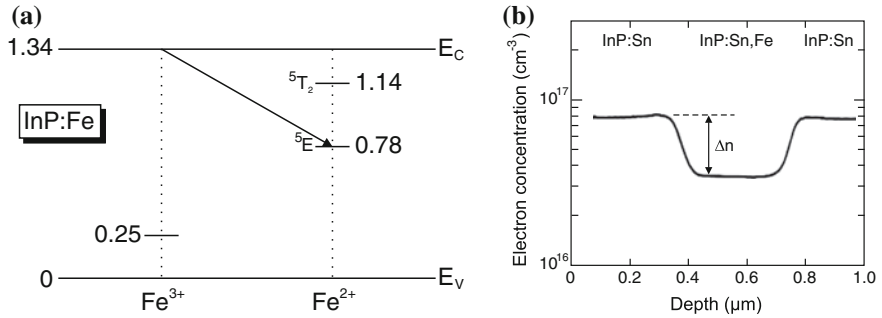


Fig. 7.40 (**a**) Schematic band diagram of InP with levels of Fe impurities in the 3+ and 2+ charge states at low temperature. All energies are given in eV. The *arrow* denotes capture of an electron (from the conduction band or a shallow donor) on the deep acceptor. Compare this figure also with Figs. 9.33 and 10.25. (**b**) Depth profile of electron concentration in an InP:Sn/InP:Sn,Fe/InP:Sn structure. The change $\Delta n \approx 4.5 \times 10^{16}\,\mathrm{cm}^{-3}$ of electron concentration is due to the compensation by Fe and corresponds to the chemical iron concentration determined by SIMS, $[\mathrm{Fe}] = 4.9 \times 10^{19}\,\mathrm{cm}^{-3}$. Part (**b**) adapted from [610]

example, a deep acceptor compensates all electrons if $N_A > N_D$. Since the acceptor is deep ($E_A^b \gg kT$), it does not release holes for reasonable temperatures. Examples of suitable impurities for compensation of electrons are Si:Au [599], GaAs:Cr [600] and InP:Fe [601]. A deep donor, e.g. InP:Cr [602], is necessary to compensate p-type conductivity.

Figure 7.40a shows the terms of Fe in InP [603, 604]. An overview of transition metals in III–V semiconductors can be found in [605]. The electron configuration of neutral Fe atoms is $3d^6 4s^2$ (cf. Table 16.2). The Fe is incorporated on the In site and thus has a Fe^{3+} state as a neutral acceptor (A^0). The Fe^{3+} state has the electron configuration $3d^5$. The arrow in Fig. 7.40a represents the capture of an electron from the conduction band or from a shallow donor. The charge state of the Fe becomes Fe^{2+} (charged acceptor, A^-) with the electron configuration $3d^6$. The cubic crystal field (T_d symmetry) splits this ^{5}D Fe state[7] into two terms [606] that exhibit further fine structure [604]. The large thermal activation energy of 0.64 eV found in the Hall effect on semi-insulating InP:Fe [601] corresponds to the energy separation of the ^{5}E level and the conduction band.

The maximum electron concentration that can be compensated in this way is limited by the solubility of Fe in InP [607], about $1 \times 10^{17}\,\mathrm{cm}^3$. Higher Fe incorporation leads to the formation of Fe (or FeP) precipitates and degrades the crystal quality. Only a fraction of the incorporated Fe may then be electrically active and contribute to the compensation. The maximum electrically active Fe concentration is found

[7]The notation is ^{2S+1}J (multiplicity), with S being the total spin and J being the total angular momentum.

to be $5–6\times10^{16}\,\mathrm{cm}^{-3}$ [608]. The compensation can be directly visualized via the depth profile of the electron concentration in a n-si-n structure (Fig. 7.40b). The poor thermal stability of Fe, i.e. high diffusion coefficient, has evoked proposals for more stable dopants such as InP:Ru [609].

7.7.9 Isoelectronic Impurities

Isoelectronic impurities, generally represent a deep level with a short range potential. The isoelectronic trap introduces a bound state for an electron or a hole. Once a carrier has been captured, the defect becomes charged. The other carrier type is then easily trapped, forming a bound exciton (Sect. 10.3.2). The theory of isoelectronic impurities is outlined in [611]. A detailed theoretical treatment of N in GaAs and GaP is given in [472].

In GaP:N, an electron is spatially localized on the N impurity. Most of the wave function is at the X-point. The nitrogen-bound electron level in GaP (A_1 symmetry) is close to the conduction band edge and within the band gap. Important for the energy position is the lattice relaxation, leading to an inward relaxation of the surrounding Ga atoms (Fig. 7.42). Due to the spatial localization of the wave function it is delocalized in k-space (Fig. 7.41a) and obtains a sizeable component at the Γ-point, facilitating zero-phonon absorption from the valence band. This effect is present only when the lattice relaxation around the impurity is considered; without relaxation the Γ-component is zero, with relaxation about 1 % [472]. The Γ-component of the wavefunction is larger for localization at an isoelectronic impurity than at a shallow donor such as sulfur [612]. This way a large oscillator strength for optical transitions occurs

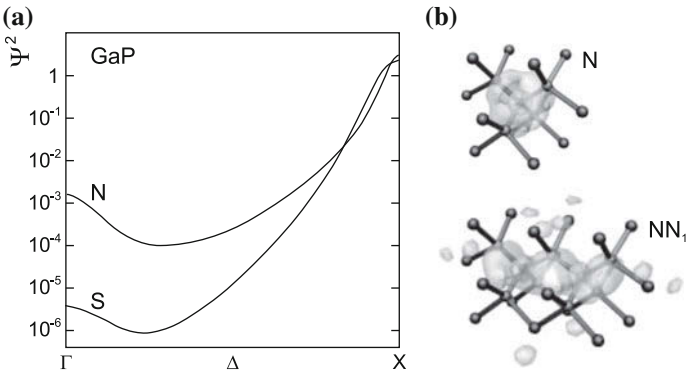


Fig. 7.41 (**a**) Model calculation of the wave-vector dependence of the probability density of an electron bound to a 10 meV deep isoelectronic trap (N) and to a 100 meV deep shallow donor (S) in GaP. Adapted from [612]. (**b**) Wavefunction (isosurface at 20 % of maximum) of isolated nitrogen (N) and neighboring N–N pair (NN_1) in GaP. Adapted from [472]

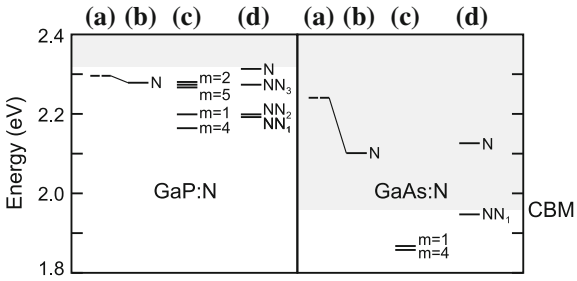


Fig. 7.42 Energy levels of nitrogen impurity states in GaP (*left*) and GaAs (*right*). The energy scale is relative to the bulk GaP valence band maximum, the conduction band minima (CBM) are thus shown relative to the vacuum level. The conduction band is shown in *grey*. For both materials, (a) denotes the isolated N impurity level calculated *without* lattice relaxation (*dashed line*), and (b) *with* lattice relaxation. (c) denotes the position of N–N pair levels, *m* denoting the neighbor. (d) shows selected experimental data. NN_1 denotes the direct neighbor NN-pair. The other NN_n follow the usual nomenclature as in [616]. Data taken from [472] (color figure online)

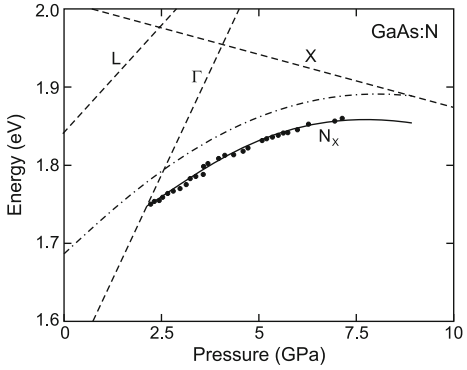


Fig. 7.43 Pressure dependence of the energy of excitons bound to isolated nitrogen impurities in GaAs (*circles*), measured from the top of the GaAs valence band. The *dashed lines* are the pressure dependent GaAs bulk band gaps (cmp. Fig. 6.42). The *solid* (*dash-dotted*) *line* is a theoretical model for the nitrogen-bound exciton (electron) level. Adapted from [615]

(Sects. 9.6.9 and 10.3.2). The wavefunction of an isolated single N impurity and a neighboring N–N pair (NN_1) in GaP are shown in Fig. 7.41b.

Isolated nitrogen impurities in (unstrained) GaAs introduce states only within the conduction band (Fig. 7.42). The reason is that the GaAs conduction band edge is further from the vacuum level than that of GaP (see Fig. 12.21). Only the NN_1 and NN_4 pair levels are theoretically expected to be within the GaAs band gap. The index denotes the *n*-th neighbor position. The NN_1 level has been experimentally observed [613, 614]. The isolated nitrogen impurity level is forced into the GaAs band gap upon hydrostatic pressure [614, 615] (Fig. 7.43). Further levels deeper within the band gap are due to clusters containing more than two nitrogen atoms.

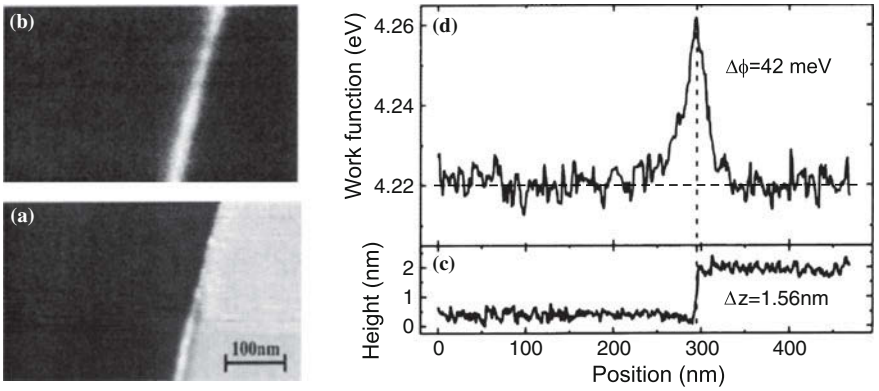


Fig. 7.44 Image of (**a**) topography ($\Delta z = 2.8\,\text{nm}$) and (**b**) work function ($\Delta\phi = 4.21$–$4.26\,\text{eV}$) of a surface step along [111] on a n-GaP(110) surface cleaved in UHV. Adapted from [618]. (**c**) and (**d**) show the corresponding linescans. Adapted from [618]

7.7.10 Surface States

The investigation of (semiconductor) surfaces is a large field with sophisticated methods that allow real-space imaging with atomic resolution by scanning probe microscopy and highly depth resolved electronic studies. The surface represents first of all a break in the periodic crystal potential and thus a defect of the bulk crystal. The unsatisfied bonds partly rearrange, e.g. by building dimers, forming a surface reconstruction or remain as dangling bonds. The surface exhibits a surface density of states. Such states can lie in the band gap and capture electrons, leading to recombination and a depletion layer. For a brief introduction on semiconductor surface physics see Chap. 11; for more details we refer to [617].

As an example of the formation of electronic states at surface defects we show in Fig. 7.44 the comparison of topography and work function (measured by Kelvin probe force microscopy [618]) at a surface step on a GaP(110) surface that has been prepared by cleaving in-situ in ultrahigh vacuum (UHV). The depletion-type band bending of the surface is about 0.4 eV. The further increase of the position of the vacuum level at the step edge shows the presence of trap states in the band gap causing the conduction band to bend upwards (cf. Sect. 21.2.1). Modeling of the effect shows that the charge density at the surface is $6 \times 10^{11}\,\text{cm}^{-2}$ and at the step edge $1.2 \times 10^{6}\,\text{cm}^{-1}$.

7.8 Hydrogen in Semiconductors

The role of hydrogen in semiconductors was first recognized in studies of ZnO [619]. It is now clear that hydrogen plays an important role in the passivation of defects. As a ‘small’ atom, it can attach easily to dangling bonds and form an electron-pair

bond. Thus, surfaces, grain boundaries, dislocations and shallow (donor and acceptor) and deep impurity levels become passivated. A good overview and many details of the physics and technological use of hydrogen in semiconductors can be found in [620, 621]. The hydrogen must be typically introduced as atomic species into semiconductors, e.g. from a plasma in the vicinity of the surface or by ion irradiation.

With regard to silicon it is important to note that the Si–H bond is stronger than the Si–Si bond. Thus a silicon surface under atomic hydrogen exhibits Si–H termination rather than Si–Si dimers [622]. Due to the stronger bond, the hydrogenation leads to an increase of the silicon band gap, which can be used for surface passivation [623], leading to reduced reverse diode current.

The hydrogen concentration in amorphous Si (a-Si) can be as high as 50 % [624]. Electronic grade a-Si contains typically 10–30 atomic % hydrogen and is thus rather a silicon–hydrogen alloy.

Hydrogen in crystalline silicon occupies the bond-center interstitial position (see Fig. 3.18b) as shown in Fig. 7.45a. The complexes formed by hydrogen with shallow acceptors and donors have been studied in detail. It is now generally accepted that for acceptors (e.g. boron) in silicon the hydrogen is located close to the bond-center position of the Si–B pair (BM, bond minimum) as sketched in Fig. 7.46a. The boron atom forms an electron-pair bond with three silicon atoms of the tetrahedra, the fourth silicon bonds to the hydrogen atom. The complex therefore no longer acts as an acceptor. The silicon atoms and the acceptor relax their positions. The adiabatic potential energy surface of hydrogen in Si:B is shown in Fig. 7.45b. The hydrogen can sit on four equivalent sites (BM) along the ⟨111⟩ directions of the initial B–Si_4 tetrahedron. This reduces the symmetry, e.g. of H–B vibrations [626].

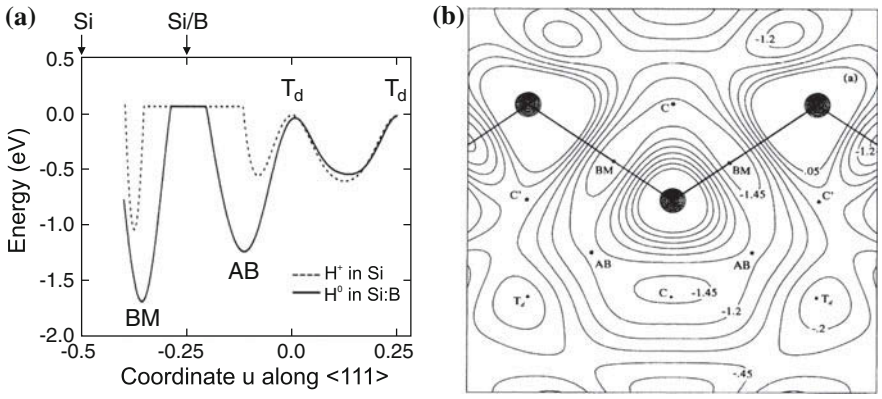


Fig. 7.45 (**a**) Energy for positions u of the hydrogen atom along the ⟨111⟩ direction for H^+ in pure Si (Si atom at $u = -0.25$) and neutral hydrogen (B atom at $u = -0.25$). u is measured in units of $\sqrt{3}a_0$. For all positions of the hydrogen atom the positions of the other atoms have been relaxed in the calculation. Data from [625]. (**b**) Adiabatic potential energy in the (110) plane for hydrogen in Si:B. 'BM' denotes the bond minimum site (high valence electron density), C and C' are equivalent for pure Si. Reprinted with permission from [625], ©1989 APS

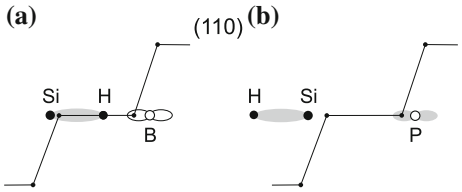


Fig. 7.46 Schematic model for hydrogen in silicon forming a complex with (**a**) a shallow acceptor (boron, empty orbital) and (**b**) a shallow donor (phosphorus, double-filled orbital)

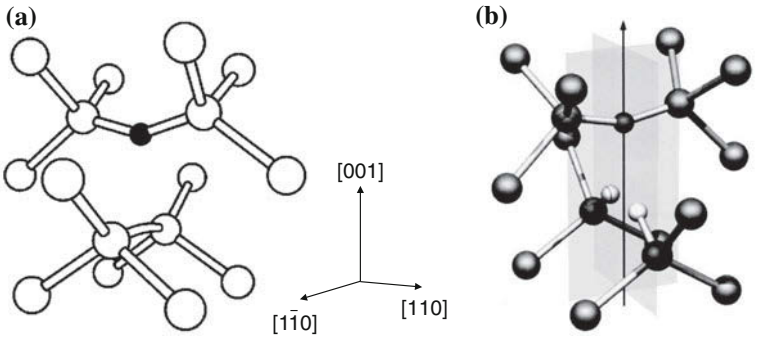


Fig. 7.47 (**a**) Structure of the V–O complex (A center) in silicon. The *black sphere* represents the oxygen atom. Reprinted with permission from [628], ©2004 APS. (**b**) Calculated ground-state structure for the V–O–H_2 center in silicon. Oxygen is over the C_2 axis, and the two *white* spheres represent hydrogen. Reprinted with permission from [629], ©2000 APS

The energetic barrier for the hydrogen orientation has been determined to be 0.2 eV theoretically [625] for a hydrogen motion along the path BM–C–BM in Fig. 7.45b. Stress (along [100] and [112]) reduces the symmetry and leads to splitting of the local vibrational modes, now showing axial symmetry [627]. However, this preferential orientation disappears with an activation energy of 0.19 eV, close to the theoretical value.

Hydrogen has experimentally been found to also passivate shallow donors. The microscopic configuration is sketched in Fig. 7.46b. The hydrogen atom sits on the Si–AB (antibonding) position and forms an electron-pair bond with the silicon atom. The donor, e.g. phosphorus, is left with a double-filled p-orbital (lone pair) whose level is in the valence band and thus no longer contributes to conductivity. Molecular hydrogen can passivate the so-called A center in Si, an oxygen–vacancy complex [629]. The atomistic configuration of the V–O–H_2 complex is shown in Fig. 7.47. The deep double donor S in Si with a level at 0.3 eV below the conduction-band edge can also be passivated by two hydrogen atoms [630].

Chapter 8
Transport

Abstract The physics of transport in semiconductors is treated foremost for charge transport. Band transport and scattering, mobility, low field and high field effects as well as polarons and hopping transport are covered. A short section mentions ionic transport before heat conduction and coupled heat and charge transport including thermopower and Peltier effect are discussed.

8.1 Introduction

Charge and heat energy can be transported through the semiconductor in the presence of appropriate (generalized) forces. Such a force can be an electric field or a temperature gradient. Both transport phenomena are coupled since electrons transport energy and charge simultaneously through the crystal. First, we will treat the charge transport as a consequence of a gradient in the Fermi level, then the heat transport upon a temperature gradient and finally the coupled system, i.e. the Peltier and Seebeck effects. Detailed treatments of carrier transport can be found in [631, 632].

Practically all important semiconductor devices are based on the transport of charge, such as diode, transistor, photodetector, solar cell and laser.

Carriers move in the semiconductor driven by a gradient in the Fermi energy. We distinguish

- drift, as a consequence of an electric field $\mathbf{E}$,
- diffusion, as a consequence of a concentration gradient ∇n or ∇p.

In inhomogeneous semiconductors for which the position of the band edges is a function of position, another force occurs. This will not be treated here, since later (cf. Chap. 12) it will be included as an additional, internal electric field.

In Sects. 8.2–8.5 we treat band conductivity, i.e. the transport of charge carriers in extended states, the conduction and valence bands characterized by an effective mass. Conductivity is then determined by the carrier concentration (free electrons and holes) and scattering mechanisms (mobility). In disordered semiconductors such as amorphous material, the charge transport due to *hopping* between localized states close to the Fermi level dominates the conductivity which is discussed in Sect. 8.7.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_8

Many semiconductor properties, such as the carrier concentration and the band gap, depend on the temperature. Thus, device properties will also depend on temperature. During operation of a device typically heat is generated, e.g. by Joule heating due to finite resistivity. This heat leads to an increase of the device temperature that subsequently alters the device performance, mostly for the worse. Ultimately, the device can be destroyed. Thus cooling of the device, in particular of the active area of the device, is essential. Mostly the thermal management of device heating limits the achievable performance (and lifetime) of the device. In high-power devices quite high energy densities can occur, e.g. the facet of a high-power semiconductor laser has to withstand an energy density beyond $10\,\mathrm{MW\,cm^{-2}}$.

8.2 Conductivity

Under the influence of an electric field the electrons accelerate according to (cf. (6.32))

$$\mathbf{F} = m^* \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \hbar \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = q\,\mathbf{E} = -e\,\mathbf{E}. \tag{8.1}$$

In the following, q denotes a general charge, while e is the (positive) elementary charge. We also consider an isotropic effective mass m^* at first. After the time δt the $\mathbf{k}$ vector of all conduction electrons (and the center of the Fermi sphere) has been shifted by $\delta\mathbf{k}$

$$\delta\mathbf{k} = -\frac{e\,\mathbf{E}}{\hbar}\,\delta t. \tag{8.2}$$

In the absence of scattering processes this goes on further (similar to an electron in vacuum). This regime is called *ballistic* transport. In a (periodic) band structure, the electron will perform a closed cycle as indicated in Fig. 8.1. Such motion is called a Bloch oscillation. However, in a bulk crystal the period T of such an oscillation $eET/\hbar = 2\pi/a_0$ is of the order of $10^{-10}\,\mathrm{s}$ for $E = 10^4\,\mathrm{V/cm}$. This time is much longer than a typical scattering time of $10^{-14}\,\mathrm{s}$. Thus, in bulk material the Bloch electron cannot reach the zone boundary. However, in artificial superlattices (cf. Chap. 12) with larger periodicity ($\approx$10 nm), high electric fields ($\approx 10^6\,\mathrm{V/cm}$) and high quality (reduced collision time) such motion is possible. We note that in the absence of scattering, electrons also perform a periodic oscillation in a magnetic field (cyclotron motion).

In a real semiconductor, at finite temperatures, impurities, phonons and defects (finally also the surface) will contribute to scattering. In the relaxation-time approximation it is assumed that the probability for a scattering event, similar to friction, is proportional to the (average) carrier velocity. The average relaxation time τ is introduced via an additional term $\dot{\mathbf{v}} = -\mathbf{v}/\tau$ that sums up all scattering events.[1]

[1] Going beyond the relaxation time approximation is discussed in Sect. I.

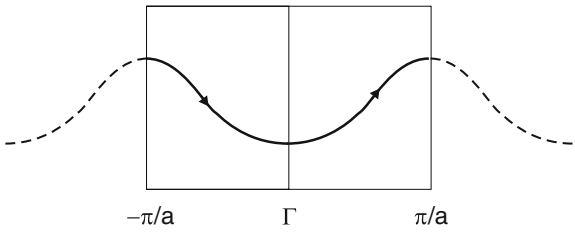


Fig. 8.1 Schematic representation of a Bloch oscillation

Thus, the maximum velocity that can be reached in a static electric field is given by (steady-state velocity)

$$\mathbf{v} = -\frac{e\,\mathbf{E}\,\tau}{m^*}. \tag{8.3}$$

The current density per unit area is then linear in the field, i.e. fulfills Ohm's law

$$\mathbf{j} = n\,q\,\mathbf{v} = \frac{n\,e^2\,\mathbf{E}\,\tau}{m^*} = \sigma\,\mathbf{E}. \tag{8.4}$$

The conductivity σ in the relaxation-time approximation is given by

$$\sigma = \frac{1}{\rho} = \frac{n\,e^2\,\tau}{m^*}. \tag{8.5}$$

In the case of a cylindrically symmetric mass such as for electrons in silicon or germanium, for the effective mass in (8.5) the effective conductivity mass must be used,

$$\frac{1}{m^*_\sigma} = \frac{1}{3}\left(\frac{2}{m_\mathrm{t}} + \frac{1}{m_\mathrm{l}}\right). \tag{8.6}$$

The specific resistivity is the inverse of the conductivity. Metals have a high conductivity (see Table 8.1), e.g. for Cu at room temperature $\sigma = 5.8 \times 10^5\,\Omega^{-1}\,\mathrm{cm}^{-1}$. At low temperatures (4 K) the conductivity is even a factor of 10^5 higher. The mean free path $d = \tau\,v_\mathrm{F}$ is

$$d = \frac{\sigma\,m^* v_\mathrm{F}}{n\,e^2}, \tag{8.7}$$

v_F being the Fermi velocity ($E_\mathrm{F} = m^*\,v_\mathrm{F}^2/2$). For copper, $d = 3\,\mathrm{mm}$ at low temperature (and thus susceptible to the sample geometry) while at room temperature the mean free path is only about 40 nm. However, this becomes an issue when the metal line width and height of interconnects in integrated circuits approaches this length scale [633] (see Sect. 24.5.5).

In semiconductors, the carrier concentration depends strongly on the temperature. At zero temperature the conductivity is zero. Also, the scattering processes and thus the relaxation time constant exhibit a temperature dependence. The conductivity spans a large range from insulating to almost metallic conduction (see Table 8.1).

Table 8.1 Conductivity at room temperature for various metals, semiconductors, insulators and liquids

Material	σ ($\Omega^{-1}\,\mathrm{cm}^{-1}$)
Ag	6.25×10^5
Al	3.6×10^5
Au	4.35×10^5
Cu	5.62×10^5
Fe	1.1×10^5
Pt	1.02×10^5
Ge pure ($N_\mathrm{D} \sim 10^{13}\,\mathrm{cm}^{-3}$)	10^{-2}
Ge ($N_\mathrm{D} \sim 10^{15}\,\mathrm{cm}^{-3}$)	1
Ge ($N_\mathrm{D} \sim 10^{17}\,\mathrm{cm}^{-3}$)	2×10^1
Ge ($N_\mathrm{D} \sim 10^{18}\,\mathrm{cm}^{-3}$)	2×10^2
Si pure	4.5×10^{-6}
Si:As ($N_\mathrm{D} \sim 3 \times 10^{19}\,\mathrm{cm}^{-3}$)	4×10^2
Si:B ($N_\mathrm{A} \sim 1.5 \times 10^{19}\,\mathrm{cm}^{-3}$)	1.2×10^2
GaAs pure	1.4×10^{-7}
ZnO:Al (highly doped)	$\approx 1 \times 10^4$
Pentacene	10^{-8}–10^{-4}
SiO_2	$\approx 10^{-15}$
Al_2O_3	$\approx 10^{-16}$
H_2O pure	4×10^{-8}
Hexane	$\approx 10^{-18}$

8.3 Low-Field Transport

First we consider only *small* electric fields. The real meaning of this will only become clear in Sect. 8.4 on high-field transport. In the low-field regime the velocity is proportional to the electric field.

8.3.1 Mobility

The mobility is defined (scalar terms) as

$$\mu = \frac{v}{E}. \tag{8.8}$$

By definition, it is a negative number for electrons and positive for holes. However, the numerical value is usually given as a positive number for both carrier types. In an intrinsic semiconductor the mobility is determined by scattering with phonons.

Table 8.2 Mobilities of electrons and holes at room temperature for various semiconductors

Material	$-\mu_n$(cm^2/Vs)	μ_p(cm^2/Vs)
Si	1300	500
Ge	4500	3500
GaAs	8800	400
GaN	300	180
InSb	77 000	750
InAs	33 000	460
InP	4600	150
ZnO	230	8

Further scattering is introduced by impurities, defects or alloy disorder. The conductivity is (8.4)

$$\sigma = q\, n\, \mu \tag{8.9}$$

for each carrier type. Using (8.5) the mobility in the relaxation time approximation is

$$\mu = \frac{q\,\tau}{m^*}. \tag{8.10}$$

In the presence of both electrons and holes,

$$\sigma = \sigma_e + \sigma_h = -e\, n\, \mu_n + e\, p\, \mu_p, \tag{8.11}$$

where μ_n and μ_p are the mobilities for electrons and holes, respectively. These are given by $\mu_n = -e\,\tau_n/m_e^*$ and $\mu_p = e\,\tau_p/m_p^*$.

As the unit for mobility, usually cm^2/Vs is used. While Cu at room temperature has a mobility of 35 cm^2/Vs, semiconductors can have much higher values. In two-dimensional electron gases (cf. Chap. 12), the mobility can reach several 10^7 cm^2/Vs at low temperature (Fig. 12.36). In bulk semiconductors with small band gap, a high electron mobility is caused by its small effective mass. Some typical values are given in Table 8.2.

8.3.2 Microscopic Scattering Processes

The relaxation time constant summarizes all scattering mechanisms. If the relaxation times τ_i of various processes are independent, the Matthiesen rule can be used to obtain the mobility ($\mu_i = q\,\tau_i/m^*$)

$$\frac{1}{\mu} = \sum_i \frac{1}{\mu_i}. \tag{8.12}$$

The various scattering mechanisms have quite different temperature dependences such that the mobility is a rather complicated function of temperature. In [634] the mechanisms determining the low and high-field transport properties of (cubic) semiconductors are reviewed.

8.3.3 *Ionized Impurity Scattering*

Theoretically, this problem is treated similar to Rutherford scattering. A screened Coulomb potential is assumed, as the scattering potential

$$V(r) = -\frac{Z\,e}{4\pi\epsilon_0\epsilon_\mathrm{r}}\,\frac{1}{r}\,\exp\left(-\frac{r}{l_\mathrm{D}}\right), \tag{8.13}$$

where l_D is the screening length. The problem has been treated classically by Conwell and Weisskopf [635] and quantum mechanically by Brooks [636] and Herring. An expression for the mobility that encompasses the Conwell–Weisskopf and Brooks–Herring results is derived in [637]. Further details are given in [638, 639]. For the mobility it is found that

$$\mu_\mathrm{ion.imp.} = \frac{2^{7/2}\,(4\pi\epsilon_0\epsilon_\mathrm{r})^2}{\pi^{3/2}\,Z^2\,e^3\,\sqrt{m^*}}\,\frac{(kT)^{3/2}}{N_\mathrm{ion}}\,\frac{1}{\ln(1+b)-1/(1+1/b)}, \tag{8.14}$$

with $b = 4\,(k/l_\mathrm{D})^2 = 8\,m^*\,E\,(l_\mathrm{D}/\hbar)^2$. In the Thomas- Fermi screening model

$$l_\mathrm{D}^2 = 4\pi\,\frac{e^2}{\epsilon_0\epsilon_\mathrm{r}}\,N(E_\mathrm{F}) = \left(\frac{3}{\pi}\right)^{1/3}\,\frac{4\,m^*\,e^2}{\epsilon_0\epsilon_\mathrm{r}\,\hbar^2}\,n^{1/3}. \tag{8.15}$$

The formula (8.14) is valid only for $b \gg 1$, i.e. small carrier densities. A similar formula from [638] is

$$\mu_\mathrm{ion.imp.} = \frac{128\sqrt{2\pi}\,(\epsilon_0\epsilon_\mathrm{r})^2\,(kT)^{3/2}}{m^{*1/2}\,Z^2\,N_\mathrm{ion}\,e^3}\,\left[\ln\frac{24\,m^*\,\epsilon_0\epsilon_\mathrm{r}\,(kT)^2}{n\,e^2\,\hbar^2}\right]^{-1}. \tag{8.16}$$

For large ionized impurity (and carrier) density ($b \ll 1$), the mobility is given by [483]

$$\mu_\mathrm{ion.imp.} = \frac{4\,e}{3^{1/3}\,\pi^{2/3}\,h}\,n^{-2/3}, \tag{8.17}$$

the value of the pre-factor being about $3 \times 10^{14}\,(\mathrm{Vs})^{-1}$.

The scattering time depends like $\tau \propto (E/kT)^s$ on the kinetic energy; for moderate or weak scattering $s = 3/2$, for very strong scattering, $s = -1/2$ [632].

For typical substitutional impurities, the charge of the scattering center is $|Z| = 1$; in oxides, oxygen vacancies may have $Z = 2$. At high impurity densities, impurity

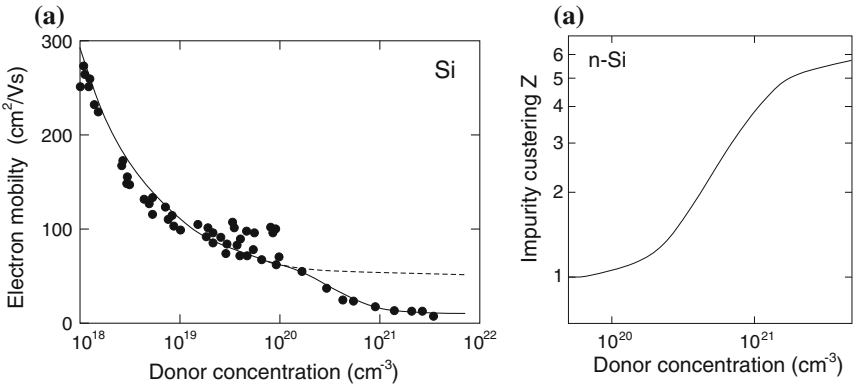


Fig. 8.2 (**a**) Electron mobility in highly doped silicon. Experimental data (*symbols*) from various sources and modeling with ionized impurity scattering with (*solid line*) and without (*dashed line*) considering impurity clustering. (**b**) Effective impurity cluster charge Z_D. Adapted from [640]

clusters may form with $|Z| > 1$; this will have a strong influence on the scattering rate since it proportional to Z^2. The decrease of mobility for $N_D > 10^{20}\,\mathrm{cm}^{-3}$ (Fig. 8.2a) is attributed to such effect which can be described with an effective impurity clustering charge Z_D (Fig. 8.2b) [640, 641].

8.3.4 Deformation Potential Scattering

Acoustic phonons with small wavevector, i.e. a wavelength large compared to the unit cell, can have TA or LA character. The TA phonons represent a shear wave (with zero divergence), the LA phonons are a compression wave (with zero rotation). The LA is a plane wave of displacement $\delta\mathbf{R}$ parallel to the k-vector $\mathbf{q}$,

$$\delta\mathbf{R} = \mathbf{A}\,\sin\left(\mathbf{q}\cdot\mathbf{R} - \omega t\right). \tag{8.18}$$

The strain tensor is given by

$$\epsilon_{ij} = \frac{1}{2}\left(\mathbf{q}_i\,\mathbf{A}_j + \mathbf{q}_j\,\mathbf{A}_i\right)\cos\left(\mathbf{q}\,\mathbf{R} - \omega t\right). \tag{8.19}$$

It has a diagonal form $\epsilon_{ij} = \mathbf{q}_i\mathbf{A}_j$ for $\mathbf{q}$ and $\omega \to 0$. Therefore, the LA phonon creates an oscillatory volume dilatation (and compression) with amplitude $\mathbf{q}\cdot\mathbf{A}$. This volume modulation affects the position of the band edges. For the conduction-band edge the energy change is related to the volume change by the hydrostatic deformation potential $E_{\mathrm{ac.def.}} = V\partial E_C/\partial V$. Since the modulation is small compared to the energy of the charge carriers, it is mostly an elastic scattering process. The Hamilton operator for the LA scattering is

$$\hat{H} = E_{\text{ac.def.}} (\mathbf{q} \cdot \mathbf{A}). \tag{8.20}$$

The size of the LA amplitude is given by the number of phonons in the mode that is given by the Bose–Einstein distribution, $N_{\text{ph}}(\hbar\omega) = \left[\exp\left(\frac{\hbar\omega}{kT}\right)\right]^{-1}$. The mobility due to acoustic deformation potential scattering is found to be

$$\mu_{\text{ac.def.}} = \frac{2\sqrt{2\pi}\, e\, \hbar^4\, c_{\text{l}}}{3\, m^{*5/2}\, E_{\text{ac.def.}}^2} (kT)^{-3/2}, \tag{8.21}$$

where $c_{\text{l}} = \rho c_{\text{s}}^{\text{LA}}$, ρ being the density and c_{s} being the sound velocity. The scattering time increases like $\tau \propto E^{-1/2}$ with the kinetic energy [632].

The acoustical deformation potential scattering is important at high temperatures. It is dominating in nonpolar semiconductors (Ge, Si) at high temperatures (typically at and above room temperature).

8.3.5 Piezoelectric Potential Scattering

In piezoelectric crystals (see Sect. 15.4), i.e. crystals that show an electric polarization upon strain, certain acoustic phonons lead to piezoelectric fields. In GaAs, with $\langle 111 \rangle$ being the piezoelectric directions, this is the case for shear waves. In strongly ionic crystals, e.g. II–VI semiconductors, the piezoelectric scattering can be stronger than the deformation potential scattering. The mobility due to piezoelectric potential scattering is

$$\mu_{\text{pz.el.}} = \frac{16\sqrt{2\pi}}{3} \frac{\hbar\, \epsilon_0 \epsilon_{\text{r}}}{m^{*3/2}\, e\, K^2} (kT)^{-1/2}, \tag{8.22}$$

with $K = \frac{e_{\text{p}}^2/c_{\text{l}}}{\epsilon_0 \epsilon_{\text{r}} + e_{\text{p}}^2/c_{\text{l}}}$, e_{p} being the piezoelectric coefficient.

8.3.6 Polar Optical Scattering

LO phonons are connected with an electric field antiparallel to the displacement (5.52). In the scattering mechanism the absorbed or emitted phonon energy $\hbar\omega_0$ is comparable to the thermal energy of the carriers. Therefore, the scattering is inelastic and the relaxation-time approximation does not work. The general transport theory is complicated. If the temperature is low compared to the Debye temperature, $T \ll \Theta_{\text{D}}$

$$\mu_{\text{pol.opt.}} = \frac{e}{2\, m^*\, \alpha\, \omega_0} \exp\left(\frac{\Theta_{\text{D}}}{T}\right), \tag{8.23}$$

where $\alpha = \frac{1}{137}\sqrt{\frac{m^* c^2}{2\,k\,\Theta_D}}\left(\frac{1}{\epsilon(\infty)} - \frac{1}{\epsilon(0)}\right)$ is the dimensionless polar constant.

8.3.7 Dislocation Scattering

Dislocations can contain charge centers and thus act as scattering centers [642]. This has been first demonstrated for n-Ge crystals that have been deformed [643, 644]. The deformation has introduced acceptor-type defects reducing the mobility in particular at low temperatures (similar to ionized impurity scattering). The mobility due to dislocation scattering in an n-type semiconductor is given by [645, 646]

$$\mu_{\text{disl.}} = \frac{30\sqrt{2\pi}\,\epsilon^2\,d^2\,(kT)^{3/2}}{N_{\text{disl}}\,e^3\,f^2\,L_D\,\sqrt{m^*}} \propto \frac{\sqrt{n}}{N_{\text{disl}}}\,T, \tag{8.24}$$

d being the average distance of acceptor centers along the dislocation line, f their occupation rate, N_{disl} the area density of dislocations and $L_D = (\epsilon kT/(e^2 n))^{1/2}$ the Debye screening length. The relation $\mu \propto \sqrt{n}/N_{\text{disl}}$ has been confirmed for various n-type GaN samples [647].

8.3.8 Grain Boundary Scattering

The lowering of mobility due to transport across grain boundaries is an important effect in polycrystalline materials, such as poly-silicon for solar cells or thin film transistors [648–651]. Grain boundaries contain electronic traps whose filling depends on the doping of the bulk of the grains. Charges will be trapped in the grain boundaries and a depletion layer will be created.[2] At low doping the grains are fully depleted and all free carriers are trapped in the grain boundaries. This means low conductivity, however, no electronic barrier to transport exists. At intermediate doping, traps are partially filled and the partial depletion of the grain leads to the creation of an electronic barrier ΔE_b (Fig. 8.3a) hindering transport since it must be overcome via thermionic emission. At high doping the traps are completely filled and the barrier vanishes again. Accordingly the mobility goes through a minimum as a function of the doping concentration (Fig. 8.3b) [648]. In [652] these data have been modeled with a 20 nm grain size, the value found in [648] from TEM analysis.

[2] The following arguments may only be followed once the concept of depletion layers and band bending is understood, see Sect. 21.2.1.

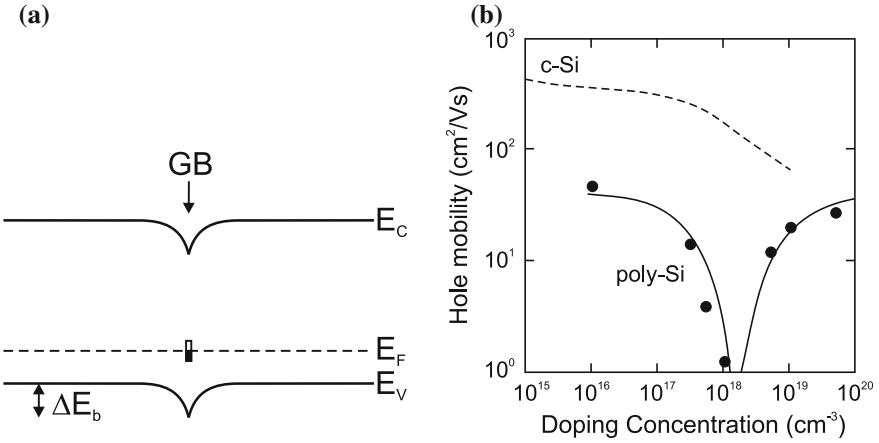


Fig. 8.3 (**a**) Electronic barrier (ΔE_b) for (hole) transport at a grain boundary (GB). (**b**) Average hole mobility in poly-silicon, experimental data (*symbols*) and theoretical model (*solid line*). The dependence for monocrystalline silicon is shown for comparison as *dashed line*. Adapted from [648]

The expression for the limitation of the mobility due to scattering at grain boundaries is given by [651, 653]

$$\mu_\mathrm{GB} = \frac{e\,L_\mathrm{G}}{\sqrt{8m^*\pi k}}\,T^{-1/2}\,\exp\left(-\frac{\Delta E_\mathrm{b}}{kT}\right), \tag{8.25}$$

where L_G is the grain size.

8.3.9 Temperature Dependence

The sum of all scattering processes leads to a fairly complicated temperature dependence of the mobility $\mu(T)$. In covalent semiconductors (Si, Ge) the most important processes are the ionized impurity scattering ($\mu \propto T^{3/2}$) at low temperatures and the deformation potential scattering ($\mu \propto T^{-3/2}$) at high temperatures (Fig. 8.4a). In polar crystals (e.g. GaAs) at high temperatures the polar optical scattering is dominant (Fig. 8.4b).

In Fig. 8.5 the electron mobility of bulk and thin-film ZnO is compared. Since ZnO is polar the mobility at room temperature is limited by polar optical phonon scattering. In the thin film, grain-boundary scattering (Sect. 8.3.8) additionally occurs and limits the mobility.

Since the carrier concentration increases with increasing temperature and the mobility decreases, the conductivity has a maximum, typically around 70 K

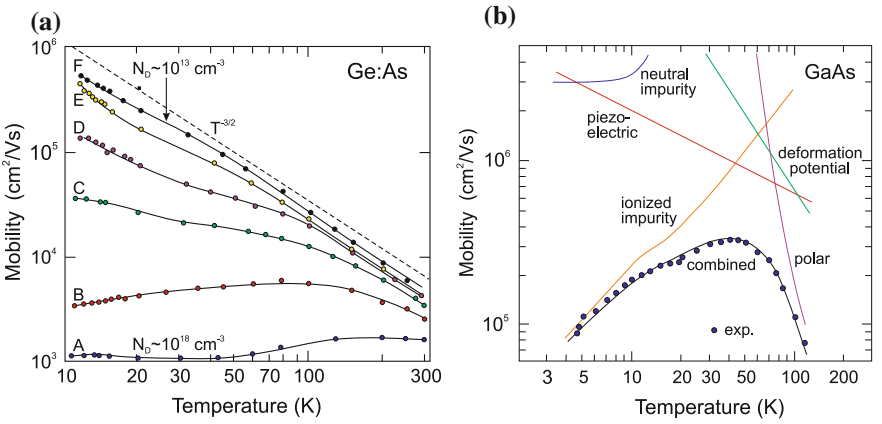


Fig. 8.4 (**a**) Temperature dependence of the electron mobility in n-doped Ge (for various doping levels from $N_D \approx 10^{18}$ for sample A to $10^{13}\,\mathrm{cm}^{-3}$ for sample F in steps of a factor of ten). *Dashed line* indicates $T^{-3/2}$ dependence of deformation potential scattering, *solid lines* are guides to the eye. Adapted from [515]. (**b**) $\mu_n(T)$ for n-type GaAs ($N_D \approx 5 \times 10^{13}\,\mathrm{cm}^{-3}$, $N_A \approx 2 \times 10^{13}\,\mathrm{cm}^{-3}$). *Solid lines* are theoretical mobilities for various scattering mechanisms and combined mobility according to (8.12). Adapted from [654]

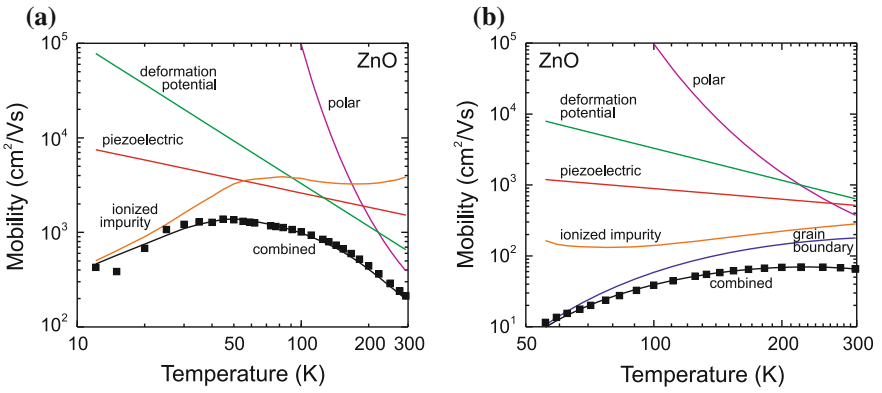


Fig. 8.5 Temperature dependence of the electron mobility in n-type (**a**) bulk ZnO and (**b**) a PLD-grown ZnO thin film on sapphire. In the latter, grain-boundary scattering is limiting the mobility. *Squares* are experimental data, *solid lines* are theoretical mobilities for various scattering mechanisms and combined mobility according to (8.12). Experimental data from [655]

(see Fig. 8.6). At very high temperature, when intrinsic conduction starts, σ shows a strong increase due to the increase in n.

At low temperature, the disorder due to doping (random positions of the impurity atoms) leads to a temperature driven metal–insulator transition as depicted in Fig. 7.29.

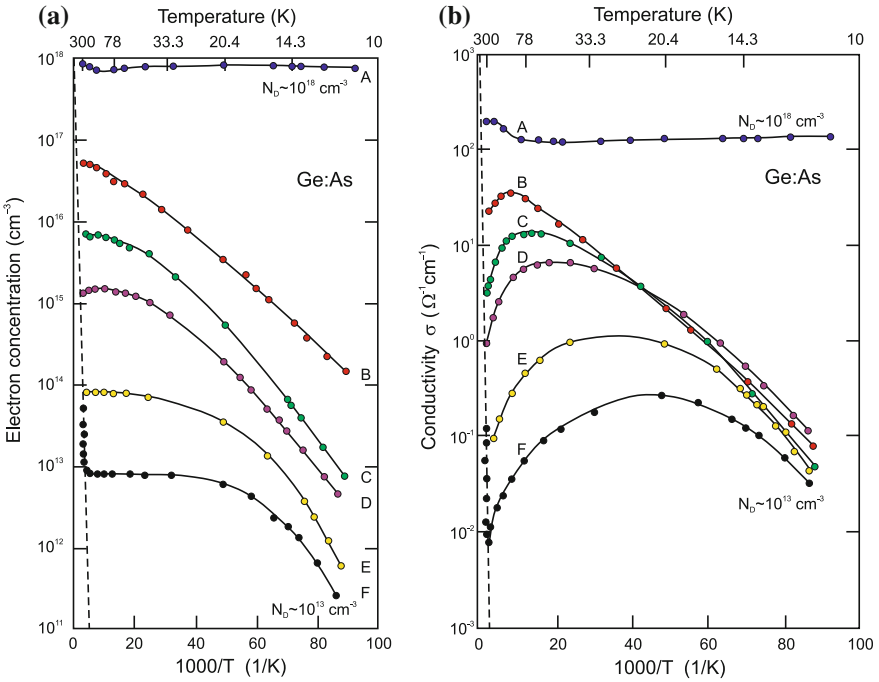


Fig. 8.6 (**a**) Carrier concentration and (**b**) conductivity of n-type Ge as a function of temperature. The doping level varies from $N_D \approx 10^{13}$–10^{18} (samples A–F as in Fig. 8.4a where the mobility of the same samples is shown). The *dashed lines* are for intrinsic Ge. The *solid lines* are guides to the eye. Adapted from [515]

8.3.10 Doping Dependence

The mobility decreases with increasing dopant concentration as already shown in Figs. 8.2 and 8.4a. In Fig. 8.7a the low doping limit is due to deformation potential scattering; the decrease with doping is due to ionized impurity scattering. At high doping level, it becomes more important at room temperature than (acoustical or optical) phonon scattering [656]. The mobility of carriers in n- and p-type silicon with very high carrier concentrations is depicted in Fig. 8.7b.

Thus, for bulk material high carrier density and high mobility are contrary targets and cannot be achieved simultaneously. A solution will be provided with the concept of *modulation* doping where the dopants and the (two-dimensional) carrier gas will be spatially separated in a heterostructure (cf. Sect. 12.3.4).

At high doping, the substitutional character of the impurities may be lost and secondary phases can arise, e.g. as observed for highly doped ZnO:Ga, exhibiting octahedral coordination of gallium in a parasitic $ZnGa_2O_4$ spinel phase for [Ga]=4 % [553]. The onset of such segregation phenomena is accompanied with the decrease of mobility and conductivity.

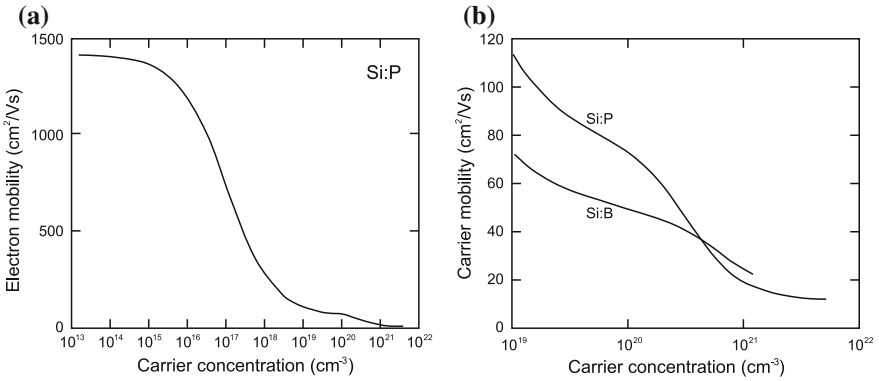


Fig. 8.7 (**a**) Electron mobility in Si:P at room temperature over a wide range of carrier concentrations. (**b**) Electron mobility in Si:P and hole mobility in Si:B for various high carrier concentrations. Adapted from [657]

8.3.11 Piezoresistivity

The dependence of resistivity on stress or strain is known as piezoresistive effect, first described in [658]. It is a consequence of the modification of the band structure upon stress and the change of effective masses (Sect. 6.10.2). In a cubic material, the resistivity ρ_i for transport in cartesian direction i changes compared to the unstrained state in a phenomenological description according to

$$\frac{\Delta\rho_i}{\rho_i} = \pi_{ij}\,\sigma_j, \tag{8.26}$$

where $\boldsymbol{\pi}$ is the piezoresistivity tensor (8.27) and the σ_j form the six-component stress tensor (5.59),

$$\boldsymbol{\pi} = \begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{12} & 0 & 0 & 0 \\ \pi_{12} & \pi_{11} & \pi_{12} & 0 & 0 & 0 \\ \pi_{12} & \pi_{12} & \pi_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & \pi_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \pi_{44} \end{pmatrix}. \tag{8.27}$$

Values for the piezoelectric coefficients are given in Table 8.3 for Si, Ge and GaAs.

The piezoelectric effect has been discussed in detail [659] and modeled for p-type Si [660]. We shall only give a simple example which is particularly relevant for advanced CMOS design (Sect. 24.5.5); the directional dependence of the piezoresistive coefficient of silicon is shown for uniaxial stress within in the (001) plane in Fig. 8.8. Uniaxial tensile stress increases hole resistivity along ⟨110⟩ stress directions, compressive stress thus increases hole conductivity.

Table 8.3 Piezoresistivity coefficients (in $10^{-11}\,\mathrm{Pa}^{-1}$) for Si, Ge and GaAs at room temperature

Material	ρ (Ω cm)	π_{11}	π_{12}	π_{44}	Reference
p-Si	7.8	6.6	−1.1	138.1	[658]
n-Si	11.7	−102.2	53.4	−13.6	[658]
p-Ge (Ge:Ga)	15.0	−10.6	5.0	98.6	[658]
n-Ge (Ge:As)	9.9	−4.7	−5.0	−137.9	[658]
p-GaAs	$\sim 10^{-3}$	−12.0	−0.6	46	[662]
n-GaAs	$\sim 10^{-3}$	−3.2	−5.4	−2.5	[662, 663]

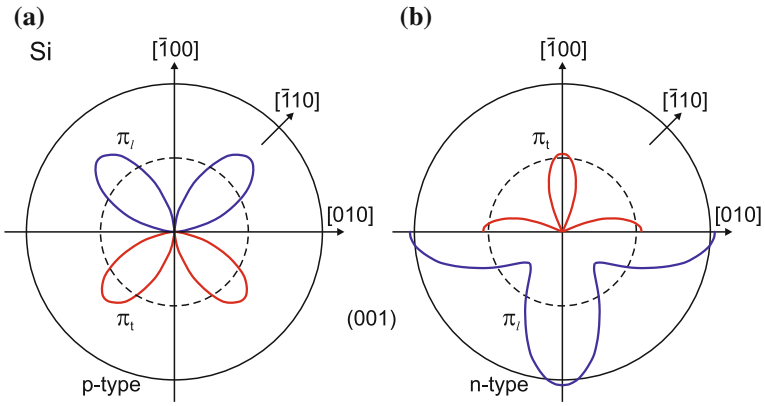


Fig. 8.8 Piezoresistive coefficient for current parallel (perpendicular) to the stress π_l as *blue lines* (π_t, *red lines*) for uniaxially stressed Si (001) at room temperature, (**a**) for p-type Si, (**b**) for n-type Si. The *upper* (*lower*) halves of the graphs show positive (negative) values of the piezoresistive coefficient, i.e. resistivity increases (decreases) with tensile stress. The *solid circle* indicates the value of $|\pi| = 10^{-9}\,\mathrm{Pa}^{-1}$, the *dashed circle* half that value. Adapted from [661]

8.4 High-Field Transport

In the case of small electric fields the scattering events are elastic. The drift velocity is linearly proportional to the electric field. The average thermal energy is close to its thermal value $3kT/2$ and the carriers are close to their band edges (Fig. 8.9a). The scattering efficiency, however, is reduced already at moderate fields. Then, the electron temperature [664] becomes larger than the lattice temperature. With increasing electrical field the carriers can gain more and more energy and will on average populate higher states, assuming a non-Boltzmann (and non-Fermi) statistical distribution [665]. The electron distribution in **k**-space is depicted for silicon for three different electric fields in Fig. 8.9b,c. Hot carriers suffer additional scattering processes that are discussed in the following, namely optical phonon emission, intervalley scattering and impact ionization.

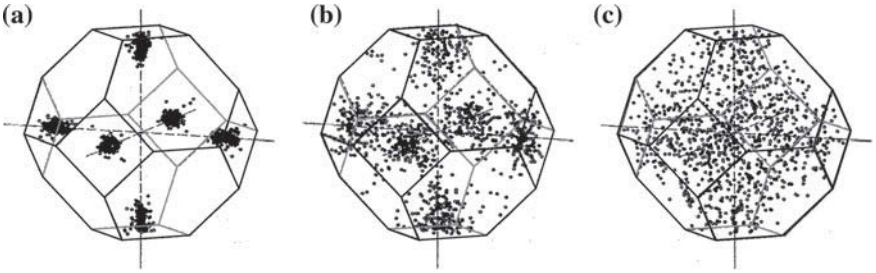


Fig. 8.9 Distribution of electrons in silicon in momentum space (cmp. Fig. 6.29c) for electric fields of (**a**) 10 kV/cm, (**b**) 10^2 kV/cm and (**c**) 10^3 kV/cm. Adapted from [665]

8.4.1 Drift-Saturation Velocity

If the carrier energy is large enough it can transfer energy to the lattice by the emission of an optical phonon. This mechanism is very efficient and limits the maximum drift velocity. Such behavior is non-ohmic. The limiting value for the drift velocity is termed the *drift-saturation velocity*. It is given by [666]

$$v_\mathrm{s} = \sqrt{\frac{8}{3\pi}} \sqrt{\frac{\hbar\omega_\mathrm{LO}}{m^*}}. \tag{8.28}$$

This relation can be obtained from an energy-balance consideration. The energy gain per unit time in the electric field is equal to the energy loss by the emission of an optical phonon.

$$q\,\mathbf{v}\cdot\mathbf{E} = \frac{\hbar\omega_\mathrm{LO}}{\tau}, \tag{8.29}$$

where τ is the typical relaxation time constant for LO phonon emission. Together with (8.3) we find (8.29) except for the pre-factor, which is close to 1. The exact pre-factor results from a more exact quantum-mechanical treatment. For Ge the drift-saturation velocity at room temperature is 6×10^6 cm/s, for Si it is 1×10^7 cm/s (Fig. 8.10a). The carrier velocity also depends on the crystallographic direction [667].

8.4.2 Negative Differential Resistivity

In GaAs there is a maximum drift velocity of about 2×10^7 cm/s and following a reduction in velocity with increasing field (1.2×10^7 cm/s at 10 kV/cm, 0.6×10^7 cm/s at 200 kV/cm) as shown in Fig. 8.10a. This regime, above the threshold field of $E_\mathrm{thr} = 3.2$ kV/cm in GaAs, is called *negative differential resistivity* (NDR) and was predicted in [673]. This phenomenon can be used in microwave oscillators, e.g. the Gunn element (Sect. 21.5.11).

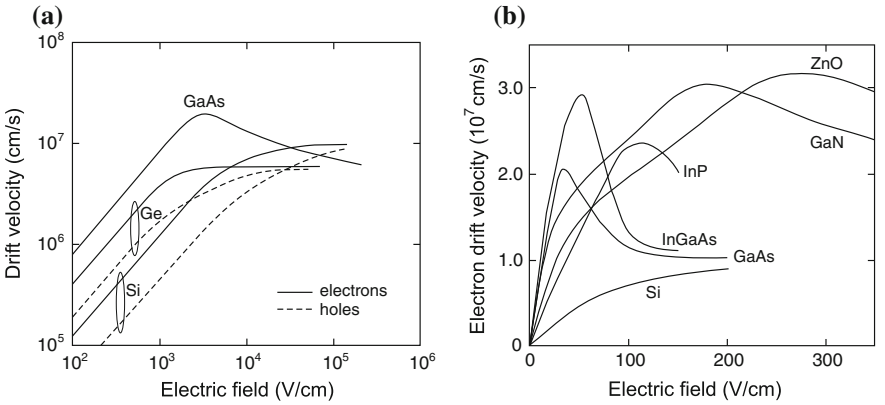


Fig. 8.10 Drift velocity at room temperature as a function of applied electric field for (**a**) high-purity Si, Ge, and GaAs on a double-logarithmic plot and (**b**) on linear plots for Si [668], Ge [669], GaAs [598], InP [670], InGaAs [671], GaN and ZnO [672]

Table 8.4 Material parameters for multivalley bandstructure of GaAs and InP

Material	E_g (eV)	ΔE (eV)	E_{thr} (kV/cm)	v_P (10^7 cm/s)	Lower valley (Γ)		Upper valley (L)	
					m^* (m_0)	μ_n (cm^2/Vs)	m^* (m_0)	μ_n (cm^2/Vs)
GaAs	1.42	0.36	3.2	2.2	0.068	≈8000	1.2	≈180
InP	1.35	0.53	10.5	2.5	0.08	≈5000	0.9	≈100

ΔE denotes the energetic separation of the two lowest valleys of the conduction band, E_{thr} the threshold field for NDR and v_P the peak velocity (at E_{thr}). Most values from [675]

The effect occurs in a multivalley band structure (see Fig. 8.11, for values cf. Table 8.4), e.g. in GaAs or InP, when the carrier energy is high enough to scatter (Fig. 8.11c, d) from the Γ minimum (small mass and high mobility) into the L valley (large mass and low mobility) [674].

The temperature dependence of the saturation velocity is shown in Fig. 8.12. With increasing temperature the saturation velocity decreases since the coupling with the lattice becomes stronger.

8.4.3 Velocity Overshoot

When the electric field is switched on, the carriers are at first in the Γ minimum (Fig. 8.11a). Only after a few scattering processes are they scattered into the L minimum. This means that in the first moments transport occurs with the higher mobility of the lowest minimum (Fig. 8.11e). The velocity is then larger than the (steady-state) saturation velocity in a dc field. This phenomenon is called *velocity overshoot* and is

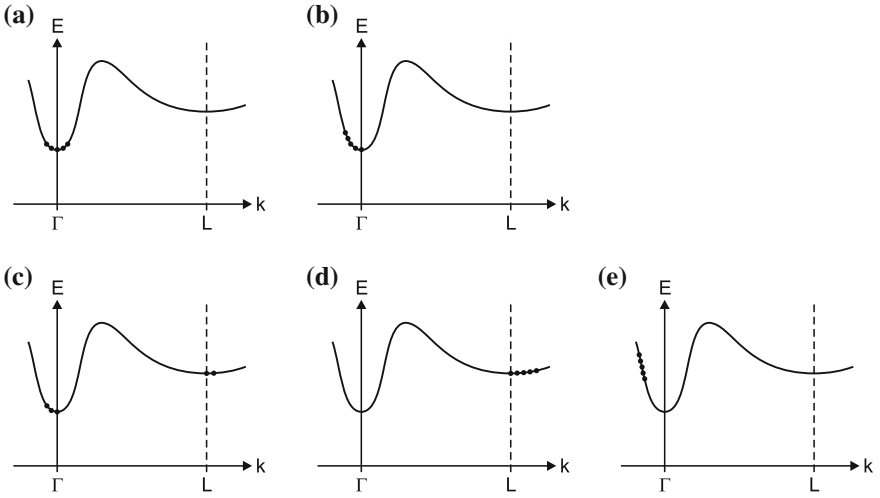


Fig. 8.11 Charge-carrier distribution in a multivalley band structure (e.g. GaAs, InP) for (**a**) zero, (**b**) small ($E < E_a$), (**c**) intermediate and (**d**) large ($E > E_b$) field strength. The situation shown in (**e**) is reached temporarily during velocity overshoot (see also Fig. 8.13)

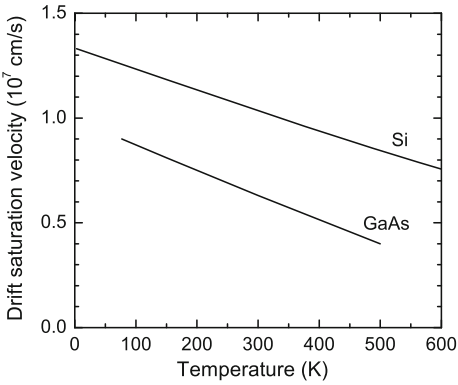


Fig. 8.12 Temperature dependence of the saturation velocity for Si (following $v_s = v_{s_0}(1 + 0.8\exp(T/600\,\mathrm{K}))^{-1}$ with $v_{s_0} = 2.4 \times 10^7$ cm/s from [668]) and GaAs [598, 676, 677]

a purely dynamic effect (Fig. 8.13). Velocity overshoot in GaN is discussed in [678]. It is an important effect in small transistors.

8.4.4 Impact Ionization

If the energy gain in the field is large enough to generate an electron–hole pair, the phenomenon of impact ionization occurs. The kinetic energy is $\propto v^2$. Momentum and energy conservation apply. Thus, at small energies (close to the threshold for impact ionization) the vectors are short and collinear to fulfill momentum conservation. At

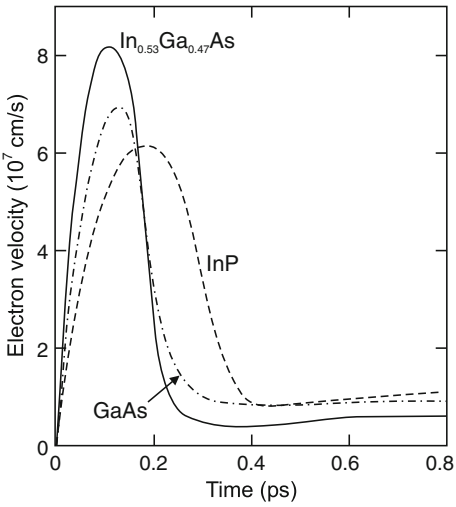


Fig. 8.13 Time dependence of the electron velocity at room temperature upon a step-like electric field (40 kV/cm) for GaAs (*dash-dotted line*), InP (*dashed line*) and $In_{0.53}Ga_{0.47}As$ (*solid line*)

higher energy, larger angles between the velocity vectors of the impact partners can also occur. If the process is started by an electron (Fig. 8.14a) the threshold energy is given by [679]

$$E_{\mathrm{e}}^{\mathrm{thr}} = \left(1 + \frac{m_{\mathrm{e}}}{m_{\mathrm{e}} + m_{\mathrm{hh}}}\right) E_{\mathrm{g}}. \tag{8.30}$$

If the process starts with a heavy hole, the threshold [679],

$$E_{\mathrm{hh}}^{\mathrm{thr}} = \left(1 + \frac{m_{\mathrm{hh}}}{m_{\mathrm{e}} + m_{\mathrm{hh}}}\right) E_{\mathrm{g}}, \tag{8.31}$$

is larger because of the larger hole mass.

The threshold for impact ionization triggered by a split-off hole (shown schematically in Fig. 8.14b) is [680]

$$E_{\mathrm{h}}^{\mathrm{thr}} = \left(1 + \frac{m_{\mathrm{so}}\,(1 - \Delta_0/E_{\mathrm{g}})}{2\,m_{\mathrm{hh}} + m_{\mathrm{e}} - m_{\mathrm{so}}}\right) E_{\mathrm{g}}. \tag{8.32}$$

Thus so-holes have typically the smaller threshold.[3] At energies where impact ionization occurs, non-parabolicities are typically important, thus (8.30)–(8.32) are only indicative. The threshold behavior and the dependence of the scattering rate as a function of the primary carrier energy in Si, calculated considering the detailed band structure, is shown in Fig. 8.15.

[3] Assuming $m_{\mathrm{so}} = m_{\mathrm{e}}$, $m_{\mathrm{e}} \ll m_{\mathrm{hh}}$ and $\Delta_0 \ll E_{\mathrm{g}}$, $E_{\mathrm{so}}^{\mathrm{thr}}/E_{\mathrm{e}}^{\mathrm{thr}} \approx 1 - (m_{\mathrm{e}}/m_{\mathrm{hh}})(1 + \Delta/E_{\mathrm{g}})/2 < 1$.

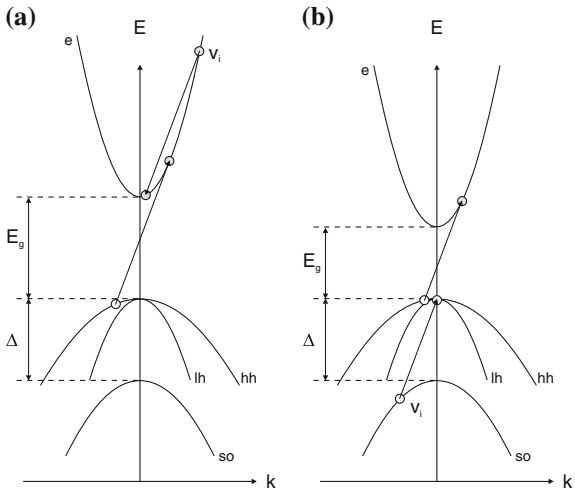


Fig. 8.14 Electron and hole transitions for impact ionization close to the threshold energy. Ionization is triggered by (**a**) an electron and (**b**) a split-off hole of velocity v_i

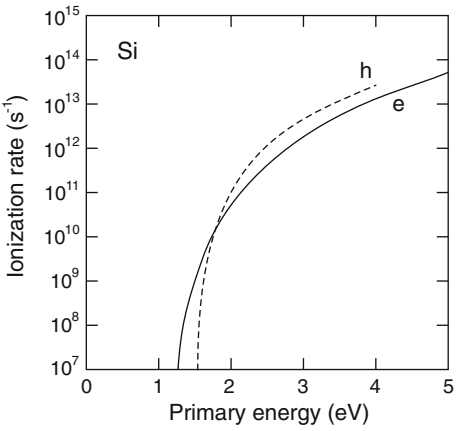


Fig. 8.15 Impact ionization rate as a function of primary carrier energy for electrons (*solid line*) and holes (*dashed line*) in silicon at room temperature. The curves are fit to results from a Monte-Carlo simulation. Adapted from [681, 682]

The generation rate G of electron–hole pairs during impact ionization is given by

$$G = \alpha_n \, n \, v_n + \alpha_p \, p \, v_p, \tag{8.33}$$

where α_n is the electron ionization coefficient. It describes the generation of electron–hole pairs per incoming electron per unit length. α_p denotes the hole ionization coefficient. The coefficients depend strongly on the applied electric field. They are shown in Fig. 8.16. They also depend on the crystallographic direction.

The impact ionization initiated by electrons and holes in silicon has been calculated considering the full band structure using a Monte Carlo technique in [681] and [682], respectively. In both cases the impact ionization rate is anisotropic for excess energies smaller than 3 eV and become isotropic above. The average energies

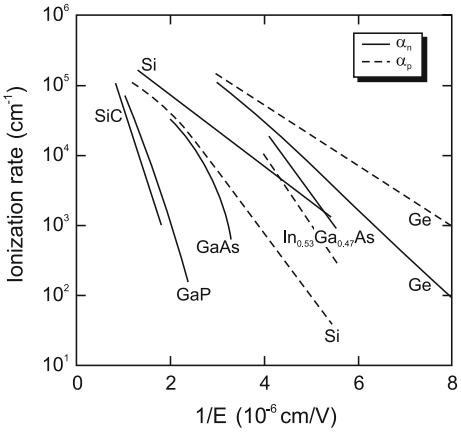


Fig. 8.16 Impact ionization rates for electrons and holes as a function of the inverse electric field for Si, Ge and other compound semiconductors at 300 K. Adapted from [500]

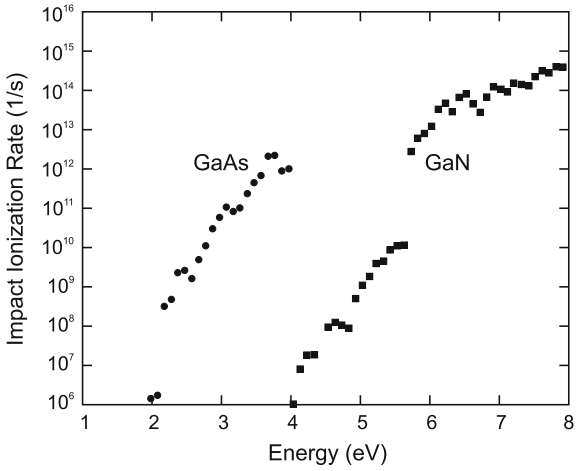


Fig. 8.17 Averaged rates for electron initiated impact ionization in GaAs (*circles*) and GaN (*squares*). Adapted from [683]

at the moment of generation of secondary generated carriers depends linearly on the primary electron or hole energy.

The energy dependence of the electron initiated impact ionization rate has been calculated for GaAs, GaN and ZnS considering details and anisotropy of the band structure in [683]. The rates averaged over the Brillouin zone are compared in Fig. 8.17. Because of the large band gap of GaN, impact ionization can only be generated by electrons in higher conduction bands. The sharp increase of ionization rate for GaN around 5.75 eV correlates with a large valence band DOS from hole states at the zone boundary.

8.5 High-Frequency Transport

The above consideration pertained to dc (or slowly varying) fields. Now, we consider an ac field. It accelerates the carriers but at the same time the dissipative force in the relaxation-time approximation is present, i.e. (for electrons)

$$m^*\dot{\mathbf{v}} = -e\,\mathbf{E} - m^*\,\frac{\mathbf{v}}{\tau}. \tag{8.34}$$

For a harmonic field $E \propto \exp(-\mathrm{i}\omega t)$ the complex conductivity ($\mathbf{j} = \sigma\mathbf{E} = nq\mathbf{v}$) is

$$\sigma = \frac{n\,e^2\,\tau}{m^*}\,\frac{1}{1-\mathrm{i}\,\omega\tau} = \frac{n\,e^2}{m^*}\,\frac{\mathrm{i}}{\omega+\mathrm{i}\gamma}, \tag{8.35}$$

with $\gamma = 1/\tau$ being the damping constant. Splitting into real and imaginary parts yields

$$\sigma = \frac{n\,e^2\,\tau}{m^*}\left(\frac{1}{1+\omega^2\tau^2} + \mathrm{i}\frac{\omega\tau}{1+\omega^2\tau^2}\right). \tag{8.36}$$

For small frequencies ($\omega \to 0$) the dc conductivity from (8.5) is recovered, i.e. $\sigma = ne^2\tau/m^*$. For high frequencies ($\omega\tau \gg 1$)

$$\sigma = \frac{n\,e^2\,\tau}{m^*}\left(\frac{1}{\omega^2\tau^2} + \mathrm{i}\frac{1}{\omega\tau}\right). \tag{8.37}$$

8.6 Polarons

In an ionic lattice, the electron polarizes the ions and causes a change of their equilibrium position. Depending of the severity of this effect, the lattice polarization leads to a modification of carrier (electron or hole) mass during band transport (Sect. 8.6.1) (*large* polarons) or the lattice deformation is so strong that it leads to carrier localization on the length scale of the lattice constant. Such self-trapped carriers are termed *small* polarons and discussed in Sect. 8.6.1. Reviews are given in [684, 685].

8.6.1 Large Polarons

When the electron moves through the ionic crystal and must drag an ion displacement with it, the effective electron mass changes to the 'polaron mass' m_p,[4]

[4] For the calculation, many-particle theory and techniques are needed; the best solution is still given by Feynman's path integral calculation [686–688].

Table 8.5 Fröhlich coupling constant α for various semiconductors

GaSb	GaAs	GaP	GaN	InSb	InAs	InP	InN
0.025	0.068	0.201	0.48	0.022	0.045	0.15	0.24
3C-SiC	ZnO	ZnS	ZnSe	ZnTe	CdS	CdSe	CdTe
0.26	1.19	0.63	0.43	0.33	0.51	0.46	0.35

Data from [691]

$$m_\mathrm{p} = m^* \left(1 + \frac{\alpha}{6} + 0.025\,\alpha^2 + \cdots\right), \tag{8.38}$$

for $\alpha \leq 1$, with m^* being the band mass as defined in Sect. 6.8.2 and α the Fröhlich coupling constant[5]

$$\alpha = \frac{1}{2}\,\frac{e^2}{\hbar}\sqrt{\frac{2\,m^*}{\hbar\omega_\mathrm{LO}}}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right). \tag{8.39}$$

This process it called the *polaronic* effect and requires additional energy [686, 689]. Often, the polaron mass is given as $m_\mathrm{p} = m^*/(1 - \alpha/6)$ which is the result of perturbation theory [689] and an approximation to (8.38) for small α.

For large coupling parameter, $\alpha \gg 1$, the polaron mass is given by [687]

$$m_\mathrm{p} = m^*\,\frac{16}{81\,\pi^4}\,\alpha^4. \tag{8.40}$$

The energy of the electron is lowered due to the interaction with the lattice. The energy E_0 for $k = 0$ is given, relative to the uncoupled case, by

$$E_0 = -\left(\alpha + 0.0098\,\alpha^2 + \cdots\right)\,\hbar\omega_0 \quad , \alpha \leq 1 \tag{8.41a}$$

$$E_0 = -\left(2.83 + 0.106\,\alpha^2\right)\,\hbar\omega_0 \quad , \alpha \gg 1 \tag{8.41b}$$

Numerical results are reported in [690].

Polarons in semiconductors are typically 'large' or Fröhlich-type polarons, i.e. the coupling constant is small (Table 8.5). The dressing with phonons (as the ion displacement is called in a quantum-mechanical picture) is then only a perturbative effect and the number of phonons per electron ($\approx\alpha/2$) is small. If α becomes large ($\alpha > 1$, $\alpha \sim 6$), as is the case for strongly ionic crystals such as alkali halides, the polaron becomes localized by the electron–phonon interaction[6] and hopping occurs infrequently from site to site.

[5] This constant is part of the matrix element in the Hamiltonian of the electron–phonon interaction and is related to the electric field created by LO phonons, as given in (5.52).

[6] One can think about it in the way that the electron strongly polarizes the lattice and digs itself a potential hole out of which it can no longer move.

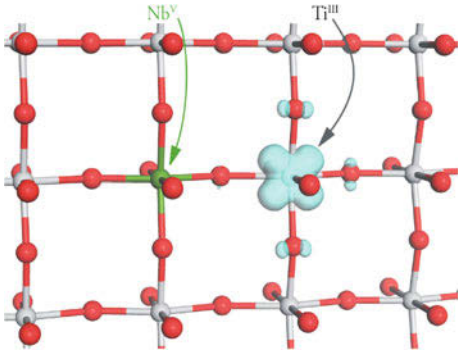


Fig. 8.18 Hole from Nb acceptor localized on Ti site (small polaron) in rutile TiO_2. Adapted from [694]

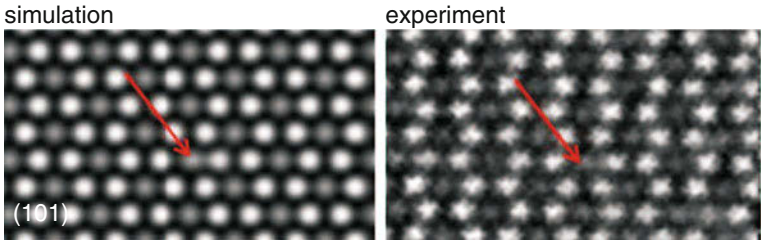


Fig. 8.19 Simulated and experimental TEM images of β-Ga_2O_3 in (101) projection. The *arrow* denotes the position of a polaron. Adapted from [693]

8.6.2 Small Polarons

In a polaron, the charge carrier (electron or hole) sits in a potential well resulting from the ionic displacements it created. In some materials, the shape and strength of this potential well is such that the charge is confined to a volume of approximately one unit cell or less. In this case, the polaron is called a small polaron. An example of a hole polaron in rutile TiO_2:Nb is depicted in Fig. 8.18. In oxides often the hole from an acceptor is bound to oxygen, e.g. in $BaTiO_3$:Na, as reviewed in [692]. In Fig. 8.19 the lattice relaxation due to a hole bound to oxygen in the monoclinc unit cell of β-Ga_2O_3 is depicted directly using aberration corrected TEM. The bonding of the hole to the oxygen atom breaks the bond to a Ga atom which moves by 0.1 nm from its equilibrium position [693].

A proper theoretical analysis of a small polaron requires ab initio techniques that account for the motion of each atom in the few unit cells nearest the electron.[7]

The transport of small polarons occurs generally via thermally-activated hopping (cmp. Sect. 8.7). Under certain conditions the following mobilities for drift and Hall

[7]This paragraph has been taken from the concise tutorial by S.J.F. Byrnes [695].

effect have been given [684]:

$$\mu_{\mathrm{d}} \propto T^{-1} \exp(-W/(2kT)), \tag{8.42}$$

$$\mu_{\mathrm{H}} \propto T^{-1/2} \exp(-W/(6kT)), \tag{8.43}$$

W being the polaron binding energy. Generally, materials with small polar on transport exhibit high carrier density, often due to structural defects, and low mobility.

8.7 Hopping Transport

Disordered solids such as amorphous semiconductors, films containing quantum dots or material with many defects are characterized by a large density of localized states which can form band tails or a large density of states within the band gap. Hopping conduction is the tunneling between localized states and has been treated with various models [696–698].

A commonly observed phenomenon is the variable range hopping with a conductivity given by

$$\sigma = \sigma_0 \exp\left(-(T_0/T)^s\right) \tag{8.44}$$

with $s = 1/4$. Such law is fulfilled for amorphous silicon (Fig. 8.20). Mott has derived [700] and the exponent $s = 1/4$ using the following argument: The probability p to

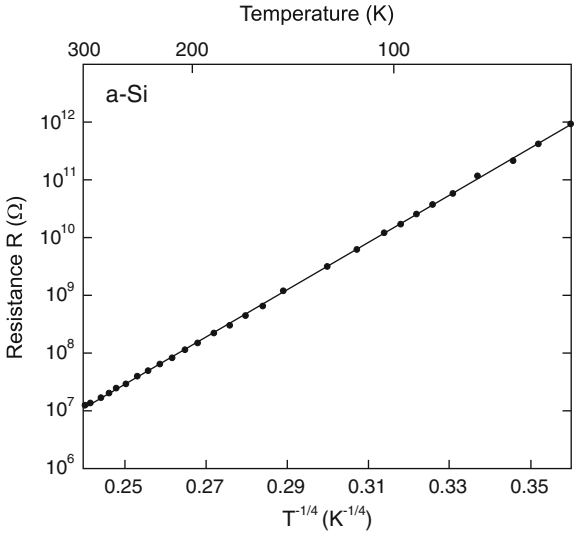


Fig. 8.20 Temperature dependence of the planar resistance for Si films deposited at room temperature. *Solid line* is linear fit with $T_0 = 8 \times 10^7$ K according to (8.44) ($s = 1/4$). Adapted from [699]

hop from one localized site to another is proportional to

$$p \propto \exp(-2\alpha R - W/kT). \tag{8.45}$$

The first term stems from the probability to find the electron within radius R from its initial site, α being the decay constant of its wave function, $\Psi(r) \propto \exp(-\alpha r)$. The second term is the Boltzmann factor for bridging the energy mismatch W between localized states with a phonon-assisted process, assuming a low temperature limit ($kT \ll W$). There is a trade-off between hopping to levels closer in energy but spatially further away (on average), preferred at low temperature and the hopping to energy levels with larger W but spatially closer at higher temperatures. Thus the hopping range changes with temperature, giving the mechanism its name.

$D(E_\mathrm{F})$ shall be the (constant) density of localized states around the Fermi level. Within a radius R, there is on average one state of energy between 0 and $W(R)$ when (for three-dimensional bulk material)

$$W(R) = \frac{1}{D(E_\mathrm{F})\,(4\pi/3)\,R^3}. \tag{8.46}$$

Substituting (8.46) in (8.44) and searching for the maximum yields the most probable hopping distance

$$R \approx (\alpha\, kT\, D(E_\mathrm{F}))^{-1/4}, \tag{8.47}$$

showing again, the varying range of hopping with temperature. Thus we find for T_0 in (8.44),

$$T_0 \approx \frac{\alpha^3}{k\, D(E_\mathrm{F})}. \tag{8.48}$$

Other types of hopping mechanisms are the Efros–Shklovskii variable range hopping ($s = 1/2$), emerging for an energy dependent density of states $D(E) \propto (E - E_\mathrm{F})^2$ due to Coulomb interaction between hopping sites [701], or the next neighbor hopping ($s = 1$).

From (8.44) one can rewrite for $\xi = \mathrm{d}(\ln \sigma(T))/\mathrm{d} \ln T$,

$$\ln \xi = \ln s + s\, \ln T_0 - s\, \ln T. \tag{8.49}$$

Thus in a plot of $\ln \xi$ versus $\ln T$, the exponent s can be determined from the slope. As can be seen in Fig. 8.21, for the conductivity of a hydrogenated amorphous silicon thin film the transition of hopping mechanism from Efros–Shklovskii variable range hopping ($s \approx 1/2$) to next neighbor hopping ($s \approx 1$) takes place around $T = 220\,\mathrm{K}$, as discussed in detail in [702].

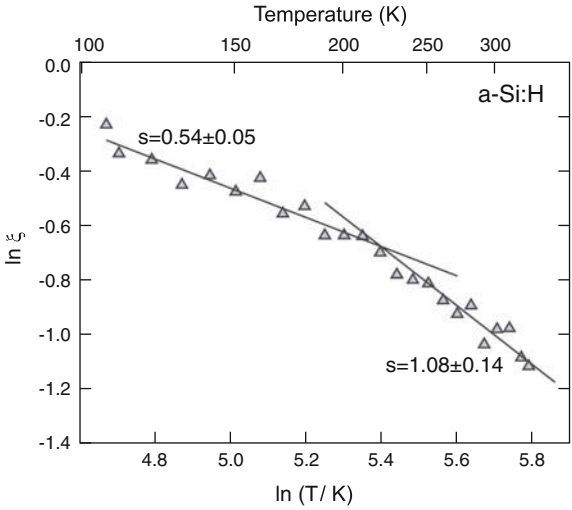


Fig. 8.21 Temperature dependence of conductivity of a hydrogenated amorphous Si thin film, plotted as $\ln \xi$ versus $\ln T$ (8.49). *Solid lines* are linear fits for constant s according to (8.49) as labelled. Adapted from [702]

8.8 Transport in Amorphous Semiconductors

Many models have been presented for the carrier transport in amorphous semiconductors [178]. The most important concept is that of a *mobility edge*, an energy separating localized from delocalized states [475, 476, 703]. This is schematically depicted in Fig. 8.22. The carrier transport between localized states is mediated via tunneling (hopping) which has been described in the previous section (Sect. 8.7). The localization of carriers in random lattices has been treated by Anderson [560] and reviewed in [558]. If the degree of disorder surpasses a certain value, diffusion is suppressed (at $T = 0$) and conductivity vanishes altogether (Anderson metal–insulator transition).

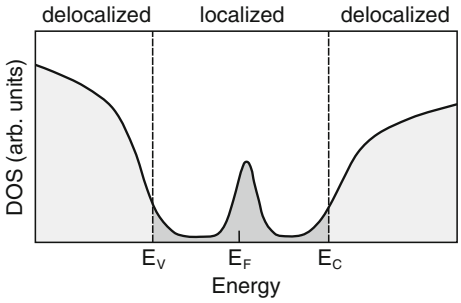


Fig. 8.22 Schematic density of states of amorphous semiconductor with band tails and deep levels. The localized (delocalized) states are shown in *dark* (*light*) *grey*. The mobility edges for electrons and holes are indicated by *dashed lines*

The transport in delocalized states is similar to band transport. The conductivity (for electrons) is given as

$$\sigma = -e \int_{E_C}^{\infty} D_e(E)\, \mu_e(E)\, f_e(E)\, \mathrm{d}E. \tag{8.50}$$

If the Fermi energy is close to the middle of the gap, pinned to deep states, the Fermi-Dirac distribution can be replaced by the Boltzmann factor. Assuming a constant density of states and mobility for the delocalized states,

$$\sigma = -e\, D_e(E_C)\, \mu_e(E_C)\, kT\, \exp\left(\frac{E_C - E_F}{kT}\right). \tag{8.51}$$

Charge carriers from localized states in the tails can be thermally excited into delocalized states and contribute to conductivity (thermally activated hopping). The mobility then contains an exponential thermal activation term [178].

8.9 Ionic Transport

Ionic transport is the movement of ions upon application of a voltage. Here, we discuss only solid electrolytes. The transport can include the motion of one or several of the constituents of the lattice and the transport of other ions (e.g. hydrogen ions (protons), oxygen ions) through the crystal. Related to this is the diffusive ionic movement of impurities or defects through the crystal (cmp. Sect. 4.2.3). Ionic conduction of the lattice constituents under dc voltage will eventually destroy the crystal.

In typical semiconductors like silicon or gallium arsenide, the conductivity is entirely due to electronic conduction. A typical solid electrolyte is zirconia (ZrO_2) doped with yttria, so-called yttria-stabilized zirconia (YSZ) that takes on a cubic fluorite lattice (see Sect. 3.4.8). It can conduct oxygen ions via the mobility of oxygen vacancies for use in solid-oxide fuel cells (SOFC) [704]. The conductivity is about 0.01 S/cm at a temperature around 1000 K, almost entirely due to ionic transport. Doping with calcium oxide results in an oxygen conductor that is used in oxygen sensors in automobiles (lambda sensor). The ionic conductivity can be significantly increased, compared to bulk material, along interfaces [705, 706].

Other typical solid electrolytes are cuprous iodide (CuI) [494] and also AgI. In the high temperature cubic phase (α-polymorph), the iodide ions form a fairly rigid cubic framework and the metal ions are mobile; the copper diffusion pathways have been discussed [707, 708]. The temperature dependence of conductivity of CuI is shown in Fig. 8.23.

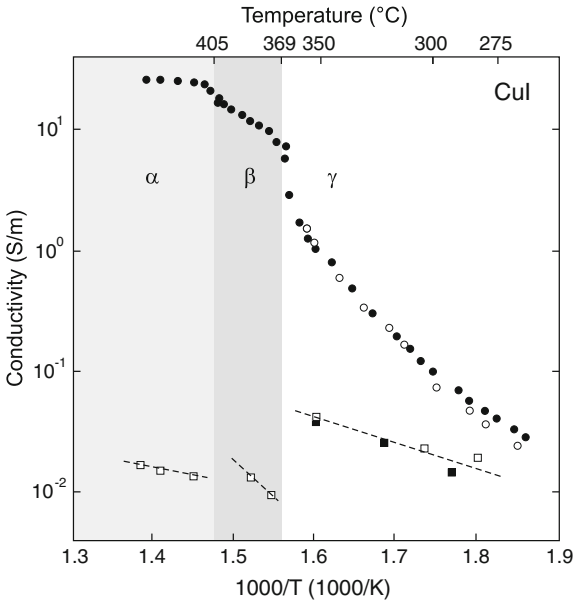


Fig. 8.23 Total (*circles*) and electronic (*squares*) conductivity of CuI coexisting with copper. *Filled* (*empty*) symbols refer to polycrystalline (single crystal) samples. The different structural phases (α (cubic), β (wurtzite), γ (zincblende)) are indicated by shaded areas as labeled. *Dashed lines* are guides to the eyes. Adapted from [709]

8.10 Diffusion

A gradient of a particle concentration n leads to a particle current proportional to $-\nabla n$. This diffusion law (Fick's law) corresponds microscopically to a random walk. The gradients of the semiconductor carrier densities ∇n or ∇p thus lead to electron and hole currents, respectively:

$$\mathbf{j}_\mathrm{n} = eD_\mathrm{n}\nabla n \tag{8.52a}$$

$$\mathbf{j}_\mathrm{p} = -eD_\mathrm{p}\nabla p. \tag{8.52b}$$

The coefficients D_n and D_p are called the electron and hole diffusion coefficient, respectively. Thus the total electron and hole currents in the presence of an electric field $\mathbf{E}$ and diffusion are

$$\mathbf{j}_\mathrm{n} = -e\mu_\mathrm{n} n\,\mathbf{E} + eD_\mathrm{n}\,\nabla n \tag{8.53a}$$

$$\mathbf{j}_\mathrm{p} = e\mu_\mathrm{p} p\,\mathbf{E} - eD_\mathrm{p}\,\nabla p. \tag{8.53b}$$

This relation can also be deduced more generally from the gradient of the Fermi level as

$$\mathbf{j}_\mathrm{n} = -e\mu_\mathrm{n} n\,\mathbf{E} - n\mu_\mathrm{n}\,\nabla E_\mathrm{F} \tag{8.54a}$$

$$\mathbf{j}_\mathrm{p} = e\mu_\mathrm{p} p\,\mathbf{E} - p\mu_\mathrm{p}\,\nabla E_\mathrm{F}. \tag{8.54b}$$

Using (7.6) and (7.7) for the concentrations (valid also in the case of degeneracy) and using $\mathrm{d}F_j(x)/\mathrm{d}x = F_{j-1}(x)$ we obtain

$$\mathbf{j}_\mathrm{n} = -e\mu_\mathrm{n} n\,\mathbf{E} - kT\mu_\mathrm{n}\frac{F_{1/2}(\eta)}{F_{-1/2}(\eta)}\,\nabla n \tag{8.55a}$$

$$\mathbf{j}_\mathrm{p} = e\mu_\mathrm{p} p\,\mathbf{E} - kT\mu_\mathrm{p}\frac{F_{1/2}(\zeta)}{F_{-1/2}(\zeta)}\,\nabla p, \tag{8.55b}$$

with $\eta = (E_\mathrm{F} - E_\mathrm{C})/kT$ and $\zeta = -(E_\mathrm{F} - E_\mathrm{V})/kT$. If the pre-factor of the density gradient is identified as the diffusion coefficient we find the (generalized) so-called 'Einstein relations' ($\beta = e/(kT)$) [529, 710]:

$$D_\mathrm{n} = -\beta^{-1}\mu_\mathrm{n}\frac{F_{1/2}(\eta)}{F_{-1/2}(\eta)} \tag{8.56a}$$

$$D_\mathrm{p} = \beta^{-1}\mu_\mathrm{p}\frac{F_{1/2}(\zeta)}{F_{-1/2}(\zeta)}. \tag{8.56b}$$

Useful analytical approximations have been discussed in [711]. We note that, e.g., (8.56a) can also be written as [712, 713]

$$D_\mathrm{n} = -\beta^{-1}\mu_\mathrm{n} n\,\frac{\partial\eta}{\partial n}. \tag{8.57}$$

In the case of nondegeneracy, i.e. when the Fermi level is within the band gap and not closer than about $4kT$ to the band edges, $\eta = \ln(n/N_\mathrm{C})$. Then $\partial\eta/\partial n = 1/n$, and the equation simplifies to $D = (kT/q)\,\mu$, i.e. the 'regular' Einstein-relations,

$$D_\mathrm{n} = -\beta^{-1}\mu_\mathrm{n} \tag{8.58a}$$

$$D_\mathrm{p} = \beta^{-1}\mu_\mathrm{p}. \tag{8.58b}$$

In this case, (8.53a,b) read

$$\mathbf{j}_\mathrm{n} = -e\mu_\mathrm{n} n\,\mathbf{E} - kT\mu_n\,\nabla n \tag{8.59a}$$

$$\mathbf{j}_\mathrm{p} = e\mu_\mathrm{p} p\,\mathbf{E} - kT\mu_p\,\nabla p. \tag{8.59b}$$

We recall that both diffusion coefficients are positive numbers, since μ_n is negative. Generally, the diffusion coefficient depends on the density. A Taylor series of the Fermi integral yields

$$D_\mathrm{n} = -\beta^{-1}\mu_\mathrm{n}\left[1 + 0.35355\left(\frac{n}{N_\mathrm{C}}\right) - 9.9\times10^{-3}\left(\frac{n}{N_\mathrm{C}}\right)^2 + \cdots\right]. \tag{8.60}$$

8.11 Continuity Equation

The balance equation for the charge is called the continuity equation. The temporal change of the charge in a volume element is given by the divergence of the current and any source (generation rate G), e.g. an external excitation, or drain (recombination rate U). Details about recombination mechanisms are discussed in Chap. 10. Thus, we have

$$\frac{\partial n}{\partial t} = G_\mathrm{n} - U_\mathrm{n} - \frac{1}{q}\,\nabla\cdot\mathbf{j}_\mathrm{n} = G_\mathrm{n} - U_\mathrm{n} + \frac{1}{e}\,\nabla\cdot\mathbf{j}_\mathrm{n} \tag{8.61a}$$

$$\frac{\partial p}{\partial t} = G_\mathrm{p} - U_\mathrm{p} - \frac{1}{e}\,\nabla\cdot\mathbf{j}_\mathrm{p}. \tag{8.61b}$$

In the case of nondegeneracy we find, using (8.53a,b)

$$\frac{\partial n}{\partial t} = G_\mathrm{n} - U_\mathrm{n} - \mu_\mathrm{n} n\,\nabla\cdot\mathbf{E} - \mu_\mathrm{n}\,\mathbf{E}\,\nabla n + D_\mathrm{n}\Delta n \tag{8.62a}$$

$$\frac{\partial p}{\partial t} = G_\mathrm{p} - U_\mathrm{p} - \mu_\mathrm{p} p\,\nabla\cdot\mathbf{E} - \mu_\mathrm{p}\,\mathbf{E}\,\nabla p + D_\mathrm{p}\Delta p. \tag{8.62b}$$

In the case of zero electric field these read

$$\frac{\partial n}{\partial t} = G_\mathrm{n} - U_\mathrm{n} + D_\mathrm{n}\Delta n \tag{8.63a}$$

$$\frac{\partial p}{\partial t} = G_\mathrm{p} - U_\mathrm{p} + D_\mathrm{p}\Delta p, \tag{8.63b}$$

and if the stationary case also applies:

$$D_\mathrm{n}\Delta n = -G_\mathrm{n} + U_\mathrm{n} \tag{8.64a}$$

$$D_\mathrm{p}\Delta p = -G_\mathrm{p} + U_\mathrm{p}. \tag{8.64b}$$

8.12 Heat Conduction

We consider here the heat transport [714] due to a temperature gradient. The heat flow $\mathbf{q}$, i.e. energy per unit area per time in the direction $\hat{\mathbf{q}}$, is proportional to the local gradient of temperature. The proportionality constant κ is called, heat conductivity,

$$\mathbf{q} = -\kappa\,\nabla T. \tag{8.65}$$

In crystals, the heat conductivity can depend on the direction and thus κ is generally a tensor of rank 2. In the following, κ will be considered as a scalar quantity. The quite generally valid Wiedemann–Franz law connects the thermal and electrical

conductivities

$$\kappa = \frac{\pi^2}{3}\left(\frac{k}{e}\right)^2 T\sigma. \tag{8.66}$$

The balance (continuity) equation for the heat energy Q is

$$\nabla\cdot\mathbf{q} = -\frac{\partial Q}{\partial t} = -\rho C\frac{\partial T}{\partial t} + A, \tag{8.67}$$

where ρ denotes the density of the solid and C the heat capacity. A denotes a source or drain of heat, e.g. an external excitation. Combining (8.65) and (8.67), we obtain the equation for heat conductivity

$$\Delta T = \frac{\rho C}{\kappa}\frac{\partial T}{\partial t} - \frac{A}{\kappa}, \tag{8.68}$$

which simply reads $\Delta T = 0$ for a stationary situation without sources.

The random mixture of various atoms in natural elements represents a perturbation of the perfectly periodic lattice. Since the mass of the nuclei varies, in particular lattice vibrations will be perturbed. Thus we expect an effect on the heat conductivity. In Fig. 8.24, the thermal conductivity of crystals from natural Ge and enriched ^{74}Ge are compared [715], the latter having, as expected, the higher heat conductivity, i.e. less scattering. The thermal conductivity of isotopically pure ^{28}Si thin films has been measured to be 60 % greater than natural silicon at room temperature and at least 40 % greater at 100 °C, a typical chip operating temperature [716, 717].

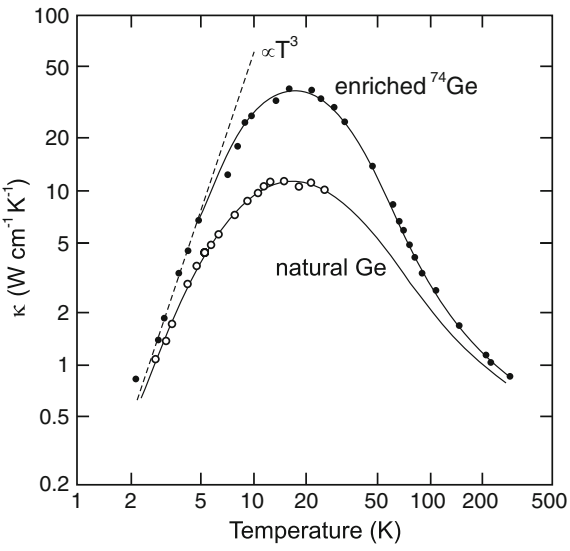


Fig. 8.24 Thermal conductivity of Ge versus temperature. The enriched Ge consists of 96 % ^{74}Ge while the natural isotope mix is 20 % ^{70}Ge, 27 % ^{72}Ge, 8 % ^{73}Ge, 27 % ^{74}Ge and 8 % ^{76}Ge. The *dashed line* shows the $\kappa \propto T^3$ dependence at low temperatures (Debye's law). Adapted from [715]

8.13 Coupled Heat and Charge Transport

The standard effect of coupled charge and heat transport is that a current heats its conductor via Joule heating. However, more intricate use of thermoelectric effects can also be employed to cool certain areas of a device. For further details see [718, 719].

For the analysis of coupled charge and heat transport we first sum the electric field and the concentration gradient to a new field $\hat{\mathbf{E}} = \mathbf{E} + \nabla E_{\mathrm{F}}/e$. Then, the heat flow and charge current are

$$\mathbf{j} = \sigma \hat{\mathbf{E}} + L \nabla T \tag{8.69}$$

$$\mathbf{q} = M \hat{\mathbf{E}} + N \nabla T, \tag{8.70}$$

where $\hat{\mathbf{E}}$ and ∇T are the stimulators for the currents. From the experimental point of view there is interest to express the equations in $\mathbf{j}$ and ∇T since these quantities are measurable. With new coefficients they read

$$\hat{\mathbf{E}} = \rho \mathbf{j} + S \nabla T \tag{8.71}$$

$$\mathbf{q} = \Pi \mathbf{j} - \kappa \nabla T, \tag{8.72}$$

where ρ, S and Π are the specific resistance, thermoelectric power and Peltier coefficient (transported energy per unit charge), respectively. The relations with the coefficients σ, L, M, and N are given by

$$\rho = \frac{1}{\sigma} \tag{8.73a}$$

$$S = -\frac{L}{\sigma} \tag{8.73b}$$

$$\Pi = \frac{M}{\sigma} \tag{8.73c}$$

$$\kappa = \frac{ML}{\sigma} - N. \tag{8.73d}$$

8.13.1 Thermopower and Seebeck Effect

A semiconductor shall have two ends at different temperatures T_2 and T_1 and a temperature gradient in between in an open circuit, i.e. $\mathbf{j} = 0$. Then a field $\hat{\mathbf{E}} = S \nabla T$ and a voltage $U = S/(T_2 - T_1)$ will arise. This effect is called the thermoelectric or Seebeck effect. S is termed the Seebeck coefficient or the *thermoelectric power*, often also denoted as Q in the literature. The voltage can be measured and used to determine the temperature at one end if the temperature at the other end is known, forming a thermometer. The Seebeck coefficient is positive if the electric field is in the same direction as the temperature gradient.

A famous relation from irreversible thermodynamics connects it to the Peltier coefficient via

$$S = \frac{\Pi}{T}. \tag{8.74}$$

The Seebeck coefficient is related to the energy transport by charge carriers. The heat (energy) flow is obviously from the hot to the cold end (assuming here $T_2 > T_1$), so is the flow of charge carriers. In a simple picture, if the energy is carried by (hot) holes, the current (by definition the direction of positive charge carriers) is from the hot to the cold end ($2 \rightarrow 1$); if the energy flow is carried by electrons, the current flows from the cold to the hot end ($1 \rightarrow 2$). Accordingly, energy transport by electrons and holes gives rise to different signs of the thermoelectric coefficient (Fig. 8.25). If the cold (unheated) substrate is grounded, the sign of the voltage at a hot solder tip pressed (carefully) on the surface of the semiconductor yields the conductivity type, n-type (p-type) for a negative (positive) voltage.

However, the semiconductor should not be heated so strongly that intrinsic conduction arises. In this case the conductivity and the thermoelectric power is determined by the carrier type with the higher mobility; typically, and for the case of silicon shown in Fig. 8.25, these are the electrons thus yielding a negative Seebeck coefficient in the intrinsic regime.

For band conduction the thermopower (I.29) is given for electrons (S_n) and holes (S_p) by [722] (for a derivation see Sect. I.4)

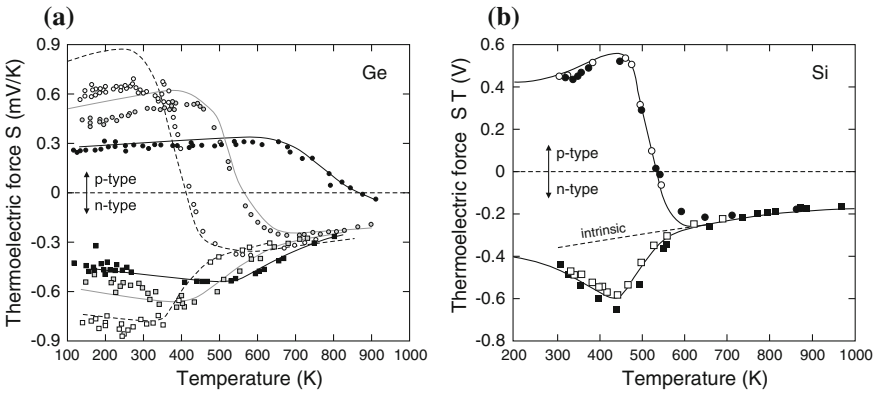


Fig. 8.25 (**a**) Seebeck coefficient S for n- and p-doped germanium. Experimental data (*symbols*) and theory (*lines*). $N_A - N_D$ is $5.7 \times 10^{15}\,cm^{-3}$ (*white circles*), $1.7 \times 10^{17}\,cm^{-3}$ (*grey*) and $7.2 \times 10^{18}\,cm^{-3}$ (*black*); $N_D - N_A$ is $3.3 \times 10^{15}\,cm^{-3}$ (*white squares*), $1.1 \times 10^{17}\,cm^{-3}$ (*grey*) and $6.2 \times 10^{17}\,cm^{-3}$ (*black*) Adapted from [720]. (**b**) Thermoelectric force Π of lowly doped n- and p-silicon as a function of temperature. *Solid line* is from simple model calculation and *symbols* represent data from silicon samples with the approximate doping of *circles*: $1 \times 10^{15}\,cm^{-3}$ B, $2 \times 10^{14}\,cm^{-3}$ donors, *squares* $4 \times 10^{14}\,cm^{-3}$ P, $9 \times 10^{13}\,cm^{-3}$ acceptors. Adapted from [721]

$$S_n = -\frac{k}{e}\left(\frac{E_C - E_F}{kT} + A_C\right) \tag{8.75a}$$

$$S_p = \frac{k}{e}\left(\frac{E_F - E_V}{kT} + A_V\right), \tag{8.75b}$$

where A_i are constants (I.31a) depending on the energy dependence of the density of states and the mobility. The sign of the thermopower tells whether conduction takes place above (negative sign) or below (positive sign) the Fermi level.

If the Fermi level is fixed and both electrons and holes contribute (two-band conduction), the thermopower is (evaluating (I.32), $b = \sigma_n/\sigma_p$ and the gap center energy $E_M = (E_C - E_V)/2$)

$$S = \frac{k}{e}\left(\frac{1-b}{1+b}\frac{E_g}{2kT} + \frac{E_F - E_M}{kT} + \frac{A_V - b\,A_C}{1+b}\right). \tag{8.76}$$

In the case of intrinsic conduction from (7.18) $E_F - E_M = (kT/2)\,\ln(N_V/N_C)$.

The thermoelectric power from some highly doped n-type silicon samples is depicted in Fig. 8.26a. At low temperature the (low) conductivity is due to conduction in a donor impurity band (cmp. Sect. 7.5.7). At high compensation of about 90 % (grey data points in Fig. 8.26a), the band is only 10 % filled and acts like a valence band with positive thermopower at sufficiently low temperature when the free carrier density is small. Without compensation, the thermopower remains negative since the almost completely filled impurity band acts conduction band like. The dependence of thermopower on doping has been simulated in [723] (Fig. 8.26); the

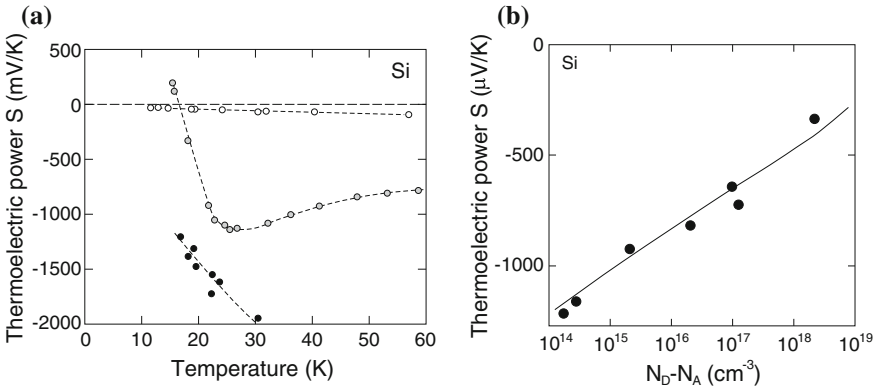


Fig. 8.26 (**a**) Thermoelectric power S of highly doped n-type silicon as a function of temperature. *Circles* are experimental data and *dashed lines* guides to the eyes. The approximate doping of the samples is *white* $2.7 \times 10^{19}\,\mathrm{cm}^{-3}$ As, *grey* $2.2 \times 10^{18}\,\mathrm{cm}^{-3}$ As, *black* $1.1 \times 10^{18}\,\mathrm{cm}^{-3}$ As and $1.0 \times 10^{18}\,\mathrm{cm}^{-3}$ B with $N_D - N_D = 1.25 \times 10^{17}\,\mathrm{cm}^{-3}$ at room temperature. Adapted from [721]. (**b**) Thermopower of doped n-type silicon at room temperature as a function of doping concentration. Experimental data (*symbols*) from [721] and theory (*solid line*) from [723]

decrease with increasing doping is mostly attributed to the reduced mobility due to ionized impurity scattering. The increase of thermopower at low temperatures is due to the *phonon-drag* effect which is discussed for the samples from [721] in [724].

As a figure of merit for the production of thermoelectric power the ZT-value is used, $ZT = \sigma S^2 T/\kappa$.

8.13.2 Peltier Effect

In a semiconductor with a temperature difference at its ends a current flow will be allowed now (short circuit). The current leads via the charge transport also to a heat (or energy) transport. This effect is called the Peltier effect. The Peltier coefficient is negative (positive) for electrons (holes). The total amount of energy P that is transported consists of the generation term and the loss due to transport:

$$P = \mathbf{j} \cdot \hat{\mathbf{E}} - \nabla \cdot \mathbf{q}. \tag{8.77}$$

With (8.71) and (8.72) we find

$$P = \frac{\mathbf{j} \cdot \mathbf{j}}{\sigma} + S\,\mathbf{j} \cdot \nabla T - \Pi\,\nabla \cdot \mathbf{q} + \kappa\,\Delta T. \tag{8.78}$$

The first term is Joule heating, the second term is Thomson heating. The third exists only when carriers are generated or when they recombine. The fourth term is the heat conduction. In the Thomson term $S\,\mathbf{j} \cdot \nabla T$ heat is generated in an n-type semiconductor if $\mathbf{j}$ and ∇T are in the same direction. This means that electrons that move from the hotter to the colder part transfer energy to the lattice. The effect can be used to construct a thermoelectric cooler, as shown in Fig. 8.27, that generates a temperature difference due to a current flow. For optimal performance σ should be large to prevent excess Joule heating and κ should be small such that the generated temperature difference is not rapidly equalized.

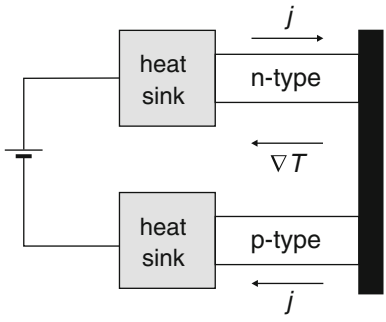


Fig. 8.27 Schematic Peltier cooler. The heat sinks (*grey*) and the cold junction (*black*) on the left are metals that make ohmic contacts with the semiconductors. The current flow is such that electrons move through the n-type semiconductor from right to left

Chapter 9
Optical Properties

Abstract After introduction of the complex dielectric function, reflection, diffraction are briefly discussed. The focus lies on absorption mechanisms; several transition types (direct and indirect band-band transitions, impurity-related transitions, lattice absorption) are discussed including the effects of excitons, polaritons and high carrier density. Also the various effects of the presence of free carriers are given.

9.1 Spectral Regions and Overview

The interaction of semiconductors with light is of decisive importance for photonic and optoelectronic devices as well as for the characterization of semiconductor properties. When light hits a semiconductor, reflection, transmission and absorption are considered, as for any dielectric material. The response of the semiconductor largely depends on the photon energy (or wavelength) of the light and various processes contribute to the dielectric function.

An overview of the electromagnetic spectrum in the optical range is given in Table 9.1. The energy and wavelength of a photon are related by[1] $E = h\nu = hc/\lambda$, i.e.

$$E\ [\mathrm{eV}] = \frac{1240}{\lambda\ [\mathrm{nm}]}. \tag{9.1}$$

In the infrared regime, energy is often measured in wave numbers (cm^{-1}) for which the conversion $1\,\mathrm{meV} = 8.056\,\mathrm{cm}^{-1}$ holds.

[1] The more exact numerical value in (9.1) is 1239.84.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_9

Table 9.1 Spectral ranges with relevance to semiconductor optical properties

Range		Wavelengths	Energies
Deep ultraviolet	DUV	<250 nm	>5 eV
Ultraviolet	UV	250–400 nm	3–5 eV
Visible	VIS	400–800 nm	1.6–3 eV
Near infrared	NIR	800 nm–2 μm	0.6–1.6 eV
Mid-infrared	MIR	2–20 μm	60 meV–0.6 eV
Far infrared	FIR	20–80 μm	1.6–60 meV
THz region	THz	>80 μm	<1.6 meV

9.2 Complex Dielectric Function

The dielectric function ϵ fulfills

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} = \epsilon_0 \epsilon \mathbf{E}, \tag{9.2}$$

and is generally a rank 2 tensor since **D** and **E** must not be collinear. This is the case, e.g. for birefringence as discussed in Sect. 13.2.2. The general form of the dielectric tensor for various crystal symmetries is compiled in Table 9.2. In most cases in the following ϵ will be used as scalar. The dielectric function is frequency dependent $\epsilon(\omega)$ due to the various oscillators playing a role and decreases (non-monotonically) from its static value (for $\omega = 0$) to 1 for $\omega \to \infty$. In some cases also its **k**-dependence is important, known as spatial dispersion (cmp. Sect. 9.6.8). The dielectric function is generally complex and written as

$$\epsilon = \epsilon' + \mathrm{i}\,\epsilon''. \tag{9.3}$$

The real and the imaginary part of the dielectric function are related to each other via the Kramers–Kronig relations (Appendix C).

The complex index of refraction is

$$n^* = n_\mathrm{r} + \mathrm{i}\,\kappa = \sqrt{\epsilon}. \tag{9.4}$$

From n^{*2} follows

$$\epsilon' = n_\mathrm{r}^2 - \kappa^2 \tag{9.5}$$

$$\epsilon'' = 2\, n_\mathrm{r}\, \kappa. \tag{9.6}$$

From $\epsilon\,\bar{\epsilon} = (n_\mathrm{r}^2 + \kappa^2)^2$ and (9.5) follows

Table 9.2 General form of the dielectric tensor for various crystals

crystal system	optical symmetry		examples
cubic	isotropic	$\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix}$	Si, GaAs, MgO, ZnSe, CuI
tetragonal hexagonal trigonal	uniaxial	$\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & c \end{pmatrix}$	$CuGaSe_2$, GaN, Bi_2Se_3
orthorhombic	biaxial	$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$	Sb_2Se_3
monoclinic	biaxial	$\begin{pmatrix} a & d & 0 \\ d & b & 0 \\ 0 & 0 & c \end{pmatrix}$	β-Ga_2O_3, anthracene
triclinic	biaxial	$\begin{pmatrix} a & d & e \\ d & b & f \\ e & f & c \end{pmatrix}$	tetracene

$$n_\mathrm{r}^2 = \frac{\epsilon' + \sqrt{\epsilon'^2 + \epsilon''^2}}{2} \tag{9.7}$$

$$\kappa = \frac{\epsilon''}{2\, n_\mathrm{r}}. \tag{9.8}$$

The real part of the complex index of refraction n_r is responsible for the dispersion, the imaginary part κ is named extinction coefficient and is related to the absorption coefficient of the plane wave (damping of $\mathbf{E}^2$) by

$$\alpha = 2\,\frac{\omega}{c}\,\kappa = \frac{4\pi}{\lambda}\,\kappa = 2\,k\,\kappa. \tag{9.9}$$

Here, k and λ denote the respective values in vacuum.

9.3 Reflection and Diffraction

From Maxwell's equations and the boundary conditions at a planar interface between two media with different index of refraction for the components of the electric and magnetic fields the laws for reflection and diffraction are derived. We denote the index of refraction as n and also n_r in the following. The interface between two media with refractive indices n_1 and n_2 is depicted in Fig. 9.1. In the following we assume first that no absorption occurs.

Snellius' law for the angle of diffraction is

$$n_1 \sin\phi = n_2 \sin\psi. \tag{9.10}$$

When the wave enters the denser medium, it is diffracted towards the normal. If the wave propagates into the less-dense medium (reversely to the situation shown in Fig. 9.1), a diffracted wave occurs only up to a critical angle of incidence

$$\sin\phi_{\mathrm{TR}} = \frac{n_2}{n_1}. \tag{9.11}$$

For larger angles of incidence, total internal reflection occurs and the wave remains in the denser medium. Thus, the angle in (9.11) is called the critical angle for total reflection. For GaAs and air the critical angle is rather small, $\phi_{\mathrm{TR}} = 17.4°$.

The reflectance depends on the polarization (Fresnel formulas). The index 'p' ('s') denotes parallel polarized/TM (perpendicular polarized/TE) waves.

$$R_{\mathrm{p}} = \left(\frac{\tan(\phi-\psi)}{\tan(\phi+\psi)}\right)^2 \tag{9.12}$$

$$R_{\mathrm{s}} = \left(\frac{\sin(\phi-\psi)}{\sin(\phi+\psi)}\right)^2. \tag{9.13}$$

The situation for GaAs and air is shown for both polarization directions and unpolarized radiation in Fig. 9.2 for a wave going into and out of the GaAs.

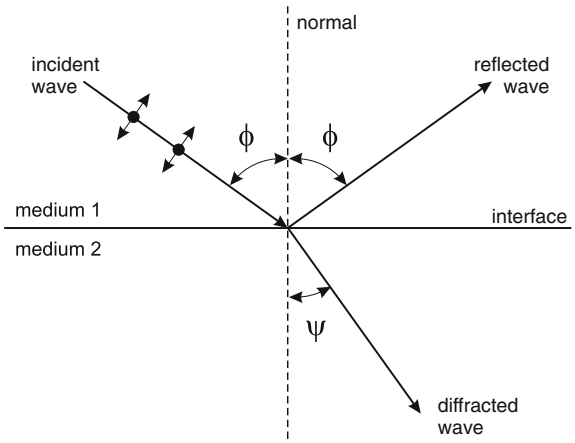


Fig. 9.1 Reflection and diffraction of an electromagnetic wave at the transition between medium '1' and '2', $n_2 > n_1$. The polarization plane is defined by the surface normal and the k-vector of the light (plane of incidence). The parallel ('p') polarized wave (TM-wave, electric field vector oscillates in the plane) is shown as '↔'; perpendicular ('s') polarization (TE-wave, electric field vector is perpendicular to plane) is depicted as '·'

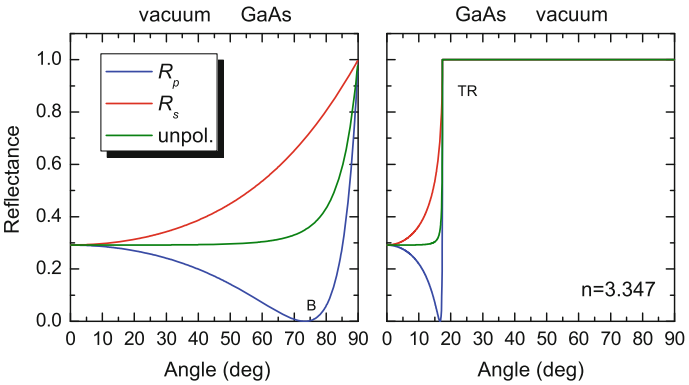


Fig. 9.2 Reflectance of the GaAs/vacuum interface (close to the band gap, $n_\mathrm{r} = 3.347$) for radiation from vacuum/air (*left panel*) and from the GaAs (*right panel*), respectively, as a function of incidence angle and polarization

When the reflected and the diffracted wave are perpendicular to each other, the reflectance of the p-polarized wave is zero. This angle is the Brewster angle ϕ_B,

$$\tan \phi_\mathrm{B} = \frac{n_2}{n_1}. \tag{9.14}$$

If a wave has vertical incidence from vacuum on a medium with index of refraction n_r, the reflectance is given (both polarizations are degenerate) as

$$R = \left(\frac{n_\mathrm{r} - 1}{n_\mathrm{r} + 1}\right)^2. \tag{9.15}$$

For GaAs, the reflectance for vertical incidence is 29.2 %.

9.4 Absorption

In the absorption process, energy is transferred from the electromagnetic field to the semiconductor. In the case of a linear absorption process, when the probability of light absorption is proportional to the incoming intensity, the decrease of intensity in the absorbing medium is exponential (Lambert–Beer's law [725, 726]),[2]

$$I(x) = I(0)\, \exp(-\alpha\, x). \tag{9.16}$$

[2] In [726], the absorption coefficient μ was defined via $I(d)/I(0) = \mu^d$, i.e. $\mu = \exp -\alpha$.

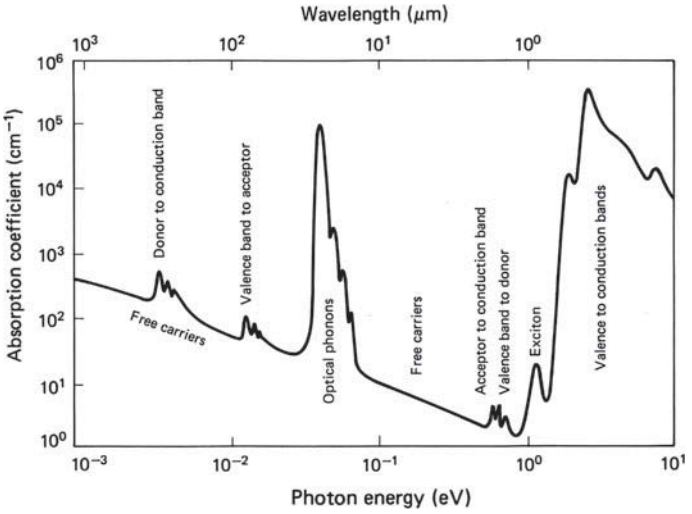


Fig. 9.3 Schematic absorption spectrum of a typical semiconductor. From [727]

The quantity α is the *absorption coefficient*, its reverse the absorption depth. The spectral dependence $\alpha(E)$, the absorption spectrum, contains the information of the possible absorption processes, their energy, momentum and angular momentum selection rules, and their oscillator strength.

In Fig. 9.3 a schematic absorption spectrum of a semiconductor is depicted. The transition of electrons from the valence to the conduction band begins at the band gap energy. The band gaps of Si, Ge, GaAs, InP, InAs, InSb are in the IR, those of AlAs, GaP, AlP, InN in the VIS, those of GaN and ZnO in the UV, MgO and AlN are in the deep UV. The Coulomb correlation of electrons and holes leads to the formation of excitons that leads to absorption below the band gap. The typical exction binding energy is in the range of 1–100 meV (see Fig. 9.16). Optical transitions from valence-band electrons into donors and from electrons on acceptors into the conduction band lead to band–impurity absorption. In the region from 10–100 meV the interaction with lattice vibrations (phonons) leads to absorption if the phonons are infrared active. Further in the FIR lie transitions from impurities to the closest band edge (donor to conduction and acceptor to valence band). A continuous background is due to free-carrier absorption.

If absorption is considered, the reflectance (9.15) needs to be modified. Using the complex index of refraction $n^* = n_\mathrm{r} + \mathrm{i}\kappa$, it is given as

$$R = \left|\frac{n^* - 1}{n^* + 1}\right|^2 = \frac{(n_\mathrm{r} - 1)^2 + \kappa^2}{(n_\mathrm{r} + 1)^2 + \kappa^2}. \tag{9.17}$$

9.5 Electron–Photon Interaction

The absorption process is quantum mechanically described by the coupling of electrons and photons. The process is described with time-dependent perturbation theory. If $\mathbf{H}_{\mathrm{em}}$ is the perturbation operator (electromagnetic field), the transition probability per time w_{fi} for electrons from (unperturbed) state 'i' (initial) to state 'f' (final) is given (with certain approximations) by Fermi's golden rule

$$w_{\mathrm{fi}}(\hbar\omega) = \frac{2\pi}{\hbar} \left|H'_{\mathrm{fi}}\right|^2 \delta(E_{\mathrm{f}} - E_{\mathrm{i}} - \hbar\omega), \tag{9.18}$$

where $\hbar\omega$ is the photon energy, E_{i} (E_{f}) is the energy of the initial (final) state. H'_{fi} is the matrix element

$$H'_{\mathrm{fi}} = \langle \Psi_{\mathrm{f}} \left|\mathbf{H}'\right| \Psi_{\mathrm{i}} \rangle, \tag{9.19}$$

where Ψ_{i} (Ψ_{f}) are the wavefunctions of the unperturbed initial (final) state.

$\mathbf{A}$ is the vector potential for the electromagnetic field, i.e. $\mathbf{E} = -\dot{\mathbf{A}}$, $\mu\mathbf{H} = \nabla \times \mathbf{A}$, $\nabla \cdot \mathbf{A} = 0$ (Coulomb gauge). The Hamiltonian of an electron in the electromagnetic field is

$$\mathbf{H} = \frac{1}{2m} \left(\hbar\mathbf{k} - q\mathbf{A}\right)^2 . \tag{9.20}$$

When terms in $\mathbf{A}^2$ are neglected (i.e. two-photon processes), the perturbation Hamiltonian is thus

$$\mathbf{H}_{\mathrm{em}} = -\frac{q}{m}\,\mathbf{A}\,\mathbf{p} = \frac{\mathrm{i}\,\hbar q}{m}\,\mathbf{A} \cdot \nabla \approx q\,\mathbf{r} \cdot \mathbf{E}. \tag{9.21}$$

The latter approximation is valid for small wavevectors of the electromagnetic wave and is termed the *electric dipole approximation*.

In order to calculate the dielectric function of the semiconductor from its band structure we assume that $\mathbf{A}$ is weak and we can apply (9.18). The transition probability R for the photon absorption rate at photon energy $\hbar\omega$ is then given by[3]

$$R(\hbar\omega) = \frac{2\pi}{\hbar} \int_{\mathbf{k}_{\mathrm{c}}} \int_{\mathbf{k}_{\mathrm{v}}} |\langle c|\mathbf{H}_{\mathrm{em}}|v\rangle|^2\, \delta\left(E_{\mathrm{c}}(\mathbf{k}_{\mathrm{c}}) - E_{\mathrm{v}}(\mathbf{k}_{\mathrm{v}}) - \hbar\omega\right) \mathrm{d}^3\mathbf{k}_{\mathrm{c}}\, \mathrm{d}^3\mathbf{k}_{\mathrm{v}}, \tag{9.22}$$

with the Bloch functions $|c\rangle$ and $|v\rangle$ of the conduction and valence band, respectively, as given in (6.36b).

The vector potential is written as $\mathbf{A} = A\hat{\mathbf{e}}$ with a unit vector $\hat{\mathbf{e}}$ parallel to $\mathbf{A}$. The amplitude is connected to the electric-field amplitude E via

[3] Here we assume that the valence-band states are filled and the conduction-band states are empty. If the conduction-band states are filled and the valence-band states are empty, the rate is that of stimulated emission.

$$A = -\frac{E}{2\,\omega}\left[\exp\left(\mathrm{i}(\mathbf{qr} - \omega t)\right) + \exp\left(-\mathrm{i}(\mathbf{qr} - \omega t)\right)\right]. \tag{9.23}$$

In the electric-dipole approximation the momentum conservation $\mathbf{q} + \mathbf{k}_\mathrm{v} = \mathbf{k}_\mathrm{c}$, $\mathbf{q}$ being the momentum of the light wave is approximated by $\mathbf{k}_\mathrm{v} = \mathbf{k}_\mathrm{c}$. The matrix element is then given by

$$|\langle c|\mathbf{H}_\mathrm{em}|v\rangle|^2 = \frac{e^2\,|A|^2}{m^2}\,\left|\langle c|\hat{\mathbf{e}}\cdot\mathbf{p}|v\rangle\right|^2, \tag{9.24}$$

with

$$\langle c\,|\hat{\mathbf{e}}\cdot\mathbf{p}|v\rangle\big|^2 = \frac{1}{3}\,|\mathbf{p}_\mathrm{cv}|^2 = M_\mathrm{b}^2, \tag{9.25}$$

and the momentum matrix element $\mathbf{p}_\mathrm{cv}$ given in (6.35). A $\mathbf{k}$-independent matrix element $|\mathbf{p}_\mathrm{cv}|^2$ is often used as an approximation. In Fig. 9.4 the matrix elements for valence to conduction band transitions in GaN are shown as a function of $\mathbf{k}$.

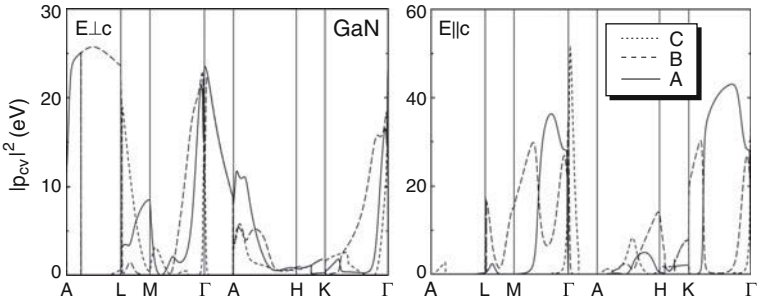


Fig. 9.4 Theoretical momentum matrix elements $|p_\mathrm{cv}|^2$ along high-symmetry directions in the Brillouin zone (see Fig. 3.34d) for transitions between valence and conduction bands in GaN and light polarized perpendicular (*left panel*) and parallel (*right panel*) to the c-axis. The transitions are A: $\Gamma_9(\mathrm{A}) \rightarrow \Gamma_{7\mathrm{c}}$, B: $\Gamma_7(\mathrm{B}) \rightarrow \Gamma_{7\mathrm{c}}$, C: $\Gamma_7(\mathrm{C}) \rightarrow \Gamma_{7\mathrm{c}}$ (see Fig. 6.38). Adapted from [728]

In terms of the electric-field amplitude $E(\omega)$ the transition probability is

$$R(\hbar\omega) = \frac{2\pi}{\hbar}\left(\frac{e}{m\,\omega}\right)^2\left|\frac{E(\omega)}{2}\right|^2 |\mathbf{p}_\mathrm{cv}|^2 \int_\mathbf{k} \delta\left(E_\mathrm{c}(\mathbf{k}) - E_\mathrm{v}(\mathbf{k}) - \hbar\omega\right)\,\mathrm{d}^3\mathbf{k}. \tag{9.26}$$

If the integration over $\mathbf{k}$ is restricted to those values allowed in unit volume, the power that is lost from the field in unit volume is given by $R\,\hbar\omega$, leaving a $1/E$ factor. The dielectric function $\epsilon = \epsilon_\mathrm{r} + \mathrm{i}\epsilon_\mathrm{i}$ is given by

$$\epsilon_\mathrm{i} = \frac{1}{4\pi\epsilon_0}\left(\frac{2\pi\, e}{m\,\omega}\right)^2 |\mathbf{p}_\mathrm{cv}|^2 \int_\mathbf{k} \delta\left(E_\mathrm{c}(\mathbf{k}) - E_\mathrm{v}(\mathbf{k}) - \hbar\omega\right)\, \mathrm{d}^3\mathbf{k} \tag{9.27a}$$

$$\epsilon_\mathrm{r} = 1 + \int_\mathbf{k} \frac{e^2}{\epsilon_0\, m\, \omega_\mathrm{cv}^2}\, \frac{2\,|\mathbf{p}_\mathrm{cv}|^2}{m\,\hbar\omega_\mathrm{cv}}\, \frac{1}{1-\omega^2/\omega_\mathrm{cv}^2}\, \mathrm{d}^3\mathbf{k}, \tag{9.27b}$$

with $\hbar\omega_\mathrm{cv} = E_\mathrm{c}(\mathbf{k}) - E_\mathrm{v}(\mathbf{k})$. The second equation has been obtained via the Kramers–Kronig relations[4] (see Appendix C).

Comparison with (D.7) yields that the oscillator strength of the band–band absorption is given by

$$f = \frac{e^2}{\epsilon_0\, m\, \omega_\mathrm{cv}^2}\, \frac{2\,|\mathbf{p}_\mathrm{cv}|^2}{m\,\hbar\omega_\mathrm{cv}}, \tag{9.28}$$

with

$$N_\mathrm{cv} = \frac{2\,|\mathbf{p}_\mathrm{cv}|^2}{m\,\hbar\omega_\mathrm{cv}} \tag{9.29}$$

being the classical 'number' of oscillators with the frequency ω_cv.

9.6 Band–Band Transitions

9.6.1 Joint Density of States

The strength of an allowed optical transitions between valence and conduction bands is proportional to the joint density of states (JDOS) $D_\mathrm{j}(E_\mathrm{cv})$ (cf. (6.59), (6.60) and (9.27a))

$$D_\mathrm{j}(E_\mathrm{cv}) = 2\int_{S(\tilde{E})} \frac{\mathrm{d}^2 S}{(2\pi/L)^3}\, \frac{1}{|\nabla_\mathbf{k} E_\mathrm{cv}|}, \tag{9.30}$$

where E_cv is an abbreviation for $E_\mathrm{c}(\mathbf{k}) - E_\mathrm{v}(\mathbf{k})$ and $\mathrm{d}^2 S$ is a surface element of the constant energy surface with $\tilde{E} = E_\mathrm{cv}$. The spin is assumed to generate doubly degenerate bands and accounts for the pre-factor 2. Singularities of the JDOS (*van-Hove singularities* or *critical points*) appear where $\nabla_\mathbf{k} E_\mathrm{cv}$ vanishes. This occurs when the gradient for both bands is zero or when both bands are parallel. The latter generates particularly large JDOS because the condition is valid at many points in $\mathbf{k}$-space.

[4] The real and imaginary parts of the dielectric function are generally related to each other via the Kramers–Kronig relations.

Table 9.3 Functional dependence of the joint density of states for critical points in 3, 2 and 1 dimensions

Dim.	Label	Type	D_j for $E < E_0$	D_j for $E > E_0$
3D	M_0	Min.	0	$\sqrt{E - E_0}$
	M_1	Saddle	$C - \sqrt{E_0 - E}$	C
	M_2	Saddle	C	$C - \sqrt{E - E_0}$
	M_3	Max.	$\sqrt{E_0 - E}$	0
2D	M_0	Min.	0	C
	M_1	Saddle	$-\ln(E_0 - E)$	$-\ln(E - E_0)$
	M_2	Max.	C	0
1D	M_0	Min.	0	$\sqrt{E - E_0}$
	M_1	Max.	$\sqrt{E_0 - E}$	0

E_0 denotes the energy (band separation) at the critical point, C stands for a constant value. The type of critical point is given (min.: minimum, saddle: saddle point, max.: maximum)

Generally, the (three-dimensional) energy dispersion $E(\mathbf{k})$ around a three-dimensional critical point (here developed at $\mathbf{k} = 0$) can be written as

$$E(\mathbf{k}) = E(0) + \frac{\hbar^2 k_x^2}{2m_x} + \frac{\hbar^2 k_y^2}{2m_y} + \frac{\hbar^2 k_z^2}{2m_z}. \tag{9.31}$$

The singularities are classified as M_0, M_1, M_2 and M_3 with the index being the number of masses m_i in (9.31) that are negative. M_0 (M_3) describes a minimum (maximum) of the band separation. M_1 and M_2 are saddle points. For a two-dimensional $\mathbf{k}$-space there exist M_0, M_1 and M_2 points (minimum, saddle point and maximum, respectively). For a one-dimensional $\mathbf{k}$-space there exist M_0 and M_1 points (minimum and maximum, respectively). The functional dependence of the JDOS at the critical points is summarized in Table 9.3. The resulting shape of the dielectric function is visualized in Fig. 9.5.

9.6.2 Direct Transitions

Transitions between states at the band edges at the Γ point are possible (Fig. 9.6). The $\mathbf{k}$ conservation requires (almost) vertical transitions in the $E(\mathbf{k})$ diagram because the length of the light $\mathbf{k}$ vector, $k = 2\pi/\lambda$, is much smaller than the size of the Brillouin zone $|k| \leq \pi/a_0$. The ratio of the lengths of the $\mathbf{k}$ vectors is of the order a_0/λ and typically about 10^{-3} for NIR wavelengths.

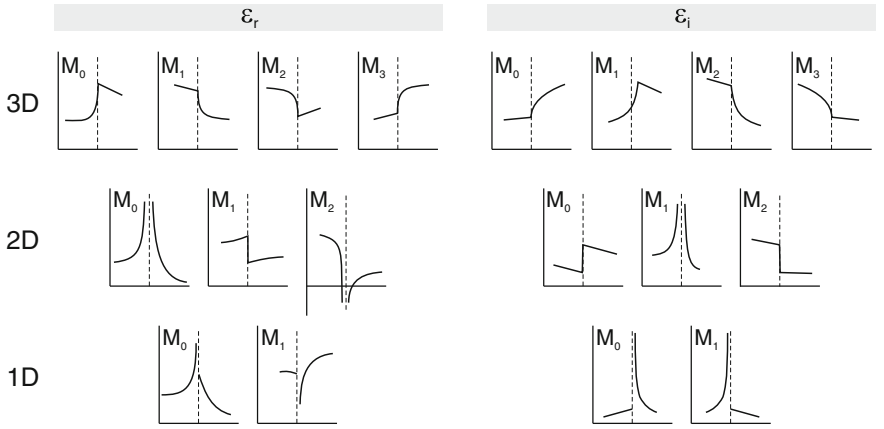


Fig. 9.5 Shape of the real (*left panel*) and imaginary (*right panel*) parts of the dielectric function in the vicinity of critical points in 3, 2 and 1 dimensions (for labels see Table 9.3). The *dashed line* in each graph indicates the energy position of the critical point E_0. Adapted from [729]

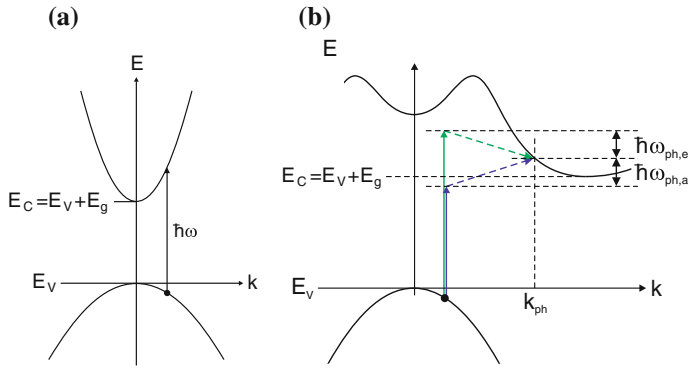


Fig. 9.6 (**a**) Direct optical transition and (**b**) indirect optical transitions between valence and conduction bands. The indirect transition involves a phonon with energy $\hbar\omega_{\mathrm{ph}}$ (index *a*: phonon absorption, *e*: phonon emission) and wavevector k_{ph}

For isotropic parabolic bands the band-band transition energy versus wavevector relation is

$$E_{\mathrm{cv}}(k) = E_{\mathrm{g}} + \frac{\hbar^2}{2}\left(\frac{1}{m_{\mathrm{e}}^*} + \frac{1}{m_{\mathrm{h}}^*}\right)k^2. \tag{9.32}$$

When the energy dependence of the matrix element is neglected, the absorption coefficient is determined by the corresponding square-root joint density of states (M_0 critical point):

$$\alpha(E) \propto \frac{\sqrt{E - E_{\mathrm{g}}}}{E} \approx \propto \sqrt{E - E_{\mathrm{g}}}. \tag{9.33}$$

The approximation is valid if the considered energy interval, e.g. around a band edge, is small.

Absorption spectra of $(In_xGa_{1-x})_2O_3$ alloy thin films at room temperature are shown in Fig. 9.7a. The α^2 versus photon energy plots shows a linear dependence with broadening and additional states at the band edge due to disorder effects. The extrapolation of the linear parts yields the absorption edge (Fig. 9.7b).

Absorption spectra of GaAs are shown in Fig. 9.8a for photon energies close to the band gap at various temperatures. The rapid increase, typical for direct semiconductors, is obvious. In particular at low temperatures, however, the absorption lineshape close to the band gap is dominated by an excitonic feature, discussed in Sect. 9.6.6.

Due to the increasing density of states, the absorption increases with the photon energy (Fig. 9.8c). At 1.85 eV there is a step in the absorption spectrum of GaAs due to the beginning of the contribution of transitions between the s-o hole band and the conduction band (see $E_0 + \Delta_0$ transition in Fig. 9.8e). When bands run in parallel, i.e. with the same separation, in the $E(\mathbf{k})$ diagram, the absorption processes accumulate at the same transition energy. In this way higher peaks in the absorption spectrum due to the E_1 or E_0' transitions originate as indicated in the band structure in Fig. 9.8e.

The selection rules for transitions from valence to conduction band must take into account the angular momentum and spin states of the wavefunctions. The optical transitions for circularly polarized light are shown in Fig. 9.9a, fulfilling the selection rule $\Delta m_j = \pm 1$. A lifting of the energetic degeneracies of these states occurs, e.g. by magnetic fields or spatial confinement (cmp. Fig. 12.29). For two-photon absorption (Sect. 9.6.14), the selection rule is $\Delta m_j = \pm 2$ as shown in Fig. 9.9b [732].

We note that in some materials the direct transition between certain bands is forbidden. An example is SnO_2 where the direct transition from the topmost valence

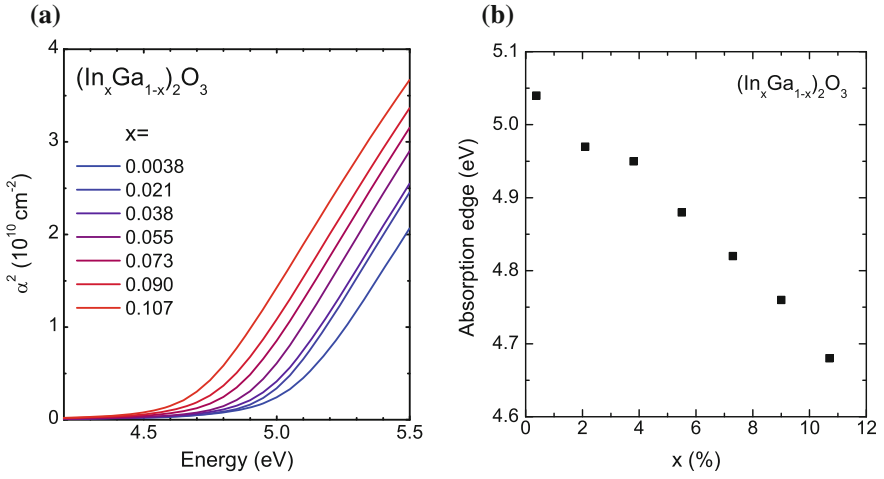


Fig. 9.7 (**a**) Absorption spectra of $(In_xGa_{1-x})_2O_3$ alloy thin films on Al_2O_3, plotted as α^2 versus photon energy. (**b**) Band edge determined from extrapolation of linear parts

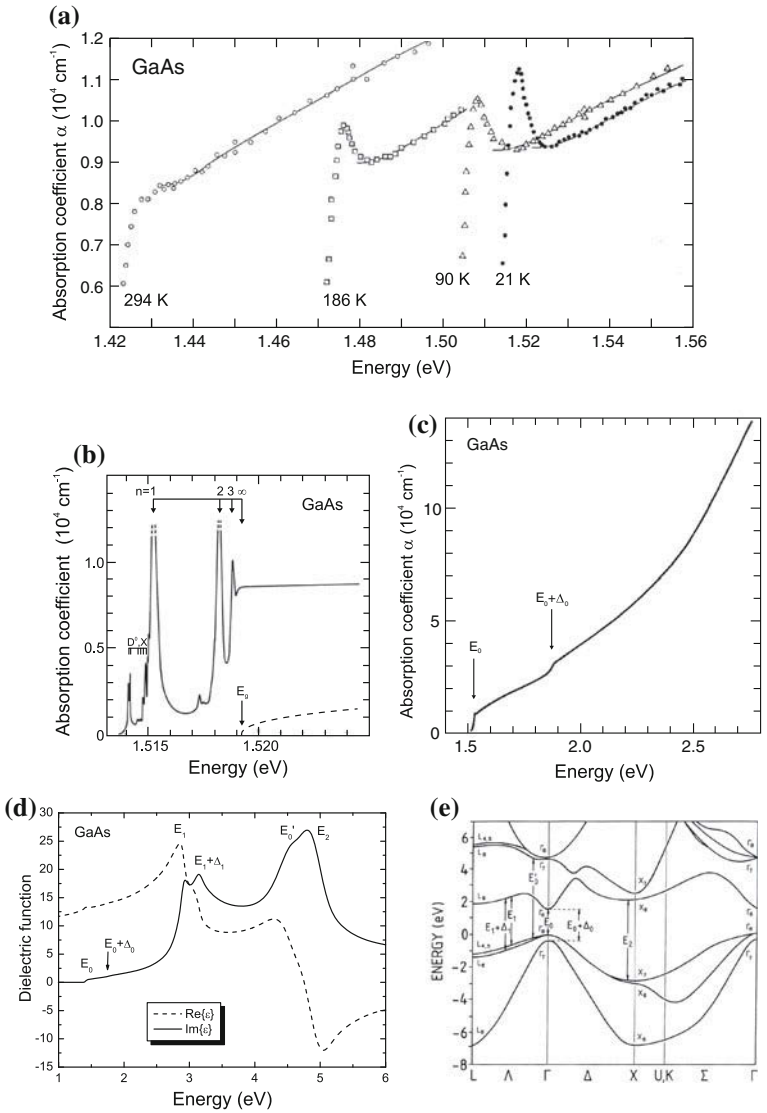


Fig. 9.8 (**a**) Absorption of GaAs close to the band gap at different temperatures. Adapted from [730]. (**b**) High-resolution absorption spectrum of highly pure GaAs at $T = 1.2\,\mathrm{K}$ in the exciton region. *Dashed line* is theory without excitonic correlation. Adapted from [731]. (**c**) Absorption spectrum of GaAs at $T = 21\,\mathrm{K}$ in the vicinity of the band gap. Adapted from [730]. (**d**) Complex dielectric function of GaAs at $T = 300\,\mathrm{K}$, *dashed* (*solid*) *line* real (imaginary) part of dielectric constant. Peak labels relate to transitions shown in part (**e**). (**e**) Band structure of GaAs with band gap transition (E_0) and higher transitions ($E_0 + \Delta_0$, E_1, $E_1 + \Delta_1$, E_0', and E_2) indicated

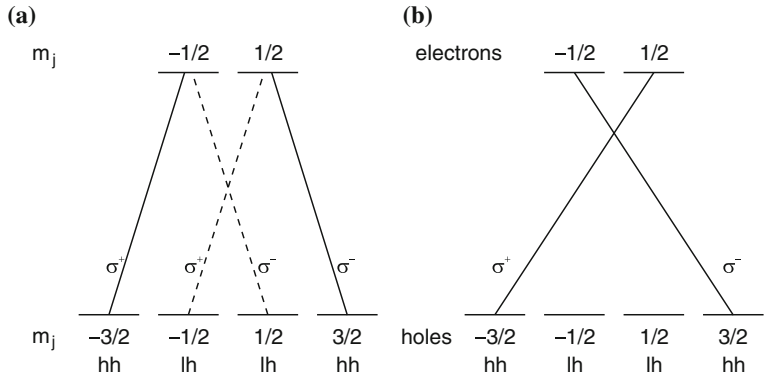


Fig. 9.9 Optical selection rules for band–band transitions in bulk material for (**a**) single photon transitions and (**b**) two-photon transitions (with photon energy equal to half the transition energy)

band into the lowest conduction band (at Γ) is forbidden (cmp. Fig. 9.46). If the matrix element increases linearly with $E - E_g$, the absorption coefficient varies like

$$\alpha(E) \propto (E - E_g)^{3/2}. \tag{9.34}$$

9.6.3 Indirect Transitions

In an indirect band structure the missing **k** difference (across the Brillouin zone) between valence- and conduction-band state needs to be provided by a second quantum. A phonon can provide the necessary momentum and additionally contributes a small amount of energy $\hbar\omega_{ph}$. There are several steps in the absorption spectrum due to various involved phonons (or combinations of them). At low temperature ($T = 1.6\,K$, Fig. 9.10) phonons can only be generated and the absorption starts at energies *above* the band gap. At higher temperatures (typically above 40 K [733], Fig. 9.10), acoustical phonons assisting the optical absorption transition can also be absorbed from the crystal; in this case due to energy conservation the absorption starts already at an energy $E_g - \hbar\omega_{ph}$ *below* the band gap. At even higher temperatures (>200 K, Fig. 9.10), also optical phonons can be absorbed.

The perturbation calculation yields an absorption coefficient with a quadratic dependence on energy (9.35a) [734]. Essentially, for the absorption into a specific (empty) conduction band state (with square-root density of states) various initial (filled) valence band states (also with square-root density of states) are possible, making the probability depend on the product of the DOS and thus on the energy to the first power. Integrating over all energy states with energy separation $E \pm \hbar\omega_{ph}$, yields an E^2-dependence.[5] Considering the temperature dependent population of

[5] A flat optical phonon dispersion is assumed.

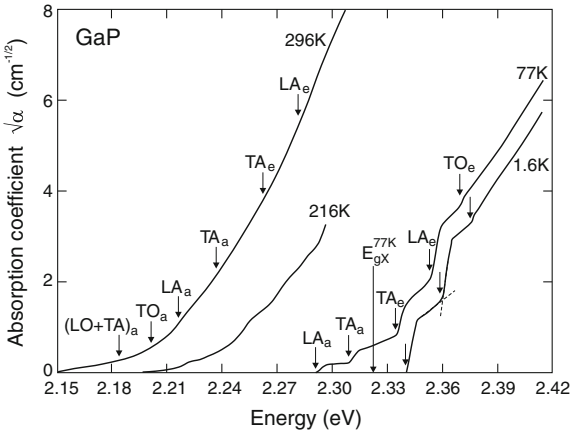


Fig. 9.10 Absorption edge of GaP ($\sqrt{\alpha}$ vs. E) at various temperatures. The index 'e' ('a') indicates phonon emission (absorption) during the optical absorption process. The theoretical excitonic gap (E_{gX}) at $T = 77\,\mathrm{K}$ is indicated. Adapted from [733]

the phonon density of states (Bose statistics, (E.3)) the absorption coefficients for transitions with phonon emission (α_e) and phonon absorption (α_a) are:

$$\alpha_e(E) \propto \frac{(E - (E_g + \hbar\omega_{ph}))^2}{1 - \exp(-\hbar\omega_{ph}/kT)} \tag{9.35a}$$

$$\alpha_a(E) \propto \frac{(E - (E_g - \hbar\omega_{ph}))^2}{\exp(\hbar\omega_{ph}/kT) - 1}. \tag{9.35b}$$

The two-particle process is less probable than the direct absorption that only involves one photon. The strength of indirect absorption close to the band gap is about 10^{-3} smaller than for the direct transition.

An 11-parameter formula based on terms like (9.35a) can describe the room temperature absorption spectrum of silicon in the visible with a precision of a few percent [735].

The absorption spectra close to the absorption edge are shown for GaP (Fig. 9.10) and Si (Fig. 9.11a). According to (9.35a), the plot of $\sqrt{\alpha}$ versus energy (Macfarlane–Roberts plot [736]) yields a straight line beyond the spectral region of phonon effects. The complicated form close to the (indirect) gap energy is due to the contribution of different phonons. The phonon energies found to contribute to the silicon absorption edge [737] agree with the TA and TO energy at the X minimum [738] (Fig. 9.11b). Also multiple phonons can contribute (Fig. 9.10). The momentum conservation can also be achieved by impurity scattering or electron-electron scattering [739].

We note also that the indirect semiconductors have an optical transition between Γ valence- and conduction-band states. However, this transition is at higher energies than the fundamental band gap, e.g. for Si ($E_g = 1.12\,\mathrm{eV}$) at 3.4 eV (see Fig. 6.6a). In

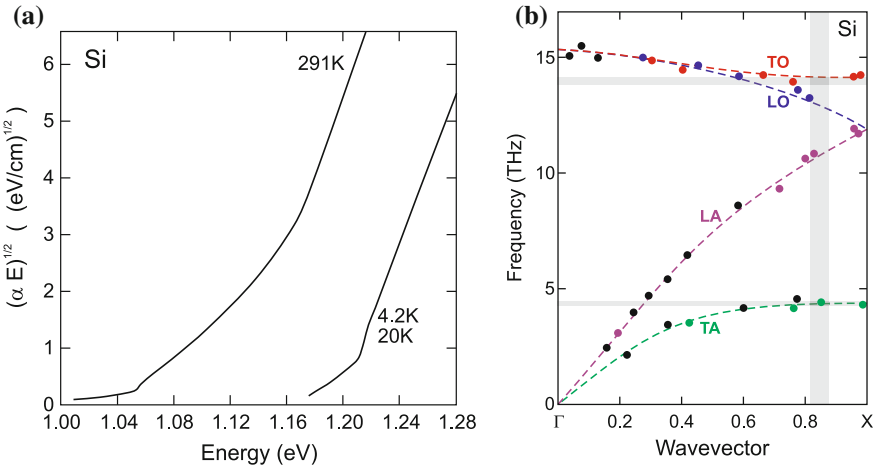


Fig. 9.11 (**a**) Absorption edge of Si at two different temperatures. Adapted from [737]. (**b**) Phonon energies in silicon along [001] obtained from neutron scattering (*black* unidentified, *green* TA, *purple* LA, *blue* LO, *red* TO). The *vertical grey bar* indicates the position of the conduction band minimum, the *horizontal grey bars* the energies of the phonons observed at the indirect optical absorption edge. The *dark grey* overlap areas indicate that TO and TA phonons contribute. Adapted from [738]

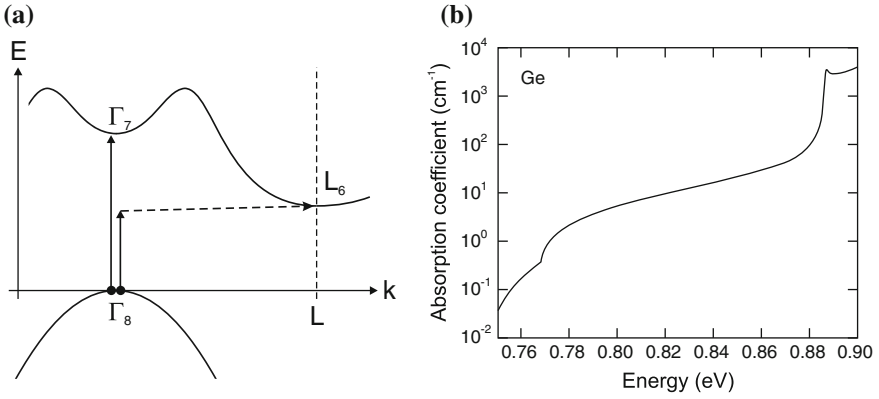


Fig. 9.12 (**a**) Scheme of indirect and direct optical transitions starting at the top of the valence band in Ge. *Vertical solid lines* represent the involved photon, the *horizontal dashed line* the involved phonon. (**b**) Experimental absorption spectrum of Ge ($T = 20\,\mathrm{K}$). Adapted from [737]

Fig. 9.12, the absorption scheme for indirect and direct absorbtion processes starting with an electron at the top of the valence band is shown together with an experimental absorption spectrum for Ge with the direct transition ($\Gamma_8 \rightarrow \Gamma_7$) at 0.89 eV, 0.136 eV above the fundamental band gap.

In Fig. 9.13, the absorption edge of $BaTiO_3$ is shown. An indirect transition with an increase of (weak) absorption $\propto E^2$ and an indirect gap of $E_\mathrm{i} = 2.66\,\mathrm{eV}$ and a

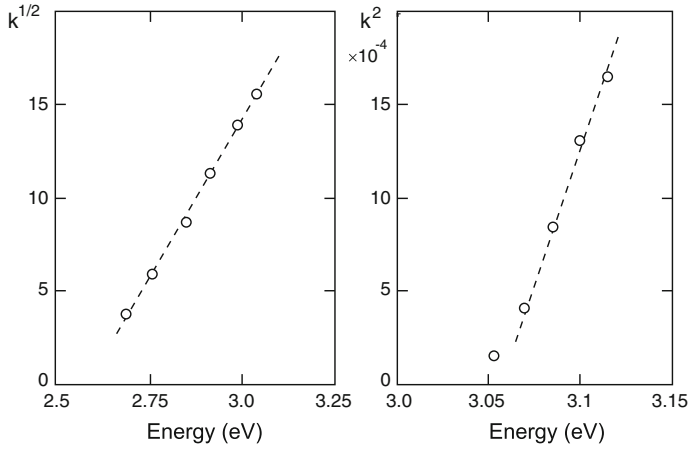


Fig. 9.13 Absorption of $BaTiO_3$ at room temperature. Experimental data (*circles*) from [740] with fits (*dashed lines*) $\propto E^2$ and $\propto E^{1/2}$, respectively

direct transition with an increase of (strong) absorption $\propto E^{1/2}$ and a direct gap of $E_\mathrm{d} = 3.05\,\mathrm{eV}$ are observed. These transitions could be due to holes at the M (indirect gap) and Γ (direct gap) points (cf. Sect. 6.3.11), respectively.

9.6.4 Urbach Tail

Instead of the ideal $(E - E_\mathrm{g})^{1/2}$ dependence of the direct band-edge absorption, often an exponential tail is observed (see Fig. 9.14). This tail is called the Urbach tail [741] and follows the functional dependence (for $E < E_\mathrm{g}$)

$$\alpha(E) \propto \exp\left(\frac{E - E_\mathrm{g}}{E_0}\right), \tag{9.36}$$

where E_0 is the characteristic width of the absorption edge, the so-called Urbach parameter.

The Urbach tail is attributed to transitions between band tails below the band edges. Such tails can originate from disorder of the perfect crystal, e.g. from defects or doping, and the fluctuation of electronic energy bands due to lattice vibrations. The temperature dependence of the Urbach parameter E_0 is thus related to that of the band gap as discussed in [742, 743].

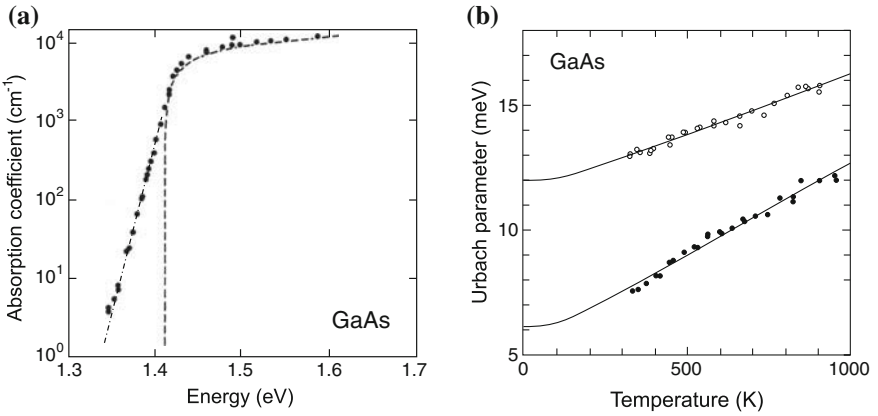


Fig. 9.14 (**a**) Experimental absorption spectrum (*circles*) of GaAs at room temperature on a semi-logarithmic plot. The exponential tail below the band gap is called the Urbach tail (the *dash-dotted* line corresponds to $E_0 = 10.3\,\text{meV}$ in (9.36)). The *dashed line* is the theoretical dependence from (9.33). Adapted from [744]. (**b**) Temperature dependence of Urbach parameter E_0 for two GaAs samples. Experimental data for undoped (*solid circles*) and Si-doped ($n = 2 \times 10^{18}\,\text{cm}^{-3}$, *empty circles*) GaAs and theoretical fits (*solid lines*) with one-phonon model. Adapted from [742]

9.6.5 Amorphous Semiconductors

The sharp features in the dielectric function due to critical points in the band structure of crystalline semiconductors are washed out in amorphous material. As an example the spectra of the imaginary part of the dielectric function for crystalline (trigonal) and amorphous selenium are shown in Fig. 9.15.

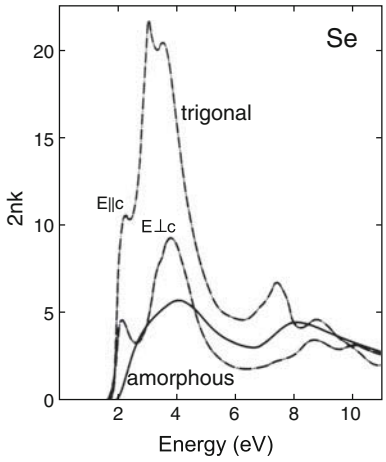


Fig. 9.15 Imaginary part of the dielectric function of amorphous (*solid line*) and crystalline (trigonal) selenium (*dash-dotted lines* for two different polarization directions). From [745]

9.6.6 Excitons

An electron in the conduction band and a hole in the valence band form a hydrogen-like state due to the mutual Coulomb interaction. Such a state is called an exciton. The center-of-mass motion is separated and has a dispersion $E = \frac{\hbar^2}{2M}\mathbf{K}^2$, where $M = m_\mathrm{e} + m_\mathrm{h}$ is the total mass and $\hbar\mathbf{K}$ is the center-of-mass momentum

$$\mathbf{K} = \mathbf{k}_\mathrm{e} + \mathbf{k}_\mathrm{h}. \tag{9.37}$$

The relative motion yields hydrogen-like quantized states $E_n \propto n^{-2}$ ($n \geq 1$):

$$E_\mathrm{X}^n = -\frac{m_\mathrm{r}^*}{m_0}\,\frac{1}{\epsilon_\mathrm{r}^2}\,\frac{m_0 e^4}{2(4\pi\epsilon_0\hbar)^2}\,\frac{1}{n^2}, \tag{9.38}$$

where m_r^* denotes the reduced effective mass $m_\mathrm{r}^{*-1} = m_\mathrm{e}^{*-1} + m_\mathrm{h}^{*-1}$. The third factor is the atomic Rydberg energy (13.6 eV). The exciton binding energy $E_\mathrm{X}^\mathrm{b} = -E_\mathrm{X}^1$ is scaled by $(m^*/m_0)\,\epsilon_\mathrm{r}^{-2} \approx 10^{-3}$. A more detailed theory of excitons beyond the simple hydrogen model presented here, taking into account the valence-band structure, can be found in [746] for direct and [747] for indirect cubic and in [748] for wurtzite semiconductors. The exciton binding energies for various semiconductors are listed in Table 9.4 and shown in Fig. 9.16a versus the band gap.

The radius of the exciton is

$$r_\mathrm{X}^n = n^2\,\frac{m_0}{m_\mathrm{r}^*}\,\epsilon_\mathrm{r}\,a_\mathrm{B}, \tag{9.39}$$

where $a_\mathrm{B} = 0.053\,\mathrm{nm}$ denotes the hydrogen Bohr radius.[6] The Bohr radius of the exciton is $a_\mathrm{X} = r_\mathrm{X}^1$ (14.6 nm for GaAs, ~2 nm for ZnO). The exciton moves with the center-of-mass $\mathbf{K}$-vector through the crystal. The complete dispersion is (see Fig. 9.16b)

Table 9.4 Exciton (E_X^b) and biexciton (E_XX^b, see Sect. 9.6.10) binding energies in various bulk semiconductors

Material	E_X^b (meV)	E_XX^b (meV)	$E_\mathrm{XX}^\mathrm{b}/E_\mathrm{X}^\mathrm{b}$
GaAs	4.2		
GaAs QW	9.2	2.0	0.22
ZnSe	17	3.5	0.21
GaN	25	5.6	0.22
CdS	27	5.4	0.20
ZnS	37	8.0	0.22
ZnO	59	15	0.25

Values for 10 nm GaAs/15 nm $Al_{0.3}Ga_{0.7}As$ quantum well (QW) are taken from [749]

[6] Cf. (7.20); an electron bound to a donor can be considered as an exciton with an infinite hole mass.

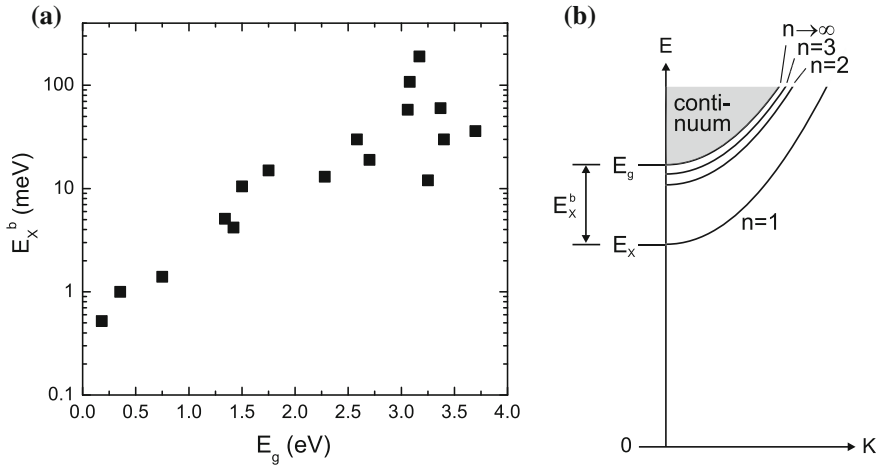


Fig. 9.16 (**a**) Exciton binding energy versus band gap for various semiconductors. (**b**) Schematic dispersion of excitonic levels. The K-vector refers to the center-of-mass motion

$$E = E_{\mathrm{g}} + E_{\mathrm{X}}^{n} + \frac{\hbar^2}{2M}\,\mathbf{K}^2. \tag{9.40}$$

The oscillator strength of the exciton states decays $\propto n^{-3}$. The absorption due to excitons is visible in Fig. 9.8a for GaAs at low temperatures. If inhomogeneities are present, typically only the $n = 1$ transition is seen. However, under special conditions also higher transitions of the exciton Rydberg series are seen (e.g. $n = 2$ and 3 in Fig. 9.8b).

The exciton concept was introduced first for absorption in Cu_2O [750]. The $J = 1/2$ absorption spectrum ('yellow series') is shown in Fig. 9.17. In this particular

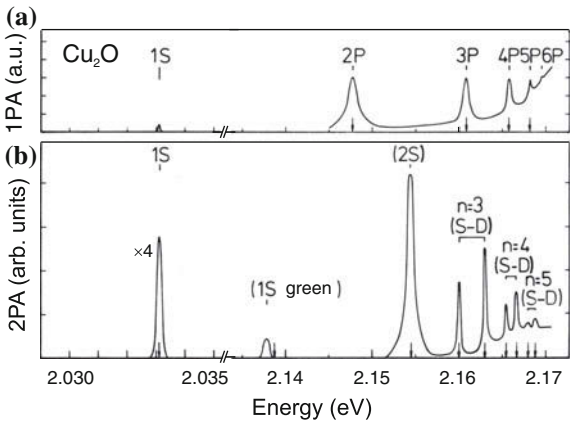


Fig. 9.17 One-photon (*top*) and two-photon (*bottom*) absorption spectra of Cu_2O at $T = 4.2$ K. *Arrows* denote theoretical peak positions. Adapted from [752]

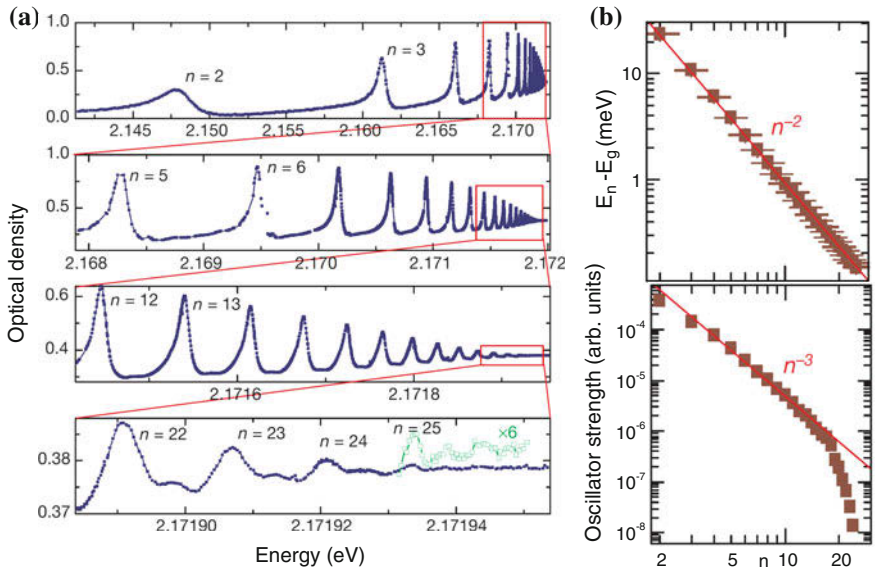


Fig. 9.18 (One photon) Absorption spectrum of Cu_2O (thickness 34 μm) at $T = 1.2$ K with transitions labelled $n = 2 \ldots 25$. Adapted from [751]

material both the valence and conduction bands have s character, thus the 1s transition of the exciton is forbidden and the np transitions are observed in normal (one-photon) absorption. With two-photon absorption also the s (and d) transitions can be excited. On a piece of natural Cu_2O, the Rydberg series has been measured up to $n = 25$ [751] (Fig. 9.18a). The peak energy and the oscillator strength follow the n^{-2} ($E_X^b = 92$ meV, $E_g = 2.17208$ eV) and n^{-3} laws, respectively, expected from a hydrogen model (Fig. 9.18b). The deviation from the n^{-3}-dependence for the oscillator strength at large n is due to interaction effects of excitons with large radius at finite exciton density.

The scattering (unbound) states of the exciton [753] for $E > E_g$ contribute to absorption above the band gap. The factor by which the absorption spectrum is changed is called the Sommerfeld factor. For bulk material it is

$$S(\eta) = \eta \frac{\exp(\eta)}{\sinh(\eta)}, \tag{9.41}$$

with $\eta = \pi \left[E_X^b/(E - E_g)\right]^{1/2}$. The change of the absorption spectrum due to the Coulomb correlation is shown in Fig. 9.19. There is a continuous absorption between the bound and unbound states. At the band gap there is a finite absorption ($S(E \to E_g) \to \infty$). The detail to which exciton peaks can be resolved depends on the spectral broadening.

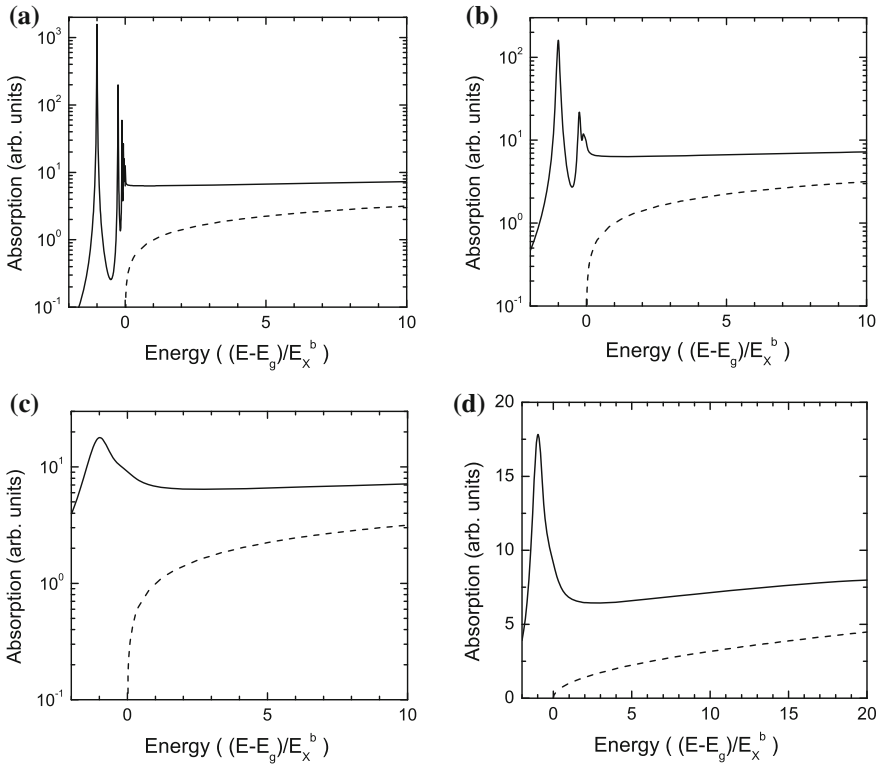


Fig. 9.19 Modification of the absorption edge of a direct transition by excitonic effects for different spectral (Lorentzian) broadening ($\propto (E^2 + \Gamma^2/4)^{-1}$), (**a**) $\Gamma = 0.01E_{\mathrm{X}}^{\mathrm{b}}$, (**b**) $\Gamma = 0.1E_{\mathrm{X}}^{\mathrm{b}}$, (**c**) $\Gamma = E_{\mathrm{X}}^{\mathrm{b}}$. (**d**) is (**c**) in linear scale. *Dashed lines* are electron–hole plasma absorption according to (9.33)

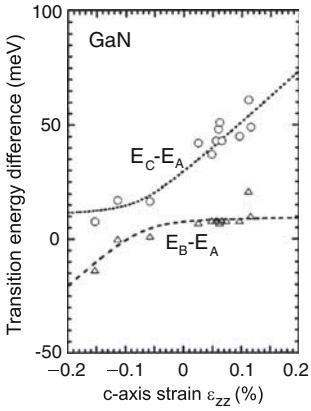


Fig. 9.20 Theoretical dependency (*lines*) for the the differences of the C-line and A-line as well as B-line and A-line exciton transition energies in GaN as a function of the *c*-axis strain. *Symbols* are experimental data from [754]. Adapted from [468]

Table 9.5 Phonon broadening parameters (FWHM) of various bulk semiconductors

Material	$\hbar\omega_{\mathrm{LO}}$ (meV)	Γ_0 (meV)	γ_{AC} (μeV/K)	γ_{LO} (meV)
GaAs	36.8	0	4 ± 2	16.8 ± 2
ZnSe	30.5	1.9	0 ± 7	84 ± 8
GaN	92	10	15 ± 4	408 ± 30
ZnO	33	1.2	32 ± 26	96 ± 24

Values from fits with (9.42) to experimental data for GaAs [759], ZnSe [757], GaN [756], ZnO [760] (phonon energy fitted) as shown in Fig. 9.22b

In Fig. 9.20 the energy separations of the A-, B-, and C-excitons in GaN are shown [468]. Thus, the ordering of the valence bands depends on the strain state of the semiconductor.

9.6.7 Phonon Broadening

The scattering with phonons and the related dephasing leads to homogeneous broadening Γ_{hom} of absorption (and recombination) lines. Acoustic and optical phonons contribute to the broadening according to the dependence [755]

$$\Gamma_{\mathrm{hom}}(T) = \Gamma_0 + \gamma_{\mathrm{AC}}\, T + \gamma_{\mathrm{LO}}\, \frac{1}{\exp(\hbar\omega_{\mathrm{LO}}/kT) - 1}, \tag{9.42}$$

where $\hbar\omega_{\mathrm{LO}}$ is the optical phonon energy and the last factor is the Bose function (E.24). Γ_0 is a temperature-independent contribution, $\Gamma_0 = \Gamma(T = 0)$. The increasing broadening with increasing temperature is obvious, e.g., in absorption spectra (Fig. 9.21a). In Fig. 9.21b experimental data for GaAs, ZnSe and GaN are assembled. The data have been fitted with (9.42); the resulting phonon broadening parameters are listed in Table 9.5.[7] The optical transitions in polar semiconductors exhibit stronger coupling to optical phonons. The phonon coupling parameters from different measurements on GaN are discussed and compared in [758].

9.6.8 Exciton Polariton

Electrons and holes are particles with spin 1/2. Thus, the exciton can form states with total spin $S = 0$ (para-exciton, singlet) and $S = 1$ (ortho-exciton, triplet). The

[7] Such parameter can be directly determined from spectroscopic broadening (as in [756]) or a time-resolved measurement of the decay of the coherent polarization (four-wave mixing) as in [757]. In the latter, the decay constant of the dephasing T_2 is related to the decay constant τ of the FWM-signal by $T_2 = 2\tau$ for homogeneous broadening. The Fourier transform of $\exp -t/(2\tau)$ is a Lorentzian of the type $\propto ((E - E_0)^2 + \Gamma^2/4)^{-1}$ with $\Gamma = 1/\tau$ being the FWHM.

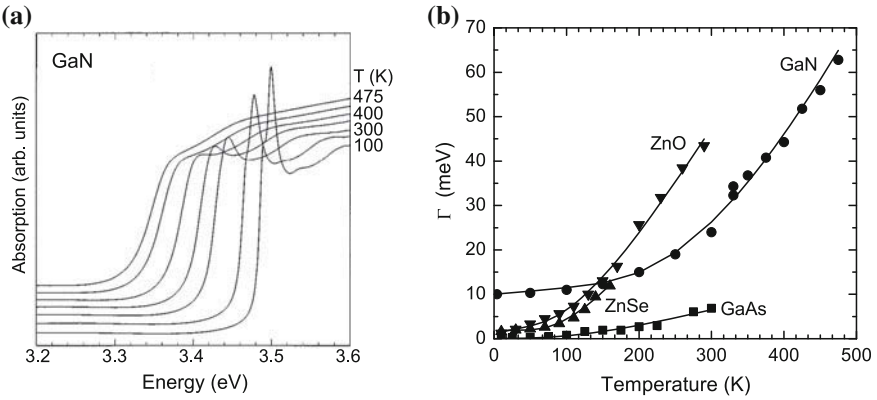


Fig. 9.21 (**a**) Absorption spectra of GaN bulk (0.38 μm thick epilayer on sapphire) for various temperatures $T = 100, 200, 300, 350, 400, 450$, and 475 K. Adapted from [756]. (**b**) Homogeneous broadening as a function of temperature, *symbols* are experimental data, *solid lines* are fits, rf. Table 9.5

exchange interaction leads to a splitting of these states, the singlet being the energetically higher. The singlet state splits into the longitudinal and transverse exciton with respect to the orientation of the polarization carried by the Bloch functions and the center-of-mass motion **K** of the exciton. Dipole transitions are only possible for singlet excitons (bright excitons). The triplet excitons couple only weakly to the electromagnetic field and are thus also called dark excitons.

The coupling of these states to the electromagnetic field creates new quasiparticles, the exciton polaritons [761, 762]. The dielectric function of the exciton (with background dielectric constant ϵ_b) is

$$\epsilon(\omega) = \epsilon_b \left[1 + \frac{\beta}{1 - (\omega^2/\omega_X)^2}\right] \cong \epsilon_b \left[1 + \frac{\beta}{1 - (\omega^2/\omega_T)^2 + \hbar\, K^2/(M\, \omega_T)}\right], \tag{9.43}$$

where β is the oscillator strength and the energy is $\hbar\omega_X = \hbar\omega_T + \hbar^2\, K^2/2M$. $\hbar\omega_T$ is the energy of the transverse exciton at $K = 0$. With this dispersion the wave dispersion must be fulfilled, i.e.

$$c^2\, k^2 = \omega^2\, \epsilon(\omega), \tag{9.44}$$

where k is the k-vector of the light that needs to be $k = K$ due to momentum conservation. The dependence of the dielectric function on k is called *spatial dispersion* [763]. Generally, up to terms in k^2 it is written as

$$\epsilon(\omega) = \epsilon_b \left[1 + \frac{\beta}{1 - (\omega^2/\omega_0)^2 + D\, k^2}\right]. \tag{9.45}$$

The term k^2 with curvature D (for the exciton polariton $D = \hbar/(M\,\omega_\mathrm{T})$) plays a role in particular when $\omega_\mathrm{T}^2 - \omega^2 = 0$. For $\mathbf{k} \neq 0$ even a cubic material is anisotropic. The dimensionless curvature $\hat{D} = Dk'^2$ should fulfill $\hat{D} = \hbar/(Mc) \ll 1$ in order to make k^4 terms unimportant. For exciton polaritons[8] typically $\hat{D} = \hbar\omega_\mathrm{T}/(m\,c^2) \approx 2 \times 10^{-5}$ for $\hbar\omega_\mathrm{T} = 1\,\mathrm{eV}$ and $m^* = 0.1$.

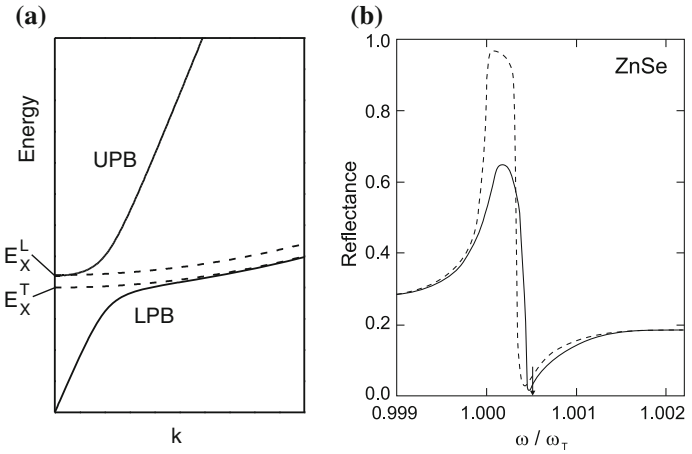


Fig. 9.22 (**a**) Schematic dispersion of exciton polaritons. The lower polariton branch ('LPB') is at small k photon-like, at large k exciton-like. The upper branch ('UPB') is exciton-like at small k and photon-like at larger k. The limit of the UPB for $k \to 0$ is the energy of the longitudinal exciton. The *dashed lines* represent the pure exciton dispersions. (**b**) Theoretical effect of spatial dispersion on the reflectance at the fundamental exciton resonance at normal incidence for ZnSe material parameters ($\hbar\omega_\mathrm{T} = 2.8\,\mathrm{eV}$, $\beta = 1.0 \times 10^{-3}$ and a background dielectric constant of $\epsilon_\mathrm{b} = 8.1$, damping was set to $\Gamma = 10^{-5}\omega_\mathrm{T}$). The *arrow* denotes the position of ω_L. The *solid* (*dashed*) *line* is with (without) spatial dispersion for $\hat{D} = 0.6 \times 10^{-5}$ ($\hat{D} = 0$). Data from [763]

From (9.44) together with (9.45) two solutions result:

$$2\omega^2 = c^2k^2 + (1 + \beta + Dk^2)\omega_0^2 \pm \left[-4c^2k^2(1 + Dk^2)\omega_0^2 + (c^2k^2 + (1 + \beta + Dk^2)\omega_0^2)^2\right]^{1/2}. \qquad (9.46)$$

The two branches are shown schematically in Fig. 9.22a. Depending on the k value they have a photonic (linear dispersion) or excitonic (quadratic dispersion) character. The anticrossing behavior at $k' \approx \omega_\mathrm{T}/c$ (for $\hbar\omega_\mathrm{T} = 1\,\mathrm{eV}$, $k' \approx 0.5 \times 10^{-5}\,\mathrm{cm}^{-1}$) creates a bottleneck region in the lower polariton branch. This name stems from the small emission rate of acoustic phonons (i.e. cooling) in that region, as predicted in [764] and experimentally found, e.g. in CdS [765]. The polaritons decay into a

[8] The dependence of the optical-phonon energies on k is typically too small to make spatial dispersion effects important. According to (5.19) $\hat{D} = -(a_0\omega_\mathrm{TO}/4c)^2 \approx 4 \times 10^{-11}$ for typical material parameters (lattice constant $a_0 = 0.5\,\mathrm{nm}$, TO phonon frequency $\omega_\mathrm{TO} = 15\,\mathrm{THz}$).

photon when they hit the surface. The effect of the oscillator strength of the dispersion is shown in Fig. 9.23 for two-exciton resonance. In the case of several excitons (9.45) reads

$$\epsilon(\omega) = \epsilon_\mathrm{b} \left[1 + \sum_{i=1}^{n} \frac{\beta_i}{1 - (\omega^2/\omega_{0,i})^2 + D_i\, k^2} \right]. \tag{9.47}$$

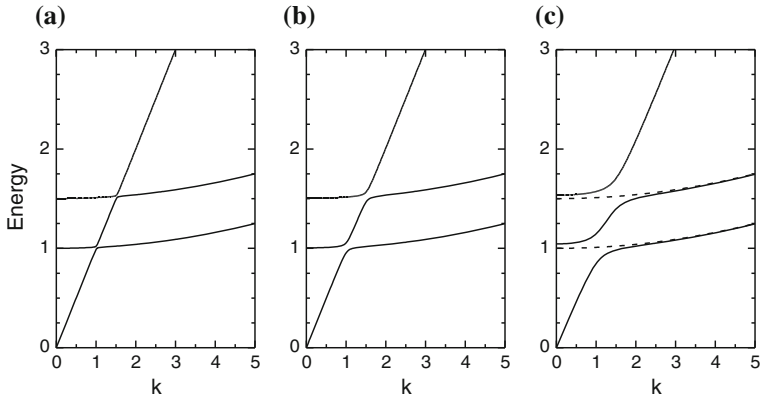


Fig. 9.23 Schematic polariton dispersion for a two-exciton resonance (curvature of exciton dispersion greatly exaggerated, $\hat{D} = 10^{-2}$) at $\omega_{\mathrm{T},1} = 1$ and $\omega_{\mathrm{T},2} = 1.5$ for three different oscillator strengths (**a**) $f = 10^{-3}$, (**b**) $f = 10^{-2}$, (**c**) $f = 10^{-1}$. The *dashed lines* in (**c**) represent the pure exciton dispersions

For $k = 0$ either $\omega = 0$ (lower polariton branch) or $\epsilon(\omega_\mathrm{L}) = 0$. For the latter we find from (9.45)

$$\omega_\mathrm{L} = \sqrt{1 + \beta}\, \omega_\mathrm{T}. \tag{9.48}$$

Therefore, the energy splitting ΔE_LT, mostly denoted as Δ_LT, between the L- and T-exciton energy given by

$$\Delta E_\mathrm{LT} = \hbar(\omega_\mathrm{L} - \omega_\mathrm{T}) = \left[\sqrt{1 + \beta} - 1 \right] \hbar\omega_\mathrm{T} \approx \frac{1}{2}\, \beta\, \hbar\omega_\mathrm{T} \tag{9.49}$$

is proportional to the exciton oscillator strength (for experimental values see Table 9.6). We note that if (D.9) is used for the dielectric function, β in (9.49) needs to be replaced by $\beta/\epsilon_\mathrm{b}$.

The effect of spatial dispersion on the reflection at the fundamental exciton resonance is depicted in Fig. 9.22b. For non-normal incidence an additional feature due to the longitudinal wave is observed for p-polarization [763]. For a detailed discussion additional effects due to anisotropy in wurtzite crystals, an exciton free layer at the semiconductor surface, additional boundary conditions and damping need to

Table 9.6 Exciton energy (low temperature), LT splitting and exciton polariton oscillator strength for various semiconductors

		CdS A	CdS B	ZnO A	ZnO B	ZnSe	GaN A	GaN B	GaAs
$\hbar\omega_T$	(eV)	2.5528	2.5681	3.3776	3.3856	2.8019	3.4771	3.4816	1.5153
Δ_{LT}	(meV)	2.2	1.4	1.45	5	1.45	1.06	0.94	0.08
β	(10^{-3})	1.7	1.1	0.9	3.0	1.0	0.6	0.5	0.11

Values for ZnO from [766], values for GaAs from [767], all other values from [768]

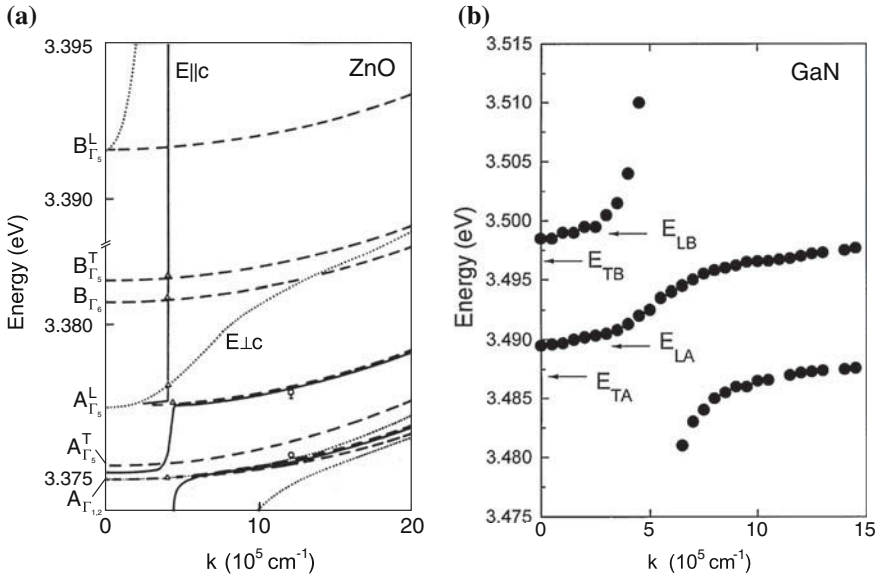


Fig. 9.24 (**a**) Exciton polariton dispersion ($\mathbf{k} \perp c$) of ZnO with experimental data ($T = 1.8\,\mathrm{K}$). *Solid* (*dotted*) *lines* are for polaritons with $\mathbf{E} \parallel c$ ($\mathbf{E} \perp c$). The *dashed lines* refer to excitons. Adapted from [771]. (**b**) Exciton polariton dispersion ($T = 2\,\mathrm{K}$) in GaN (on sapphire) for $\mathbf{E} \perp c$. Adapted from [772]

be considered [769, 770]. The polariton dispersions of ZnO and GaN are shown in Fig. 9.24.

9.6.9 Bound-Exciton Absorption

Excitons can localize at impurities or inhomogeneities. Such excitons are called *bound excitons*. Here, the absorption due to such complexes is discussed. The recombination is discussed in Sect. 10.3.2. In GaP:N excitons are bound to isoelectronic

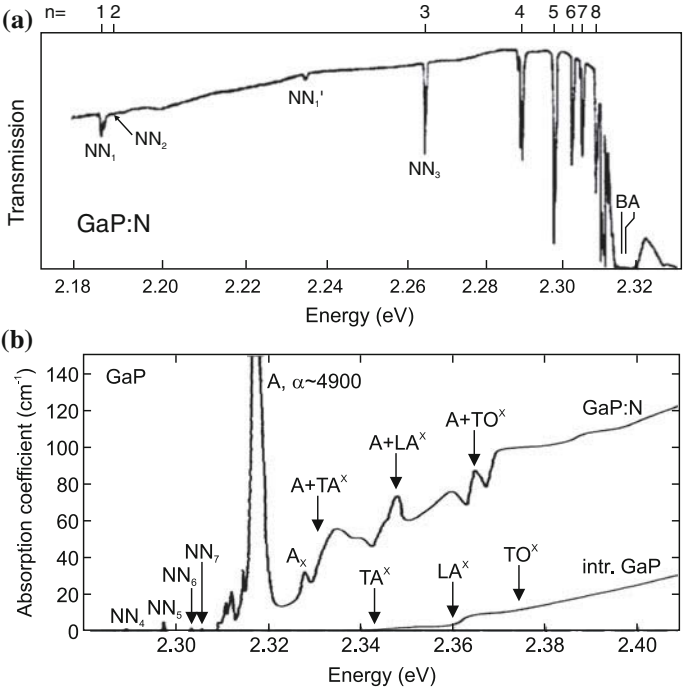


Fig. 9.25 (**a**) Transmission spectrum of GaP:N with a nitrogen concentration of about $10^{19}\,\mathrm{cm}^{-3}$ at 1.6 K (thickness: 1.1 mm). n is indicated for the first eight transitions due to excitons bound to nitrogen pairs. NN_n' indicate phonon replica. The 'A' line denotes the position of the transition due to excitons bound to a single nitrogen atom (observable for samples with low N doping). The 'B' line is forbidden and due to the $J = 2$ exciton. Adapted from [616]. (**b**) Absorption spectra of N-doped ($N_\mathrm{N} = 7 \times 10^{18}\,\mathrm{cm}^{-3}$) and intrinsic GaP ($T = 2\,\mathrm{K}$). Adapted from [612]

N impurities (substituting P), resulting in the 'A' line at 2.3171 eV (at $T = 4.2\,\mathrm{K}$).[9] The absorption due to A excitons is well resolved in the spectrum of Fig. 9.25b. At sufficiently high nitrogen doping, there exist nitrogen pairs, i.e. a complex where a nitrogen impurity has a second nitrogen impurity in the vicinity. The pairs are labeled NN_n. It was believed that the second nitrogen atom is in the n-th shell around the first one. However, the proper level asignment is probably different in the view of modern theory [472]. Also clusters with more than two nitrogen atoms may exist. NN_1 is a prominent level and relates to a N–Ga–N complex having 12 equivalent sites for the second N atom on the next neighbor anion site. The transitions due to excitons bound to NN_n, as shown in Fig. 9.25a, give a series of lines (see Table 9.7) that fulfill $\lim_{n\to\infty} \mathrm{NN}_n = \mathrm{A}$. Although GaP has an indirect band structure, the absorption coefficient of N-related transitions is large, about $10^5\,\mathrm{cm}^{-1}$ for a nitrogen doping

[9] The A line is due to excitons with $J = 1$, resulting of coupling of the electron spin 1/2 with the hole angular momentum of 3/2. The B-line is a dipole forbidden line due to 'dark' excitons with $J = 2$.

Table 9.7 Index of nitrogen pairs NN_n and energy separation ΔE of bound-exciton transitions from the free-exciton line for $n = 1 \ldots 10$ and the 'A' line

n	1	2	3	4	5	6	7	8	9	10	∞ (A)
ΔE (meV)	143	138	64	39	31	25	22	20	18	17	11

level of $10^{19}\,\mathrm{cm}^{-3}$.[10] This is due to the fact that the electron spatially localized at the nitrogen isoelectronic trap (Sect. 7.7.9) has a sizeable $k = 0$-component of its wave-function (Fig. 7.41), leading to a large transition probability for Γ-point holes with an oscillator strength of 0.09 [773].

9.6.10 Biexcitons

Similar to two hydrogen atoms forming a hydrogen molecule, two excitons can also form a bound complex, the biexciton involving two electrons and two holes. The biexciton binding energy is defined as

$$E_{\mathrm{XX}}^{\mathrm{b}} = 2\,E_{\mathrm{X}} - E_{\mathrm{XX}}. \tag{9.50}$$

Biexcitons are binding in bulk material. Accordingly, the biexciton recombination or absorption occurs at lower energy than that of the exciton. Values of the biexciton binding energy are listed in Table 9.4 for various semiconductors. The ratio of biexciton and exciton binding energies is fairly constant about 0.2. In semiconductors with small exciton binding energy, such as GaAs, biexcitons are hard to observe in bulk material but show up in heterostructures that provide additional carrier confinement (see also Sect. 14.3.4). While the exciton density increases linearly with external excitation, the density of biexcitons increases quadratically.

9.6.11 Trions

The complexes 'eeh' and 'ehh' are called trions. Also, the notation X^- and X^+ is common. X^- is typically stable in bulk material but hard to observe. In quantum wells or dots, trions are easier to observe. In quantum dots excitons with higher charge, e.g. X^{2-}, have also been observed (see Fig. 14.34).

[10] Also the recombination (Sect. 10.3.2) is efficient and allows green GaP:N and yellow GaAsP:N light emitting diodes.

9.6.12 Band Gap Renormalization

The band structure theory has been developed so far for small carrier densities. If the carrier density is large the interaction of free carriers has to be considered. The first step was exciton formation. However, at high temperatures (ionization) and at large carrier density (screening) the exciton is not stable. Exchange and correlation energy leads to a decrease of the optical absorption edge that is called *band gap renormalization* (BGR).

An effect due to significant carrier density is to be expected when the density is of the order of the exciton volume, i.e. $n \sim a_B^{-3}$. For $a_B \sim 15\,\text{nm}$ (GaAs) this means $n \sim 3 \times 10^{17}\text{cm}^{-3}$. The dimensionless radius r_s is defined via

$$\frac{4\pi}{3} r_s^3 = \frac{1}{n\, a_B^3}. \tag{9.51}$$

The sum of exchange and correlation energies E_{xc} is found to be mostly independent of material parameters [774] (Fig. 9.26a) and follows the form

$$E_{xc} = \frac{a + b\, r_s}{c + d\, r_s + r_s^2}, \tag{9.52}$$

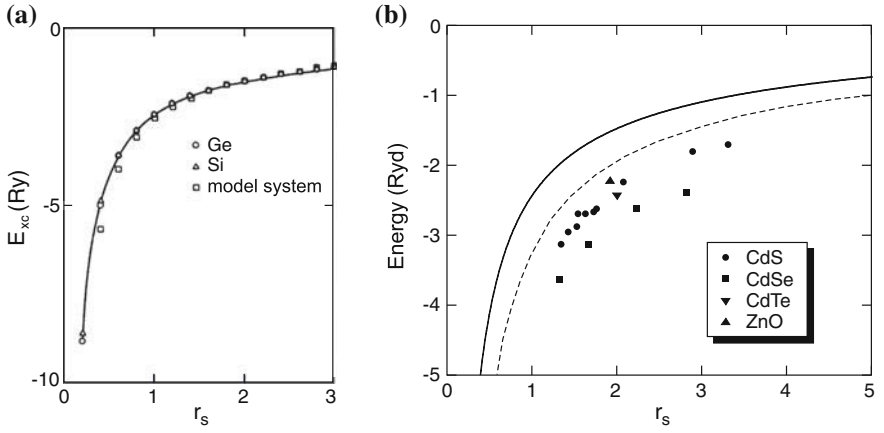


Fig. 9.26 (**a**) Theoretical exchange and correlation energies in units of the exciton Rydberg energy as a function of the dimensionless variable r_s for Ge, Si and a model system (with one isotropic conduction and valence band each). The *solid line* is a fit according to (9.52). Adapted from [774]. (**b**) Band gap renormalization in terms of the excitonic Rydberg for various II–VI semiconductors. *Solid line* is the relation according to (9.52), *dashed line* is the dependence predicted in [775] for $T = 30\,\text{K}$. Data are compiled in [776]

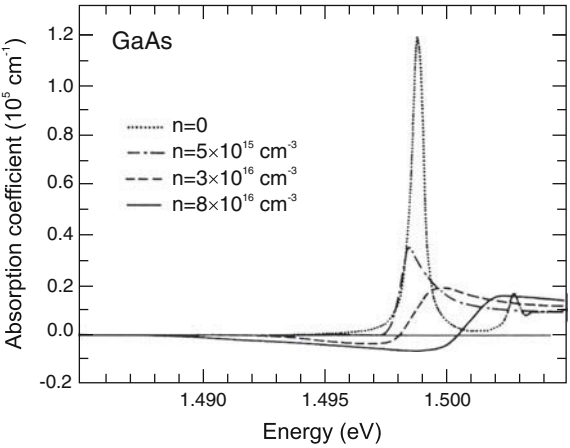


Fig. 9.27 Absorption of GaAs (low temperature, $T = 10\,\mathrm{K}$) as a function of the electron–hole density n (theory). Adapted from [777]

with $a = -4.8316$, $b = -5.0879$, $c = 0.0152$ and $d = 3.0426$. Thus the density dependence of the band gap at small carrier density is $\propto n^{1/3}$. Experimental data for a number of II–VI semiconductors roughly follow such a dependence (Fig. 9.26b).

In Fig. 9.27, a theoretical calculation of the absorption spectrum of bulk GaAs for various carrier densities (n = p) [777] is shown. With increasing density, the excitonic resonance broadens and vanishes. The shape approaches the electron–hole plasma shape. The absorption edge shifts to smaller energies. At high carrier density, the absorption becomes negative in a spectral range before absorption sets in. In this spectral region, the material exhibits gain and an incoming light wave is amplified (cmp. Sect. 10.2.6).

9.6.13 Electron–Hole Droplets

At low temperature and high density, electron–hole pairs in Ge and Si can undergo a phase transition into a liquid state. This electron–hole liquid (EHL) was suggested in [778] and is a Fermi liquid exhibiting the high conductivity of a metal and the surface and density of a liquid. The condensation is due to exchange interaction and correlation. The formation is fostered by the band structure of Ge [779] and the long lifetime of carriers in the indirect band structure. In unstressed Ge typically a cloud of electron–hole droplets with diameter in the μm range exists. The phase diagram is shown in Fig. 9.28a. In suitably stressed Ge electron–hole droplets with several hundred μm diameter form around the point of maximum shear strain in inhomogeneously strained crystals, as shown in Fig. 9.28b. The pair density in such a liquid is of the order of $10^{17}\,\mathrm{cm}^{-3}$.

We note that the metallic EHL state hinders observation of the Bose–Einstein condensation (BEC) of (bosonic) excitons. The light-exciton mass offers a high

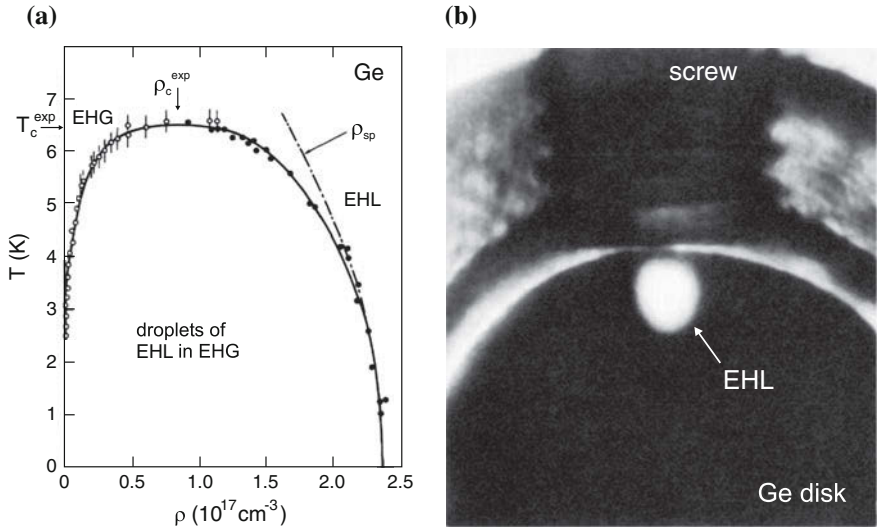


Fig. 9.28 (**a**) Temperature–density phase diagram of electrons and holes in Ge. The regions of electron–hole gas (EHG) and liquid (EHL) and the droplet phase are labeled. *Solid line* is theoretical calculation, *symbols* are experimental data from [780]. The *dash-dotted line* denoted ρ_{sp} is the experimentally obtained temperature dependence of the liquid density due to single-particle excitations. ρ_c^{exp} and T_c^{exp} denote the experimental critical density and temperature, respectively. Adapted from [781]. (**b**) Photographic image of radiative recombination (at 1.75 μm wavelength) from a 300-μm diameter droplet of electron–hole liquid (EHL) in a stressed (001) Ge disk (diameter 4 mm, thickness 1.8 mm) at $T = 2$ K. The stress is applied from the top by a nylon screw along a ⟨110⟩ direction. Adapted from [782], reprinted with permission, © 1977 APS

condensation temperature in the 1 K range (compared to the mK range for atoms). Recent experiments with spatially indirect excitons in coupled quantum wells lead towards BEC [783, 784]. A sufficiently long lifetime ensures cooling of the excitons close to the lattice temperature. Another potential candidate for BEC are long-living excitons (ms-range) in Cu_2O [785]. The condensation of polaritons (cf. Sect. 9.6.8) in microcavities to well-defined regions of **k**-space has been discussed in [786] and compared to bosonic condensation in bulk.

9.6.14 Two-Photon Absorption

So far, only absorption processes that involve one photon have been considered. The attenuation of the intensity I of a light beam (of frequency ω_0) along the z direction can be written as

$$\frac{\mathrm{d}I}{\mathrm{d}z} = -\alpha I - \beta I^2, \tag{9.53}$$

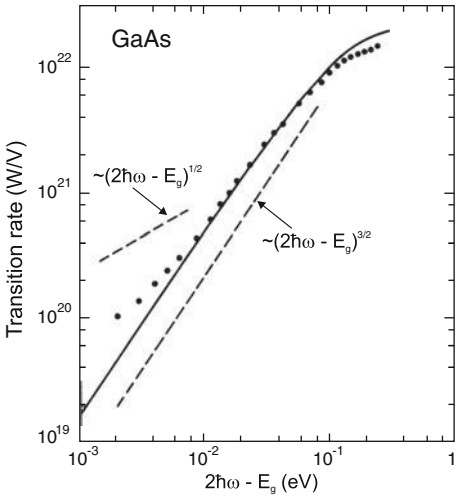


Fig. 9.29 Experimental two-photon absorption spectrum of GaAs ($T = 4\,\mathrm{K}$) (*dots*) plotted as a function of the difference of the double-photon energy $2\hbar\omega$ from the GaAs band edge E_g. The *solid line* is a theoretical calculation, the *dashed lines* represent slopes with exponent $1/2$ and $3/2$, respectively. Adapted from [789]

where α is due to the (linear) absorption coefficient (and possibly scattering) and β is the two-photon absorption coefficient. A two-photon process can occur in two steps, e.g. via a midgap level, which is not considered any further here. Here, we consider two-photon absorption (TPA) via the population of a state at $2\hbar\omega_0$ higher energy than the initial state with a nonlinear optical process. The TPA coefficient is related to the nonlinear third-order electric dipole susceptibility tensor [787] χ_{ijkl}. Within the two-band approximation theory predicts [788]

$$\beta \propto \left(2\,\hbar\omega_0 - E_\mathrm{g}\right)^{3/2}. \tag{9.54}$$

The exponent $3/2$ is indeed found experimentally, as shown in Fig. 9.29 for GaAs. The strength of absorption depends on the relative orientation of the light polarization with respect to the main crystallographic directions, e.g. TPA for polarization along $\langle 110\rangle$ is about 20 % larger than for the $\langle 100\rangle$ orientation.

9.7 Impurity Absorption

9.7.1 Shallow Levels

For charge carriers bound to shallow impurities long range Coulomb forces are most important and they exhibit a hydrogen-like term scheme

$$E_n = \frac{m^*}{m_0}\frac{1}{\epsilon_\mathrm{r}^2}\frac{1}{n^2} \times 13.6\,\mathrm{eV}, \tag{9.55}$$

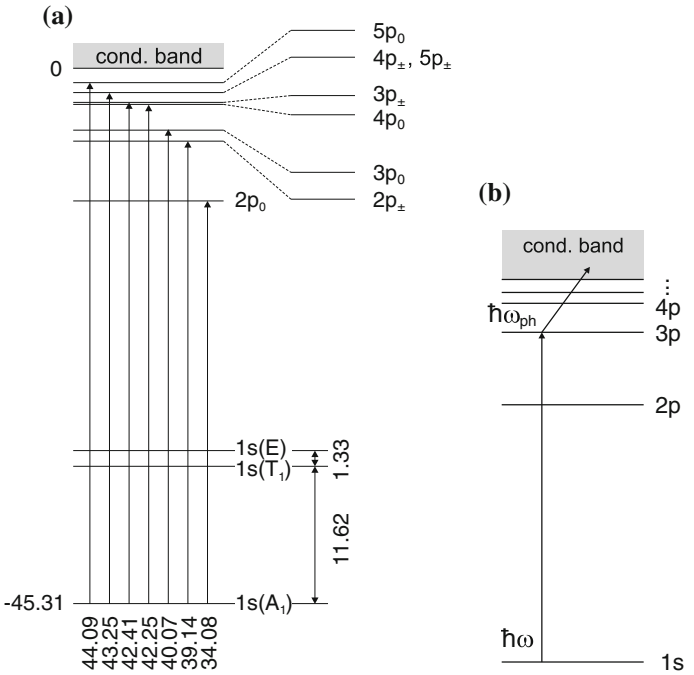


Fig. 9.30 (**a**) Term scheme of phosphorus donor in silicon, all energies in meV. After [791]. (**b**) Schematic sequence for photothermal ionization, here absorption of a photon with $\hbar\omega = E_{3p} - E_{1s}$ and subsequent absorption of a phonon with energy $\hbar\omega_{ph} \geq E_\infty - E_{3p}$

with the ionization limit E_∞ being the conduction (valence) band edge for donors (acceptors), respectively. They can be excited by light to the nearest band edge. Such absorption is typically in the FIR region and can be used for photodetectors in this wavelength regime. The optical absorption cross section of impurity absorption can be related to the carrier capture cross section [513, 514].

The actual transition energies can deviate from (9.55) due to deviation of the potential close to the impurity from the pure Coulomb potential. Such an effect is known as the chemical shift or central cell correction (cf. Sect. 7.5.5) and is characteristic of the particular impurity. In GaAs such shifts are small ($\sim$100 μeV) [790].

The term scheme for P in Si is shown in Fig. 9.30a. The ground state (1s) is split because of a reduction of the tetrahedral symmetry due to intervalley coupling. The anisotropic mass at the X-valley in Si causes the p states (and states with higher orbital momentum) to split into p_0 and $p_\pm$ states. Such an effect is absent in a direct semiconductor with an isotropic conduction-band minimum such as GaAs (Fig. 9.31). Optical transitions between the 1s and various p states can be directly observed in absorption, e.g. for Si:P in [792]. These transitions are also observed in photoconductivity because the missing energy to the ionization into the continuum is supplied by a phonon at finite temperature (photothermal ionization) (Fig. 9.30b) [791]. The

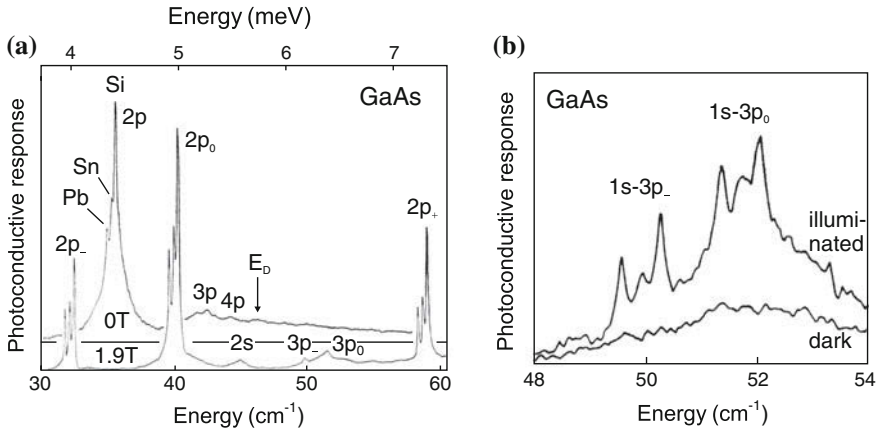


Fig. 9.31 (**a**) Far-infrared photoconductivity response (Lyman-type s→p series) of not intentionally doped GaAs with residual donors Pb, Sn, and Si, $N_A = 2.6\times10^{13}\,cm^{-3}$, $N_D - N_A = 8\times10^{12}\,cm^{-3}$. The *upper* (*lower*) *curve* is for a magnetic field of 0 (1.9) T. Measurement temperature is 4.2 K. (**b**) Photoconductive response of a (different) GaAs sample with the same impurities ($N_D = 1\times10^{13}\,cm^{-3}$) with (*upper curve*) and without (*lower curve*) illumination with above-bandgap light ($B = 1.9\,T$, $T = 4.2\,K$). Adapted from [793]

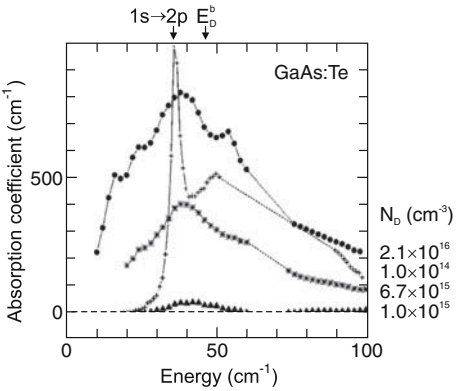


Fig. 9.32 Low-temperature ($T = 1.35\,K$) absorption spectra of highly doped n-type GaAs:Te with doping concentrations as labeled (*circles* $N_D = 2.1\times10^{16}\,cm^{-3}$, *stars* 6.7×10^{14}, *triangles* 1.0×10^{15}). A sharp photoconductivity spectrum (in arbitrary units) from low-doped GaAs:Te (*crosses*, $N_D = 1.0\times10^{14}\,cm^{-3}$) is shown for comparison (cf. Fig. 9.31a). The energy of the 1s→2p transition and the donor binding energy (onset of continuum absorption) are indicated. Adapted from [794]

splitting of the 2p transition in Fig. 9.31a is the chemical shift due to different donors incorporated in the GaAs (Si, Sn, and Pb). Peak broadening is mostly due to Stark broadening due to neighboring charged impurities. The application of a magnetic field induces Zeeman-like splittings and increases the sharpness of the peaks. The peak width can be further increased by illuminating the sample with light having a

higher energy than the band gap. The additional charge carriers neutralize charged impurities and allow higher resolution (Fig. 9.31b).

In Fig. 9.32 absorption spectra of highly doped n-type GaAs are shown. For doping concentrations larger than the critical concentration of $\sim 1 \times 10^{16}\,\mathrm{cm}^{-3}$ (cf. Table 7.6) significant broadening is observed due to the formation of an impurity band.

9.7.2 Deep Levels

The absorption of deep levels is typically in the infrared. In Fig. 9.33a the possible optical absorption processes involving the Fe levels in InP (cf. Sect. 7.7.8) during the charge transfer $Fe^{3+} \rightarrow Fe^{2+}$ are shown. These transitions and their fine structure (Fig. 9.33b) have been observed in calorimetric absorption spectroscopy (CAS) experiments [604].

In Fig. 9.34 the photoconductivity of is shown. The sharp peaks are due to transitions of interstitial, singly ionized Mg, Mg_i^+ [795]. Mg in Si is a double donor [575] (see Sect. 7.7.2). Above the ionization limit of about 256 meV, the peaks are replicated, shifted by the LO phonon energy of 59.1 meV. However, now they rather appear as dips. This behavior is typical for a discrete state interacting with a continuum, also called Fano resonance [796, 797] with its characteristic lineshape, going *below* the continuum level.

The absorption spectra due to various deep acceptors in GaAs are compared in Fig. 9.35. The density of states in the band increases with k (proportional to $\sqrt{E - E_c}$). The carrier on the impurity is strongly localized and described with a wave packet centered around Γ, its k-components decreasing with increasing k. Thus the maximum absorption will be at an intermediate k-value and an associ-

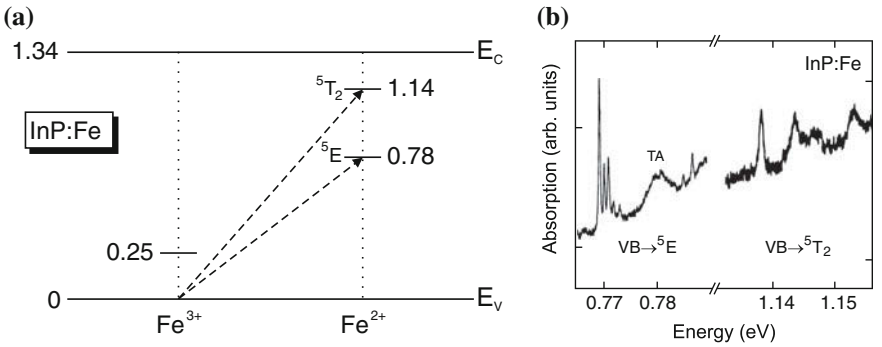


Fig. 9.33 (**a**) Schematic band diagram of InP with levels of Fe impurities in the 3+ and 2+ charge states at low temperature. All energies are given in eV. The *arrows* denote the optical transition of a valence-band electron to the Fe center, $Fe^{3+} + \hbar\omega \rightarrow Fe^{2+} + h$. (**b**) Calorimetric absorption spectra (at $T = 1.3\,\mathrm{K}$) of InP:Fe, $[\mathrm{Fe}] = 5 \times 10^{16}\,\mathrm{cm}^{-3}$. Part (**b**) adapted from [604]

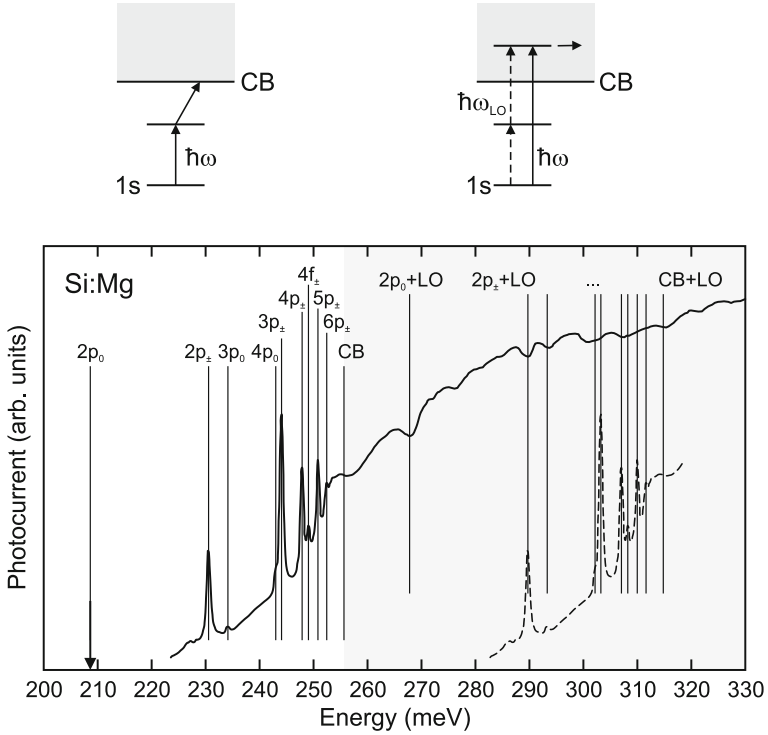


Fig. 9.34 Photocurrent spectrum of Si:Mg. Transitions are due to Mg_i^+ from its 1s state to excited states as labeled and indicated by *vertical lines*. CB denotes the conduction band edge (ionization limit). Above the CB edge (*shaded area*) phonon-assisted absorption occurs (Fano resonances). For comparison the absorption spectrum below CB is shown shifted by the phonon energy (*dashed line*). Above the plot, the transition mechanisms (photothermal ionization and Fano resonance) are schematically shown. Adapted from [795]

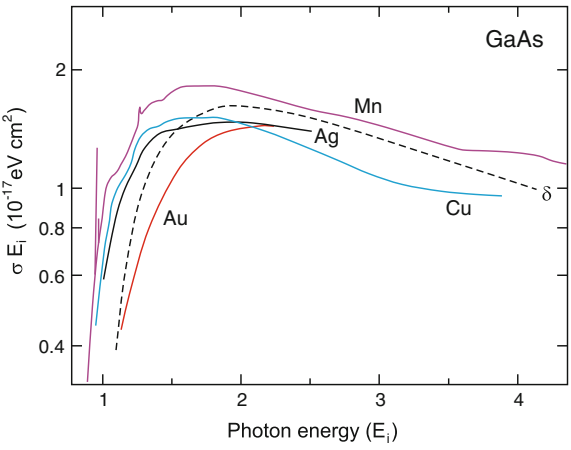


Fig. 9.35 Absorption spectra ($\sigma = \alpha/p$) due to various deep impurities in GaAs as labeled. The *dashed line* is a theoretical lineshape assuming a hole bound to a δ-potential. The energy axis is scaled by the ionization energy. The kink for Mn at 3.5 $E_i \approx 450$ meV is due to the onset of absorption into the split-off valence band. Adapted from [279]

ated energy larger than the ionization energy E_i (lowest transition to continuum at for $k = 0$). The lineshapes in Fig. 9.35 fit to a model with a δ-potential (zero range model, neglecting long range Coulomb terms) [798] with maximum absorption close to 2 E_i,

$$\alpha(E) \propto \frac{E_\mathrm{i}^{1/2}\,(E - E_\mathrm{i})^{3/2}}{E^3}. \tag{9.56}$$

9.8 Absorption in the Presence of Free Charge Carriers

In the presence of charge carriers, various absorption processes can occur. First, the dissipative motion of carriers leads to infrared absorption, termed the free carrier absorption (Sect. 9.8.1). Filling of a band with carriers leads to a shift of the band-band absorption edge, the Burstein–Moss shift (Sect. 9.8.2). Besides the free-carrier absorption, free carriers present in the semiconductor can lead to further absorption processes with transition energies below the band gap. These processes are due to transitions within the band structure and can be

- inter-valence band transitions of holes (Sect. 9.8.3),
- phonon-assisted inter-valley transitions of electrons (Sect. 9.8.4),
- phonon-assisted intra-band transitions of electrons (Sect. 9.8.5).

9.8.1 Free-Carrier Absorption

The absorption due to free carriers in the infrared spectral range (away from phonon resonances) can be described with the Drude model [799].

A time-dependent electric field accelerates the charge carriers within a band. The excess energy is subsequently transferred to the lattice via scattering with phonons. A review of the effect of free carriers on optical properties can be found in [800]. In the relaxation-time approximation energy is relaxed with a time constant τ. Thus energy is absorbed from the electromagnetic wave and dissipated. Effectively, this process represents an intra-band excitation.

The complex conductivity (8.36) is given by

$$\sigma^* = \sigma_\mathrm{r} + \mathrm{i}\sigma_\mathrm{i} = \frac{n\,e^2\,\tau}{m^*}\left(\frac{1}{1+\omega^2\tau^2} + \mathrm{i}\frac{\omega\tau}{1+\omega^2\tau^2}\right). \tag{9.57}$$

We note that a static magnetic field introduces birefringence as discussed in more detail in Sect. 13.2.2. The wave equation for the electric field is

$$\nabla^2\mathbf{E} = \epsilon_\mathrm{r}\,\epsilon_0\,\mu_0\,\ddot{\mathbf{E}} + \sigma^*\,\mu_0\,\dot{\mathbf{E}}. \tag{9.58}$$

For a plane wave $\propto \exp[\mathrm{i}(\mathbf{kr} - \omega t)]$ the wavevector obeys

$$k = \frac{\omega}{c}\sqrt{\epsilon_\mathrm{r} + \mathrm{i}\,\frac{\sigma^*}{\epsilon_0\,\omega}}, \tag{9.59}$$

where $c = (\epsilon_0\mu_0)^{-1/2}$ is the velocity of light in vacuum, ϵ_r is the background dielectric constant (for large ω).

The part ϵ_FC of the dielectric function due to free carriers is

$$\epsilon_\mathrm{FC} = \frac{\mathrm{i}}{\epsilon_0\,\omega}\,\sigma^*. \tag{9.60}$$

The complex index of refraction is

$$n^* = n_\mathrm{r} + \mathrm{i}\,\kappa = \sqrt{\epsilon_\mathrm{r} + \mathrm{i}\frac{\sigma^*}{\epsilon_0\,\omega}}. \tag{9.61}$$

Taking the square of this equation yields

$$n_\mathrm{r}^2 - \kappa^2 = \epsilon_\mathrm{r} + \mathrm{i}\,\frac{\sigma_\mathrm{i}}{\epsilon_0\,\omega} = \epsilon_\mathrm{r} - \frac{n\,e^2}{\epsilon_0\,m^*}\,\frac{\tau^2}{1+\omega^2\tau^2} \tag{9.62a}$$

$$2\,n_\mathrm{r}\,\kappa = \frac{\sigma_\mathrm{r}}{\epsilon_0\omega} = \frac{n\,e^2}{\epsilon_0\,\omega\,m^*}\,\frac{\tau}{1+\omega^2\tau^2}. \tag{9.62b}$$

The absorption coefficient is related to κ by (9.9). For the case of higher frequencies, i.e. $\omega\tau \gg 1$, the absorption is

$$\alpha = \frac{n\,e^2}{\epsilon_0\,c\,n_\mathrm{r}\,m^*\,\tau}\,\frac{1}{\omega^2}. \tag{9.63}$$

The absorption decreases with increasing frequency like ω^{-r}. The classical Drude treatment as followed here results in an exponent of $r = 2$. This is the case for neutral impurity scattering and also for small frequencies $\hbar\omega \ll E_\mathrm{F}$. A more detailed discussion of the energy dependence of free-carrier absorption can be found in [801]. Other exponents have been derived for scattering by acoustical phonons ($r = 3/2$), LO phonons ($r = 5/2$) and ionized impurities ($r = 7/2$). More detailed quantum mechanical treatments of free-carrier absorption in the presence of impurities and phonons can be found in [802, 803].

For semiconductors free-carrier absorption is particularly important in the mid- and far-infrared regions when carriers are present due to doping or thermal excitation. In Fig. 9.36a absorption spectra of n-type Ge for various doping concentrations are shown. The absorption coefficient in the transparency regime varies proportionally to $\lambda^2 \propto \omega^{-2}$ as predicted in (9.63). In Fig. 9.36a, the absorption can be seen to rise for photon energy above 0.7 eV due to absorption in the band structure. Electrons are

excited from the valence band across the fundamental band gap into the conduction band (cmp. Sect. 9.6.3), which is an indirect transition in Ge.

In Fig. 9.36b the absorption coefficient due to free carrier absorption at fixed wavelength is shown as a function of dopant concentration.[11] The slope is slightly overlinear, indicating a weak dependence $\tau(n)$. A sub-linear relation has been found for heavily p-doped GaAs [804].

The index of refraction is given by (also for $\omega\tau \gg 1$)

$$n_r^2 = \epsilon_r - \frac{ne^2}{\epsilon_0 m^* \omega^2} + \kappa^2 = \epsilon_r \left[1 - \left(\frac{\omega_p}{\omega}\right)^2\right] + \frac{\epsilon_r^2}{4n_r^2}\left(\frac{\omega_p}{\omega}\right)^4 \frac{1}{\omega^2\tau^2}$$
$$\approx \epsilon_r \left[1 - \left(\frac{\omega_p}{\omega}\right)^2\right], \tag{9.64}$$

where

$$\omega_p = \sqrt{\frac{n\,e^2}{\epsilon_r\,\epsilon_0\,m^*}} \tag{9.65}$$

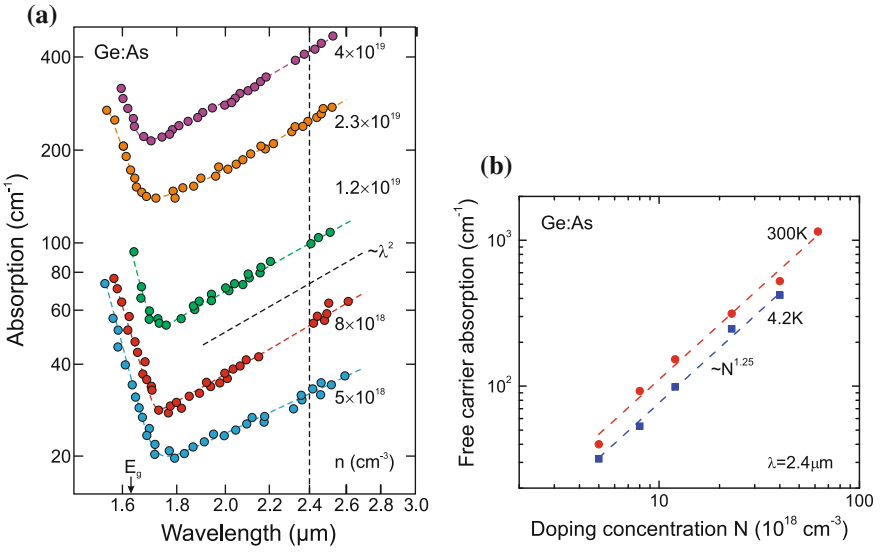


Fig. 9.36 (**a**) Optical absorption spectra (at $T = 4.2\,\mathrm{K}$) of n-type Ge for various As dopant concentrations as labeled. The *arrow* denotes the band edge of undoped Ge, the *vertical dashed line* the energy for which the free-carrier absorption is measured in part (**b**). The inclined *dashed line* visualizes the slope $\propto \lambda^2$. *Curved dashed lines* are guides to the eye. Adapted from [739]. (**b**) Free-carrier absorption at $\lambda = 2.4\,\mu\mathrm{m}$ as determined from part (**a**) of the figure (*blue squares*) as a function of As dopant concentration. Additionally data at 300 K (*red circles*) from the same samples are included [739]. The *dashed lines* visualizes the slope $\propto N_D^{1.25}$

[11] Even at low temperature, $n \approx N_D$ since $N_D \gg N_c$ (cf. [515] and Sect. 7.5.7).

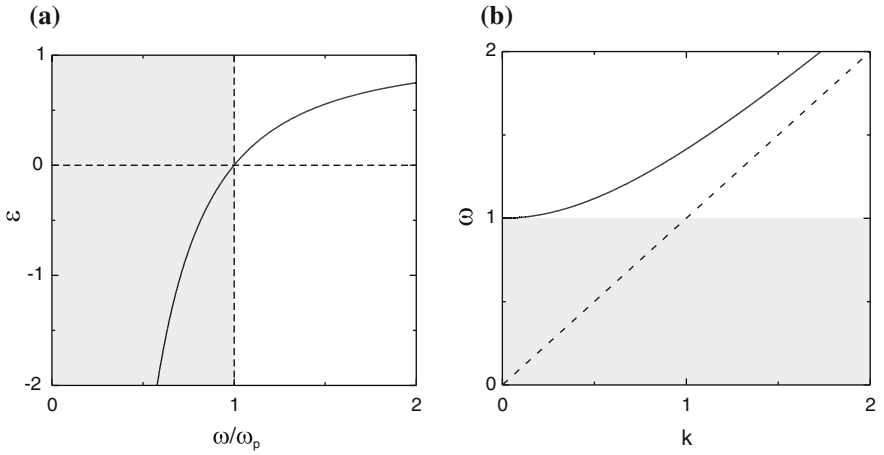


Fig. 9.37 (**a**) Dielectric constant for plasmon oscillations. *Shaded area* represents region of attenuation (negative ϵ). (**b**) Dispersion relation (k in units of ω_p/c, ω in units of ω_p) in the presence of free carriers (9.67, for $\epsilon_r = 1$). *Shaded area* represents forbidden frequency range for propagating solutions. *Dashed line* is photon dispersion $\omega = ck$

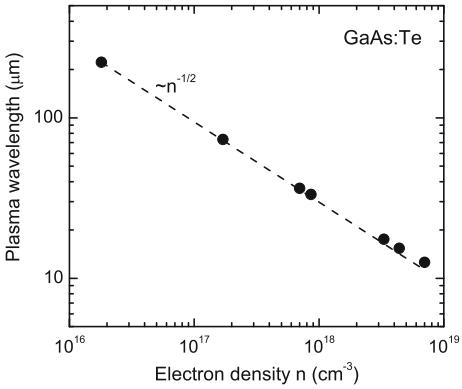


Fig. 9.38 Plasma wavelength λ_p for n-type GaAs with various electron concentrations due to different doping levels. *Filled circles*: experimental values, *dashed line*: $n^{-1/2}$ dependence; the deviation is due to nonparabolicity of the electron mass (cf. Fig. 9.50b). Data from [805]

is the plasma frequency. The approximation is valid for small absorption and when $(\omega\tau)^{-2}$ can be neglected. A graphical representation is given in Fig. 9.37a. For coupling to electromagnetic waves (still $\omega\tau \gg 1$)

$$\epsilon(\omega) = \epsilon_r \left[1 - \left(\frac{\omega_p}{\omega}\right)^2\right] = \frac{c^2 k^2}{\omega^2} \tag{9.66}$$

must be fulfilled. It follows that the dispersion relation in the presence of free carriers (Fig. 9.37b) is

$$\omega^2 = \omega_p^2 + \frac{c^2 k^2}{\epsilon_r}. \tag{9.67}$$

Fig. 9.39 Czochralski-grown Ga_2O_3 crystals with different electron concentration, courtesy of Z. Galazka, IKZ Berlin

For $\omega > \omega_p$, $\epsilon > 0$, thus waves can propagate. For $\omega < \omega_p$, however, the dielectric constant is negative, i.e. $\epsilon < 0$. For such frequencies waves are exponentially damped and cannot propagate or penetrate a layer. This effect can be used in a plasmon waveguide. The expected dependence of the plasmon wavelength on the carrier density $\lambda_p = 2\pi c/\omega_p \propto n^{-1/2}$ is depicted in Fig. 9.38 for GaAs. For semiconductors the plasmon frequency is in the mid-or far-infrared spectral region.[12]

The effect of free-carrier absorption is visualized in Fig. 9.39. The blue hue of the boule with the higher electron concentration is due to free carrier absorption in the long wavelength part of the visible spectrum.

9.8.2 Burstein–Moss Shift

In the discussion so far it has been assumed that all target states in the conduction band are empty. In the presence of free carriers the absorption is modified by the

- change of the distribution function
- many-body effects (band gap renormalization)

The latter is discussed in the next section. For a degenerate electron distribution all states close to the conduction-band edge are populated. Thus a transition from the valence band cannot take place into such states. This shift of the absorption edge to higher energies is called the Burstein–Moss shift [806, 807]. Originally, the Burstein–Moss shift was evoked to explain the absorption shift in InSb with varying carrier concentration (Fig. 9.40).

k-conserving optical transitions between parabolic hole and electron bands have the dependence

$$E = E_g + \frac{\hbar^2 k^2}{2\, m_e^*} + \frac{\hbar^2 k^2}{2\, m_h^*} = E_g + \frac{\hbar^2 k^2}{2\, m_r}, \tag{9.68}$$

[12] The much higher free-electron density in metals shifts the plasma frequency to the UV, explaining the reflectivity of metals in the visible and their UV transparency.

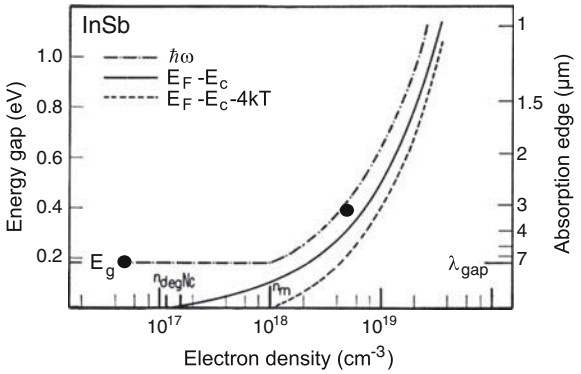


Fig. 9.40 Burstein–Moss effect at InSb ($E_g = 0.18\,\text{eV}$) at room temperature. Theoretical dependence and data points for intrinsic InSb and $5 \times 10^{18}\,\text{cm}^{-3}$ n-type. Data from [806]

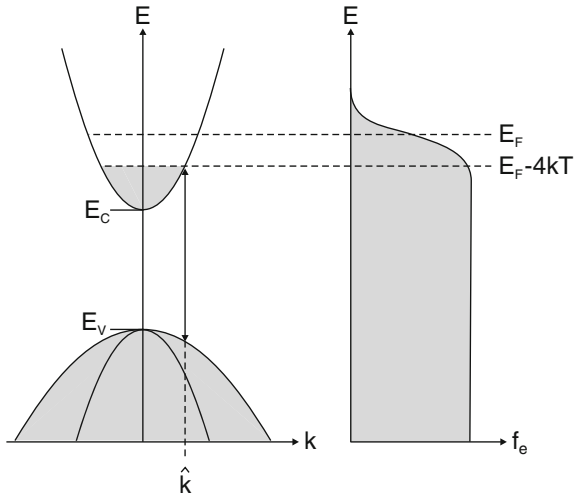


Fig. 9.41 Principle of Burstein–Moss shift. *Left panel*: Schematic band structure with completely filled electron states shown in *grey*. The k-vector for the lowest photon energy optical absorption process is indicated as $\hat{k}$. *Right panel*: Electron distribution function for a degenerate electron gas with Fermi level in the conduction band

where m_r is the reduced mass of electron and hole. About $4kT$ below the Fermi level all levels in the conduction band are populated (Fig. 9.41). Thus the k value at which the absorption starts is given as

$$\hat{k} = \sqrt{\frac{2\,m_r}{\hbar^2}\,(E_F - E_C - 4kT)}. \tag{9.69}$$

Besides the energy shift in the conduction band, the corresponding energy shift in the valence band $\hbar k^2/(2m_h)$ must be considered. Thus, the Burstein–Moss shift of the absorption edge is

$$\Delta E = \hbar\omega - E_g = (E_F - 4kT - E_C)\left(1 + \frac{m_e}{m_h}\right). \tag{9.70}$$

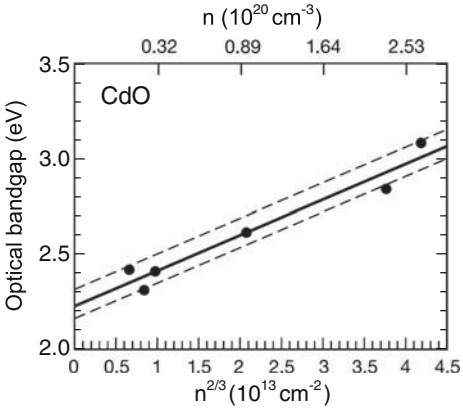


Fig. 9.42 Burstein–Moss effect for CdO. The linear fit is for $E_g = 2.22(8)$ eV and $m_r = 0.113(11)\, m_e$. The *dashed lines* indicate the confidence interval of ± 0.08 eV. Adapted from [808]

The relation between n and the Fermi level is given by (7.6). If $E_F - E_C \gg kT$ the Fermi integral can be approximated by $\frac{2}{\sqrt{\pi}} \frac{2}{3} \left(\frac{E_F - E_C}{kT}\right)^{3/2}$. Using (7.8) for N_C, the Burstein–Moss shift can be written for this case as

$$\Delta E = n^{2/3} \frac{h^2}{8\, m_e} \left(\frac{3}{\pi}\right)^{2/3} \left(1 + \frac{m_e}{m_h}\right) \approx 0.97 \frac{h^2}{8\, m_r} n^{2/3}. \tag{9.71}$$

The $n^{2/3}$ dependence of the energy shift is found, e.g., for CdO[13] with different carrier concentrations (due to different deposition temperature, no intentional doping) [808] and depicted in Fig. 9.42.

9.8.3 Inter-Valenceband Transitions

Transitions within the valence band can occur between three bands, i.e. lh→hh, so→hh, and so→lh, as schematically depicted in Fig. 9.43. Theoretical treatments have been given in [809, 810]. For GaAs, such intravalence-band absorption occurs at photon energies close to Δ_0 as shown in Fig. 9.44a for p-type GaAs:Zn [811]. For p-type GaSb, the absorption coefficient below the fundamental band gap is found almost entirely due to inter-valence band transitions, as shown in Fig. 9.44b for a hole concentration of $p = 3.2 \times 10^{16}\,\mathrm{cm}^{-3}$ [812].

[13] CdO is an indirect semiconductor, the optical band gap is the energy of the direct transition at the Γ-point. The indirect transitions involve holes from other points in the Brillouin zone (cmp. Fig. 6.10).

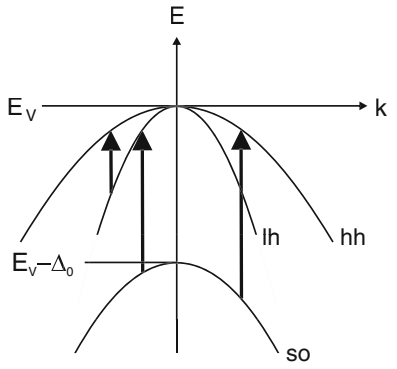


Fig. 9.43 Schematic optical transitions within the valence band

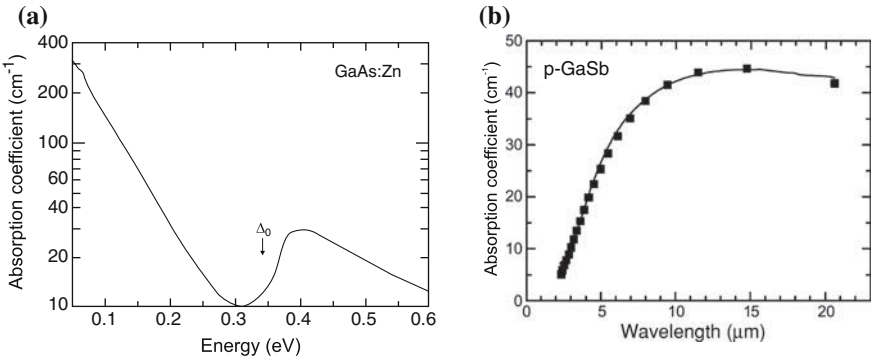


Fig. 9.44 (**a**) Optical absorption spectrum of GaAs:Zn with $p = 2.7 \times 10^{17}\,\mathrm{cm}^{-3}$ at $T = 84\,\mathrm{K}$. The absorption above the split-off energy Δ_0 is due to the hh/lh → s-o process . Adapted from [811]. (**a**) Optical absorption coefficient of GaSb with $p = 3.2 \times 10^{16}\,\mathrm{cm}^{-3}$. Experimental data (*solid line*) and calculation of inter-valence band contribution (*squares*). The free carrier contribution is less than $5\,\mathrm{cm}^{-1}$ in the considered spectral range. Adapted from [812]

9.8.4 Inter-Valley Transitions

Electrons at the conduction band minimum can undergo optical transitions to the same band at a different point of the Brillouin zone. Such intervalley transition, as sketched in Fig. 9.45a, is phonon-assisted to fulfill momentum conservation and occurs around the energy difference ΔE between the two valleys (cmp. Table 8.4).

For InP with an electron concentration of $n = 1.65 \times 10^{18}\,\mathrm{cm}^{-3}$, below the fundamental band edge at 1.4 eV, an additional contribution starting around 0.8–0.9 eV is found besides the free carrier absorption (Fig. 9.45b) [813]. Taking into account the filling of the bottom of the conduction band, an energy separation for the two valleys of $\Delta E = 0.90 \pm 0.02\,\mathrm{eV}$ was found for various values of the electron

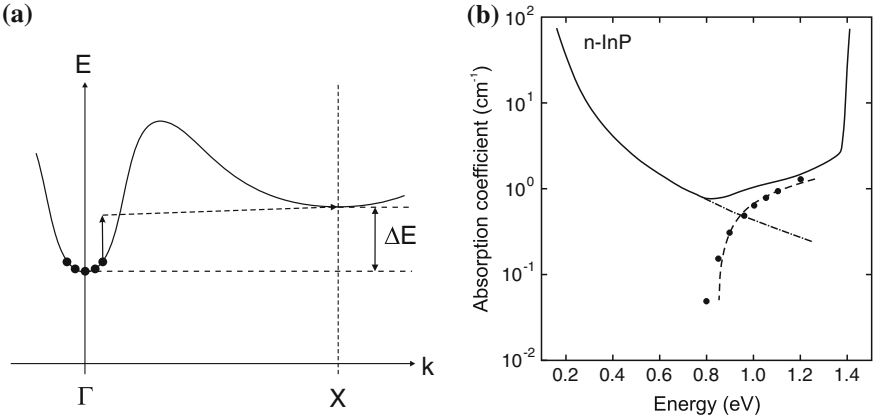


Fig. 9.45 (**a**) Schematic of inter-valley conduction band transitions involving a photon (*solid line arrow*) and a phonon (*dashed line arrow*). (**b**) Optical absorption coefficient of InP with $n = 1.65 \times 10^{18}\,\mathrm{cm}^{-3}$. Experimental data (*solid line*) and calculation of inter-valley band contribution (*dashed line*). The extrapolated free carrier contribution is shown as *dash-dotted line* and the difference of experimental absorption and extrapolated free carrier contribution as *circles*. Adapted from [813]

concentrations. This energy corresponds to the energy difference of conduction band minima at Γ and X in InP. The lineshape of this absorption processes can be modeled and fits well the difference of measured absorption and extrapolated free-carrier absorption spectra. Transitions to the lower minimum at L ($\Delta E = 0.6\,\mathrm{eV}$) are not observed, possibly masked by the free-carrier absorption.

9.8.5 Intra-Band Transitions

Phonon-assisted transitions within the lowest conduction band (not to a different valley), as indicated schematically in Fig. 9.46a for the SnO_2 band structure [814], can cause absorption at photon energies below the fundamental absorption edge. Actually in SnO_2, the optical transition across the fundamental band gap is only weakly dipole-allowed and leads to small absorption coefficient below $100\,\mathrm{cm}^{-1}$ directly above the fundamental band gap of about 3.6 eV. The strong dipole-allowed transition with absorption coefficient around $10^5\,\mathrm{cm}^{-1}$ begins at about 4.3 eV and stems from electrons in a lower valence band [815]. The free-carrier absorption due to transitions within the lowest conduction band are calculated to start already at 2.8 eV (Fig. 9.46b) and thus can also impact transparency in the visible spectral range.

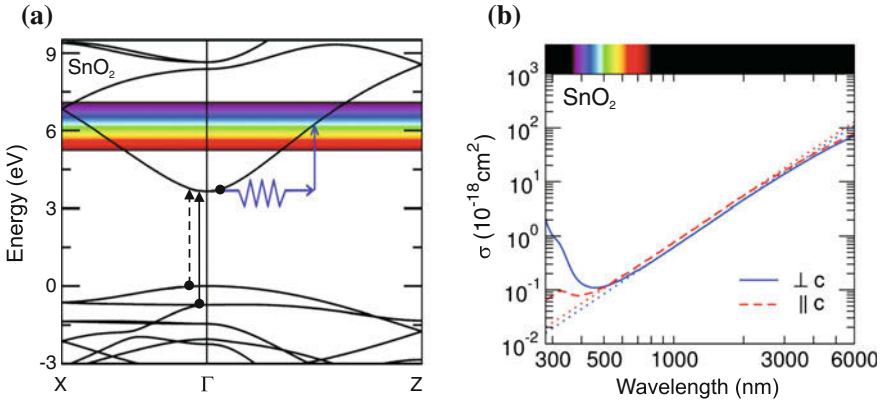


Fig. 9.46 (**a**) Band structure of SnO_2 and absorption processes; the transition from the topmost valence band (*dashed arrow*) is forbidden. (**b**) Calculated free-carrier absorption ($\sigma = \alpha/n$) for SnO_2. The *solid* and *dashed lines* are results including phonon-assisted transitions for two light polarizations. The *dotted lines* are fits of the Drude model to the infrared regime. Adapted from [814]

9.9 Lattice Absorption

While there is no interaction of optical phonons and (infrared) light in Si and Ge due to crystal structure symmetry [816], strong effects are present for compound semiconductors. A review can be found in [817].

9.9.1 Dielectric Constant

The dielectric constant (with damping parameter Γ) in the vicinity of the optical phonon energies is given by (cf. (5.50))

$$\epsilon(\omega) = \epsilon(\infty)\left(\frac{\omega_{\mathrm{LO}}^2 - \omega^2 - \mathrm{i}\omega\Gamma}{\omega_{\mathrm{TO}}^2 - \omega^2 - \mathrm{i}\omega\Gamma}\right). \tag{9.72}$$

The dispersion relation (without damping) can be written as

$$\epsilon(\omega) = \epsilon(\infty) + \frac{\epsilon(0) - \epsilon(\infty)}{1 - (\omega/\omega_{\mathrm{LO}})^2} = \epsilon(\infty)\left[1 + \frac{f}{1 - (\omega/\omega_{\mathrm{LO}})^2}\right]. \tag{9.73}$$

Thus the oscillator strength (compare with (D.10)) is $f = \frac{\epsilon(0)-\epsilon(\infty)}{\epsilon(\infty)}$. With the LST relation (5.49) the oscillator strength is

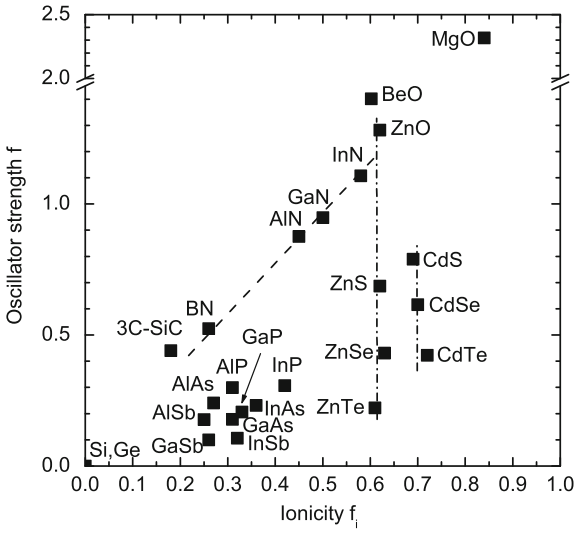


Fig. 9.47 Lattice absorption oscillator strength f from (9.74) for various elemental, III–V and II–VI semiconductors as a function of their ionicity f_i (cf. Table 2.1). *Dashed line* is linear dependence on ionicity for similar (reduced) mass, *dash-dotted lines* are guides to the eye for similar ionicity and varying mass

$$f = \frac{\omega_{LO}^2 - \omega_{TO}^2}{\omega_{TO}^2} \approx 2\,\frac{\omega_{LO} - \omega_{TO}}{\omega_{TO}}, \tag{9.74}$$

and thus proportional to the splitting $\Delta_{LT} = \omega_{LO} - \omega_{TO}$ between the longitudinal and transverse optical phonon frequency. The approximation in (9.74) is valid for $\Delta_{LT} \ll \omega_{TO}$.

The oscillator strength increases with the ionicity, i.e. the electronegativity difference of the atoms in the base (Fig. 9.47). Additionally, the oscillator strength depends on the reduced mass and the high-frequency polarizability; this can be seen, e.g., for the series of the Zn compounds that all have similar ionicity. For the series of the nitrides, the mass effect is small since the reduced mass is dominated by the light N mass.

9.9.2 *Reststrahlenbande*

The absorption of electromagnetic radiation by optical phonons is governed by the dielectric function that has been derived in (9.72). For small damping, i.e. $\Gamma \ll \Delta_{LT}$, the dielectric constant is negative between ω_{TO} and ω_{LO}. From $\epsilon_r = n_r^2 - \kappa^2$ it follows that κ^2 is much larger than n_r^2. Therefore, the reflectance (9.17) will be close to 1. This energy range is the so-called *reststrahlenbande*. This term stems from multiple reflections in this wavelength regime that suppresses neighboring spectral regions and thus achieves a certain monochromatization in the far-infrared spectral region. Within the semiconductor the absorption is large in the reststrahlenbande (Fig. 9.48).

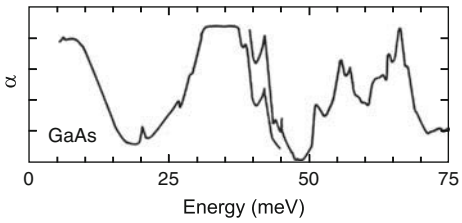


Fig. 9.48 Far-infrared absorption of GaAs. In the region around 35 meV is the reststrahlenbande with high absorption due to optical phonons. Adapted from [146], based on [818]

9.9.3 Polaritons

The coupled propagation of phonons and electromagnetic radiation is related to the equation (without phonon damping)

$$\epsilon(\omega) = \epsilon(\infty)\left(\frac{\omega_{\mathrm{LO}}^2 - \omega^2}{\omega_{\mathrm{TO}}^2 - \omega^2}\right) = \frac{c^2\,k^2}{\omega^2}. \tag{9.75}$$

There are two branches of propagating waves (real k):

$$\omega^2 = \frac{1}{2}\left(\omega_{\mathrm{LO}}^2 + \frac{c^2k^2}{\epsilon(\infty)}\right) \pm \sqrt{\frac{1}{4}\left(\omega_{\mathrm{LO}}^2 + \frac{c^2k^2}{\epsilon(\infty)}\right)^2 - \left(\frac{c^2k^2\omega_{\mathrm{TO}}^2}{\epsilon(\infty)}\right)^2}. \tag{9.76}$$

For $k = 0$ we find the solutions $\omega = \omega_{\mathrm{LO}}$ and $\omega = kc/\sqrt{\epsilon(0)}$. For large k we find $\omega = \omega_{\mathrm{TO}}$ and $\omega = kc/\sqrt{\epsilon(\infty)}$. These solutions are shown in Fig. 9.49. Both branches have a phonon- and a photon-like part. The coupled state between the phonon and the photon field is called the (phonon-) polariton.

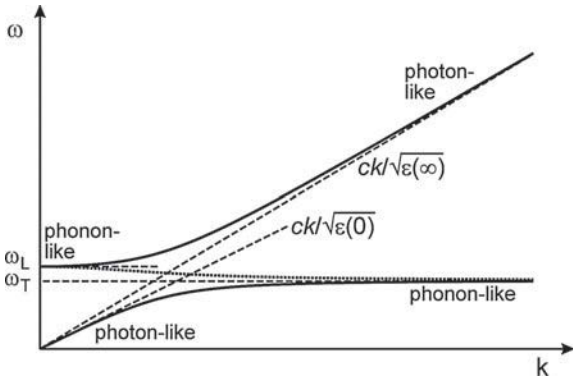


Fig. 9.49 Dispersion of the polariton. The *dotted line* displays the dispersion for a purely imaginary wavevector with the absolute value k

In the interval $[\omega_{TO}, \omega_{LO}]$ the wavevector is purely imaginary, i.e. $k = i\tilde{k}$ with real $\tilde{k}$. For this case there is only one solution that is also depicted in Fig. 9.49,

$$\omega^2 = \frac{1}{2}\left(\omega_{LO}^2 + \frac{c^2\tilde{k}^2}{\epsilon(\infty)}\right) + \sqrt{\frac{1}{4}\left(\omega_{LO}^2 + \frac{c^2\tilde{k}^2}{\epsilon(\infty)}\right)^2 + \left(\frac{c^2\tilde{k}^2\omega_{TO}^2}{\epsilon(\infty)}\right)^2}. \tag{9.77}$$

9.9.4 Phonon–Plasmon Coupling

The coupling of phonons and plasmons in the spectral region of the reststrahlenbande leads to the development of two new branches, the longitudinal phonon plasmon modes (LPP+ and LPP−), in the common dispersion. The dielectric function is

$$\epsilon(\omega) = \epsilon(\infty)\left(1 + \frac{\omega_{LO}^2 - \omega^2}{\omega_{TO}^2 - \omega^2} - \frac{\omega_p^2}{\omega^2}\right). \tag{9.78}$$

For $\epsilon(\omega) = 0$ for $k = 0$ (coupling to photons) the two solutions ω_{LPP+} and ω_{LPP-} do not cross as a function of ω_p (Fig. 9.50),

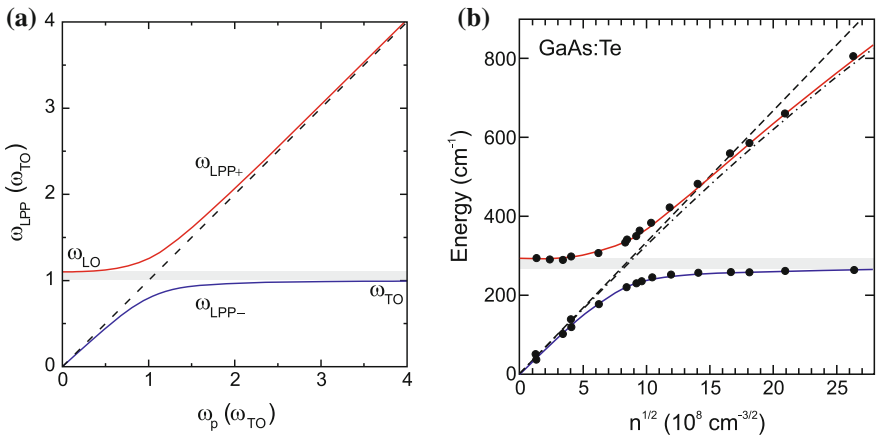


Fig. 9.50 (**a**) Frequency of the coupled longitudinal-phonon plasmon (LPP) modes (lower (upper) polariton branch in *blue* (*red*)) as a function of the plasma frequency. *Dashed line* shows uncoupled plasmon frequency ($\omega = \omega_p$), *grey area* indicates spectral region between TO and LO modes. (**b**) Experimental data on the polariton energies in n-type GaAs with different carrier concentration $\omega_p \propto \sqrt{n\,m^*}$ (9.65). *Dashed* (*dash-dotted*) *line* is plasmon frequency ω_p without (with) consideration of conduction band non-parabolicity (cf. Fig. 6.31b). Data from [805, 819]

$$\omega_{\mathrm{LPP}\pm} = \frac{1}{2}\left[\omega_{\mathrm{LO}}^2 + \omega_{\mathrm{p}}^2 \pm \sqrt{(\omega_{\mathrm{LO}}^2 + \omega_{\mathrm{p}}^2)^2 - 4\,\omega_{\mathrm{TO}}^2\,\omega_{\mathrm{p}}^2}\right]. \tag{9.79}$$

For small plasma frequencies $\omega_{\mathrm{LPP}+} = \omega_{\mathrm{LO}}$, i.e. the optical phonons couple to the electromagnetic field without change. Also $\omega_{\mathrm{LPP}-} = \omega_{\mathrm{p}}$. For large carrier density, i.e. $\omega_{\mathrm{p}} \gg \omega_{\mathrm{LO}}$, we find $\omega_{\mathrm{LPP}-} = \omega_{\mathrm{TO}}$ and $\omega_{\mathrm{LPP}+} = \omega_{\mathrm{p}}$. Thus, the carriers have effectively screened the electric field of the phonon that had led to the increase of the TO to the LO frequency.

Chapter 10
Recombination

Les hommes discutent, la nature agit.
Men argue, nature acts.

Voltaire

Abstract The various mechanisms and statistics of carrier recombination in semiconductors including band-band, excitonic, band-impurity (Shockley–Read–Hall kinetics) and Auger recombination are explained. Also recombination at extended defects and surfaces is treated. Using the diffusion-recombination theory, the one-dimensional carrier profiles for typical situations in experiments and devices are derived.

10.1 Introduction

In thermodynamic nonequilibrium excess charges can be present in the semiconductor. They can be created by carrier injection through contacts, an electron beam or the absorption of light with wavelength smaller than the band gap. After the external excitation is turned off, the semiconductor will return to the equilibrium state. The relaxation of carriers into energetically lower states (and energy release) is called recombination. The term stems from the electron recombining with the hole created after absorption of a photon. However, there are other recombination mechanisms. A dedicated textbook is [820].

In the simplest picture an excitation generates carriers with a rate G (carriers per unit volume and unit time). In the steady state (after all turn-on effects) a constant excess charge n carrier density is present. Then the generation exactly compensates the recombination processes. The principle of detailed balance even says that each microscopic process is balanced by its reverse process. If the time constant of the latter is τ, n is given by $n = G\,\tau$. This follows from the steady-state solution $\dot{n} = 0$ of

$$\frac{\mathrm{d}n}{\mathrm{d}t} = G - \frac{n}{\tau}. \tag{10.1}$$

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_10

In the literature two limiting cases have been discussed, the *relaxation* and the *lifetime* semiconductor, depending on the relation of two time constants. The one time constant τ_0 is the relaxation time constant due to recombination as discussed in the following. The smaller τ_0 is, the faster excited electrons and holes recombine and 'disappear'. Fast lifetimes are typically present in direct semiconductors (compared to indirect ones), semiconductors with high defect density and amorphous semiconductors. The other time constant is $\tau_D = \epsilon/\sigma$, the dielectric relaxation time; it describes the transport of carriers due to mobility (and diffusion). Large dielectric relaxation times are present in semiconductors with high mobility (low defect density, small carrier mass), small τ_D typically for hopping conduction. The relaxation case is given for $\tau_0 \ll \tau_D$; carriers will recombine quickly and it is hard to build up non-equilibrium carriers and separate them with an applied electric field. In the recombination case $\tau_D \ll \tau_0$, non-equilibrium carriers can assume non-uniform distributions and an applied electrical field generates separate quasi-Fermi levels for electrons and holes[1] (cmp. Sect. 7.6).

10.2 Band–Band Recombination

The band–band recombination is the relaxation from an electron in the conduction band into the valence (the empty state there is the hole). In a direct semiconductor, electrons can make an optical transition between the bottom of the conduction band to the top of the valence band. In an indirect semiconductor, this process is only possible with the assistance of a phonon and is thus much less probable.

10.2.1 Spontaneous Emission

We consider the spontaneous recombination of an electron of energy E_e and a hole of energy E_h (Fig. 10.1a). $C(E_e, E_h)$ is a constant proportional to the matrix element of the optical transition (cf. Sect. 9.5). The spontaneous recombination rate r_{sp} at photon energy $E \geq E_C - E_V = E_g$ is (assuming energy conservation, i.e. $E = E_e - E_h$, but without k-conservation in a dense plasma [821]),

$$r_{sp}(E) = \int_{E_C}^{\infty} \mathrm{d}E_e \int_{-\infty}^{E_V} \mathrm{d}E_h \; C(E_e, E_h) \times D_e(E_e)\, f_e(E_e)\, D_h(E_h)\, f_h(E_h)\, \delta(E - E_e + E_h) \tag{10.2}$$

[1] In the relaxation case, the separation of quasi-Fermi levels is $\ll kT$.

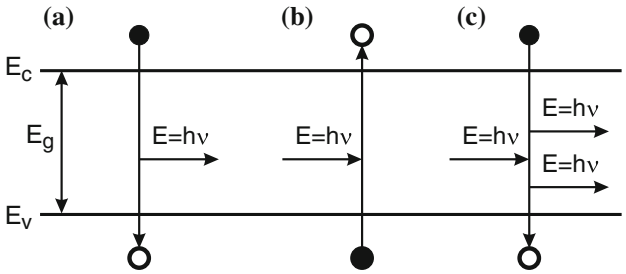


Fig. 10.1 Processes of band–band recombination: (**a**) spontaneous emission, (**b**) absorption and (**c**) stimulated emission. A *full* (*empty*) *circle* represents an occupied (unoccupied) electron state

$$= \int_{E_C}^{E+E_V} \mathrm{d}E_e \; C(E_e, E_e - E) \times D_e(E_e)\, f_e(E_e)\, D_h(E_e - E)\, f_h(E_e - E),$$

where f_h denotes the hole occupation $f_h = 1 - f_e$.

The lineshape of the band–band recombination with k-conservation[2] is proportional to the joint density of states (9.30) and the Fermi distribution function. At small excitation and at low doping it can be approximated by the Boltzmann distribution function and the lineshape is given as

$$I(E) \propto \sqrt{E - E_g}\, \exp\left(-\frac{E}{kT}\right). \tag{10.3}$$

An experimental spectrum is shown in Fig. 10.2 together with a fit according to (10.3). The expected FWHM of the peak is $1.7954\, kT$, which is about 46 meV at $T = 300$ K. At low sample temperature, the temperature of the carrier gas is typically higher than the lattice temperature, depending on the cooling mechanisms (carrier–carrier scattering, optical phonon emission, acoustic phonon emission, recombination, ...) and the excitation rate. The carrier temperature in GaAs, determined from the Boltzmann tail of spontaneous emission (photoluminescence) is depicted in Fig. 10.3 as a function of excitation density; clearly it increases with increasing excitation.

The recombination rate in indirect semiconductors is small since the transition is phonon-assisted. For silicon, an internal quantum efficiency in the 10^{-6}-range has been reported [824]. For germanium, the direct transition is energetically fairly close to the fundamental, indirect L–Γ band edge transition (Fig. 9.12).The energy difference can be reduced from its bulk value of 136 meV by tensile strain. Additionally, the direct transition can be favored by heavily n-doping and filling the L conduction band minimum states (see Sect. 9.8.2). In this case, direct recombination from

[2]Excitonic effects are neglected here, e.g. for temperatures $kT \gg E_X^b$. Such effects are discussed in Sect. 10.3.

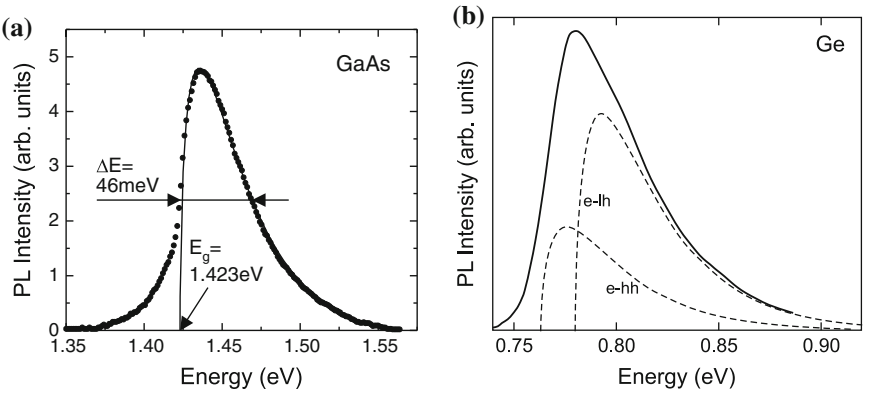


Fig. 10.2 (**a**) Photoluminescence spectrum of an undoped LPE-grown epitaxial GaAs layer at room temperature and low cw ($\lambda = 647\,\text{nm}$) excitation density ($10\,\text{W/cm}^2$). The *solid line* is a lineshape fit with (10.3) and $E_\text{g} = 1.423\,\text{eV}$ and $T = 293\,\text{K}$. (**b**) Room temperature, direct (e_Γ–h_Γ) recombination from heavily n-doped ($10^{19}\,\text{cm}^{-3}$) germanium ($1\,\mu\text{m}$ thick Ge layer on silicon (001)) with biaxial (thermal) tensile strain. The strain-split valence band edge (Fig. 6.43) causes the e–hh and e–lh transitions (individual contributions with lineshape according to (10.3) shown as *dashed lines*) to occur at different energies. Adapted from [822]

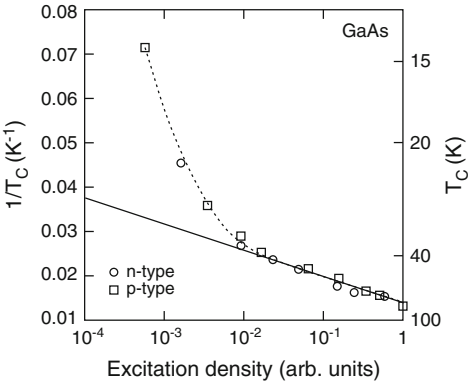


Fig. 10.3 Carrier temperature T_C in GaAs as a function of excitation density at a lattice temperature of 1.6 K. The *dashed line* is guide to the eyes, the *solid line* corresponds to an activation energy of 33 meV, similar to the GaAs optical phonon energy. Adapted from [823]

the conduction band Γ-minimum can be observed [822] and the effective energy difference has been lowered to about 100 meV.

10.2.2 Absorption

A similar consideration is made for the absorption process (Fig. 10.1b). An electron is transferred upon light absorption from a valence-band state (occupied) to a conduction-band state that must be empty. The coefficient is B_1. Also, the process

is proportional to the light intensity, represented by the density of occupied photon states $N_{\mathrm{ph}}(E)$,

$$r_{\mathrm{abs}}(E) = \int_{E_{\mathrm{C}}}^{E+E_{\mathrm{V}}} \mathrm{d}E_{\mathrm{e}}\; B_1(E_{\mathrm{e}}, E_{\mathrm{e}} - E) \times D_{\mathrm{e}}(E_{\mathrm{e}})\,(1 - f_{\mathrm{e}}(E_{\mathrm{e}}))\,D_{\mathrm{h}}(E_{\mathrm{e}} - E)\,(1 - f_{\mathrm{h}}(E_{\mathrm{e}} - E))\,N_{\mathrm{ph}}(E). \quad (10.4)$$

10.2.3 Stimulated Emission

In this case, an incoming photon 'triggers' the transition of an electron in the conduction band into an empty state in the valence band. The emitted photon is in phase with the initial photon (Fig. 10.1c). The rate is (with coefficient B_2):

$$r_{\mathrm{st}}(E) = \int_{E_{\mathrm{C}}}^{E+E_{\mathrm{V}}} \mathrm{d}E_{\mathrm{e}}\; B_2(E_{\mathrm{e}}, E_{\mathrm{e}} - E) \times D_{\mathrm{e}}(E_{\mathrm{e}})\,f_{\mathrm{n}}(E_{\mathrm{e}})\,D_{\mathrm{h}}(E_{\mathrm{e}} - E)\,f_{\mathrm{h}}(E_{\mathrm{e}} - E)\,N_{\mathrm{ph}}(E). \quad (10.5)$$

The photon density N_{ph} at a given energy is given by Planck's law and the Bose–Einstein distribution (Appendix E)

$$N_{\mathrm{ph}}(E) = N_0 \frac{1}{\exp(E/kT) - 1}. \quad (10.6)$$

The pre-factor is the density of states of the electromagnetic field[3] $N_0(E) = 8\,\pi\,E^2\,(n_{\mathrm{r}}/hc)^3$.

10.2.4 Net Recombination Rate

In thermodynamical equilibrium the rates fulfill

$$r_{\mathrm{sp}}(E) + r_{\mathrm{st}}(E) = r_{\mathrm{abs}}(E). \quad (10.7)$$

The population functions are Fermi functions with quasi-Fermi levels F_{n} and F_{p}. Since for absorption and stimulated emission the same quantum-mechanical matrix element is responsible, $B_1 = B_2$. The detailed balance (10.7) yields

$$C(E_1, E_2) = B_1(E_1, E_2)\,N_{\mathrm{ph}} \left[\exp\left(\frac{E - (F_{\mathrm{n}} - F_{\mathrm{p}})}{kT}\right) - 1\right]. \quad (10.8)$$

[3]The total number of photon states in vacuum between the frequencies zero and ν is $N(\nu) = 8\pi\nu^3/(3c^3)$. With $\nu = E/h$ and $N_0 = \mathrm{d}N(E)/\mathrm{d}E$ and considering $c \to c/n_{\mathrm{r}}$ we obtain the given value for N_0.

In thermodynamic equilibrium, i.e. $F_n = F_p$,

$$C(E_1, E_2) = N_0\, B_1(E_1, E_2) = B. \tag{10.9}$$

If the constant B, the bimolecular recombination coefficient, is independent of the energy E, the integration for the net bimolecular recombination rate r_B can be executed analytically and we find

$$\begin{aligned} r_B &= \int_{E_g}^{\infty} \left[r_{sp}(E) + r_{st}(E) - r_{abs}(E) \right] \mathrm{d}E \\ &= B\, n\, p \left[1 - \exp\left(-\frac{F_n - F_p}{kT} \right) \right]. \end{aligned} \tag{10.10}$$

In thermodynamic equilibrium, of course, $r_B = 0$. The recombination rate Bnp is then equal to the thermal generation rate G_{th}

$$G_{th} = B\, n_0\, p_0. \tag{10.11}$$

The bimolecular recombination rate typically used in Shockley–Read–Hall (SRH) [825, 826] kinetics is

$$r_B = B\, (n\, p - n_0\, p_0). \tag{10.12}$$

Values for the coefficient B are given in Table 10.1. In the case of carrier injection, np is larger than in thermodynamical equilibrium, i.e. $n\, p > n_0\, p_0$, and the recombination rate is positive, i.e. light is emitted. If the carrier density is smaller than in thermodynamical equilibrium, e.g. in a depletion region, absorption is larger than emission. This effect is also known as 'negative luminescence' [827] and plays a role particularly at elevated temperatures and in the infrared spectral region.

Table 10.1 Bimolecular recombination coefficient at room temperature for a number of semiconductors

Material	B (cm^3/s)
GaN	1.1×10^{-8}
GaAs	1.0×10^{-10}
AlAs	7.5×10^{-11}
InP	6.0×10^{-11}
InAs	2.1×10^{-11}
4H-SiC	1.5×10^{-12}
Si	1.1×10^{-14}
GaP	3.0×10^{-15}

Data for GaN from [828], Si from [829], SiC from [830], other values from [831]

10.2.5 Recombination Dynamics

The carrier densities n and p, are decomposed into the densities n_0 and p_0 in thermodynamic equilibrium and the excess-carrier densities δn and δp, respectively

$$n = n_0 + \delta n \tag{10.13a}$$

$$p = p_0 + \delta p. \tag{10.13b}$$

Here, only neutral excitations are considered, i.e. $\delta n = \delta p$. Obviously the time derivative fulfills $\frac{\partial n}{\partial t} = \frac{\partial \delta n}{\partial t}$, and correspondingly for the hole density. The equation for the dynamics

$$\dot{n} = \dot{p} = -Bnp + G_{\text{th}} = -B\,(n\,p - n_0\,p_0) = -B\,(n\,p - n_{\text{i}}^2) \tag{10.14}$$

can be written as

$$\frac{\partial\,\delta p}{\partial t} = -B\;(n_0\;\delta p + p_0\;\delta n + \delta n\;\delta p)\,. \tag{10.15}$$

The general solution of (10.15) is given by

$$\delta p(t) = \frac{(n_0 + p_0)\,\delta p(0)}{[n_0 + p_0 + \delta p(0)]\,\exp\left[B\,t\,(n_0 + p_0)\right] - \delta p(0)}. \tag{10.16}$$

In the following, we discuss some approximate solutions of (10.15). First, we treat the case of a small (neutral) excitation, i.e. $\delta n = \delta p \ll n_0, p_0$. The dynamic equation is in this case

$$\frac{\partial\,\delta p}{\partial t} = -B\,(n_0 + p_0)\,\delta p. \tag{10.17}$$

Then the decay of the excess-carrier density is exponential with a time constant (lifetime) τ given by

$$\tau = \frac{1}{B\,(n_0 + p_0)}. \tag{10.18}$$

In an n-type semiconductor additionally $n_0 \gg p_0$, and thus the minority carrier lifetime τ_p is

$$\tau_p = \frac{1}{B\, n_0}. \tag{10.19}$$

If the nonequilibrium carrier densities are large, i.e. $n \approx p \gg n_0, p_0$, e.g. for strong injection, the kinetics obeys

$$\frac{\partial\, \delta p}{\partial t} = -B\, (\delta p)^2, \tag{10.20}$$

and the transient has the form

$$\delta p(t) = \frac{\delta p(0)}{1 + B\, t\, \delta p(0)}, \tag{10.21}$$

where $\delta p(0)$ is the excess hole density at time $t = 0$. Such a decay is called *hyperbolic* and the recombination is bimolecular. The exponential decay time is formally $\tau^{-1} = B\delta p(t)$ and is thus time and density dependent. A detailed discussion of minority carrier lifetime is given in [832].

10.2.6 Lasing

The net rate for stimulated emission and absorption is

$$r_{\mathrm{st}}(E) - r_{\mathrm{abs}}(E) = \left[1 - \exp\left(\frac{E - (F_{\mathrm{n}} - F_{\mathrm{p}})}{kT}\right)\right] \times \int_{E_{\mathrm{C}}}^{E+E_{\mathrm{V}}} \mathrm{d}E_{\mathrm{e}}\, B\, D_{\mathrm{e}}(E_{\mathrm{e}})\, f_{\mathrm{e}}(E_{\mathrm{e}})\, D_{\mathrm{h}}(E_{\mathrm{e}} - E)\, f_{\mathrm{h}}(E_{\mathrm{e}} - E)\, N_{\mathrm{ph}}(E). \tag{10.22}$$

The net rate at photon energy $E = \hbar\omega$ is only larger than zero (i.e. dominating stimulated emission) when

$$F_{\mathrm{n}} - F_{\mathrm{p}} > E \geq E_{\mathrm{g}}. \tag{10.23}$$

When the difference of the quasi-Fermi levels is larger than the band gap, the carrier population is inverted, i.e. close to the band edges the conduction-band states are more strongly populated with electrons than the valence-band states, as shown in Fig. 10.4. An incoming optical wave of energy E will then be net amplified by stimulated emission. Equation (10.23) is also called the thermodynamic laser condition. We note that lasing requires further conditions as discussed in Sect. 23.4.

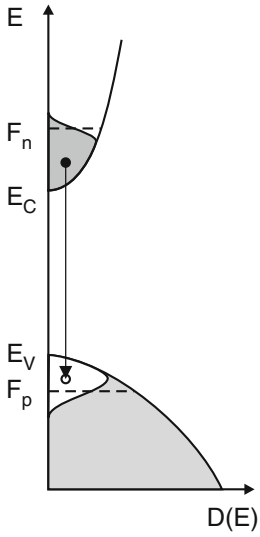


Fig. 10.4 Charge-carrier distribution during inversion, necessary for lasing. *Shaded areas* are populated with electrons. A stimulated transition between an electron and a hole is indicated

10.3 Exciton Recombination

10.3.1 Free Excitons

The observation of free-excitons is limited for semiconductors with a small exciton binding energies (such as in GaAs) to low temperatures. However, for large exciton binding energy, recombination from free-excitons is observed even at room temperature, as shown in Fig. 10.5 for ZnO.

A low temperature recombination spectrum of silicon is shown in Fig. 10.6. In pure silicon, phonon-assisted exciton recombination (cmp. Sect. 10.4) is observed involving acoustic (I^{TA}) and optical (I^{TO}) phonons. The weakly observed no-phonon line (I^0) is forbidden in perfect Si.

10.3.2 Bound Excitons

Excitons can localize at impurities, defects or other potential fluctuations and subsequently recombine [834, 835]. Excitons can be bound to neutral or ionized donors and acceptors impurities [836]. Also they can be bound to isoelectronic impurities, the most prominent example being N in GaP [837] (cmp. Sect. 9.6.9). or isoelectronic clusters [838]. The recombination of excitons localized in quantum wells (Sect. 12.4) and quantum dots (Sect. 14.3.4) is discussed later.

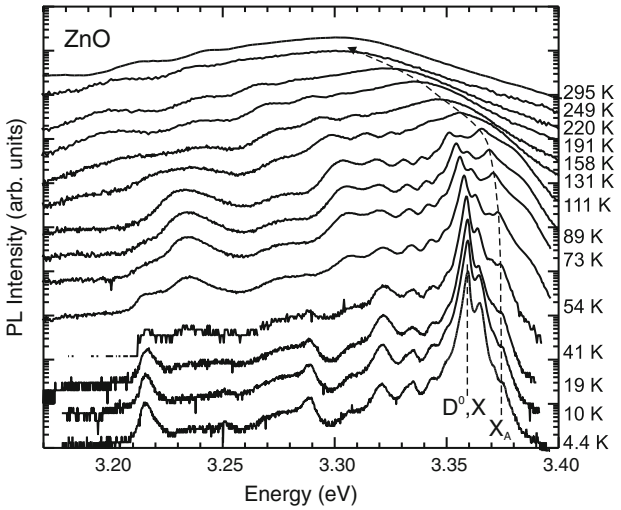


Fig. 10.5 Temperature-dependent luminescence spectra of a ZnO thin film (on sapphire). At low temperatures, the spectra are dominated by donor-bound exciton transitions (Al^0, X)). The *vertical dashed line* indicates the low-temperature position of the donor-bound exciton transition (D^0, X). The *curved dashed line* visualizes the energy position of the free-exciton transition (X_A) that becomes dominant at room temperature

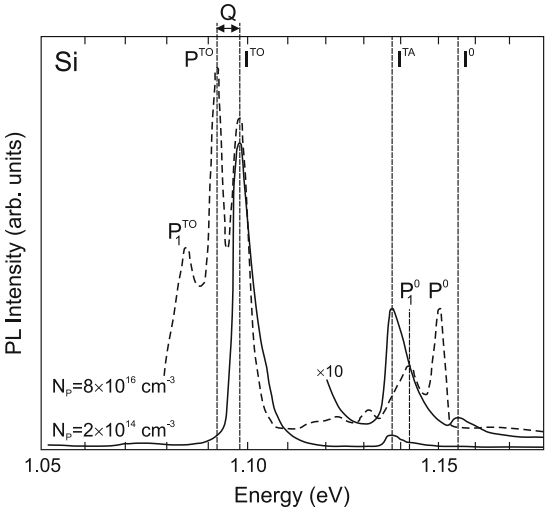


Fig. 10.6 Low temperature recombination spectra from silicon with low (*solid lines*) and sizeable (*dashed line*) phosphorus dopant concentration. Spectrum for $N_P = 2 \times 10^{14}\,cm^{-3}$ ($N_P = 8 \times 10^{16}\,cm^{-3}$) taken at 26 K (15 K). Transitions in pure Si are label with 'I', transitions involving P donors are labeled with 'P'. Q indicates the dissociation energy of the bound exciton. Adapted from [833]

Table 10.2 Localization energy Q (Q^*) of excitons on selected impurities (ionized impurities, D^+ or A^-, respectively) in various semiconductors

Host	Donor	Q (meV)	Q^* (meV)	Q^*/Q	σ	Reference
GaAs	EMD	0.88	1.8	2.0	0.28	[840]
	Zn	8.1	31.1	3.8		[841]
GaN	EMD	6.8	11.2	1.6	0.36	[842]
	Mg	20				[843]
AlN	Si	16				[843]
	Mg	40				
CdS	EMD	6.6	3.8	0.6	0.17	[844]
ZnSe	Al	4.9	5.4	1.1	0.27	[845, 846]
	Ga	5.1	6.6	1.3		
	In	5.4	7.5	1.4		
ZnO	Al	15.5	3.4	0.21	0.3	[847]
	Ga	16.1	4.1	0.25		
	In	19.2	8.5	0.44		

σ is the ratio of effective electron and hole (polaron) masses. EMD: effective mass donor

The transition energy $\hbar\omega$ of an exciton bound to a neutral impurity is

$$\hbar\omega = E_g - {E_X}^b - Q, \tag{10.24}$$

where Q is the binding (or localization) energy of the exciton to the impurity. The binding energy of an exciton to an ionized impurity is denoted with Q^*. A transition involving an exciton bound to a neutral donor is denoted (D^0, X); correspondingly (D^+, X), also denoted as (h, D^0), and (A^0, X). Values for donor-bound excitons in various semiconductors are listed in Table 10.2. The (D^0, X) complex is stable for $0 < \sigma = m_e^*/m_h^* < 0.43$ according to [839]. The (D^+, X) peak can occur on the low- or high-energy side of the (D^0, X) recombination. Whether $Q^* < Q$ or $Q^* > Q$ depends on σ being smaller or larger than 0.2, respectively [839], and is fulfilled for many semiconductors, e.g. GaAs, GaN, CdS, and ZnSe.

Recombination in silicon due to excitons involving phosphorus donors is depicted in Fig. 10.6. The (D^0, X) transition in Si:P is labeled 'P^0' ($Q = 6$ meV). Other P-related transitions are discussed in [833]. In Si, the binding energy to the impurity is about one tenth of the binding energy of the impurity (Haynes's rule [834, 848]), i.e. Q/E_D^b and $Q/E_A^b \approx 0.1$ (Fig. 10.7a). In GaP the approximate relations $Q = 0.26E_D^b - 7$ meV and $Q = 0.056E_A^b + 3$ meV have been found [837]. For donors in ZnO, the relation $Q = 0.365E_D^b - 3.8$ meV holds (Fig. 10.7b) [849]. In Fig. 10.8, the recombination spectrum of GaAs:C is shown that exhibits recombination from excitons bound to the acceptor (carbon) and shallow donors. The exciton is more strongly bound to an ionized donor (D^+) than to a neutral donor.

Varying the concentration of a specific impurity and observing the corresponding change in the intensity of the (D^0, X) transition allows to identify the chemical species to which the exciton is bound. This can be achieved via the comparison of different samples or more elegantly by introducing radioactive isotopes. This is shown in

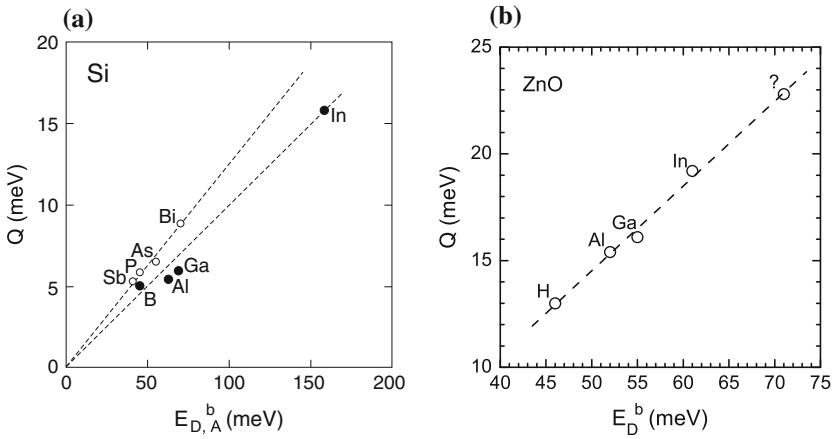


Fig. 10.7 Energy Q required to remove an exciton from a neutral impurity (10.24) as a function of the ionization energy E_D^b (*open circles*) or E_A^b (*solid circles*) of the involved impurity in (**a**) silicon (experimental data from [848]) and (**b**) ZnO (experimental data from [849]

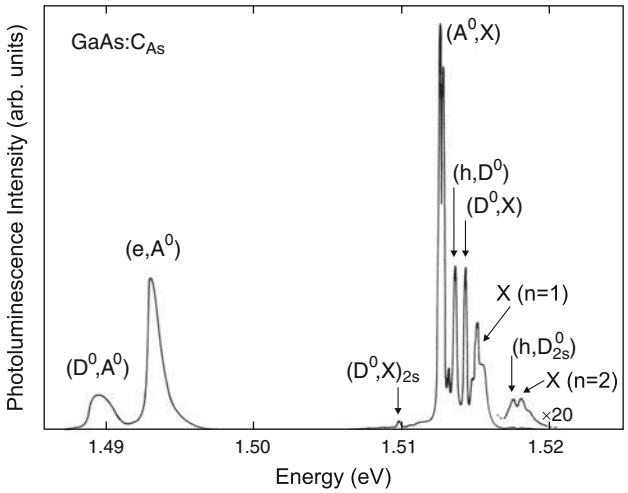


Fig. 10.8 Photoluminescence spectrum ($T = 2\,\mathrm{K}$, $D = 10\,\mathrm{mW\,cm^{-2}}$) of GaAs:$C_{As}$ ($N_A = 10^{14}\,\mathrm{cm^{-3}}$) with donor- and acceptor-related bound-exciton recombination around 1.512 eV, (e, A^0), (h, D^0) and (D^0, A^0) pair and free-exciton recombination. Adapted from [840]

Fig. 10.9 for In in ZnO; the ($^{111}In^0$, X) transition disappears with the characteristic time constant close to that (97 h) of the nuclear decay of ^{111}In into stable ^{111}Cd. However, in such experiments it should be considered that the decay product and accompanying high-energy radiation can create new electronic and structural defects, respectively.

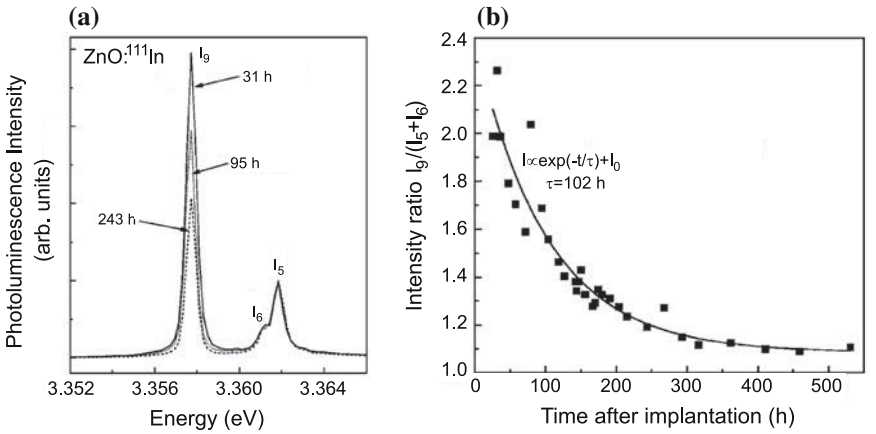


Fig. 10.9 (**a**) Low-temperature photoluminescence spectrum of ZnO implanted with ^{111}In featuring the so-called I_9-line. Spectra are recorded at various times after implantation as labeled. (**b**) Intensity of I_9-line as a function of time. Adapted from [850]

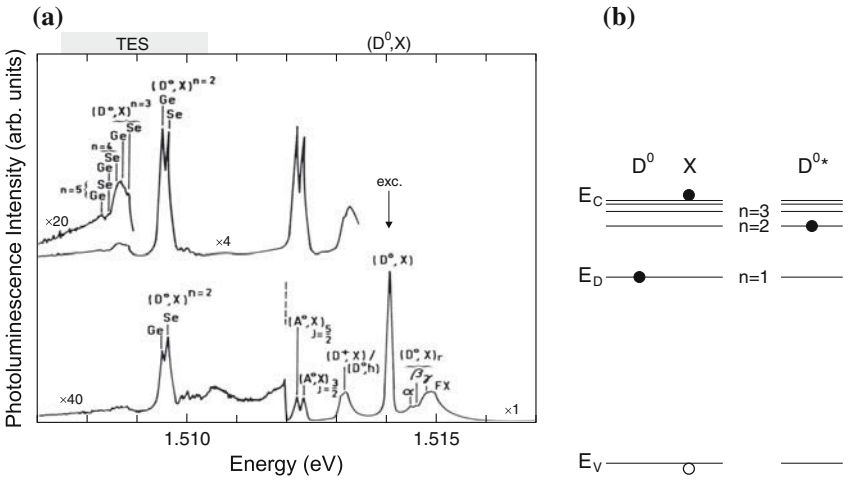


Fig. 10.10 (**a**) Photoluminescence spectrum ($T = 1.5$ K, $D = 50\,\mathrm{mW\,cm^{-2}}$) of high-purity GaAs with two donors (Ge and Se/Sn). The *lower* spectrum has been excited 6 meV above the band gap, the *upper* spectrum has been resonantly excited with the laser set to the (D^0, X) transition and exhibits $n = 2, 3, 4$, and 5 TES transitions. α, β, γ denote excited (hole rotational) states of the (D^0, X) complex. Adapted from [852]. (**b**) Schematic representation of the $n = 2$ TES process, *left* initial, *right* final state

The peak labeled $(D^0,X)_{2S}$ in Fig. 10.8 is called a two-electron satellite (TES) [851]. High-resolution spectra of the TES in GaAs [506, 852] are shown in Fig. 10.10a. The TES recombination is a (D^0,X) recombination that leaves the donor in an excited

state as schematically shown in Fig. 10.10b. Therefore a hydrogen-like series with $n = 2, 3, \ldots$ is observed with energies

$$E^n_{\mathrm{TES}} = E_{(\mathrm{D}^0,\mathrm{X})} - E^{\mathrm{b}}_{\mathrm{D}} \left(1 - \frac{1}{n^2}\right). \tag{10.25}$$

The effect of isotope disorder on the sharpness and splitting of impurity states has been investigated in [853, 854]. The recombination of excitons bound to Al, Ga and In in natural silicon (92.23 % ^{28}Si, 4.67 % ^{29}Si, 3.10 % ^{30}Si) is split into three lines due to the valley-orbit splitting [855] of electron states at the band minimum (Fig. 10.11). Each of these (A^0, X) lines is split by 0.01 cm^{-1} for Si:Al due to a symmetry reduction of the 4-fold degenerate A^0 ground state, as observed in the presence of applied axial strain or an electric field. The comparison to spectra from enriched ^{28}Si shows that the observed splitting without external perturbation is due to isotope disorder that causes random strains and splits the A^0 ground state into two doublets [854] (Fig. 10.11). Similarly, the (unsplit) phosphorus-induced (D^0, X) transition in enriched Si is found to be much sharper (<40 μeV) than in natural Si (330 μeV) [853]. At higher resolution, a hyperfine splitting of 485 neV due to the ^{31}P nuclear spin $I = 1/2$ (2×10^{12} cm^{-3}) in isotopically pure (99.991 %) ^{28}Si ($I = 0$) is observed for the (P^0, X) recombination [856]. In a magnetic field, the Zeeman-split lines have a FWHM of about 150 neV.

In Fig. 10.12 the recombination of excitons bound to the N isoelectronic impurity in lowly doped GaP is shown. The efficient recombination of nitrogen-bound electrons with holes at the Γ point is due to the wave-function component of the localized electron at $k = 0$ [612] (Fig. 7.41). The decay time of the A exciton is about 40 ns [857] and thus larger than the typical lifetime of excitons in direct semiconductors (ns-range). The forbidden B exciton has a much longer lifetime of 4 μs [857].

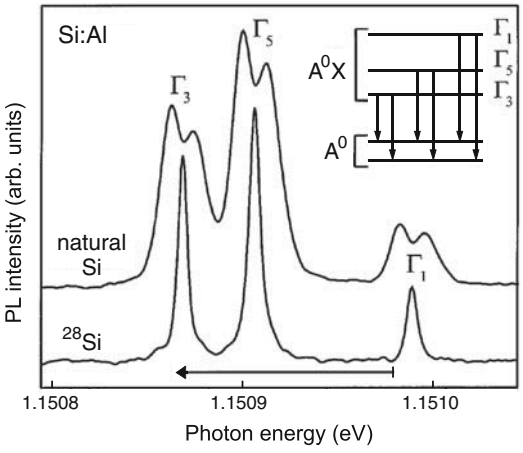


Fig. 10.11 High-resolution photoluminescence (PL) spectra of (A^0, X) recombination in natural and ^{28}Si-enriched silicon doped with aluminum ($T = 1.8$ K). The ^{28}Si PL spectrum is shifted up in energy by 0.114 meV, as indicated by the *arrow*, to compensate for the shift in band gap. The *inset* shows a level scheme for the recombination in natural silicon. Adapted from [854], reprinted with permission, ©2002 APS

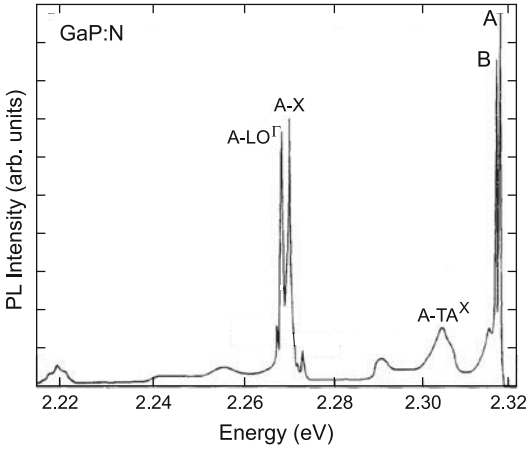


Fig. 10.12 Photoluminescence spectrum ($T = 4.2\,\mathrm{K}$) of GaP:N ($N_\mathrm{N} \approx 5 \times 10^{16}\,\mathrm{cm}^{-3}$). The A exciton is bound to an isolated nitrogen impurity, cmp. to Fig. 9.25. Adapted from [612]

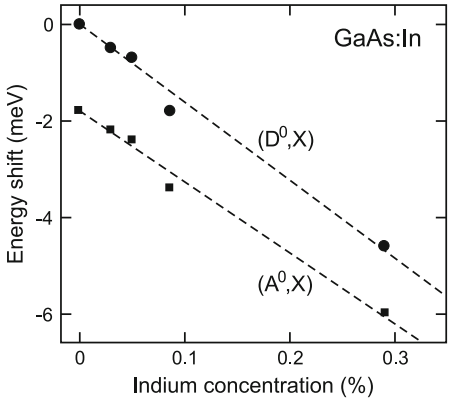


Fig. 10.13 Spectral position of neutral donor- and acceptor-bound exciton photoluminescence transition ($T = 2\,\mathrm{K}$) in GaAs doped with different amounts of indium relative to the donor-bound exciton luminescence in pure GaAs (1.5146 eV). Adapted from [858]

In the case of In in GaAs it has been found that down to the regime of $N_\mathrm{In} < 10^{19}\,\mathrm{cm}^{-3}$ the indium does not act as a substitutional isoelectronic impurity but still fully participates in the composition of a pseudo-binary system (Sect. 6.5). Recombination from excitons bound to single indium atoms or In–In pairs could not be found. The energy shift of donor- and acceptor-bound excitons in the dilute limit (Fig. 10.13) follows the band-gap dependence established for larger indium concentrations. The non-occurrence of localization effects is attributed to the small effective electron mass in InAs [472].

The luminescence intensity $I(T)$ of bound exciton lines is quenched with increasing temperature due to ionization of the excitons from the impurities. The temperature dependence can be modeled using the relation [859]

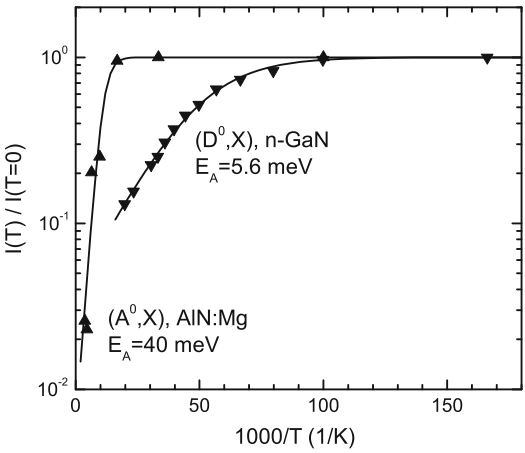


Fig. 10.14 Temperature dependent PL intensity of (D^0, X) in GaN and (A0, X) in AlN:Mg recombination. *Solid lines* are fits with (10.26). Data from [843, 861]

$$\frac{I(T)}{I(T=0)} = \frac{1}{1 + C\ \exp(-E_\mathrm{A}/kT)}, \tag{10.26}$$

E_A being the thermal activation energy and C a pre-factor. Often the activation energy is found equal to the localization energy, $E_\mathrm{A} = Q$ (Fig. 10.14, cmp. Table 10.2). If several processes contribute, additional exponential terms can be added with further activation energies. For acceptor-bound excitons in GaAs two processes are found to contribute, the ionization from the impurity into a free exciton ($E_\mathrm{A}^1 \approx Q$) and into an electron-hole pair ($E_\mathrm{A}^2 \approx Q + E_\mathrm{X}^\mathrm{b}$) [859]. In [860] the model is refined by considering the temperature dependence of the parameter C due to the ionization of the impurity itself.

So far *single* excitons bound to a center have been discussed. Also bound exciton *complexes* [862] containing up to six excitons have been observed at sufficient excitation density, e.g. for substitutional boron [863] or phosphorus [864] and interstitial Li [865] in silicon. In a multi-valley semiconductor several electrons are available to form bound excitons which follow approximately a shell model and exhibit further fine structure.

10.3.3 Alloy Broadening

The bound-exciton recombination peak in a binary compound is spectrally fairly sharp (Sect. 10.3.2), even in the presence of isotope disorder (Fig. 10.11). In an alloy (see Sect. 3.7), the random distribution of atoms (with different atomic order number Z) causes a significant broadening effect of the luminescence (and absorption) line, the so-called *alloy broadening* [866, 867]. As an example, we treat $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$. The

exciton samples, at different positions of the lattice, different coordinations of Ga and Al atoms. If the experiment averages over these configurations, an inhomogeneously broadened line is observed.

The cation concentration c_c for the zincblende lattice is given as $c_c = 4/a_0^3$, for the wurtzite lattice as $c_c = 4/(\sqrt{3}\,a^2 c)$. For example, $c_c = 2.2 \times 10^{22}\,\mathrm{cm}^{-3}$ for $Al_xGa_{1-x}As$ in the entire composition range $0 \le x \le 1$ since the lattice constant does not vary significantly, and $c_c = 4.2 \times 10^{22}\,\mathrm{cm}^{-3}$ for wurtzite $Mg_xZn_{1-x}O$ [868]. In a random alloy, the probability $p(N)$ to find exactly N Ga atoms in a given volume V (with a total of $c_c V$ cations) is given by the binomial distribution

$$p(N) = \binom{c_c V}{N} x^N (1-x)^{c_c V - N}. \tag{10.27}$$

The sampling volume for a luminescence event is the exciton volume (cf. (9.39)) that is given for the free-exciton (in 1s hydrogen state) as [866, 869]

$$V_{ex} = 10\,\pi\, a_X^3 = 10\,\pi \left(\frac{m_0}{m_r^*}\,\epsilon_s\, a_B\right)^3. \tag{10.28}$$

One should note that due to the variation of the involved material parameters V_{ex} depends itself on x. In GaAs there are about 1.2×10^6 cations in the exciton volume. In $Al_xGa_{1-x}As$, there are on average $x c_c V_{ex}$ Al atoms in the exciton volume. The fluctuation is given by the standard deviation of the binomial distribution [869]

$$\sigma_x^2 = \frac{x\,(1-x)}{c_c\, V_{ex}}. \tag{10.29}$$

The corresponding energetic broadening (full width at half-maximum) of the spectral line is given by $\Delta_E = 2.36\,\sigma$ with

$$\sigma = \frac{\partial E_g}{\partial x}\,\sigma_x = \frac{\partial E_g}{\partial x}\sqrt{\frac{x\,(1-x)}{c_c V_{ex}}}. \tag{10.30}$$

We note that instead of the quantum mechanically correct factor $10\,\pi$ [866, 869], often the factor $4\pi/3$ [867] is used, resulting in larger theoretical broadening.

Experimental data for Cd_xSe_{1-x} in Fig. 10.15a are consistent with (10.30). The theoretical dependence (10.30) is shown in Fig. 10.15b also for $Al_xGa_{1-x}As$ together with experimental data and found to disagree [870]. Since the exciton volume is much smaller (cf. 9.6.6) than in $Al_xGa_{1-x}As$, alloy broadening in $Mg_xZn_{1-x}O$ is much larger for a given for given x.

The spectral broadening due to alloy disorder masks the fine structure of recombination lines near the band edge present for binary semiconductors. Often for all temperatures only a single recombination line appears for alloys. Spectra for three different $Mg_xZn_{1-x}O$ alloys are shown in Fig. 10.16a. The increasing inhomogeneous

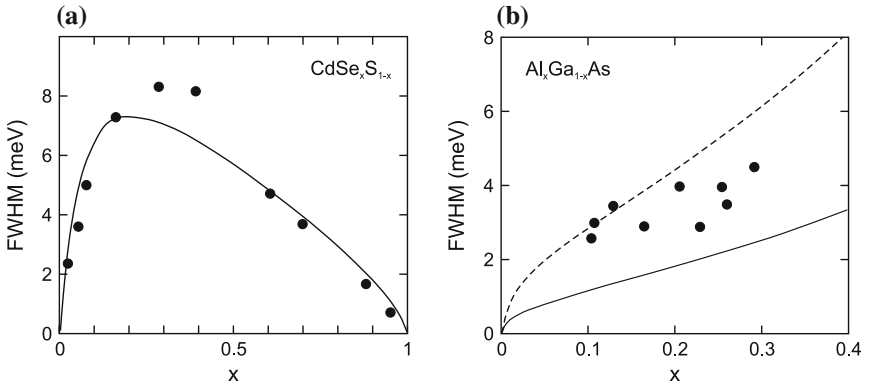


Fig. 10.15 (**a**) Spectral width of the photoluminescence from CdS_xSe_{1-x} alloys. *Solid line* is theory according to (10.30). Adapted from [866]. (**b**) Spectral width of the bound exciton recombination in $Al_xGa_{1-x}As$ with various Al content within the direct-bandgap regime. *Solid line* is (10.30) with (10.28), *dashed line* with pre-factor $4\pi/3$ instead of $10\,\pi$. Adapted from [870]

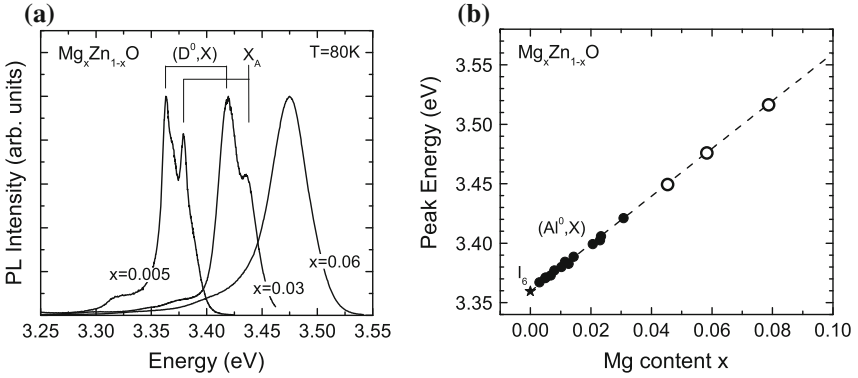


Fig. 10.16 (**a**) Photoluminescence spectra ($T = 80\,\mathrm{K}$, scaled) of three $Mg_xZn_{1-x}O$ alloy layers on sapphire with three different Mg-contents, $x = 0.005$, $x = 0.03$, and $x = 0.06$ as labeled. The energy positions of (D^0, X) and X_A peaks are marked. Adapted from [860]. (**b**) Peak energy of the photoluminescence spectrum ($T = 2\,\mathrm{K}$) of ZnO (I_6-line, *star*) and various $Mg_xZn_{1-x}O$ alloys (*circles*). For $x \leq 0.03$ (*filled circles*) the (D^0, X) recombination peak (Al donor) can be spectrally separated from the free exciton (X_A) recombination. For the samples with higher Mg content (*empty circles*) a single recombination peak is present at all temperatures. The *dashed line* is a linear least square fit for the alloys with $0 \leq x \leq 0.03$, showing that also for $x > 0.03$ the low temperature recombination peak is due to donor-bound excitons. Adapted from [871]

broadening is obvious, causing a single peak for $x > 0.03$. The temperature dependence of the peak positions is shown in Fig. 10.17 for the same samples. For $x = 0.005$ the bound exciton (Al-donor) (D^0, X) and free exciton (X_A) recombination lines can still be resolved despite the inhomogeneous broadening of $\sigma = 2.6\,\mathrm{meV}$. At

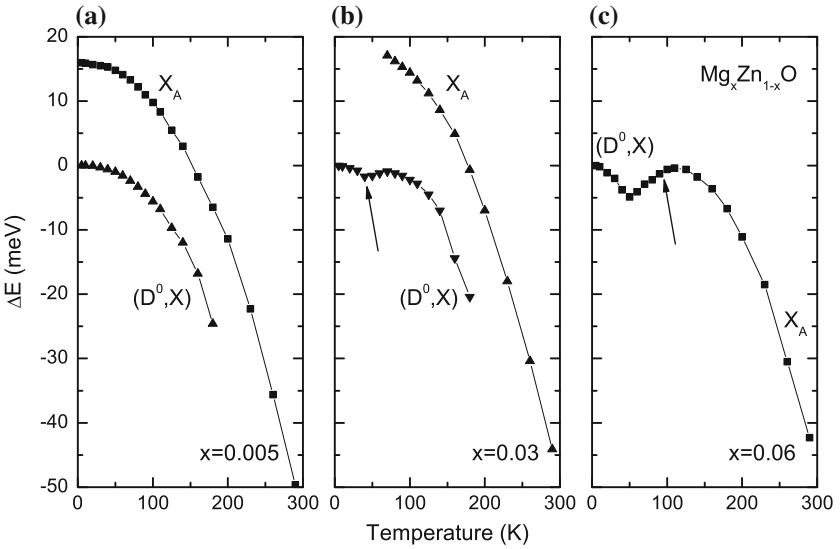


Fig. 10.17 Temperature dependence of the shift of energy position of (D^0, X) and X_A photoluminescence peak in $Mg_xZn_{1-x}O$ alloys with three different Mg-contents, (**a**) $x = 0.005$, (**b**) $x = 0.03$, (**c**) $x = 0.06$. The energy positions are given relative to the low temperature position of the respective (D^0, X) peaks. Adapted from [860]

low temperature the luminescence intensity is dominated by (D^0, X) recombination, at room temperature by free exciton (X_A) recombination. Both peaks are present at low temperatures and exhibit a red-shift with increasing temperatures due to the shrinking of the band gap (Fig. 10.17a). The (D^0, X) peak vanishes at about 180 K due to ionization of the excitons from the donors ($Q \approx 15$ meV, similar as in pure ZnO).

For larger Mg-content of $x = 0.03$ the two peaks can still mostly be separated ($\sigma = 6.0$ meV). The (D^0, X) energy position shows a small dip (about 2 meV) due to exciton localization in the alloy disorder potential (arrow in Fig. 10.17b). At low temperatures excitons are frozen in local potential minima and have a non-thermal (non-Boltzmann) population. With increasing temperature they can overcome energy barriers and thermalize, leading to a shift of the recombination peak to lower energies. Further increase of temperature populates higher levels and leads to a shift of the recombination peak to higher energies. Superimposed is the red-shift due to the band gap shrinkage. This so-called "S"-shape effect of $E(T)$ is discussed in Sect. 12.4 in detail with regard to exciton localization in a quantum well disorder potential.

For $x = 0.06$ only a single photoluminescence peak is observed for the alloy ($\sigma = 8.5$ meV). The (D^0, X) peak is the dominant for the $Mg_xZn_{1-x}O$ alloys at low temperatures even in the presence of large alloy broadening (Fig. 10.16b). The peak changes its nature from (D^0, X) at low temperatures to X_A at room temperature. In between, first exciton thermalization (red-shift) in the disorder potential and

subsequently exciton ionization from the donors (blue-shift, arrow in Fig. 10.17c) are observed [860]. Such exciton ionization from impurities has also been observed for AlGaN:Si [549, 872].

10.4 Phonon Replica

The momentum selection rule for free-exciton recombination allows only excitons with $\mathbf{K} \approx 0$ (for $\mathbf{K}$, cf. (9.37)) to recombine. The fine structure of this recombination is connected to polariton effects (cf. Sect. 9.6.8). Excitons with large $\mathbf{K}$ can recombine if a phonon or several phonons are involved [873] that provide the necessary momentum $\mathbf{q} = \mathbf{K}_1 - \mathbf{K}_2$, with $\mathbf{K}_1$ ($\mathbf{K}_2$) being the wave vector of the initial (intermediate) exciton state (Fig. 10.18). A so-called *zero-phonon* line at energy E_0 is then accompanied by *phonon replica* below E_0 at integer multiples (at low temperature) of the (LO) phonon energy $\hbar\omega_{\mathrm{ph}}$

$$E_n = E_0 - n\,\hbar\omega_{\mathrm{ph}}. \tag{10.31}$$

Phonon replicas have been observed in many polar semiconductors such as CdS [874] and ZnSe [875]. A sequence of such phonon replica, as observed in GaN [876], is depicted in Fig. 10.19a.

The lineshape of the nth phonon-assisted line is proportional to the exciton population at a given excess energy, which is proportional to the density of states and the Boltzmann distribution function [877]

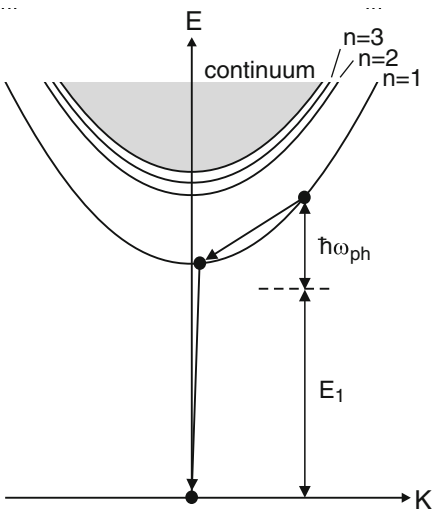


Fig. 10.18 Schematic representation of 1 LO exciton scattering of an exciton at $\mathbf{K} \neq 0$ to an intermediate state with $\mathbf{K} \approx 0$ and subsequent radiative decay. $\hbar\omega$ represents the phonon energy and E_1 the energy of the emitted photon

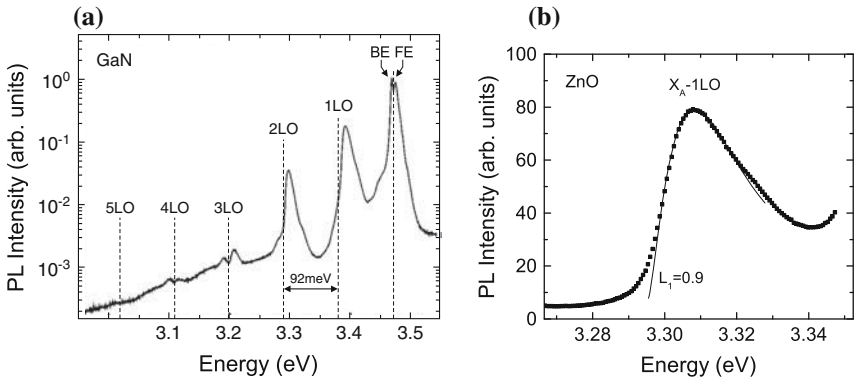


Fig. 10.19 (**a**) Photoluminescence spectrum of GaN (grown on SiC substrate) at $T = 50\,\mathrm{K}$. In addition to emission from free (FE) and bound (BE) excitons several phonon replica (labeled as 1LO–5LO) are observed. *Vertical dashed lines* indicate energy positions of multiple LO-phonon energies ($\hbar\omega_{\mathrm{LO}} = 92\,\mathrm{meV}$) below the FE peak. Adapted from [876]. (**b**) Photoluminescence spectrum of 1LO phonon-assisted recombination peak at $T = 103\,\mathrm{K}$ (from the data of Fig. 10.5). Data points (*dots*) and lineshape fit (*solid line*) according to (10.32) with the parameters $L_1 = 0.9$ and $E_1 = 3.2955\,\mathrm{eV}$ (and background)

$$I_n(E_{\mathrm{ex}}) \propto \sqrt{E_{\mathrm{ex}}}\,\exp\left(-\frac{E_{\mathrm{ex}}}{kT}\right) w_n(E_{\mathrm{ex}}). \tag{10.32}$$

Here, E_{ex} represents the exciton kinetic energy. The factor $w_n(E_{\mathrm{ex}})$ accounts for the $\mathbf{q}$-dependence of the matrix element. It is typically expressed as

$$w_n(E_{\mathrm{ex}}) \propto E_{\mathrm{ex}}^{L_n}. \tag{10.33}$$

Accordingly, as temperature dependent refinement of (10.31), the energy separation ΔE_n of the energy of the peak maximum of phonon replica from E_0 is given by

$$\Delta E_n = E_n - E_0 = -n\,\hbar\omega_{\mathrm{ph}} + \left(L_n + \frac{1}{2}\right) kT. \tag{10.34}$$

It is found theoretically that $L_1 = 1$ and $L_2 = 0$ [877]. These relations are approximately fulfilled for GaN [878]. A lineshape fit for the 1 LO phonon-assisted transition in ZnO is shown in Fig. 10.19b.

In Fig. 10.20a the 'green band' emission of ZnO is shown as presented in [879]. This band is mostly attributed to a Cu impurity; recently, evidence has grown from isotope decay and annealing studies that it is related to the zinc vacancy [880] (Fig. 10.20b). The zero phonon line is followed by many replica with a maximum at about 6 LO phonons. The intensity I_N of the N-th replica is given by [881, 882]

$$I_N \propto \exp(-S)\,\frac{S^N}{N!}, \tag{10.35}$$

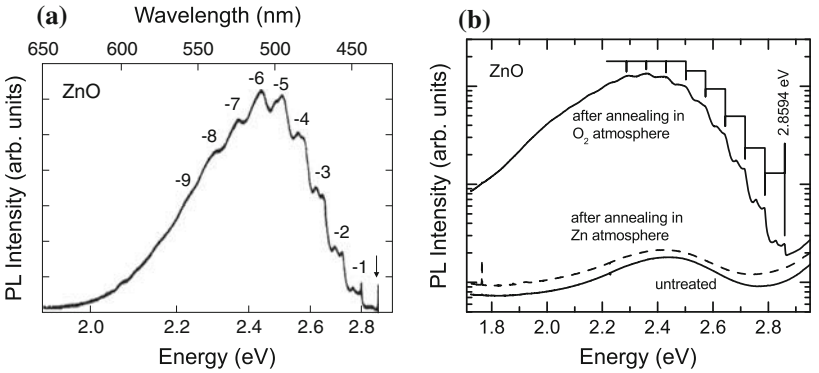


Fig. 10.20 (**a**) Luminescence spectrum of ZnO in the visible. The *arrow* denotes the zero-phonon line at 2.8590 eV. The numbers of the phonon replica are labeled. Adapted from [879]. (**b**) Luminescence spectra (*solid lines*) of a ZnO bulk crystal before ('untreated') and after annealing in O_2 atmosphere at $T = 1073$ K. After annealing in Zn atmosphere at the same temperature, the green band disappears again (*dashed line*). From [880]

where S is the so-called Huang–Rhys parameter. In [880], a coupling parameter of $S = 6.9$ has been determined.

Equation (10.35) is obtained from the consideration of transitions in the configuration diagram [881, 883] (Fig. 10.21), using the Born–Oppenheimer approximation. Here the electronic wavefunctions are separated from the vibrational wavefunctions, leading to the Franck–Condon principle, that optical transitions occur with the positions of the nuclei fixed and thus vertical in the configuration diagram Fig. 10.21. Assuming low temperatures, only the lowest state is (partially) occupied. The Huang–Rhys parameter, the average number of phonons involved in the transition, is related to the displacement $\delta q = q_1 - q_0$ of the two configurations

$$S = \frac{C\,\delta q^2}{2\,\hbar\omega_{\mathrm{ph}}}, \tag{10.36}$$

where C is the 'spring constant' of the parabola, $C = \mathrm{d}^2E/\mathrm{d}q^2$.

For small $S \ll 1$, we are in the weak coupling regime and the zero-phonon line is the strongest. In the strong coupling regime, $S > 1$, the maximum is (red-) shifted from the zero-phonon line. We note that in absorption, phonon replica occur on the high-energy side of the zero-phonon absorption. For large S the peak intensities are close to a Gaussian. The correspondence of emission and absorption is nicely seen for excitons on isoelectronic oxygen traps in ZnTe [882]. The oxygen is on substitutional Te site. Up to seven phonon replica are visible in Fig. 10.22 around the zero-phonon or A-line with a separation of about 26 meV, the optical phonon energy in ZnTe. The Huang–Rhys parameter is about 3–4. Other peaks are due to acoustic phonons.

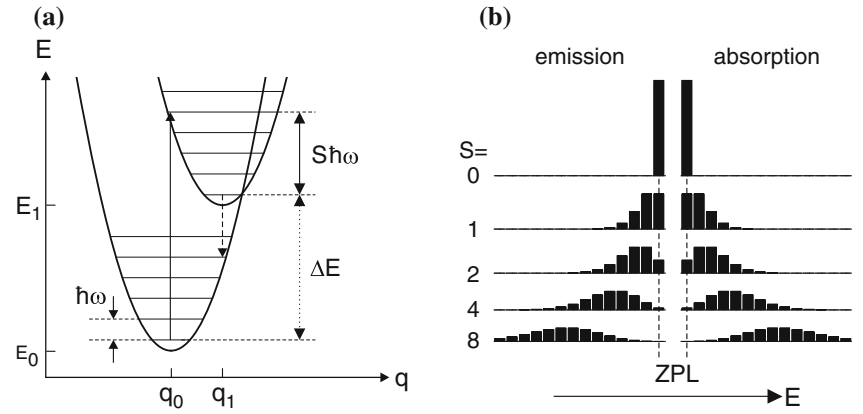


Fig. 10.21 (**a**) Configuration diagram of two states that differ in their configuration coordinate by $\delta q = q_1 - q_0$. Both are coupled to phonons of energy $\hbar\omega$. The absorption maximum (*solid vertical line*) and emission maximum (*dashed vertical line*) are shifted with respect to the zero-phonon line (*dotted vertical line*) with energy $E_1 - E_0$. The Huang–Rhys parameter is $S \sim 4$. (**b**) Intensity of zero-phonon line ('ZPL') and phonon replica (10.35) for emission and absorption processes with different values of the Huang–Rhys parameter S as labeled

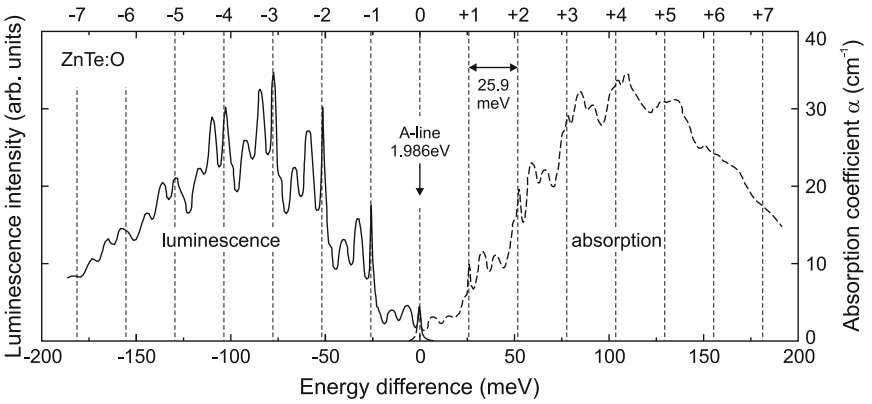


Fig. 10.22 Photoluminescence (*solid line*) and absorption (*dashed line*) spectra of excitons bound to substitutional oxygen in ZnTe at $T = 20\,\mathrm{K}$. The energy position is relative to the A-line at 1.9860 eV. The *vertical dashed lines* have a separation of 25.9 meV. Adapted from [884]

10.5 Self-Absorption

Luminescence that is emitted within the semiconductor can be (re-)absorbed before it may reach the surface and can leave the crystal. This effect is called self-absorption. It is particularly strong for radiation with an energy where the absorption $\alpha(\hbar\omega)$ is high, i.e. above the band gap of a direct semiconductor. Similarly to the penetration

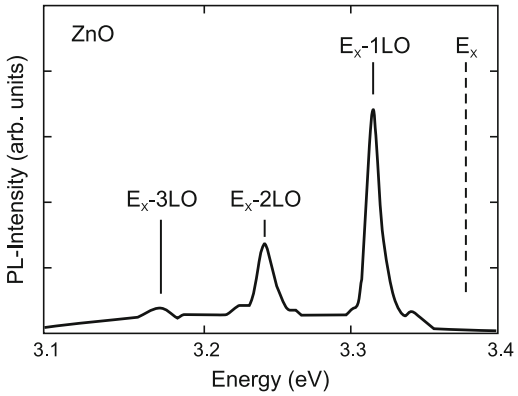


Fig. 10.23 Photoluminescence spectrum (at $T = 55\,\mathrm{K}$) from bulk ZnO excited homogeneously via two-photon excitation by a Q-switched ruby laser (pulse width 40 ns). Adapted from [885]

depth $1/\alpha$ for radiation entering the crystal, emission approximately occurs only from a layer of such thickness. For typical values of α in the range of $10^5\,\mathrm{cm}^{-1}$, the 'skin' of the semiconductor that emits light with energy above the band gap is 100 nm. For light at the low energy side of the band gap or with energy within the band gap (deep levels), the emission depth can be much larger.

After re-absorption, the energy has another chance to relax non-radiatively, thus reducing the quantum efficiency. Alternatively it can be reemitted, either at the same energy or at a lower energy. Possibly several re-absorption processes occur before a photon eventually leaves the semiconductor ('photon recycling'). Such processes are important in LED structures where photon extraction has to be optimized (Sect. 23.3.4). Emission on phonon replica (Sect. 10.4) is red-shifted from the energy range of strong absorption and thus suffers no (or only little) self-absorption. This can be seen from the spectrum of a thick ZnO crystal excited homogeneously (via two-photon absorption with a red Ruby laser), Fig. 10.23. The zero phonon line (at E_X), originating from the ≈100 nm skin of the samples and being by far the strongest in thin films (Fig. 10.5), is practically absent and emission on the phonon replica collected from the entire volume dominates the spectrum.

10.6 Donor–Acceptor Pair Transitions

Optical transitions can occur between neutral donors and acceptors. The (spatially indirect) donor–acceptor pair (DAP) recombination is present in (partially) compensated semiconductors and follows the scheme $D^0A^0 \rightarrow D^+A^-eh \rightarrow D^+A^- + \gamma$, where γ is a photon with the energy $\hbar\omega$. The energy of the emitted photon is given by

$$\hbar\omega = E_\mathrm{g} - E_\mathrm{D}^\mathrm{b} - E_\mathrm{A}^\mathrm{b} + \frac{1}{4\pi\epsilon_0}\frac{e^2}{\epsilon_\mathrm{r} R}, \tag{10.37}$$

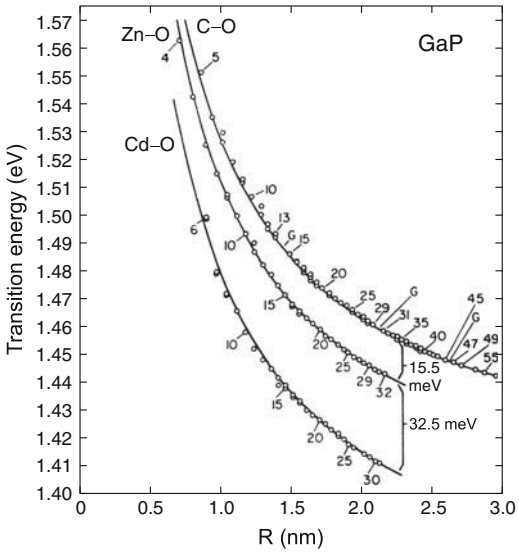


Fig. 10.24 Transition energies in GaP ($T = 1.6\,\mathrm{K}$) of the donor–acceptor recombination involving the deep oxygen donor and C, Zn, and Cd acceptors, respectively. The lines follow (10.37) for $E_g^{\mathrm{GaP}} = 2.339\,\mathrm{eV}$, $\epsilon_r = 11.1$ and $(E_D^b)_O = 893\,\mathrm{meV}$, $(E_A^b)_C = 48.5\,\mathrm{meV}$, $(E_A^b)_{Zn} = 64\,\mathrm{meV}$, and $(E_A^b)_{Cd} = 96.5\,\mathrm{meV}$. Predicted missing modes for GaP:C, O are labeled with 'G'. Adapted from [886]

where R is the distance between the donor and the acceptor for a specific pair. Since R is discrete, the DAP recombination spectrum consists of several discrete lines. If the donor and acceptor occupy the same sublattice, e.g. O and C both substituting P sites in GaP, the spatial distance of the donor and acceptor is $R(n) = a_0\sqrt{n/2}$, where a_0 is the lattice constant and n is an integer. However, for certain 'magic' numbers $n = 14, 30, 46, \ldots$ no lattice points exist and therefore the corresponding lines are missing (labeled 'G' in Fig. 10.24). No such gaps exist in DA spectra where donors and acceptors occupy different sublattices, e.g. GaP:O, Zn (see also Fig. 10.24). In this case, the spatial separation is given by $R(n) = a_0\sqrt{n/2 - 5/16}$. If significant broadening is present, the lines are washed out and a donor–acceptor pair band forms.

10.7 Inner-Impurity Recombination

The transitions of electrons between different states of an impurity level can be nonradiative or radiative. As an example, the radiative transition of electrons in the Fe^{2+} state in InP $^5T_2 \rightarrow {}^5E$ (Fig. 10.25) and its fine structure were observed first in [888] at around 0.35 eV.

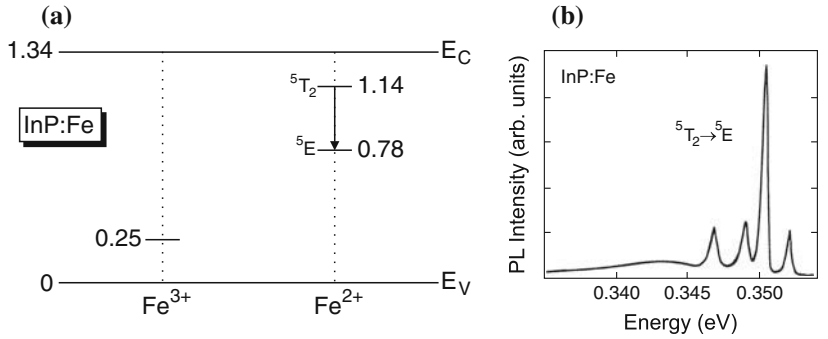


Fig. 10.25 (**a**) Schematic band diagram of InP with levels of Fe impurities in the 3+ and 2+ charge states at low temperature. All energies are given in eV. The *arrow* denotes the optical transition from an excited Fe^{2+} state to the Fe^{2+} ground state. (**b**) Photoluminescence spectrum (at $T = 4.2\,\mathrm{K}$) of InP:Fe sample with $[\mathrm{Fe}] = 5 \times 10^{16}\,\mathrm{cm}^{-3}$. Part (**b**) adapted from [887]

10.8 Auger Recombination

In competition with the radiative, bimolecular recombination is the Auger recombination (Fig. 10.26). In the Auger process, the energy that is released during the recombination of an electron and hole is not emitted with a photon but, instead, transferred to a third particle. This can be an electron (eeh, Fig. 10.26a) or a hole (hhe, Fig. 10.26b). The energy is eventually transferred nonradiatively from the hot third carrier via phonon emission to the lattice. The probability for such process is $\propto n^2 p$ if two electrons are involved and $\propto n p^2$ if two holes are involved. The Auger process is a three-particle process and becomes likely for high carrier density, either through doping, in the presence of many excess carriers, or in semiconductors with small band gap. Auger recombination is the inverse of the impact ionization (cf. Sect. 8.4.4). Phonon-assisted Auger recombination relaxes the momentum conservation rule for the involved charge carriers at the cost of an additional particle being involved in the scattering process. It has been pointed out that this process is dominating in bulk material [889, 890] and quantum wells [891].

In thermodynamic equilibrium the rates for Auger recombination and thermal Auger generation must be equal, thus

$$G_{\mathrm{th}} = C_{\mathrm{n}}\, n_0^2\, p_0 + C_{\mathrm{p}}\, n_0\, p_0^2, \tag{10.38}$$

where C_{n} and C_{p} denote the Auger recombination coefficients. The equation for the dynamics in the presence of excess carriers (if solely Auger recombination is present) is given as

$$\frac{\partial\, \delta n}{\partial t} = G_{\mathrm{th}} - R = -C_{\mathrm{n}}\,(n^2\, p - n_0^2\, p_0) - C_{\mathrm{p}}\,(n\, p^2 - n_0\, p_0^2). \tag{10.39}$$

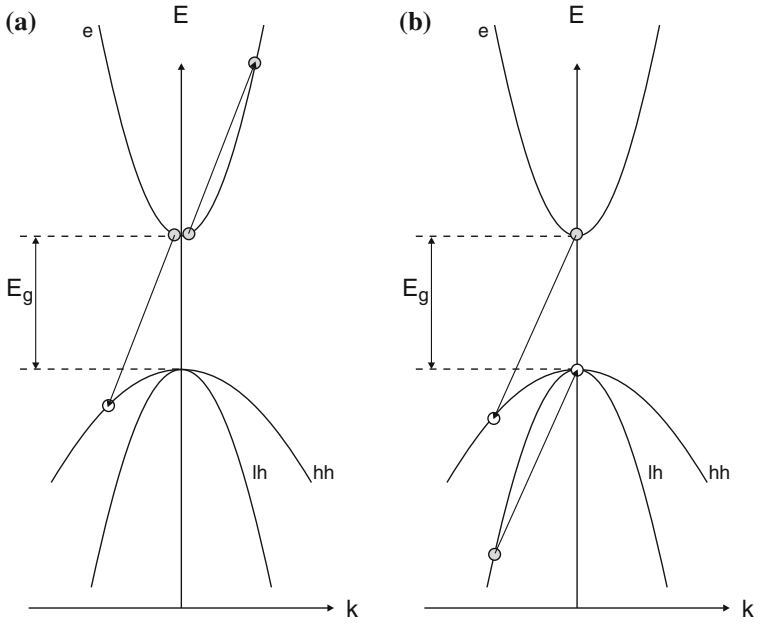


Fig. 10.26 Schematic representation of Auger recombination. An electron recombines with a hole and transfers the energy to (**a**) another electron in the conduction band, (**b**) another electron in the valence band

The Auger recombination rate typically used in SRH kinetics is

$$r_{\text{Auger}} = (C_\text{n}\, n + C_\text{p}\, p)\,(np - n_0 p_0)\ . \tag{10.40}$$

Typical values for the Auger recombination coefficients are given in Table 10.3.

In Fig. 10.27a the electron lifetime in heavily p-doped InGaAs (lattice matched to InP) is shown [893]. It follows $\tau_\text{n}^{-1} = C_\text{p}\, N_\text{A}^2$ as expected from (10.39) for p-type material. The Auger process in silicon has been discussed in detail [894]. In Fig. 10.27b experimental data for n-type and p-type Si are summarized. Auger theory can predict the lifetimes in n-type material. The predicted rate in p-type material is too small, thus a phonon-assisted process is evoked [894].

Table 10.3 Auger recombination coefficients for some semiconductors. Data for InSb from [892], SiC from [830], others from [831]

Material	C_n (cm^6/s)	C_p (cm^6/s)
4H-SiC	5×10^{-31}	2×10^{-31}
Si, Ge	2.8×10^{-31}	9.9×10^{-32}
GaAs, InP	5.0×10^{-30}	3.0×10^{-30}
InSb	1.2×10^{-26}	

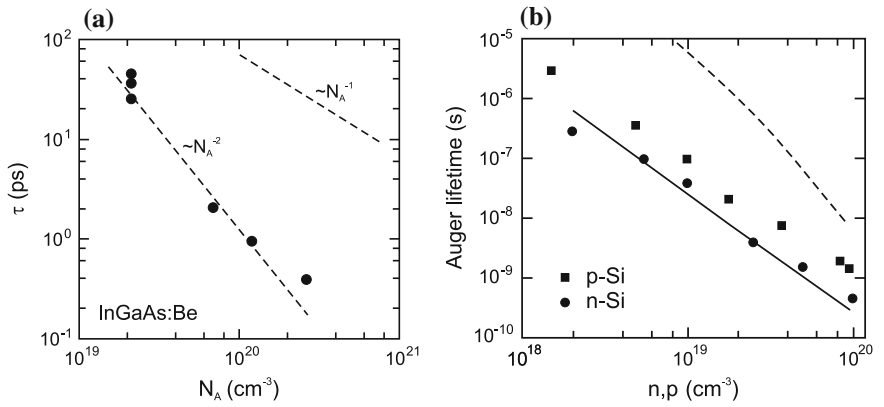


Fig. 10.27 (**a**) Experimental values of the electron lifetime in heavily p-doped InGaAs on InP at room temperature. The *dashed lines* show dependencies of Auger ($\propto N_A^{-2}$, $C_p = 8.1 \times 10^{-29}\,\mathrm{cm}^{-6}\,\mathrm{s}^{-1}$) and band-band recombination ($\propto N_A^{-1}$, $B = 1.43 \times 10^{-10}\,\mathrm{cm}^{-3}\,\mathrm{s}^{-1}$). Adapted from [893]. (**b**) Experimental Auger lifetimes in p-type (*squares*) and n-type (*circles*) silicon at 300 K. The *dashed* (*solid*) *line* is theory for p-type (n-type) material. Adapted from [894]

10.9 Band–Impurity Recombination

A very important recombination process is the capture of carriers by impurities. This process is in competition with all other recombination processes, e.g. the radiative recombination and the Auger mechanism. The band–impurity recombination is the inverse process to the carrier release from impurities and intimately related to carriers statistics (Chap. 7). It is particularly important at low carrier densities, for high dopant concentration and in indirect semiconductors since for these the bimolecular recombination is slow. This process is generally considered to be non-radiative since no photons close to the band edge are emitted.[4]

10.9.1 Shockley–Read–Hall Kinetics

The theory of capture on and recombination involving impurities is called Shockley–Read–Hall (SRH) kinetics [825]. An example of radiative band–impurity recombination (of the type shown in Fig. 10.28a) is shown in Fig. 10.8 (e, A^0) recombination at the carbon acceptor in GaAs.

[4]Depending on the energetic depth of the trap, mid or far infrared photons can be emitted.

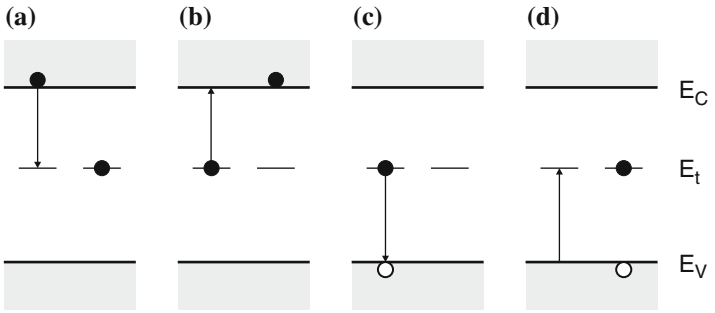


Fig. 10.28 Band-to-impurity processes at an impurity with one level (*left* initial, *right* final state in each part): (**a**) electron capture (from conduction band), (**b**) electron emission (into conduction band), (**c**) hole capture (from valence band), (**d**) hole emission (into valence band). The *arrows* indicate the transition of the electron

We consider electron traps [895] (see Fig. 10.28) with a concentration N_t with an energy level E_t. In thermodynamic equilibrium they have an electron population

$$f_t^0 = \frac{1}{\exp\left(\frac{E_t - E_F}{kT}\right) + 1}, \tag{10.41}$$

where f_t is the nonequilibrium population of the trap. Then the capture rate r_c is proportional to the unoccupied traps and the electron concentration, $r_c \propto n N_t(1-f_t)$. The proportionality factor has the form $v_{th}\sigma_n$, where v_{th} is the thermal velocity $v_{th} = \sqrt{3kT/m^*} \approx 10^7$ cm/s and σ_n is the capture cross section that is of atomic scale, typically $\sim 10^{-15}\text{cm}^2$. The capture cross section can be related to the optical absorption cross section [513, 514].

In order to make the following calculation more transparent, we put the effective-mass ratio $\sqrt{m_0/m^*}$ into σ in the following and thus have the same thermal velocity $v_{th} = \sqrt{3kT/m_0}$ for electrons and holes. The capture rate of electrons is

$$r_c = v_{th}\, \sigma_n\, n\, N_t\, (1 - f_t). \tag{10.42}$$

The emission rate from filled traps is

$$g_c = e_n\, N_t\, f_t, \tag{10.43}$$

where e_n denotes the emission probability. In a similar way, the emission and capture rates for holes can be written:

$$r_v = v_{th}\, \sigma_p\, p\, N_t\, f_t \tag{10.44}$$

$$g_v = e_p\, N_t\, (1 - f_t). \tag{10.45}$$

In thermodynamical equilibrium, capture and generation rates are equal, i.e. $r_c = g_c$ and $r_v = g_v$. Thus, the emission probability is

$$e_n = v_{th}\,\sigma_n\,n_0\,\frac{1-f_t^0}{f_t^0}. \tag{10.46}$$

Using $\frac{1-f_t^0}{f_t^0} = \exp\left(\frac{E_t-E_F}{kT}\right)$, (7.10) and (7.11) the emission probabilities can be written as

$$e_n = v_{th}\,\sigma_n\,n_t \tag{10.47}$$
$$e_p = v_{th}\,\sigma_p\,p_t, \tag{10.48}$$

with

$$n_t = N_C\,\exp\left(\frac{E_t-E_C}{kT}\right) \tag{10.49}$$
$$p_t = N_V\,\exp\left(-\frac{E_t-E_V}{kT}\right). \tag{10.50}$$

We note that $n_t\,p_t = n_0\,p_0$ (cf. (7.15)).

The temperature dependence of the thermal velocity is $\propto T^{1/2}$, the temperature dependence of the band-edge density of states is $\propto T^{3/2}$ (7.8) and (7.9). Thus,the temperature dependence of the emission rate e_n is (apart from the exponential term) $\propto T^2$ if σ is temperature independent.

Charge conservation requires in nonequilibrium (and of course in equilibrium) $r_c - r_v = g_c - g_v$. From this we obtain the population of the trap in nonequilibrium:

$$f_t = \frac{\sigma_n\,n + \sigma_p\,p_t}{\sigma_n\,(n+n_t) + \sigma_p\,(p+p_t)}. \tag{10.51}$$

The recombination rate r_{b-i} of the band–impurity recombination is then

$$\begin{aligned} r_{b-i} &= -\frac{\partial \delta n}{\partial t} \\ &= r_c - g_c \\ &= \frac{\sigma_n\,\sigma_p\,v_{th}\,N_t}{\sigma_n\,(n+n_t) + \sigma_p\,(p+p_t)}\;(n\,p - n_0\,p_0)\,. \end{aligned} \tag{10.52}$$

Using the 'lifetimes'

$$\tau_{n_0} = (\sigma_n\,v_{th}\,N_t)^{-1} \tag{10.53}$$
$$\tau_{p_0} = (\sigma_p\,v_{th}\,N_t)^{-1}, \tag{10.54}$$

this is typically written as

$$r_{\mathrm{b-i}} = \frac{1}{\tau_{p_0}\,(n + n_{\mathrm{t}}) + \tau_{n_0}\,(p + p_{\mathrm{t}})}\,(n\,p - n_0\,p_0)\,. \tag{10.55}$$

For an n-type semiconductor the Fermi level is above E_{t} and the traps are mostly full. Thus hole capture is the dominating process. The equation for the dynamics simplifies to

$$\frac{\partial \delta p}{\partial t} = -\frac{p - p_0}{\tau_{p_0}}. \tag{10.56}$$

Thus, an exponential decay with minority-carrier lifetime τ_{p_0} (or τ_{n_0} for p-type material) occurs.

A recombination center is most effective when it is close to the middle of the band gap (midgap level). The condition $\partial r_{\mathrm{b-i}}/\partial E_{\mathrm{t}} = 0$ leads to the trap energy $E_{\mathrm{t}}^{\max}$ with the maximum recombination rate being located at

$$E_{\mathrm{t}}^{\max} = \frac{E_{\mathrm{C}} + E_{\mathrm{V}}}{2} - kT\,\ln\left(\frac{\sigma_{\mathrm{n}}\,N_{\mathrm{C}}}{\sigma_{\mathrm{p}}\,N_{\mathrm{V}}}\right). \tag{10.57}$$

The curvature $\partial^2 r_{\mathrm{b-i}}/\partial E_{\mathrm{t}}^2$ at $E_{\mathrm{t}}^{\max}$ is proportional to $-(np - n_0 p_0)$ and thus indeed is negative in the presence of excess carriers. However, the maximum can be fairly broad.

The SRH kinetic presented here is valid for low densities of recombination centers. A more detailed discussion and a more general model can be found in [896].

A typical example for a recombination center is gold in silicon. The minority carrier lifetime decreases from 2×10^{-7} s to 2×10^{-10} s upon increase of the Au concentration from 10^{14} to $10^{17}\,\mathrm{cm}^{-3}$. The incorporation of recombination centers is an important measure for the design of high-frequency devices [897]. Due to importance in silicon technology the recombination properties of many metals in silicon have been investigated, in particular Fe-contamination and the role of FeB-complexes [898–900].

A reduction in minority-carrier lifetime can also be achieved by irradiation with high-energy particles and the subsequent generation of point defects with energy levels at midgap.

In Fig. 10.29 various data on minority carrier lifetime in silicon are compiled. Over some doping range, a dependence of the lifetime $\propto N^{-1}$ as in (10.53) prevails. For doping beyond the $10^{19}\,\mathrm{cm}^{-3}$ range, Auger recombination (Sect. 10.8) with $\tau \propto N^{-2}$ is dominant. A more detailed discussion can be found in [901, 902]. Generally the lifetimes are temperature dependent [903] as expected from (10.52).

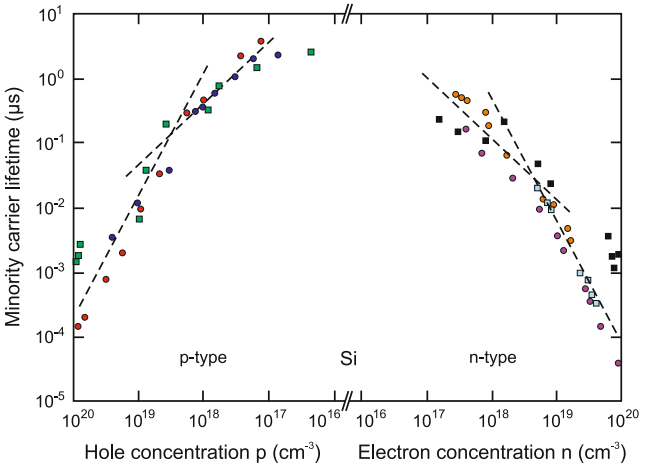


Fig. 10.29 Minority carrier lifetime at room temperature as a function of majority carrier concentration in n-type and p-type silicon. The *dashed lines* have the slopes N^{-1} and N^{-2}. Data from [902]

10.9.2 Multilevel Traps

Traps with multiple levels in the band gap have generally similar but more complicated dynamics as compared to single-level traps. Lifetimes are an average over negatively and positively charged states of the trap.

10.10 ABC Model

Summarizing the results on band-impurity recombination (Sect. 10.9), bimolecular recombination (Sect. 10.2) and Auger recombination (Sect. 10.8), the total recombination rate R can be written simplified as

$$R = A\,n + B\,n^2 + C\,n^3, \tag{10.58}$$

where A is the coefficient for the band-impurity recombination, B the bimolecular recombination coefficient and C the Auger recombination coefficient; n denotes the carrier density. This model is known as the 'ABC' model. It can be refined separating effects of electrons and holes and including higher terms. Often such model is used to investigate recombination in devices as a function of injection, e.g. [904, 905].

The internal radiative quantum efficiency η_{int} is given by ratio of the radiative recombination rate and the total recombination rate,

$$\eta_{\text{int}} = \frac{B\,n^2}{A\,n + B\,n^2 + C\,n^3} = \frac{B\,n}{A + B\,n + C\,n^2}. \tag{10.59}$$

10.11 Field Effect

The emission of electrons from a trap is thermally activated with an ionization energy $E_\mathrm{i} = E_\mathrm{C} - E_\mathrm{t}$. If the trap is in a strong electric field E, the emission probability can change. An acceptor-like trap after removal of the electron is neutral and its potential is short range. A donor has a long-range Coulomb potential after ionization. In an electric field, these potentials are modified as visualized in Fig. 10.30. Various additional processes can now occur.

10.11.1 Thermally Activated Emission

For the δ-like potential the ionization energy remains unchanged. For the Coulomb potential the barrier in the field direction is lowered by

$$\Delta E_\mathrm{i} = e\sqrt{\frac{e}{\pi\,\epsilon_0\,\epsilon_r}}\,\sqrt{E}\;. \tag{10.60}$$

The emission rate e_n is increased in the field by $\exp(\Delta E_\mathrm{i}/kT)$. This effect is called the Poole–Frenkel effect [906] and can be quite important. For silicon and $E = 2 \times 10^5\,\mathrm{V/cm}$ and $\Delta E_\mathrm{i} = 100\,\mathrm{meV}$ a 50-fold increase of the emission rate at room temperature is expected. As an example the Poole–Frenkel effect for the electron emission from (neutral) interstitial boron in silicon ($\mathrm{B_i^0} \rightarrow \mathrm{B_i^+} + e^-$) is

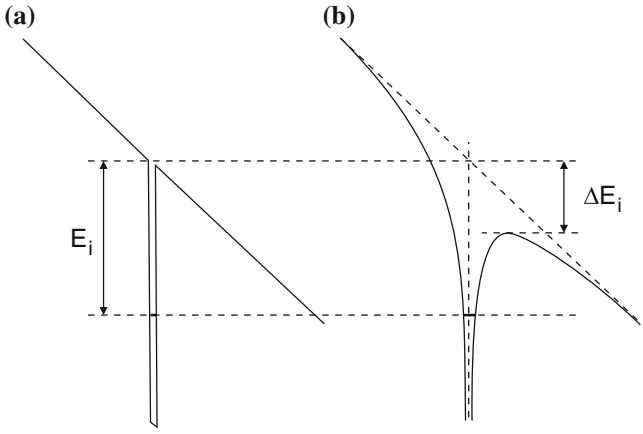


Fig. 10.30 Field effect at (**a**) a δ-like potential and (**b**) a Coulomb potential

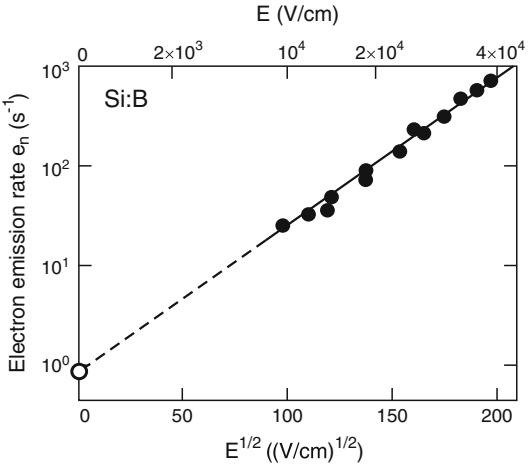


Fig. 10.31 Field effect for electron emission from interstitial boron in silicon ($T = 65\,\mathrm{K}$). The *filled circles* represent experimental data from DLTS, the data point marked with an *empty circle* is from EPR (zero field). The *line* is a linear fit and extrapolation. Adapted from [907]

shown in Fig. 10.31, following the enhancement of $e_\mathrm{n} \propto \exp(\sqrt{E})$. The extrapolation to $E = 0$ agrees with the EPR result [241, 907].[5]

10.11.2 Direct Tunneling

Carriers can tunnel from the trap level through the barrier in the field direction into the conduction band. This process is temperature independent. The transmission factor of a barrier is (in WKB approximation) proportional to $\exp[-(2/\hbar) \int \sqrt{2m\,[V(x) - E]}\,\mathrm{d}E]$. The emission probability for a triangular barrier is then

$$e_\mathrm{n} = \frac{e\,E}{4\,\sqrt{2m^* E_\mathrm{i}}}\,\exp\left(-\frac{4\,\sqrt{2m^*}\,E_\mathrm{i}^{3/2}}{3\,e\,\hbar\,E}\right). \tag{10.61}$$

In the case of a Coulomb-like potential the argument of the exponent in (10.61) needs to be multiplied by a factor $1 - (\Delta E_\mathrm{i}/E_\mathrm{i})^{5/3}$ with ΔE_i from (10.60).

10.11.3 Assisted Tunneling

In a thermally assisted tunneling process the electron on the trap level is first excited to a virtual level $E_\mathrm{t} + E_\mathrm{ph}$ by phonon absorption and then tunnels out of the trap

[5] The slope of the line in Fig. 10.31 is slightly smaller than expected from (10.60).

(photon-assisted tunneling). From the energetically higher level the tunneling rate is higher. The probability is proportional to $\exp\left(E_{\mathrm{ph}}/kT\right)$. The additional energy can also be supplied by a photon (photon-assisted tunneling).

10.12 Recombination at Extended Defects

10.12.1 Surfaces

A surface (cmp. Chap. 11) is typically a source of recombination, e.g. through midgap levels induced by the break of crystal symmetry. The recombination at surfaces is modeled as a recombination current

$$j_{\mathrm{s}} = -e\,S\,(n_{\mathrm{s}} - n_0), \tag{10.62}$$

where n_{s} is the carrier density at the surface and S is the so-called *surface recombination velocity*.

The surface recombination velocity for GaAs is shown in Fig. 10.32. For InP, if the surface Fermi level is pinned close to midgap, the surface recombination velocity increases from $\sim 5 \times 10^3$ cm/s for a doping level of $n \sim 3 \times 10^{15}\,\mathrm{cm}^{-3}$ to $\sim 10^6$ cm/s for a doping level of $n \sim 3 \times 10^{18}\,\mathrm{cm}^{-3}$ [908]. For Si, the surface recombination rate depends on the treatment of the surface and lies in the range between 1–10^4 cm/s [911]. The Si-SiO_2 interface can exhibit $S \leq 0.5$ cm/s. Time-resolved measurements and detailed modeling for Si have been reported in [911].

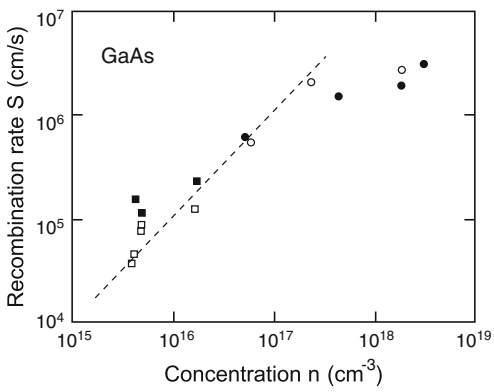


Fig. 10.32 Surface recombination velocity for GaAs as a function of n-type doping concentration. Different experimental points correspond to different surface treatment methods. *Dashed line* is a guide to the eye. Experimental data from [912]

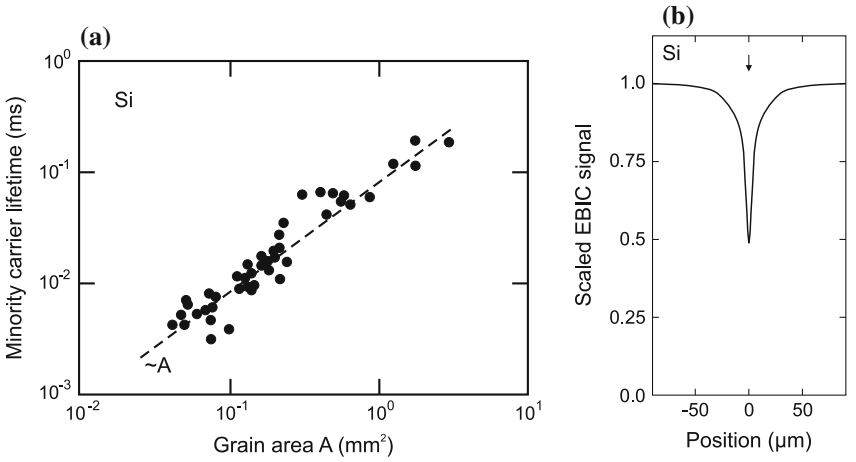


Fig. 10.33 (**a**) Minority carrier lifetime in (p-type) silicon as a function of grain boundary size. The *dashed line* has the slope $\propto A$. Data from [916]. (**b**) Linescan of the electron beam induced current (EBIC) perpendicular to a single grain boundary in silicon. The *arrow* denotes the position of the grain boundary. Compiled from [917]

10.12.2 Grain Boundaries

Grain boundaries can be a source of non-radiative recombination. This is technologically important for solar cells made from polycrystalline silicon (cf. Sect. 22.4.6). The grain boundary can be understood as an inner surface in the crystal. Modelling of recombination at a grain boundary can be done using an interface recombination velocity [913, 914] or considering deep traps [915]. The minority carrier lifetime decreases with decreasing grain boundary area A (Fig. 10.33a). The carrier loss at a grain boundary can be imaged directly via the efficiency of the collection of an electron beam induced current (EBIC) as shown in Fig. 10.33b. The minority carrier lifetime is only unaffected when the average distance to a grain boundary is much larger than the minority carrier diffusion length, $\sqrt{A} \gg L_\mathrm{D}$, otherwise the entire grain volume is subject to non-radiative recombination.

10.12.3 Dislocations

Also dislocations typically act as recombination centers, sometimes called carrier sinks. In Fig. 10.34 it can be seen that the minority carrier lifetime depends on the dislocation density n_d and follows a $\tau^{-1} \propto n_\mathrm{d}$ law, as if each dislocation is a recombination center [918]. The non-radiative recombination makes dislocations appear as 'dark line defects' in luminescence imaging [919]. In [920] also the decrease of carrier lifetime around (misfit) dislocations has been imaged. The effect of dislocations on the radiative recombination efficiency depends on the diffusion length [920].

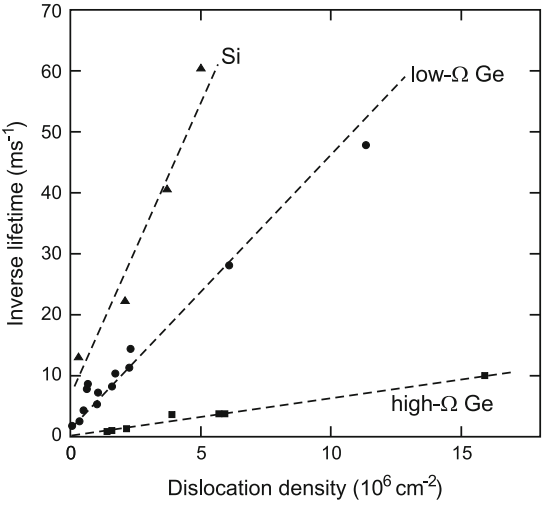


Fig. 10.34 Inverse minority carrier lifetime in n-type silicon (40 Ω cm), low resistivity Ge (3–5 Ω cm) and high resistivity Ge (30–40 Ω cm). Data from [918]

10.13 Excess-Carrier Profiles

In this section, some typical excess-carrier profiles (in one-dimensional geometry) are discussed that arise from certain excitation conditions. The excess-carrier density Δp (here holes in an n-type semiconductor, i.e. $\Delta p = p_n - p_{n_0}$) is determined by the diffusion equation (cf. (8.64a))

$$D_p \frac{\partial^2 \Delta p}{\partial x^2} = -G(x) + \frac{\Delta p}{\tau_p} . \tag{10.63}$$

10.13.1 Generation at Surface

First, the generation of excess carriers in a semi-infinite piece of semiconductor shall occur only at the surface at $x = 0$. The generation is zero everywhere else and the excitation is incorporated via the boundary condition $\Delta p(x = 0) = \Delta p_0$. The general solution for the homogeneous equation (10.63), i.e. $G = 0$, is

$$\Delta p(x) = C_1 \exp\left(-\frac{x}{L_p}\right) + C_2 \exp\left(\frac{x}{L_p}\right), \tag{10.64}$$

with the diffusion length $L_p = \sqrt{D_p \tau_p}$. Taking the boundary condition $\Delta p(x \to \infty) = 0$ the solution is

$$\Delta p(x) = \Delta p_0 \exp\left(-\frac{x}{L_p}\right). \tag{10.65}$$

In order to connect Δp_0 with the total generation rate per unit area G_{tot}, we calculate

$$G_{\text{tot}} = \int_0^\infty \frac{\Delta p(x)}{\tau_\text{p}}\,\mathrm{d}x = \frac{\Delta p_0\, L_\text{p}}{\tau_\text{p}} = \Delta p_0 \sqrt{\frac{D_\text{p}}{\tau_\text{p}}}. \tag{10.66}$$

10.13.2 Generation in the Bulk

Now, a generation rate following (9.16), realistic for photodiodes and solar cells, is considered,

$$G(x) = G_0\, \exp\left(-\alpha\, x\right), \tag{10.67}$$

i.e. due to light absorption with the (wavelength-dependent) absorption coefficient α. The total generation rate is

$$G_{\text{tot}} = \int_0^\infty G(x)\,\mathrm{d}x = \frac{G_0}{\alpha}. \tag{10.68}$$

The total generation rate is equal to the number of photons per second Φ_0 that enter the semiconductor.

The solution of (10.63) is the sum of the homogeneous solution (10.64) and a particular solution that is given by

$$\Delta p(x) = C\, \exp\left(-\alpha\, x\right). \tag{10.69}$$

The constant C is determined to be

$$C = \frac{G_0\, \tau_\text{p}}{1 - \alpha^2\, L_\text{p}^2}. \tag{10.70}$$

Therefore, the solution is

$$\Delta p(x) = C_1 \exp\left(-\frac{x}{L_\text{p}}\right) + C_2 \exp\left(\frac{x}{L_\text{p}}\right) + \frac{G_0\, \tau_\text{p}}{1 - \alpha^2\, L_\text{p}^2}\, \exp\left(-\alpha\, x\right). \tag{10.71}$$

Using again $\Delta p(x \to \infty) = 0$ (leading to $C_2 = 0$) and a recombination velocity S at the front surface, i.e.

$$-\,e\, S\, \Delta p_0 = -e\, D_\text{p}\, \left.\frac{\partial \Delta p}{\partial x}\right|_{x=0}. \tag{10.72}$$

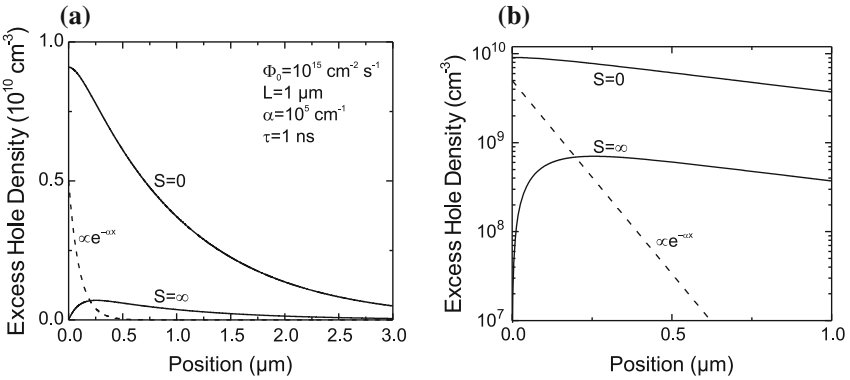


Fig. 10.35 Excess carrier density profile (10.73) in (**a**) linear and (**b**) semi-logarithmic plot for $S = 0$ and $S = \infty$. Other parameters are given in panel (**a**)

The solution is given as

$$\Delta p(x) = \frac{G_0\,\tau_\mathrm{p}}{1-\alpha^2\,L_\mathrm{p}^2}\left[\exp\left(-\alpha\,x\right) - \frac{S+\alpha\,D_\mathrm{p}}{S+D_\mathrm{p}/L_\mathrm{p}}\,\exp\left(-\frac{x}{L_\mathrm{p}}\right)\right]. \tag{10.73}$$

For vanishing surface recombination, $S = 0$, the solution is (Fig. 10.35)

$$\Delta p(x) = \frac{G_0\,\tau_\mathrm{p}}{1-\alpha^2\,L_\mathrm{p}^2}\left[\exp\left(-\alpha\,x\right) - \alpha\,L_\mathrm{p}\,\exp\left(-\frac{x}{L_\mathrm{p}}\right)\right]. \tag{10.74}$$

For $\alpha L_\mathrm{p} \gg 1$, (10.65) is recovered. This dependence is the excess-carrier profile if the absorption is strong, which is a tendency for short wavelengths. The current at the surface, $j\,(x=0) \propto \nabla\,\Delta p$, is zero.

In the case of very strong surface recombination, $S \to \infty$, (10.73) becomes

$$\Delta p(x) = \frac{G_0\,\tau_\mathrm{p}}{1-\alpha^2\,L_\mathrm{p}^2}\left[\exp\left(-\alpha\,x\right) - \exp\left(-\frac{x}{L_\mathrm{p}}\right)\right], \tag{10.75}$$

with $\Delta p(0) = 0$ (Fig. 10.35). The current at the surface is ($D_\mathrm{p}\,\tau_\mathrm{p} = L_\mathrm{p}^2$)

$$j(x=0) = -e\,D\,\left.\frac{\partial \Delta p}{\partial x}\right|_{x=0} = -e\,\frac{G_0\,L_\mathrm{p}}{1+\alpha\,L_\mathrm{p}} = -e\,\Phi_0\,\frac{\alpha\,L_\mathrm{p}}{1+\alpha\,L_\mathrm{p}}. \tag{10.76}$$

Part II
Selected Topics

Chapter 11
Surfaces

Gott schuf das Volumen, der Teufel die Oberfläche.
God created the bulk; surfaces were made by the devil.

attributed to W. Pauli

Abstract The specifics of semiconductors surfaces, their symmetry, equilibrium crystal shape, reconstructions, steps and faceting are summarized, being important for epitaxy. Physical surface properties such as vibrational and electronic states are discussed.

11.1 Introduction

Obviously every crystal has a surface all around it, connecting it to the surrounding world. This represents a brutal perturbation of the bulk periodicity with the consequence of a whole new world of physics at the surface. With regard to semiconductor technology, the surface properties are of large importance in a number of instances:

- The crystal growth of semiconductors always occurs at a surface. This subject will be discussed in Chap. 12.
- The surface is subject to interaction with the chemistry of the surrounding atmosphere. This is of essential importance for catalysis, e.g. oxidation of CO at noble metals or photocatalytic water splitting, e.g. using TiO_2, into oxygen and hydrogen [921, 922]. Photocatalysis is briefly discussed in Sect. 22.1. The interaction and reaction with the surrounding atmosphere can lead modifications of the semiconductor such as change of conductivity, e.g. in SnO_2, which can be used in the construction of gas detectors [923–925].
- Surface passivation and barriers, e.g. using photoresist, silicon oxides or nitrides, are often necessary in practical devices in order to avoid surface recombination (Sect. 10.12.1) or the interaction with atmospheric oxygen or water.

But surface physics is interesting in its own right, investigating the properties of a complex mostly two-dimensional system. 'Pure' surfaces are studied using

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_11

crystals cleaved in ultra-high vacuum or carefully prepared atomically clean surfaces. A special case are two-dimensional materials, also termed atomic sheets, such as graphene; these will be discussed in Sect. 18.1.

At the surface the atoms rearrange, compared to their bulk positions, vertically and laterally, also forming new bonds (surface reconstructions). The mechanical properties (surface phonons) and electronic properties (surface states) are different from the bulk modes. Detailed treatments of surface physics and its experimental methods can be found in [617, 926–928].

11.2 Surface Crystallography

The surface symmetry, i.e. the two-dimensional spatial periodicity of the constituent atoms, is described with the ten two-dimensional point groups (Table B.1). The point symmetries are 1-, 2-, 3-, 4-, and 6-fold rotational symmetry (Fig. 3.4) with or without mirror plane(s). The two-dimensional point symmetries of popular substrate orientations are compiled in Table 11.1; the different symmetry of the first layer and the half-space are notable. The combination of five 2D Bravais lattices (cmp. Sect. 3.3.4) with the 10 2D point groups leads to the 17 two-dimensional space groups (wallpaper groups) [929].

For the treatment of surface in reciprocal space, the three-dimensional $\mathbf{k}$-vector is split into the two-dimensional component $\mathbf{k}_{||}$ parallel and the one-dimensional component $\mathbf{k}_{\perp}$ perpendicular to the surface,

$$\mathbf{k} = \mathbf{k}_{||} + \mathbf{k}_{\perp}. \tag{11.1}$$

The three most important Brillouin zones in two-dimensional $\mathbf{k}$-space are depicted in Fig. 11.1. Often the special points of the 2D Brillouin zones are denoted with a bar over the letter.

Table 11.1 Two-dimensional point symmetries of common substrates with ideal low-index surfaces

Crystal	Surface	1st layer	1st & 2nd layers	Half space
Rocksalt	(001)	4 mm	4 mm	4 mm
	(110)	2 mm	2 mm	2 mm
	(111)	6 mm	3 m	3 m
Diamond	(001)	4 mm	2 mm	2 mm
	(110)	2 mm	2 mm	2 mm
	(111)	6 mm	3 m	3 m
Zincblende	(001)	4 mm	2 mm	2 mm
	(110)	1 m	1 m	1 m
	(111)	6 mm	3 m	3 m
Wurtzite	(00.1)	6 mm	3 m	3 m
	(10.1)	2 mm	2 mm	1 m
	(11.0)	2 mm	2 mm	1 m

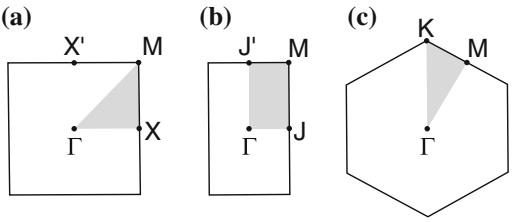


Fig. 11.1 Two-dimensional Brillouin zones for (**a**) square, (**b**) rectangular and (**c**) hexagonal surface symmetry. Special points are labelled and the *grey areas* indicate the smallest irreducible area

11.3 Surface Energy

The surface energy, more precisely the surface energy per area γ, is related to the work that is necessary to split a crystal in two parts. Such process will leave broken ('dangling') bonds. This energy will depend on the crystal orientation as is already clear from the fact that there are easy cleaving planes (Sect. 5.4.2). The surface energy will also depend on the surface reconstruction (see Sect. 11.4), i.e. the rearrangement of surface bonds and atoms. Generally, the reduction of the number of dangling bonds on a surface lowers its energy, while the distortion of bonds increases its energy.

Table 11.2 Surface energy (in J/m^2) for various C, Si and Ge surfaces

Material	{111}	{110}	{100}	{311}
C	8.12	7.48	9.72	8.34
Si	1.82	2.04	2.39	2.12
Ge	1.32	1.51	1.71	1.61

Data from [930]

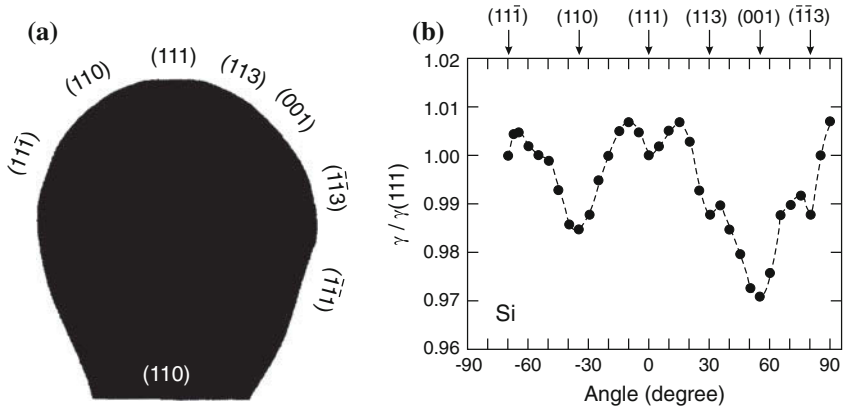


Fig. 11.2 (**a**) Equilibrium crystal shape (cross section in ⟨110⟩ azimuth) of 1.06 μm diameter Si bulb at $T = 1323$ K with facet orientations as labelled. (**b**) Anisotropy of surface energy (relative to $\gamma(111)$) for Si, the *dashed line* is guide to the eyes. Adapted from [931]

The anisotropy of the surface energy for a given orientation hkl, $\gamma(hkl)$, determines the equilibrium crystal shape (ECS) at a given temperature (below melting temperature). The crystallite is assumed to be of at least mesoscopic size such that energy effects due to edges and apexes can be neglected compared to the surface energy terms. As an example theoretically calculated surface energy values for covalent semiconductors are listed in Table 11.2. The equilibrium shape of silicon has been observed for μm-sized bulbs as shown in Fig. 11.2a for $T = 1323\,K$; the experimental anisotropy of surface energy of silicon is shown in Fig. 11.2b.

11.4 Surface Reconstruction

When in a *Gedankenexperiment* the bulk crystal is split such that a surface with defined orientation develops, many bonds are cut. These dangling bonds can be saturated with other atoms such as hydrogen. In particular under vacuum conditions, the dangling bonds will rearrange and form new bonds such as dimers along the surface, lowering the total energy. This surface reconstruction displays two-dimensional periodicity of a surface unit cell.

Since the forces from the split-away half-space are missing, the atomic planes will rearrange vertically[1] and the surface-near layers will exhibit slightly different lattice spacing than in bulk. In [932] the surface reconstructions of many semiconductors are reviewed. In Fig. 11.3a the rearrangement of surface atoms on GaAs(110) is depicted as an example. The anion is moved up, the cation moved down, preserving the bond length, rotating the bond by about $\gamma = 30°$ for various III–V semiconductors [933] (Fig. 11.3b). The height difference of the top anion and cation Δ_1 scales with the bulk lattice constant (Fig. 11.3c). The details are specific to materials and orientations.

Different reconstructions occur for different thermodynamical conditions, some of them being metastable, as depicted for GaAs(100) in Fig. 11.4. Several different surface reconstructions can also be present simultaneously at a surface within different domains.

The stable reconstruction of the Si(111) surface is the somewhat complicated 7×7 reconstruction which is schematically depicted in Fig. 11.5a as proposed in [936] ('DAS' dimer-adatom-stacking fault model). A STM image of this surface has been reported first in [937] and is depicted in Fig. 11.5b. Details of this surface are also recently a subject of interest [938].

[1] Typically the distance between the first and second layer is reduced, and the distance between the second and third layer a little bit increased.

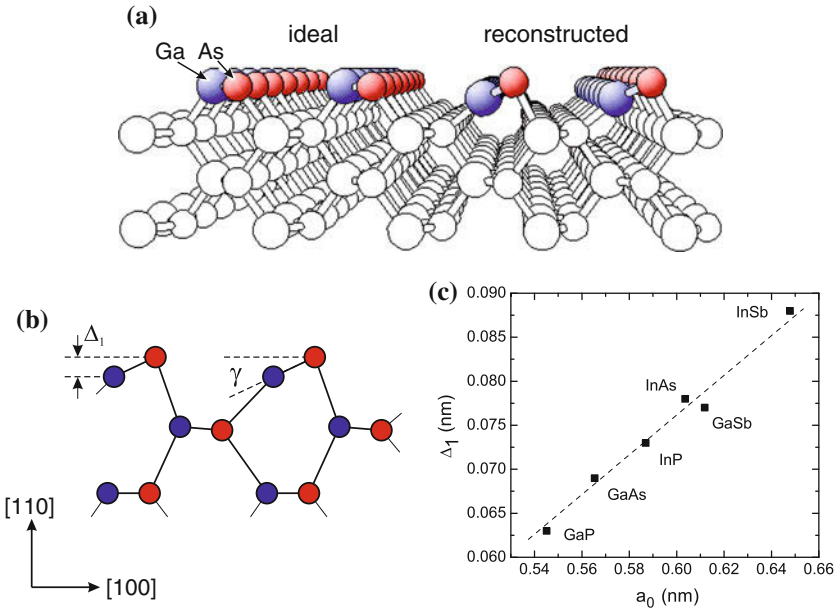


Fig. 11.3 (**a**) Schematic drawing of the GaAs(110) surface after cutting bonds ('ideal') and actual atomic rearrangement with asymmetric dimer in the reconstructed state. Adapted from [934]. (**b**) Schematic of the III–V (110) surface relaxation *red circles*: anions, *blue circles*: cations) with bond rotation ω and atom shift Δ. (**c**) Experimental values for he relaxation Δ_1 as shown in panel (**b**) for various semiconductors versus their bulk lattice constant a_0; *dashed line* is straight line as guide to the eyes. Data from [933]

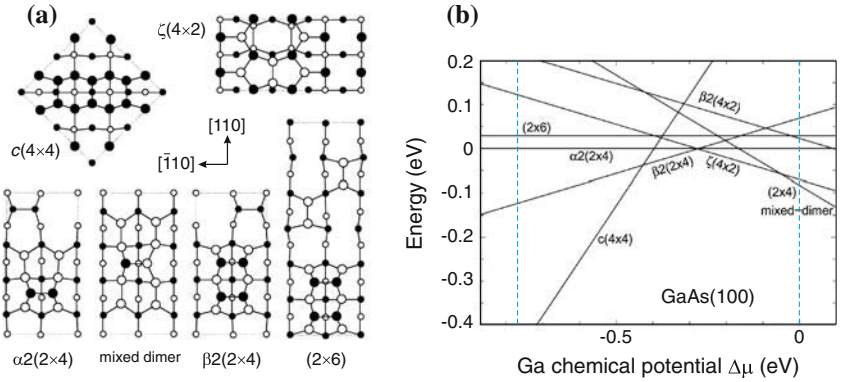


Fig. 11.4 (**a**) Various reconstructions of the GaAs(100) surface. *Filled* (*empty*) *circles* represent As (Ga) atoms. Atom positions in the top two atomic layers are indicated by *larger symbols*. (**b**) The relative formation energy per (1 × 1) unit cell of various reconstructions as a function of the Ga chemical potential. *Vertical dashed lines* mark the approximate anion- and cation-rich limits of the thermodynamically allowed range of $\Delta\mu$. Adapted from [935]

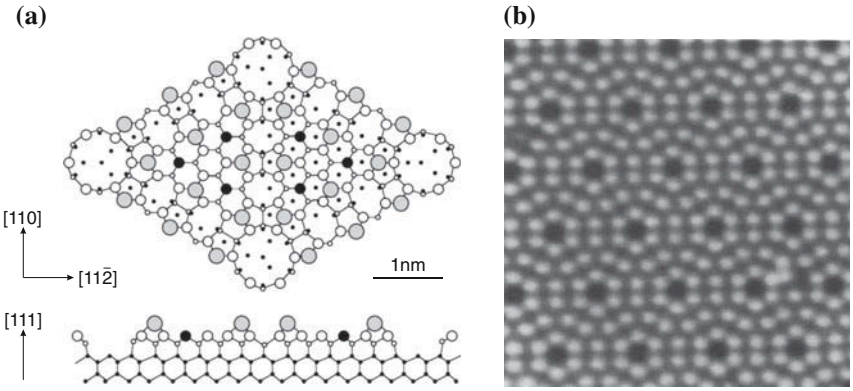


Fig. 11.5 (**a**) Schematic drawing of the 7 × 7 reconstruction of Si(111) surface. The *large grey circles* represent adatoms, the *small balck circles* rest atoms. Adapted from [936] (**b**) STM image (empty states) of such surface. Adapted from [937]

11.5 Surface Morphology

The surface reconstruction is related to the local atomic arrangement of the surface atoms. On larger length scales, surfaces can exhibit roughness in general and various specific morphologies, among them steps, step bunches, facets, hillocks or pits. When atomically flat terraces are separated by steps of equal heights, the surface is termed *vicinal*. The step height can be an atomic monolayer or more. The step edges can be straight, smoothly curved or rugged, depending on the formation energy of kinks. If the substrate surface is inclined by a small 'off-cut' angle θ with respect to a low index plane, for a step height h the average terrace width L is given by

$$L = \frac{h}{\tan\theta} \approx \frac{h}{\theta}. \tag{11.2}$$

The Si(001) surface (under certain conditions) exhibits monoatomic steps between the A-type 1 × 2 and B-type 2 × 1 terraces [939]; clearly also the step edges between A–B and B–A are quite different, one being much rougher than the other (Fig. 11.6). The surface is thus a two domain surface, the symmetry of both A- and B-type terraces being 2 mm, rotated by 90° against each other.

A similar example is the Zn(00.1) surface. The found step height of the monoatomic steps is 0.27 nm and corresponds $c/2$ [940]. The terrace width is about 12 nm (Fig. 11.7a), yielding an off-cut (11.2) of about 1.3°. The surface pattern shows triangular features in two orientations, rotated by 60°. The 3 m symmetry of the surface, although the first monolayer has 6 m point symmetry (Table 11.1), occurs in two domains rotated against each other by 60°. The LEED pattern from each individual terrace is expected to have three-fold symmetry; the mixed character of the surface yields a six-fold pattern (Fig. 11.7b).

Steps can gather and form step bunches with a height much larger than a monolayer (Fig. 11.8). Faceting of higher index planes occurs when it is energetically favorable to

Fig. 11.6 STM image of Si(001) surface with an off-cut $\theta = 0.5°$. Adapted from [939]

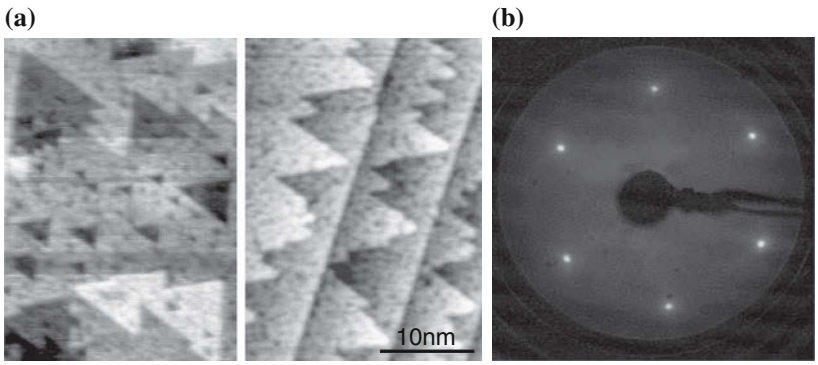


Fig. 11.7 (**a**) STM images a ZnO(00.1) O-terminated surface, left (right) panel with zero (about 1.3°) off-cut, exhibiting monolayer step heights of 0.27 nm. (**a**) LEED pattern of such surface (recorded at 70 eV). Adapted from [940]

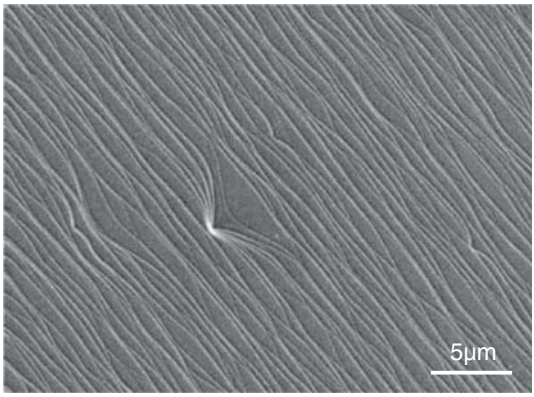


Fig. 11.8 SEM image of the surface of a 4H-SiC layer on [00.1]-oriented substrate with 8° misorientation towards [11.0]. Adapted from [942]

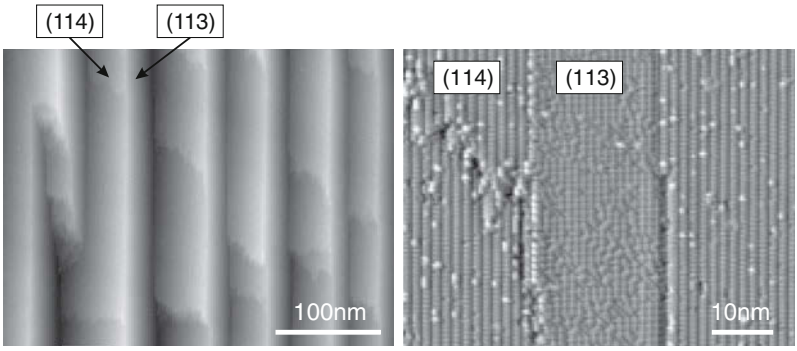


Fig. 11.9 STM images of a silicon surface close to (113), 21.5° from (001) towards (111), i.e. (113)-3.7°. Adapted from [943]

form alternating facets of low energy low index planes, for example Si(223) exhibits periodic ridges with (111) and (113) facets [941]. The faceting of a surface close to (113), exhibiting smooth (113) and rough (114) planes, is depicted in Fig. 11.9.

11.6 Surface Physical Properties

11.6.1 Surface Phonons

As a model for surface vibrational states, the diatomic one-dimensional chain model discussed in Sect. 5.2.2 can be modified to allow for a different surface spring constant C_s instead of C_1 or C_2. Such model can be solved numerically for finite chain length. The bulk dispersion is found with an extra mode in the gap or above the maximum frequency for a range of values of C_s as depicted in Fig. 11.10. When the spring constant of the surface atom is smaller than the larger of the two spring constants C_1 and C_2 (in our model calculation $C_2 > C_1$), a state in the gap forms; for a surface spring constant larger than C_2, first the surface vibration lies within the optical bulk band and then forms a state above the maximum optical frequency ω_m, given by (5.18).[2]

In order to display the dispersion of surface states together with the bulk band structure, the latter is projected to the surface $\mathbf{k}_{||}$-vector: $E_n(\mathbf{k})$ is considered as $E_{n,\mathbf{k}_\perp}(\mathbf{k}_{||}) = E_\nu(\mathbf{k}_{||})$, where $\nu = n$, $\mathbf{k}_\perp$ is a new, continuous index. In the plot of E_n^{surf} versus $\mathbf{k}_{||}$, for each value of $\mathbf{k}_{||}$ a *range* of energies reflects the bulk band structure, as shown in Fig. 11.11 for GaAs. This concept pertains to phonon dispersion as well as to electronic states.

[2]The appearance of states within the gap and above ω_m resembles that of localized vibrational modes of substitutional masses (Sect. 5.2.6).

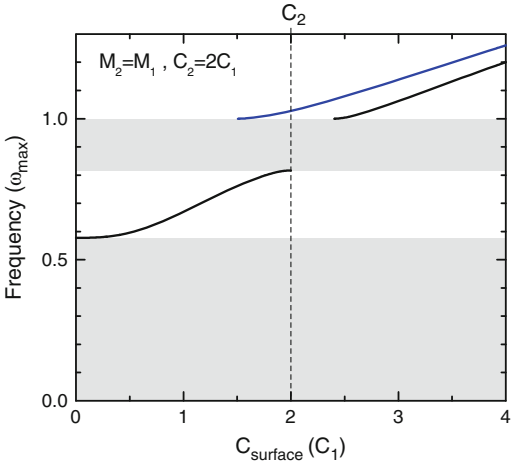


Fig. 11.10 One-dimensional model calculation of surface vibrational state frequency (in units of the maximum bulk optical phonon frequency ω_m) as a function of the spring constant $C_{surface}$ of the surface atom replacing C_1 (*blue curve*) or C_2 (*black curves*). As model parameters we use equal masses and $C_2 = 2\,C_1$ ($\gamma = 0.943$ from (5.17)). The *grey areas* represent the energy range of the acoustic and optical phonon bulk bands

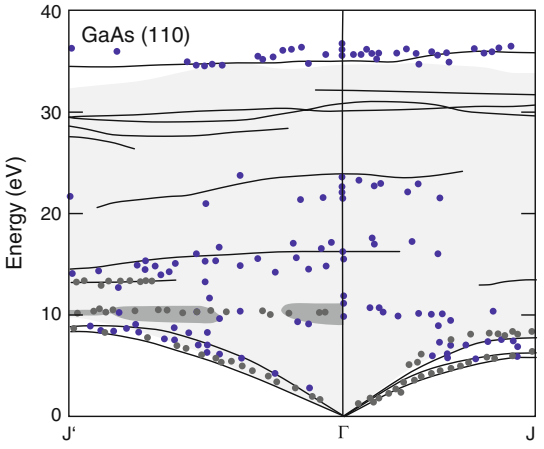


Fig. 11.11 Surface phonon dispersion for GaAs(110) (*solid lines*) with the surface projected bulk band structure (*light grey area*). The *symboles* are experimental data from two different methods. The *dark grey areas* at about 10 meV indicate regions where the A_1-peak is present in the calculated scattering cross-section. Adapted from [933]

11.6.2 Surface Plasmons

Free-carrier oscillations in the bulk had been discussed in Sect. 9.8.1. A surface plasmon is the quantum of a surface-bound plasma oscillation. Such effect has been discussed in [944] and reviewed in [945, 946]. A metal (or conductive semiconductor)

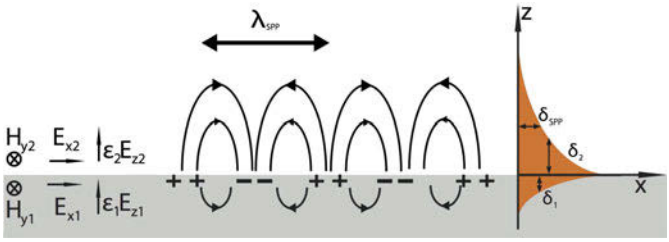


Fig. 11.12 Sketch of surface plasmon field distribution. Adapted from [947]

with dielectric constant according to (9.65) $\epsilon_1 = \epsilon_m = \epsilon_r\,(\omega_p/\omega)^2$ and a dielectric (or vacuum) with $\epsilon_2 = \epsilon_d$ are assumed (Fig. 11.12).

The surface plasmon (polariton) (SPP) is a wave localized at the surface with evanescent parts into the metal and the dielectric. The dispersion of the surface plasmon (polariton) is given by

$$k_{SPP} = \frac{\omega}{c}\sqrt{\frac{\epsilon_1\,\epsilon_2}{\epsilon_1 + \epsilon_2}}. \tag{11.3}$$

For large k, the limiting frequency is the SPP frequency (from $\epsilon_1 = \epsilon_2$)

$$\omega_{SPP} = \frac{\omega_p}{\sqrt{1 + \epsilon_d/\epsilon_r}} < \omega_p, \tag{11.4}$$

which is smaller than the plasma frequency. For a metal against vacuum, $\omega_{SPP} = \omega_p/\sqrt{2}$. The SPP dispersion for a ZnO:Ga/air interface is depicted in Fig. 11.13 for three different doping concentrations.

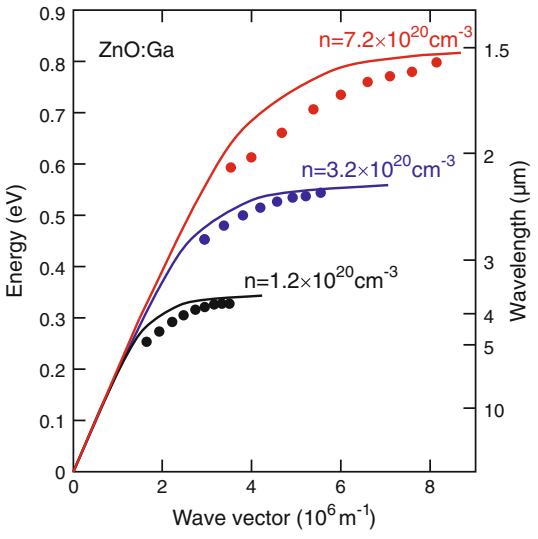


Fig. 11.13 Surface plasmon polariton dispersion for ZnO:Ga/air for three different carrier concentrations as labelled, experimental data (*symbols*) and theory (*lines*). Adapted from [948]

11.6.3 Electronic Surface States

The bulk band structure is given by the energy eigenvalues $E_n(\mathbf{k})$. The surface adds its own states $E_n^{\text{surf}}(\mathbf{k}_{||})$, many of them in the gap. The calculation of gap states has been already briefly discussed in Sect. 6.2.3. The nature of bulk states is oscillatory in the bulk and exponentially decaying in the outside (Fig. 11.14a), the nature of surface states is localized at the surface and decaying both towards the bulk and the outside (Fig. 11.14b). A third type of states is a *surface resonance* which is oscillatory in the bulk and has enhanced probability at the surface, of course also decaying towards the outside (Fig. 11.14c); such states are surface related but energetically degenerate with states of the bulk band structure.

Surface states for Si(100) in 2×1 reconstruction are shown in Fig. 11.15. The surface bands arise from (filled and unfilled) π- and π^*-orbitals from the dangling bonds on the threefold-coordinated surface atoms [949]. A calculation for the ideal (unreconstructed) (100), (110) and (111) surfaces of Si, Ge and GaAs can be found in [950]. Further work on silicon surface states (clean and with adsorbates) is compiled in [951].

For a conventional insulating material (with gap) the states of conduction and valence band have defined s- and p-type symmetry (cmp. Fig. 2.5) and the surface states can be populated with spin up and down electrons as is schematically drawn in Fig. 11.16a. Such 'normal' surface states are thus not spin-polarized. This has been shown, e.g., for the well-known surface states of Si(111)-(7×7) [952] in [953].

In a so-called *topological insulator* [954], a band inversion is present. In HgTe-like materials, s- and p-type bands are inverted at the Γ point (cmp. Sect. 6.9.4). In layered

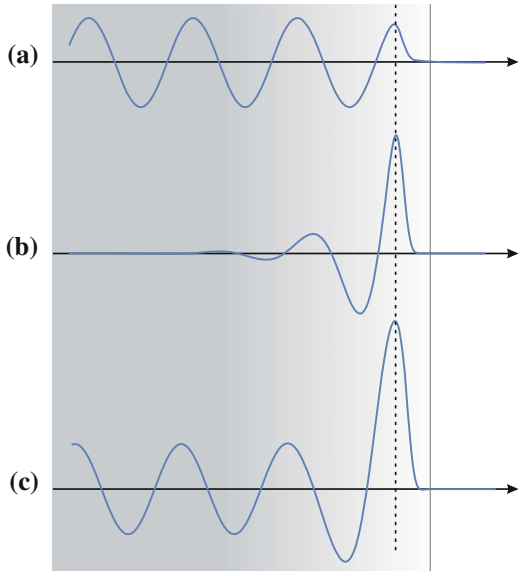


Fig. 11.14 Schematic wavefunction versus coordinate for (**a**) bulk state in the vicinity of the surface (its position indicated by the *vertical dashed line*), (**b**) surface state and (**c**) surface resonance state

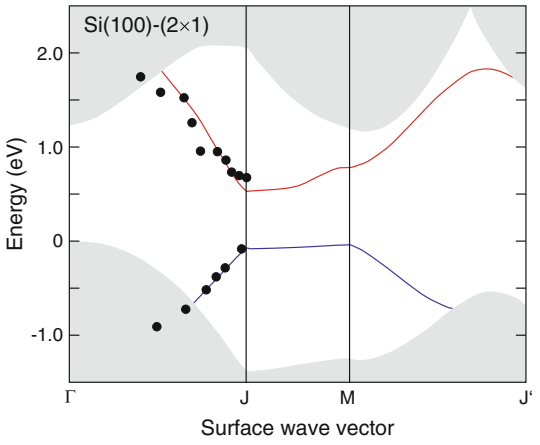


Fig. 11.15 Surface band structure of Si(100)-(2 × 1), *lines* denote a theoretical quasi-particle calculation, *symbols* represent experimental data. The *grey areas* indicate the projected bulk band structure. Adapted from [949]

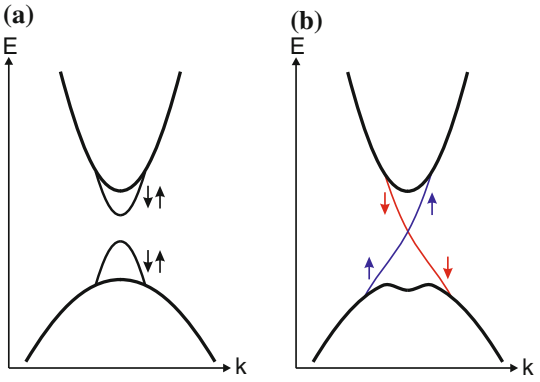


Fig. 11.16 (**a**) 'Normal' band structure with surface states, the valence band states are antisymmetric ('−') from p-states, the conduction band states are symmetric ('+') from s-states. The *arrows* denote the electron spin orientation of the states. (**b**) 'Topological' band structure with surface states crossing the gap and electrons in the two surface states having unique spin orientations

(orthorhombic) Bi_2Se_3-like tri-chalcogenide materials two p_z-orbitals with opposite parity are inverted at the Γ point. Reasons for the band inversion are spin-orbit interaction but also scalar relativistic effects and lattice distortions [955]. The spin-orbit interaction is responsible for opening a gap; the associated surface states cross the band gap (Fig. 11.16b) and are spin-polarized (Fig. 11.16c) [956, 957]. A review of various systems is given in [958]. Charge transport in such spin-polarized surface states suffers no scattering (thus these states are called 'topologically protected') unless by a center that breaks time reversal symmetry (magnetic impurity). The

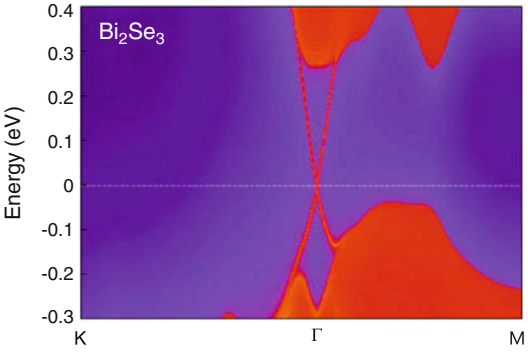


Fig. 11.17 Density of states for Bi_2Se_3 (111) surface. Adapted from [957]

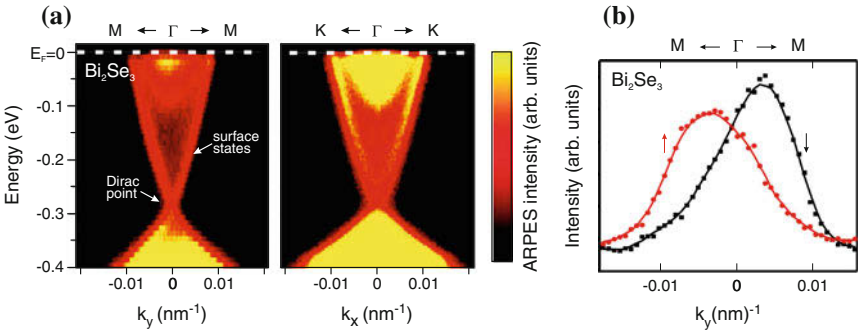


Fig. 11.18 (**a**) Angle-resolved photoemission spectra from Bi_2Se_3 (111) surface along two different **k**-space directions. The Fermi level is at energy zero. The *blue dashed line* indicates the angular scan displayed in panel (**b**). (**b**) Spin-polarized momentum distribution curve at a binding energy of −140 meV, measured along the k_y-direction. Adapted from [960]

more or less linear dispersion of these states forms a 'Dirac cone' (cmp. Sect. 18.1.2 for the band structure of graphene with six Dirac cones).

Topological surface (or interface) states have been observed in various systems [958]. As an example we show data on bulk Bi_2Se_3 [959] for which the surface states are found to cross the band gap (Figs. 11.17 and 11.18a) and spin-polarized measurements (Fig. 11.18b) show a strong spin polarization (about 50 %) of the two branches.

Chapter 12
Heterostructures

The interface is the device.

H. Kroemer [961]

Abstract Heterostructures are the most important basis for modern devices and are covered regarding various aspects including heteroepitaxy on planar and patterned substrates, surfactants, heterostructure band lineup as well as energy levels and recombination in planar confined systems (quantum wells and two-dimensional electron gases).

12.1 Introduction

Heterostructures consist of (at least two) different materials. The geometry of the interfaces between the two materials can be complicated. The simplest case is a planar interface, i.e. a layered system. A metal–semiconductor junction is generally a heterostructure. However, we will use the term mostly for structures of various semiconductors. Most of the heterostructures discussed here are epitaxial, i.e. fabricated by the successive epitaxy of the various layers on a substrate. Another method to fabricate heterostructures of different (and dissimilar) materials is wafer bonding that is briefly discussed in Sect. 12.6.

Many modern semiconductor devices rely on heterostructures, such as the heterobipolar transistor (HBT), the high electron mobility transistor (HEMT), lasers and nowadays also light-emitting diodes. Shockley had already considered heterostructures in his 1951 patent for pn-junctions. For the development and the realization of heterostructures H. Kroemer and Zh.I. Alferov were awarded the 2000 Physics Nobel Prize. The properties of charge carriers in layers that are part of heterostructures can be quite different from those in bulk material, e.g. extremely high mobility, high radiative recombination efficiency or novel states of matter, as revealed in the quantum Hall effects.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_12

12.2 Heteroepitaxy

12.2.1 Growth Methods

Since the thickness of layers in the active part of heterostructures has to be controlled to monolayer precision and the thickness of layers can go down to the single monolayer range, special growth methods have been developed [962–964]. Among these molecular beam epitaxy (MBE [965]), chemical vapor deposition (CVD [966, 967]) and metalorganic vapor phase epitaxy (MOVPE [968]) are the most common for Si, Ge, III–V and II–VI semiconductors. In particular, oxide semiconductors are also fabricated with pulsed laser deposition (PLD [969]). Liquid phase epitaxy (LPE [970]) used to be very important for the fabrication of LEDs but has lost its role largely to MOVPE.

MBE is performed in an ultrahigh vacuum (UHV) chamber, pumped by getter pumps and cryo-shrouds. The source materials are evaporated from effusion cells and directed towards the heated substrate. If the source materials are supplied as a gas stream, the method is called gas-source MBE (GSMBE). If metalorganic compounds are used as precursors, the method is denoted as MOMBE. The atoms impinge on the substrate with thermal energy and are first physisorbed. After diffusion on the surface they either desorb or they are chemisorbed, i.e. incorporated into the crystal. In order to obtain high spatial homogeneity of material properties such as composition, thickness and doping, the substrate is rotated during deposition.

During CVD and MOVPE the heated substrate is in a gaseous environment. The transport gas is typically H_2, N_2 or O_2. Precursor materials are hydrides such as silane, germane, arsine or phosphine (SiH_4, GeH_4, AsH_3, PH_3) and (for MOVPE) metalorganic compounds, such as, e.g. trimethylgallium (TMG). Due to the toxicity of the hydrides, alternative, i.e. less-toxic and less-volatile compounds are used, such as TBAs ($(CH_3)_3CAsH_2$). The crystal growth occurs after pyrolysis and catalysis of the compounds close to or on the substrate surface. All remaining C and H atoms (and whatever else that is not incorporated into the crystal) leave the reactor and are neutralized and stopped in a scrubber.

In-situ monitoring is of importance to obtain information about the growth process while it is underway. Using the information in a feedback loop it is possible to achieve in-situ control of the process, e.g. for precise determination of growth rates or layer thickness. Techniques are RHEED (reflection high-energy electron diffraction) [971] (only for UHV systems) and RAS (reflection anisotropy spectroscopy) [972, 973].

12.2.2 Substrates

Thin-film epitaxy is mostly performed on wafers, i.e. thin circular slices of substrate material. The most common substrate materials are Si (currently up to 400 mm diameter [974, 975]), Ge (also up to 300 mm [976]), GaAs (up to 6 inch), InP (up to 4 inch)

and sapphire (up to 6 inch). Typical wafer thickness is 300–500 μm. Also, very thin, flexible Si wafers (8–10 μm) have been developed [977]. A wafer is cut from a large single cylindrical crystal that is fabricated with suitable growth techniques such as Czochralski (CZ) growth (1918) modified by Teal and Little [978, 979]. In CZ growth the crystal is pulled from a seed crystal out of a melt of previously polycrystalline, pure or doped material. All dislocations stop in the narrow neck between the seed and the main body of the cylinder. The diameter of the crystal is controlled by the pulling rate and the heating power. For the growth of III–V compound semiconductors liquid encapsulated CZ (LEC) growth has been developed to counteract the high volatility of the growth-V component. During LEC growth the melt is completely covered with molten boric oxide (B_2O_3). The keys to optimization of the crystal growth process are numerical modeling and computer control. In Fig. 12.1a,c a large CZ silicon crystal and a smaller LEC GaAs crystal (boule) are shown. Over time wafer size and ingot mass have increased remarkably (Fig. 12.1b). For details on other important fabrication methods for bulk crystals, including float-zone (FZ [980]) or vertical gradient freeze (VGF), we refer to the literature [981]. Significant expertise is necessary for cutting, grinding and polishing (lapping) wafers for epitaxy.

For semiconductors, the wafer is marked with *flats* to indicate orientation and doping. In Fig. 12.2 the standard flats are shown for (100)- and (111)-oriented material. The primary (major) flat (OF, orientation flat) defines the crystallographic

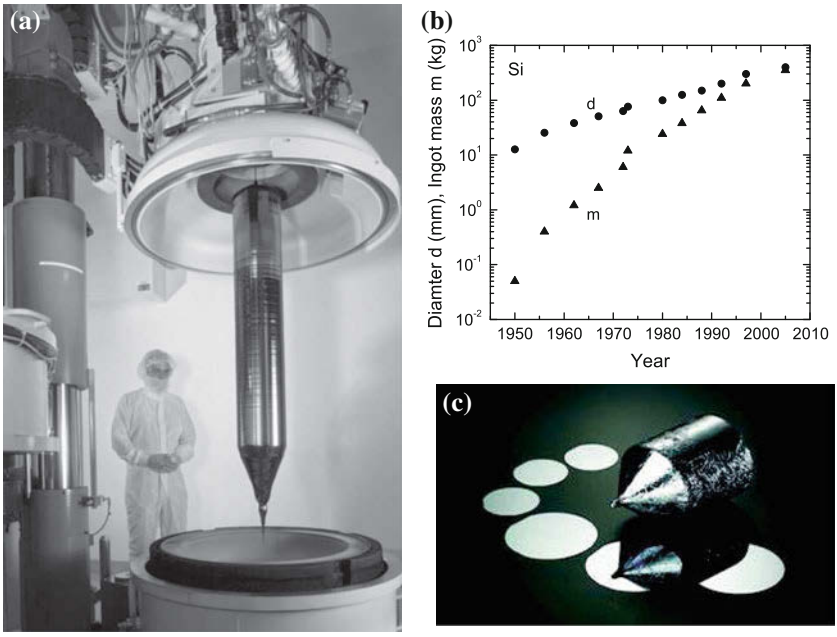


Fig. 12.1 (**a**) Silicon single crystal for 300-mm diameter wafers after opening of the crucible. From [982] with permission. (**b**) Historic development (first year of larger production) of silicon wafer diameter and ingot mass. Data from [975]. (**c**) GaAs single crystal (boule) for 4-inch wafers and some cut and polished wafers

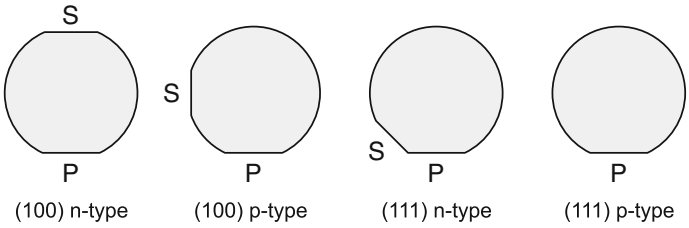


Fig. 12.2 Schematic semiconductor wafer geometry for various orientations and doping with the primary (P) and secondary (S) flats indicated

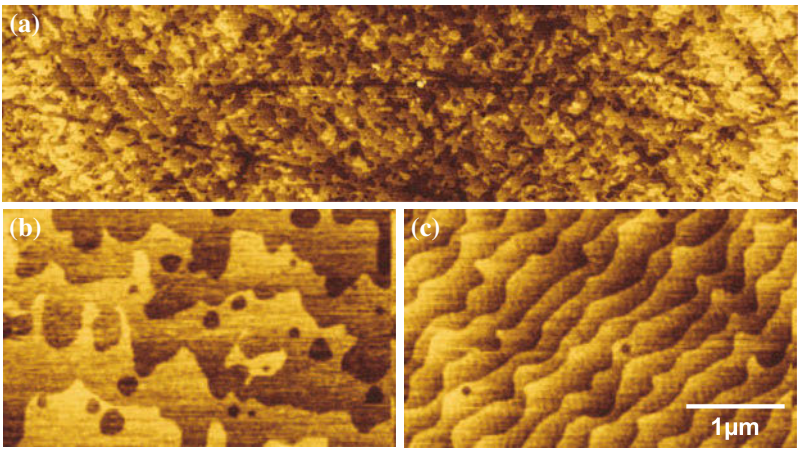


Fig. 12.3 (**a**) AFM images of an as-received ZnO wafer, exhibiting small terraces and nm deep scratches from polishing. (**b**, **c**) Two ZnO $(000\bar{1})$ wafers with vicinal surfaces after thermal treatment (1000 °C for 2 h in O_2), exhibiting atomically flat terraces with $c/2$ monoatomic step heights. Two different substrates with different off-cut (misorientation direction and angle) are shown. Adapted from [990]

orientation[1] is longer than the secondary (minor) flat which defines the conductivity type (IF, identification flat). For a 4-inch (100 mm) diameter wafer the OF is 32 and the IF 12 mm long. The front surface, on which the epitaxy is performed, typically undergoes an elaborate cleaning and polishing process. Silicon processes [981, 984] are based on the RCA cleaning procedure [985] and the related Shiraki etch [986] and can achieve clean, atomically flat surfaces [987]. III-V semiconductors are typically prepared using a polishing etch [988, 989], often solutions containing bromine. It is possible to create compound semiconductor surfaces that exhibit large, essentially monoatomically flat terraces between individual surface steps. Polishing or other surface damage can also be removed prior to epitaxy with thermal (Fig. 12.3) or ion beam treatments.

[1] In the 'US' flat definition, the primary flat is the $(01\bar{1})$ surface, in the 'EJ' definition, the primary flat is $(0\bar{1}\bar{1})$.

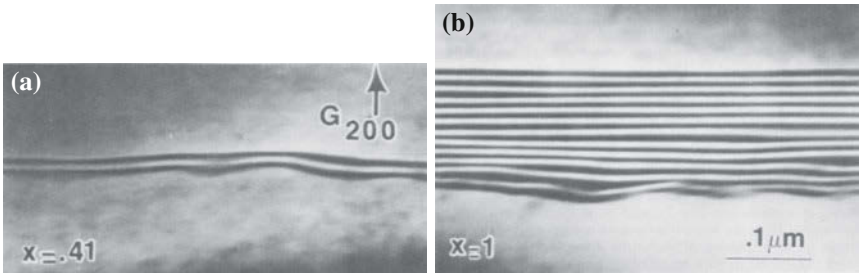


Fig. 12.4 Cross-sectional TEM of MBE-grown $Al_xGa_{1-x}As$/GaAs heterostructures for (**a**) $x = 0.41$ and (**b**) $x = 1.0$. Using an AlAs/GaAs superlattice an excellent flattening of substrate roughness is achieved. From [991]

One prerequisite for making high-quality heterostructures with thin layers is a flat surface. Even if the polished substrate is not perfect, flat interfaces can be achieved with the growth of appropriate superlattice buffer layers (Fig. 12.4). Roughness can exist on all length scales and is typically investigated using atomic force microscopy scans.

High throughput demands and the advent of multi-wafer reactors make prior cleaning and etching procedures tedious. For this purpose, substrates for well-developed material systems are offered 'epiready'. Epiready wafers are often covered with a very thin protective film that can be released using a thermal treatment at or below typical growth temperature in the growth reactor immediately prior to epitaxy. The protective film separates the polished semiconductor wafer from the ambient. Examples are a few monolayers of arsenic on GaAs or two monolayers GaN on SiC. However, the suitability of a purchased substrate may depend on the duration and conditions of its storage.

A special case is the use of *curved* substrates which allow the study of growth on different crystal orientations in a continuous fashion. Such experiments have been performed using cylindrical substrates [992] and hemispherical crystals. For the latter homoepitaxy has been reported for Si [993, 994], SiC [995] or GaAs [996]. Also heteroepitaxy has been investigated, for example for GaP and GaAs on Ge [997] (cmp. Fig. 12.12). Using such curved substrates, the growth rate can be determined as a function of crystal orientation as shown for GaAs (on GaAs) in Fig. 12.5; large differences are found for (111)A and (111)B surfaces. For growth of silicon the variations with angle are much smaller in the range of 10 % [992].

12.2.3 Growth Modes

The growth of a material A on a material B can occur via three fundamental growth modes (Fig. 12.6), the layer-by-layer or Frank-van der Merwe (FvdM) growth mode [384], the island or Volmer-Weber (VW) growth mode [998] and the

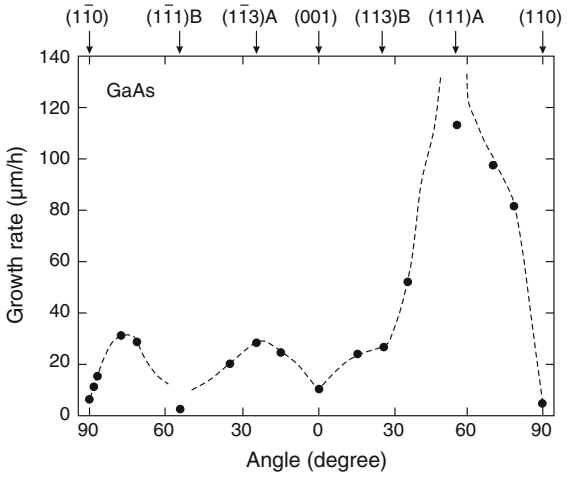


Fig. 12.5 Growth rate of GaAs in vapor phase transport at 725° as a function of crystal orientation. The *dashed line* is guide to the eyes. Adapted from [996]

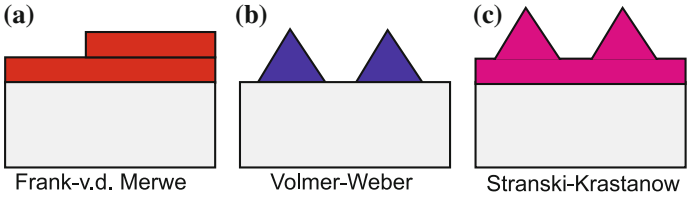


Fig. 12.6 Schematic of the three different fundamental epitaxial growth modes

Stranski-Krastanow (SK) growth mode [999]. In [999] the possibility of island formation on an initially flat heteroepitaxial surface was proposed for the growth of lattice-*matched* ionic crystals that have different charges. The term SK growth is now typically used in lattice-*mismatched* heteroepitaxy for the island formation (and related relaxation of strain energy, cmp. Fig. 14.26) on an initially two-dimensional layer (wetting layer).[2] Also growth of islands *relaxed* by misfit dislocations in strained heteroepitaxy has been termed SK growth [1000].

The growth mode is determined by the relation of the free energies (per area) of the surface σ_s, interface σ_i and film σ_f. Wetting of the substrate and growth of a coherent film (FvdM-growth) occurs for

$$\sigma_s > \sigma_i + \sigma_f. \tag{12.1}$$

If the inequality has the opposite sign, Volmer-Weber or Stranski-Krastanow growth occurs. Additionally the strain energy of the film must be considered. The SK growth

[2]This is the growth mode of self-assembled epitaxial quantum dots as discussed in Sect. 14.3.3.

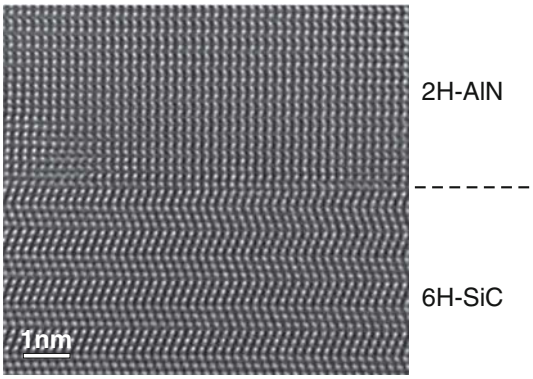


Fig. 12.7 HRTEM cross section of the interface region of MOVPE grown hexagonal AlN on 6H-SiC substrate. Adapted from [1004]

generally occurs when there is wetting of the substrate but layer strain is unfavorable and leads to islanding.

Layer-by-layer growth typically involves nucleation of two-dimensional islands and 'filling' of the remaining monolayer before the next one is started. Another growth mode resulting in smooth epitaxial layers is step-flow growth where adatoms are incorporated mainly at step edges. A more detailed discussion can be found in [1001]. Further details on crystal growth can be found in [1002].

The nucleation and the initial film growth are important and determine the film quality. Several techniques have been developed to overcome common problems. A typical strategy is the growth of a low-temperature nucleation layer.

12.2.4 Heterosubstrates

If homosubstrates are not available or very expensive, semiconductors are grown on dissimilar substrates, e.g. (Al,Ga)N and ZnO on sapphire (Al_2O_3) or SiC.[3] In Fig. 12.7 the interface region of hexagonal AlN on 6H-SiC (cmp. Fig. 3.28b) is shown. The change of crystallographic phase is obvious; the perfect atomic arrangement justifies the term 'epitaxy', literally meaning 'order on top'.

In many cases, the integration of III–V- or II–VI-based semiconductors for optoelectronic applications on silicon for electronic devices is desirable, such as GaAs/Si, InP/Si, GaN/Si or the growth on economic substrates is attractive such as GaN/Al_2O_3 and ZnO/Al_2O_3. For all such combinations the epitaxial relationship, i.e. the alignment of the crystallographic directions of both materials, which can have different space groups, has to be considered. Some examples for epitaxial relationships are given in Table 12.1. The epitaxial relation is determined by the energetically favorite formation of the interface and the early stages of growth.

[3] For ZnO, homosubstrates have recently been produced with 3-inch diameter [1003].

Table 12.1 Epitaxial relationship for various film/substrate combinations, ZnO (or GaN) on *c*-, *a*- and *r*-sapphire and Si(111)

ZnO	Al_2O_3 [00.1]	ZnO	Al_2O_3 [11.0]	ZnO	Al_2O_3 [10$\bar{1}$.2]	ZnO/GaN	Si [111]
[00.1]	[00.1]	[00.1]	[11.0]	[11.0]	[10$\bar{1}$.2]	[00.1]	[111]
[11.0]	[01.0]	[11.0]	[00.1]	[00.1]	[01$\bar{1}$.1]	–/ [21$\bar{1}$.0]	[1$\bar{1}$0]

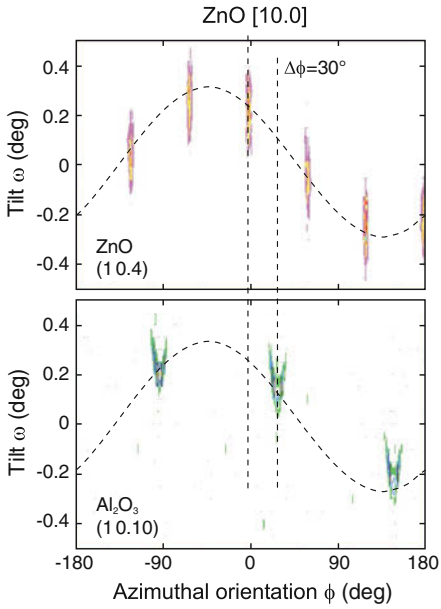


Fig. 12.8 X-ray diffraction intensity from the asymmetric ZnO (10.4) (*upper panel*) and the sapphire (10.10) (*lower panel*) reflections as a function of the azimuthal sample orientation (rotation angle ϕ around the [00.1] axis). The peaks appear at different tilt angles ω due to an overall tilt of the mounted sample (*dashed sinusoidal lines*). The ZnO [00.1] axis is not tilted with respect to the sapphire [00.1] direction

In Fig. 12.8, X-ray diffraction data are shown from a ZnO layer on *c*-oriented sapphire. The hexagonal ZnO lattice is rotated by 30° with respect to the trigonal sapphire lattice. In the case of growth of ZnO on Si(111) an amorphous SiO_2 layer can form at the interface such that the crystallographic information of the substrate is lost. The ZnO grains exhibit random in-plane orientation (Fig. 12.9).

If substrate and epilayer have different space groups, the formation of domains can occur [1006]. The two-dimensional symmetries of various substrates are listed in Table 11.1. The group theoretical minimum number of domains N_{RD} in dependence of the two-dimensional symmetries of substrate and epilayer with rotational symmetries C_n and C_m, respectively, is given by [1007]

$$N_{RD} = \frac{\mathrm{lcm}(n, m)}{m}, \tag{12.2}$$

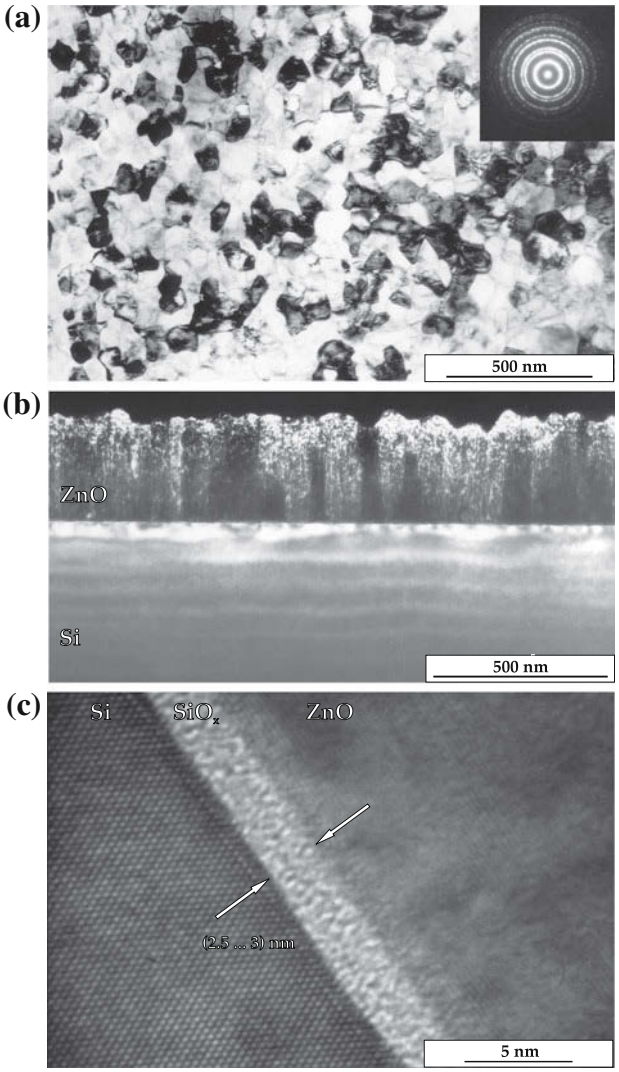


Fig. 12.9 (**a**) Plan-view TEM image (*inset*: electron diffraction diagram averaged over several grains) of ZnO on Si(111).(**b**) Cross-sectional TEM of the same sample. (**c**)High-resolution cross-sectional image of the $ZnO/SiO_2/Si$ interface region. Adapted from [1005]

where lcm denotes the least common multiple. The values are listed in Table 12.2 and visualized in Fig. 12.10.

If the main symmetry axes of substrate and epilayer are not aligned, mirror domains appear. An example is the growth of GaN(00.1) on Ge(111), for which it has been reported that the major symmetry directions do not align exactly. There is a 4° in-plane rotation of the lattices with respect to the usual [11.0] ∥ $[1\bar{1}0]$ in-plane relationship [1008] (cmp. the exact alignment for GaN/Si(111) in Table 12.1).

Table 12.2 Number of rotational (or mirror) domains N_{RD} for all 2D point groups of substrate (G_S) (rows) and epilayer (G_E) (columns). When two numbers are given ($x|y$), the first (second) number represents the number of domains if mirror symmetry planes of substrate and epilayer align (misalign).

$G_S \backslash G_E$	1	m	2	2 mm	3	3 m	4	4 mm	6	6 mm
1	1	1	1	1	1	1	1	1	1	1
m	2	1|2	2	1|2	2	1|2	2	1|2	2	1|2
2	2	2	1	1	2	2	1	1	1	1
2 mm	4	2|4	2	1|2	4	2|4	2	1|2	2	1|2
3	3	3	3	3	1	1	3	3	1	1
3 m	6	3|6	6	3|6	2	1|2	6	3|6	2	1|2
4	4	4	2	2	4	4	1	1	2	2
4 mm	8	4|8	4	2|4	8	4|8	2	1|2	4	2|4
6	6	6	3	3	2	2	3	3	1	1
6 mm	12	6|12	6	3|6	4	2|4	6	3|6	2	1|2

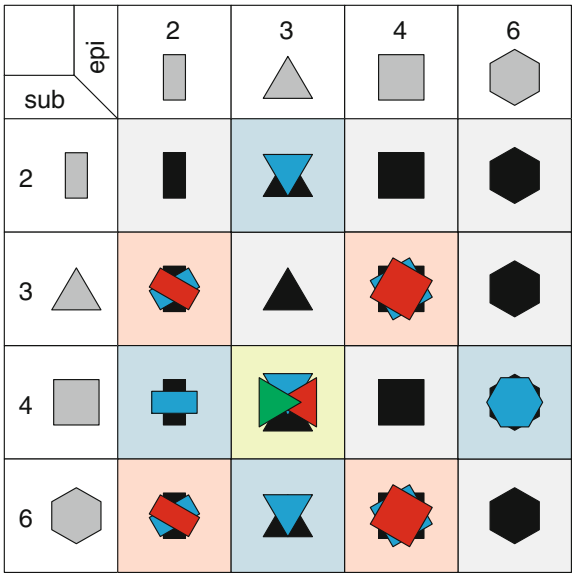


Fig. 12.10 Visualization of the minimum number of rotational domains for different substrate and epilayer symmetries according to Table 12.2

Due to 3 m mirror symmetry of the substrate, the misaligment is equivalent for clockwise and counter-clockwise rotation. Therefore two domains with an angle of 8° occur (Fig. 12.11).

Another well-known domain effect in heteroepitaxy is the occurrence of antiphase domains of zincblende semiconductors (e.g. GaAs, GaP, InP,...) grown on Si(001) with monoatomic steps; such surface actually exhibits two kind of terraces (1×2 and 2×1 reconstructions) and is thus itself not homogeneous. The formation of antiphase domains can be avoided by using an off-cut of the surface against the (001)-planes promoting double-atomic steps [1009]. In Fig. 12.12a the growth of GaP on a Ge

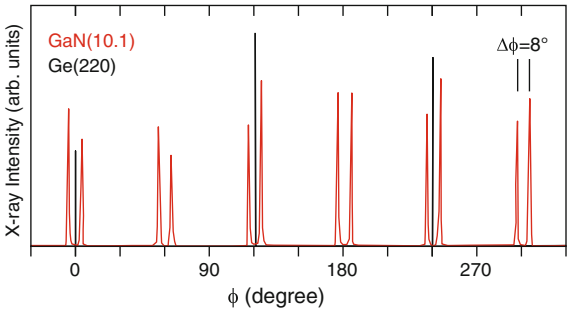


Fig. 12.11 X-ray diffraction ϕ-scans of GaN(00.1)/Ge(111). Adapted from [1008]

hemisphere with [111]-pole is depicted; spherical triangles meeting at the ⟨100⟩ poles can be seen. In a similar experiment GaAs was grown in a spherical depression in Ge(001) (Fig. 12.12b), allowing a more detailed look at the [001]-pole. Close to the it, within an approximately square area, antiphase domains (cmp. Sect. 4.4.4) form. Along the connection lines between different ⟨001⟩-poles, an antiphase domain boundary forms, microscopically broken up into domains [997].

Details of the initial deposition steps can determine the orientation in polar materials. GaN directly grown on c-Al_2O_3 grows with N-face orientation (see Fig. 3.19). The high surface mobility of Ga allows nitrogen to take its preferred position in the first atomic layer. Even under Ga-rich conditions the N atoms can kick-off the Ga from its favorite site on the surface. If an AlN buffer is used the strong bond between Al and oxygen leads to an Al atomic layer at the interface and subsequent GaN grows with a Ga-face [1010].

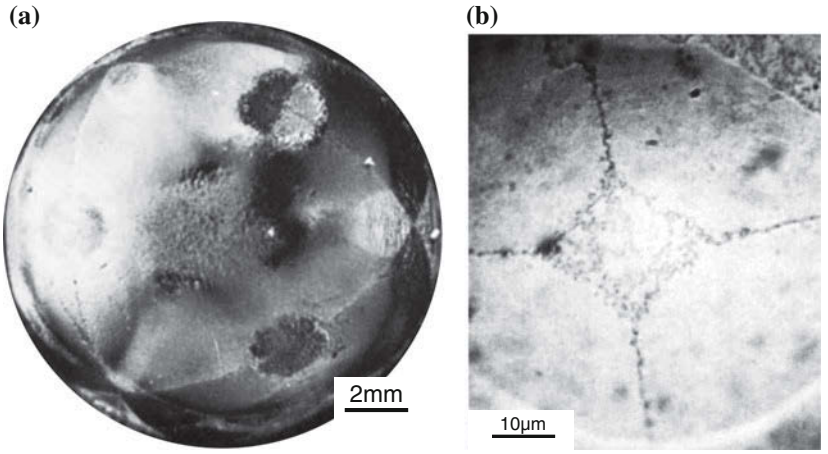


Fig. 12.12 The morphology of (**a**) GaP grown on a Ge hemisphere with [111]-pole and (**b**) GaAs grown in a Ge(001) spherical depression. Adapted from [997]

12.2.5 Patterned Substrates

Using patterning of the substrate certain growth modes and crystallographic directions of the growth front can be evoked.

Returning to the discussion of [00.1] or [00.$\bar{1}$] growth of wurtzites, on a sapphire substrate with lateral AlN patterns, laterally orientation-modulated GaN can be grown (Fig. 12.13). At the juncture of the phases an inversion domain boundary forms [1011].

The defect density can be reduced in parts of the structure using epitaxial lateral overgrowth (ELO) [1013]. The defects thread only from the limited contact area of the layer with the substrate and the part of the layer away from the mask ('wing') is free of defects (Fig. 12.14).

While the patterning of a mask may be cumbersome and require interruption of growth, the in-situ deposition of a random SiN mask with small holes has been found beneficial in the epitaxy of GaN [1014]. GaN islands nucleate selectively in the mask openings (Fig. 12.15a). The nucleation and subsequent coalescence of islands (Fig. 12.15b) leads to defect annihilation and eventually again to a flat film (Fig. 12.15c). The reduction of defect density is evident (Fig. 12.15d) and leads to improved optical properties [1015].

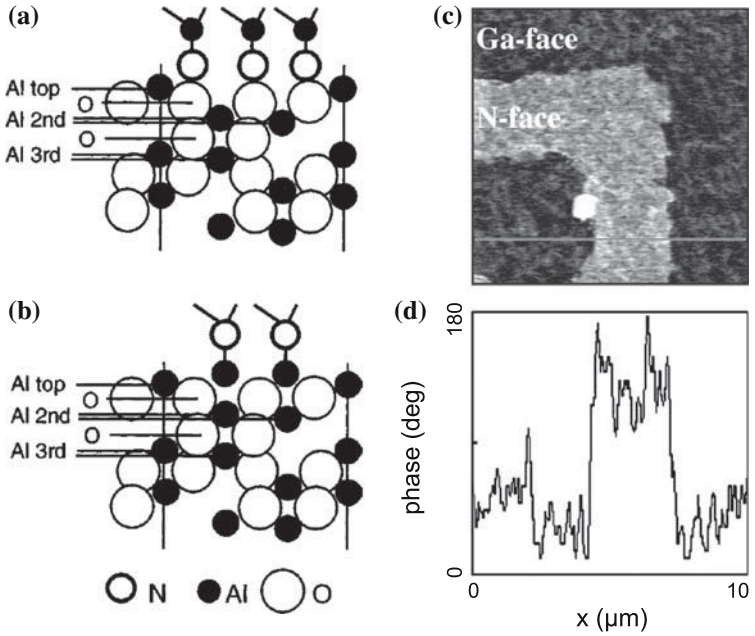


Fig. 12.13 Side view of the heterointerfaces between AlN and *c*-oriented sapphire with nitrogen (**a**) and Al (**b**) being the first layer. Adapted from [1010]. (**c**) Phase image of piezoresponse force microscopy (PFM) of lateral polarity GaN heterostructure and (**d**) linescan of phase signal along *white line* in part (**c**). Adapted from [1012], part (**c**) reprinted with permission, © 2002 AIP

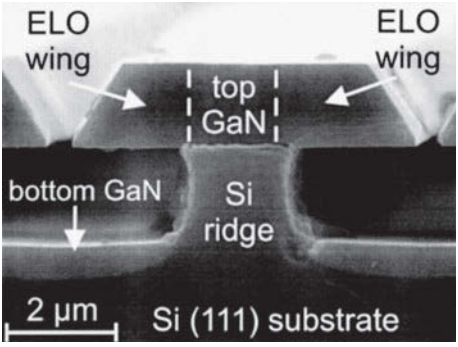


Fig. 12.14 SEM cross-sectional image of GaN grown on a structured Si(111) substrate. The laterally grown wings extend about 2.5 µm over the grooves. The thickness of the GaN layer is 0.5 µm on the bottom of the grooves, while it is 1.4 µm on top of the ridges. Reprinted with permission from [1013], © 2001 AIP

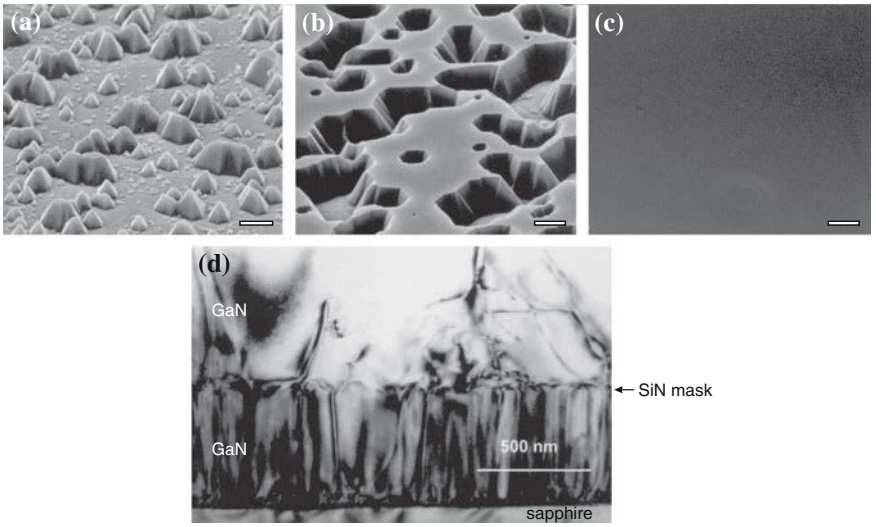


Fig. 12.15 SEM images of GaN (**a**) growing through the openings of a random SiN mask, (**b**) laterally overgrowing the mask and coalescing and (**c**) eventually forming a flat film. Bar widths are 2, 1 and 10 µm. (**d**) Cross-section TEM image. Adapted from [1014]

Using patterned substrates with stripe-like mesas and promoting the growth on a sidewall of the mesa, the direction of GaN growth front can be steered into a semipolar direction (cmp. Sect. 15.4.3). The r-plane sapphire substrate with etched mesa stripes oriented along the [11.0] direction exposes [00.1]-oriented sidewall facets on which the GaN grows along its *c*-axis; The angle of 57.6° between the r- and the c-planes of sapphire is very close to the angle (3.21c) of 58.4° between the GaN semipolar (11.2) and the (00.1) plane. The flat growth front represents a

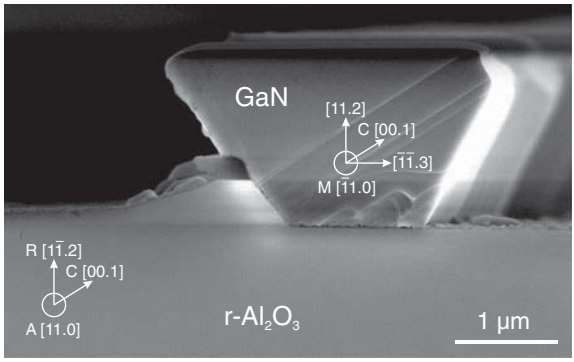


Fig. 12.16 SEM cross-sectional image of MOVPE GaN grown on patterned r-sapphire exposing a (11.2) growth front. Based on SEM image courtesy of F. Tendille

(11.2) facet [1016] (Fig. 12.16). After coalescence of grains from various mesas, a flat (11.2) surface arises, allowing the reduction of various defects including basal stacking faults [1017, 1018].

12.2.6 Pseudomorphic Structures

Heterostructures can generally be made from any sequence of materials. However a mismatch in lattice constant (or a different crystal structure) leads to strains and stresses that are of the order of 10^3 atmos for strains of 1 % ($\sigma \sim C\epsilon$, $C \approx 5 \times 10^{10}$ Pa) as discussed in Sect. 5.3.3. The total strain energy is $\propto C\epsilon^2$. Above a critical thickness $h_c \propto \epsilon^{-1}$ (cf. Sect. 5.4.1) defects, e.g. misfit dislocations (relaxing strain with their edge components), are generated (Sect. 12.2.7). There are a number of semiconductor combinations that are lattice matched and thus can be grown with arbitrary thickness. $Al_xGa_{1-x}As$ is closely lattice matched to GaAs for all Al concentrations. See Fig. 6.21 for lattice-matched pairs, e.g. $In_{0.53}Ga_{0.47}As$/InP. Often, thin layers of lattice-mismatched materials with thickness smaller than the critical thickness are used.

For many device applications the ability to grow pseudomorphic layers thicker than the critical layer thickness would be beneficial. The use of compliant substrates was proposed in [1019] to meet this demand. A recent review on compliant substrate technologies can be found in [393]. Schemes to accommodate part of the mismatch strain in the substrate include the use of cantilevered membranes, silicon-on-insulator, twist bonding, glass bonding or trenched substrates. In this sense also nanowhiskers (Sect. 14.2.3) represent a compliant substrate, enabling the growth of coherent (laterally confined) layers thicker than the critical thickness of a 2D layer (cmp. Fig. 14.9).

12.2.7 Plastic Relaxation

Above the critical thickness, a film will relax plastically by forming misfit dislocations. The cross-hatch pattern at the surface due to misfit dislocations is shown in Fig. 12.17. A TEM image of an array of misfit dislocations can be found in Fig. 5.33b.

The strain relaxation in mismatched heteroepitaxy can be determined experimentally via the wafer curvature of the heterostructure (Sect. 5.3.5). Data for the thickness dependent relaxation $\epsilon(d)$ of $In_{0.15}Ga_{0.85}As$ on GaAs (mismatch $\approx 1\%$) are shown in Fig. 12.18. Growth interruptions lead to larger relaxation at smaller thickness. Therefore relaxation at early times or small thickness (above h_c) is kinetically hindered, i.e., the available dislocation density and glide velocity are not sufficient to relieve the strain. At large thickness the strain does not go to zero (saturation regime) and the film remains in a metastable, incompletely relaxed state as discussed in more detail in [1020].

An extreme case of plastic relaxation is cracking (Sect. 4.4.1) which can occur for example in thick films during cooling due to the mismatch of the thermal expansion coefficients of substrate and epilayer (Fig. 12.19). Such cracking can be avoided by the introduction of suitable stress-relaxing layers and growth on predefined mesas [1021, 1022].

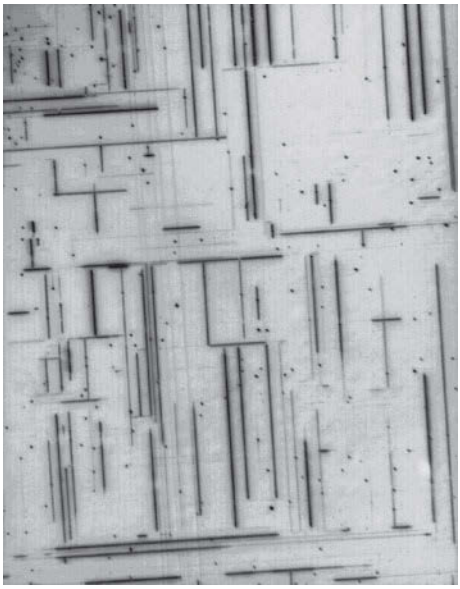


Fig. 12.17 Optical image of the surface of a supercritical, plastically relaxed $In_xGa_{1-x}As$ film on GaAs. Image width is about 1 mm. The cross-hatch pattern is due to misfit dislocations along [110] and [1$\bar{1}$0]. A pseudomorphic layer would exhibit no contrast under the given conditions

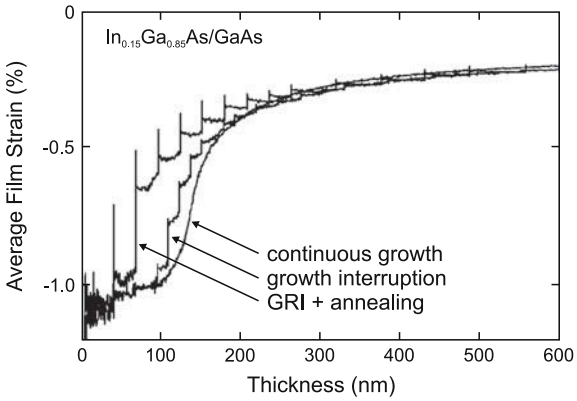


Fig. 12.18 Average film strain (measured in-situ via substrate curvature) of $In_{0.15}Ga_{0.85}As/GaAs$ grown at 450°C as a function of film thickness (determined from growth time × deposition rate). Relaxation for three different growth modes are shown (as labeled): continuous growth, employment of several growth interruptions (GRI) and GRI with annealing step (to 550°). Adapted from [1020]

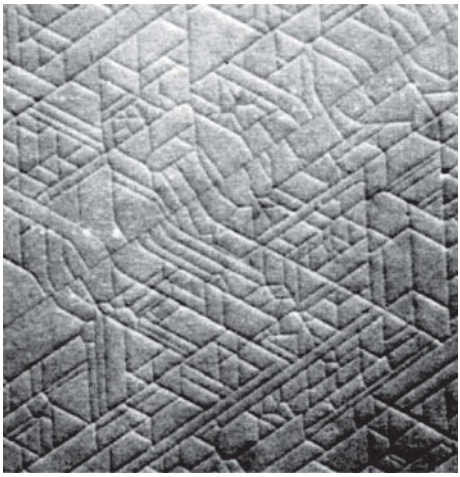


Fig. 12.19 $1 \times 1\ mm^2$ top view with a differential interference contrast microscope of a 1.3-μm thick GaN layer grown on Si(111). Reproduced with permission from [1021], © 2000 IPAP

12.2.8 Surfactants

The condition (12.1) allows layer-by-layer growth (cf. Sect. 12.2.3), i.e. the substrate surface free energy is higher than the than the total of interface and film surface free energy. This makes wetting energetically favorable. For two elements A and B, one of them must have the lower surface free energy. If A can be grown on B in Frank-van der Merwe (or Stranski-Krastanow) growth mode, then (12.1) does not hold for B on A and the growth will occur in Volmer-Weber mode, i.e. with islands. This is a serious problem for the growth of embedded layers of the type A–B–A. If the embedded film grows well, the capping does not and vice versa.

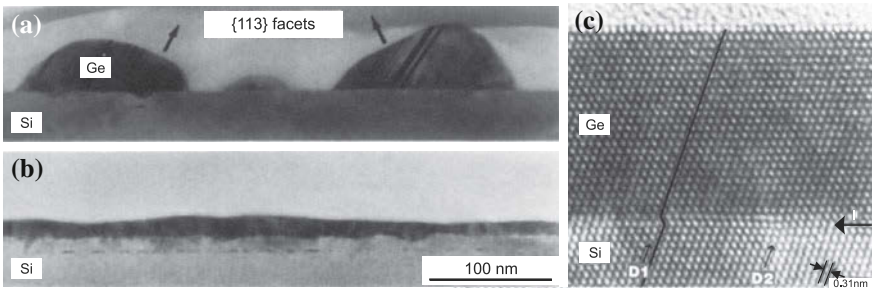


Fig. 12.20 (**a**, **b**) 10 nm thick Ge layer on Si (100) deposited at room temperature with MBE and annealed up to 770°C without (**a**) and with (**b**) Sb surfactant. Adapted from [1030]. (**c**) Cross-section TEM image of a heterostructure of 70 monolayers germanium on Si (111), grown by MBE with Sb surfactant. The *horizontal arrow* labeled 'I' denotes the position of the interface. The *arrows* labeled 'D1' and 'D2' denote the positions of partial dislocations (cmp. Fig. 4.19). Adapted from [289]

In the case of Ge on Si, Ge has the lower surface free energy than Si. Ge grows on Si in Stranski-Krastanow mode [1023] (Fig. 12.20a). Si grows both on Ge(001) and Ge/Si(001) in a Volmer-Weber mode [1024] causing severe problems for the fabrication of Si/Ge/Si quantum wells or superlattices. A substantial modification of growth mode can be achieved by using a third element C as a capping layer, saturating surface bonds. It lowers the surface free energy of both materials A and B, thus favoring wetting of the substrate. Such element C is called *surfactant* (surface-active species) [1025, 1026]. Typical examples are As [1025] or Sb [289] on Si and Ge (Fig. 12.20b). Also the surfactant modifies defect nucleation and can lead to defect-free epitaxial Ge/Si layers (Fig. 12.20c). Surfactants have also been described for the growth of compound semiconductors, e.g. In [1027] or Sb [1028, 1029] for GaAs.

12.3 Energy Levels in Heterostructures

12.3.1 Band Lineup in Heterostructures

In heterostructures, semiconductors with different band gaps are combined. The relative position of conduction and valence band (band alignment) is determined by the electron affinities χ as shown in Fig. 12.21. For a semiconductor, the electron affinity is the (positive) energy difference between vacuum level and conduction band edge. It can lead to different types of heterostructures. Early perspectives of semiconductor heterostructures are discussed in [1031].

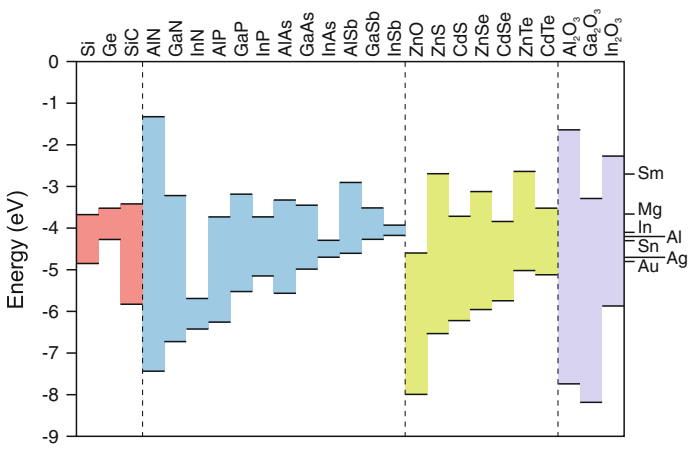


Fig. 12.21 Position of conduction and valence-band edges for a variety of semiconductors (relative to a common vacuum level at $E = 0\,\text{eV}$). Based on values from [1032], for Al_2O_3 (on InP) from [1033], for Ga_2O_3 (on GaN) from [1034] for In_2O_3 (on Si) from [1035, 1036]. On the right hand side the work functions of several metals are shown for comparison.

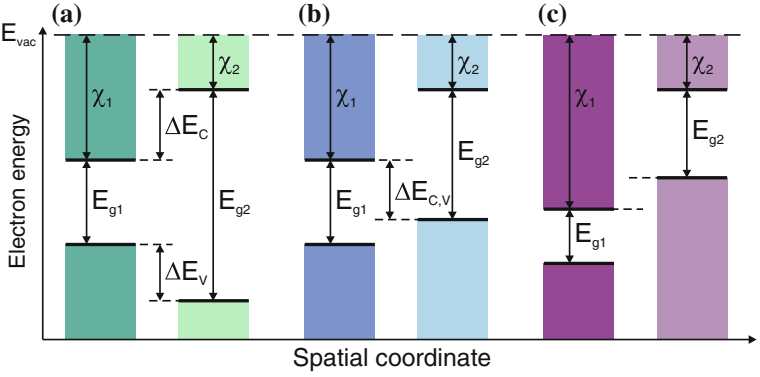


Fig. 12.22 Position of band edges (band alignment) in (**a**) type-I, (**b**) type-II and (**c**) type-III heterostructure.

In Fig. 12.22, the band alignment for type-I, type-II and type-III heterostructures are shown. In the type-I structure (straddled band lineup) the lower conduction-band edge and the higher valence-band edge are both in the material with smaller band gap. Thus, electrons and holes will localize there. In the type-II structure a staggered lineup is present and electrons and holes will localize in different materials. In the type-III structure, also termed 'broken gap' structure, the conduction band of one material is below the valence band of the other material. The technologically most relevant are type-I structures. The design of heterostructures to fulfill a certain device functionality or to have certain physical properties is called 'band gap engineering'.

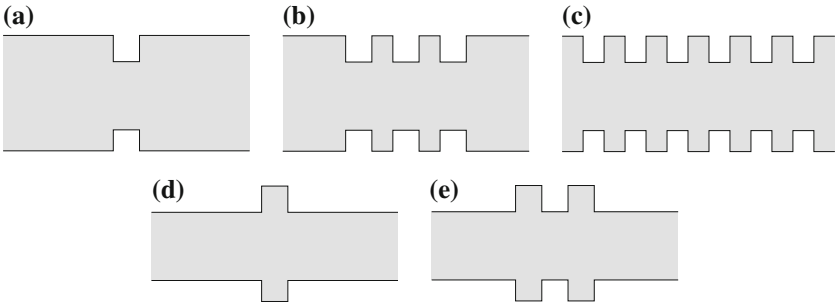


Fig. 12.23 Heterostructures with different layer sequences (band gap engineering). (**a**) quantum well (QW), (**b**) multiple quantum well (MQW), (**c**) superlattice (SL), (**d**) single-barrier tunneling structure, (**e**) double-barrier tunneling structure

In a type-I heterostructure, the conduction- and valence-band discontinuities are given, respectively, by

$$\Delta E_{\mathrm{C}} = \chi_1 - \chi_2 \tag{12.3a}$$

$$\Delta E_{\mathrm{V}} = (\chi_1 + E_{\mathrm{g}_1}) - (\chi_2 + E_{\mathrm{g}_2}). \tag{12.3b}$$

Depending on the layer sequence of high- and low-bandgap materials various configurations, as shown in Fig. 12.23 have obtained special names, such as single heterointerface, quantum well (QW), multiple quantum well (MQW), superlattice (SL). In the extreme case the layer is only one monolayer thick (Fig. 12.24) and the concept of layer and interface blurs. Such atomically precise layer sequences are mastered nowadays for a variety of material systems such as (Al,Ga)As/GaAs/InAs, InP/(In,Ga)As, Si/SiGe, ZnO/(Mg,Zn)O and also $BaTiO_3/SrTiO_3$.

The abruptness of interfaces is determined by the epitaxial machine through the switching precision of the incoming material flux and fundamentally limited by segregation phenomena which can be modeled with a segregation coefficient [1037] (cmp. Sect. 4.2.4). As shown in Fig. 12.24c, In has the tendency to be carried into the following GaAs layer. A quantitative evaluation is possible using aberration-corrected scanning transmission electron microscopy [1038].

12.3.2 Quantum Wells

The energy in a single quantum well of thickness L_z (along the growth direction z) can be calculated with the quantum-mechanical particle-in-a-box model. In the envelope function approximation (Appendix H) the wavefunction is written as a product of the Bloch function and the envelope function $\chi(z)$.

$$\Psi^{\mathrm{A,B}}(\mathbf{r}) = \exp(\mathrm{i}\,\mathbf{k}_\perp \mathbf{r})\; u_{n\mathbf{k}}(\mathbf{r})\, \chi_n(z), \tag{12.4}$$

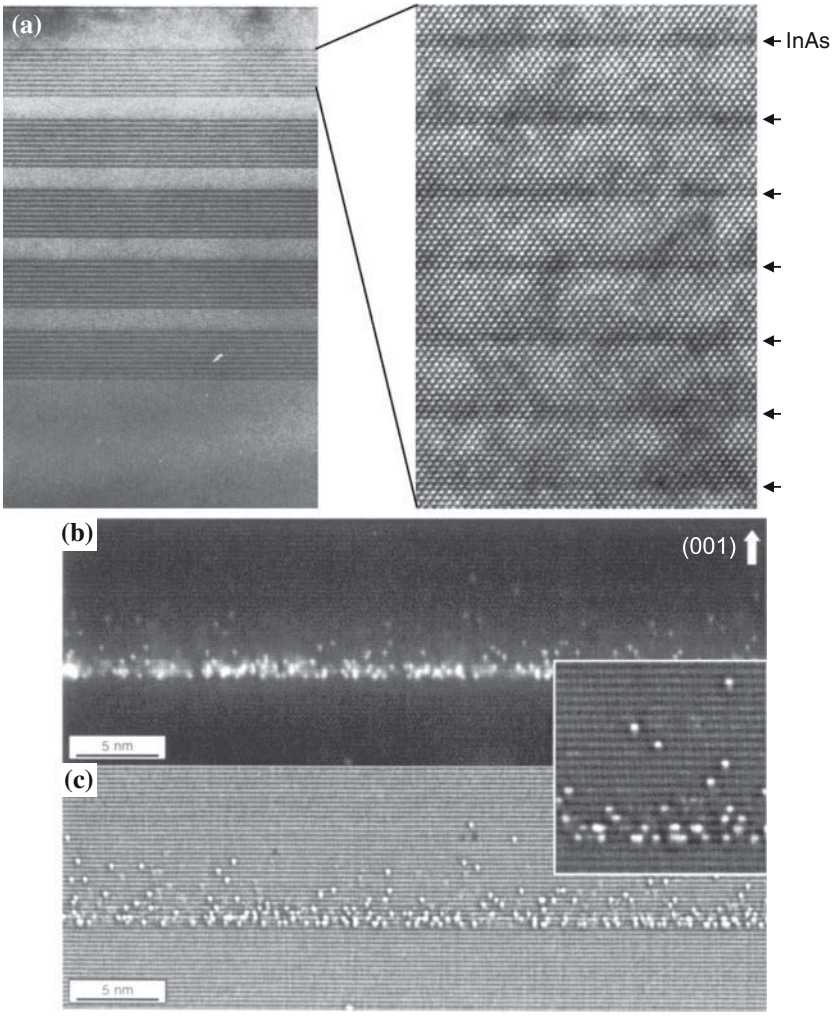


Fig. 12.24 Ultrathin heterostructures: (**a**) Cross-sectional TEM of a MOVPE-grown short-period superlattice (SPS) of InAs layers in $GaAs_{1-x}N_x$). In high resolution (*right image*), the individual rows of atoms can be seen. From [1039]. (**b**, **c**) Cross-section STM image of 2 ML InAs on GaAs; the segregation of In into the *top layer* is visible atom by atom. Adapted from [1040]

where 'A' and 'B' denote the two different materials. The envelope function χ fulfills, approximately, the one-dimensional Schrödinger-type equation,

$$\left[-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2} + V_c(z)\right]\chi_n(z) = E_n\,\chi_n(z), \tag{12.5}$$

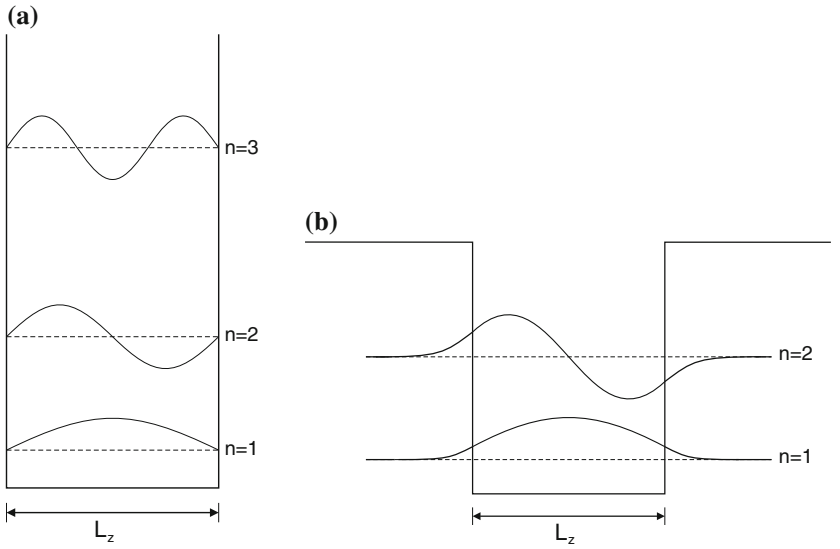


Fig. 12.25 Schematic energy levels and wavefunctions in a potential well with (**a**) infinite barriers, and (**b**) finite barrier height

where m^* denotes the effective mass. V_c is the confinement potential determined by the band discontinuities. Typically, $V_c = 0$ in the well and $V_0 > 0$ outside in the barrier. E_n are the resulting energy values of the quantized levels. In the case of infinite barriers ($V_0 \to \infty$, Fig. 12.25a) the boundary conditions $\chi(0) = \chi(L_z) = 0$ yield

$$E_n = \frac{\hbar^2}{2m^*} \left(\frac{n\,\pi}{L_z} \right)^2 \tag{12.6}$$

$$\chi_n(z) = A_n \sin\left(\frac{n\,\pi}{L_z}\, z \right), \tag{12.7}$$

where E_n is called the confinement energy.

For finite barrier height V_0 (Fig. 12.25b) the calculation leads to a transcendental equation. The wavefunction tunnels into the barrier. While for infinite barrier height the lowest level diverges for $L_z \to 0$, for finite barrier height $E_1 \to V_0$. A complication arises from the different effective mass in the well and barrier material. This is taken into account by forcing the continuity of χ and χ'/m^* across the interfaces[4] ('BenDaniel-Duke' boundary conditions [1042]). The Schrödinger equation and (semi-)analytical solutions for special and numerical methods for

[4]The kinetic energy term in (12.5) is written as $\frac{\hbar^2}{2}\,\frac{\partial}{\partial z}\,\frac{1}{m^*(z)}\,\frac{\partial \chi}{\partial z}$ for varying mass across the structure [1041].

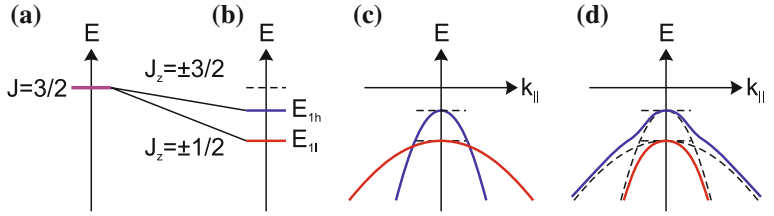


Fig. 12.26 Schematic representation of the development of hole levels in a quantum well: (**a**) degenerate bulk levels at Γ, (**b**) splitting at the subband edge (due to different quantized values of k_z), (**c**) in-plane dispersion (mass reversal), (**d**) anticrossing behavior. Based on [991]

arbitrary potential distributions are discussed in [1043]. The application of $\mathbf{k} \cdot \mathbf{p}$ theory (Appendix G) to heterostructures is discussed in [1044].

The motion of carriers in the plane is still free and has a two-dimensional dispersion. Thus, each quantized level contributes $m^*/(\pi\hbar^2)$ to the density of states at each subband edge E_n.

For holes, the situation is a little more complicated than for electrons (Fig. 12.26). First, the degeneracy of heavy and light holes is lifted since their mass enters the confinement energy. The effective hole masses along the z direction, i.e. those that enter (12.5), are

$$\frac{1}{m^z_{\mathrm{hh}}} = \gamma_1 - 2\,\gamma_2 \tag{12.8a}$$

$$\frac{1}{m^z_{\mathrm{lh}}} = \gamma_1 + 2\,\gamma_2. \tag{12.8b}$$

The light holes have the higher quantization energy. The angular momentum is quantized along the z direction. The transverse masses for the dispersion in the interface plane are

$$\frac{1}{m^{xy}_{\mathrm{hh}}} = \gamma_1 + \gamma_2 \tag{12.9a}$$

$$\frac{1}{m^{xy}_{\mathrm{lh}}} = \gamma_1 - \gamma_2. \tag{12.9b}$$

Now the heavy hole, i.e. the $J_z = \pm\frac{3}{2}$ state, has the smaller mass and the light hole ($J_z = \pm\frac{1}{2}$) the larger (Fig. 12.26c). However, this consideration is only an approximation since the lifting of degeneracy and the dispersion have to be treated on the same level. Higher terms of the perturbation calculation lead to band mixing and remove the band crossing that seems to originate from the situation at the Γ point. In reality, the bands show anticrossing behavior and are strongly deformed. The hole dispersion in a superlattice and the anticrossing behavior is shown in Fig. 12.27.

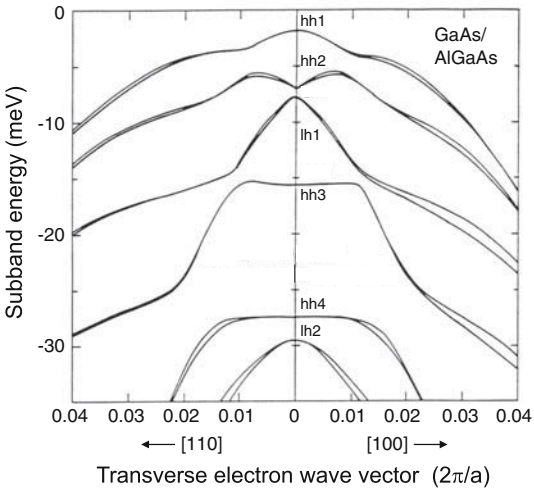


Fig. 12.27 Hole dispersion in a 68-ML GaAs/71 ML $Al_{0.25}Ga_{0.75}As$ superlattice (numerical calculation). The double curves originate from a lifting of time-reversal symmetry at $\mathbf{k} \neq 0$. Reprinted with permission from [1045], © 1985 APS

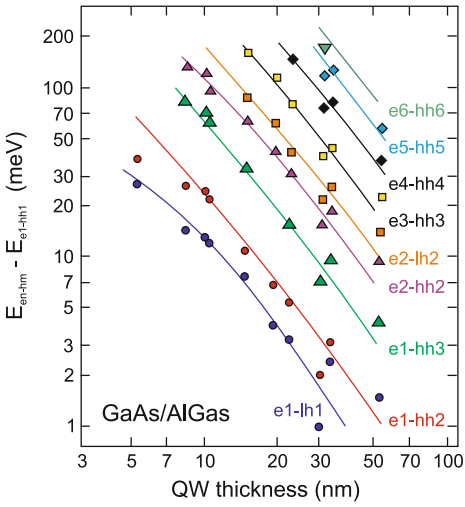


Fig. 12.28 Observed electron–hole transitions (energy difference to the first e–h transition from excitation spectroscopy) in GaAs/AlGaAs quantum wells of varying thickness. *Symbols* are experimental data, *solid lines* are theoretical model. Data from [1046]

Experimentally observed transition energies in quantum wells of varying thickness are shown in Fig. 12.28 and are in good agreement with the theoretical calculation. We note that for infinite barriers optical transitions are only allowed between confined electron and hole states with the same quantum number n. For finite barriers this selection rule becomes relaxed, and other transitions become partially allowed, e.g. e_1–hh_3. The optical matrix element from the Bloch part of the wavefunction, which was isotropic for (cubic) bulk material (9.25), is anisotropic for quantum wells. TE (TM) polarization is defined with the electromagnetic field in (perpendicular to) the plane of the quantum well (Fig. 12.29a). At the subband edge, i.e. for in-plane wavevector $k_{||} = 0$ the matrix elements for the various polarizations and propagation

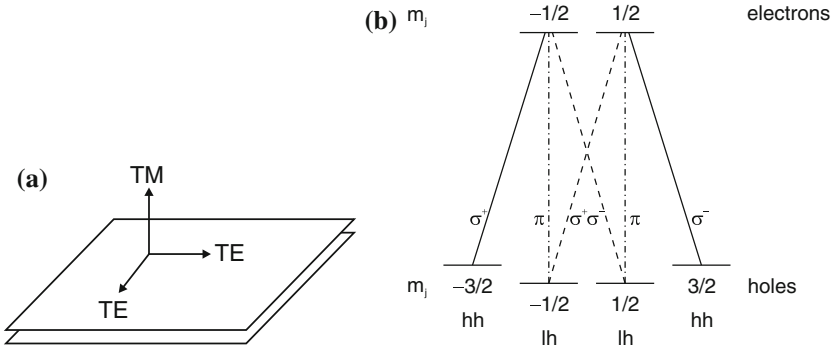


Fig. 12.29 (**a**) Directions of electric-field vector relative to the quantum-well plane for TE and TM polarization. (**b**) Optical selection rules for band–band transitions in a quantum well. If the (in-plane averaged) relative strength of the e–hh transitions (*solid lines*) is 1, the relative strength of the TE-polarized e–lh transitions (*dashed lines*) is 1/3 and that of the TM-polarized e–lh transitions (*dash-dotted lines*) is 4/3

Table 12.3 Squared momentum matrix elements $\left|\langle c|\hat{\mathbf{e}}\cdot\mathbf{p}|v\rangle\right|^2$ in a quantum well for various propagation directions in units of M_b^2. The quantum-well normal is along **z**

	Propagation	$\hat{\mathbf{e}}_\mathrm{x}$ (TE)	$\hat{\mathbf{e}}_\mathrm{y}$ (TE)	$\hat{\mathbf{e}}_\mathrm{z}$ (TM)
e–hh	**x**	–	1/2	0
	y	1/2	–	0
	z	1/2	1/2	–
	x	–	1/6	2/3
e–lh	**y**	1/6	–	2/3
	z	1/6	1/6	–
	x	–	1/3	1/3
e–so	**y**	1/3	–	1/3
	z	1/3	1/3	–

directions are given in Table 12.3. The matrix elements averaged over all in-plane directions for TE-polarization are $3/2M_\mathrm{b}^2$ ($1/2M_\mathrm{b}^2$) for the electron to heavy (light) hole transition. For TM polarization the values are 0 and $2M_\mathrm{b}^2$, respectively [1047]. The optical selection rules are shown in Fig. 12.29 (see Fig. 9.9 for bulk material). For propagation along the quantum-well plane, the ratio between the strength of the TE polarized e–hh and e–lh transitions is 3:1.

The confinement potential squeezes charge carriers bound to impurities closer to the ion. Therefore, the binding energy increases as shown in Fig. 12.30. This behavior can be modeled theoretically with good precision. It makes a difference whether the impurity is located at the center or the interface of the quantum well.

The confinement potential also squeezes electrons and holes in the exciton closer together and thus increases their Coulomb interaction. The binding energy of the quantum-well exciton is thus larger than in bulk material and depends on the well

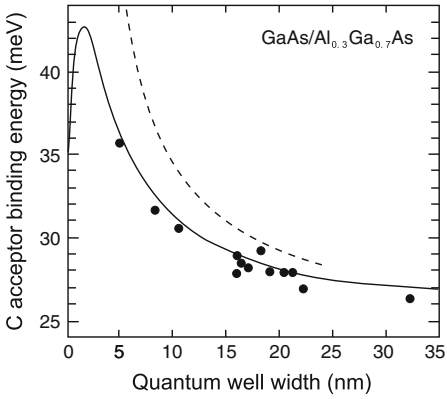


Fig. 12.30 Experimental values for the acceptor binding energy in $GaAs/Al_{0.3}Ga_{0.7}As$ quantum wells (*solid circles*) from [1048] as a function of well width. *Solid line* is theory (variational calculation) for the well-center acceptor including top four valence bands and finite barriers, *dashed line* is hydrogen-like model with infinite barrier height. Adapted from [1049]

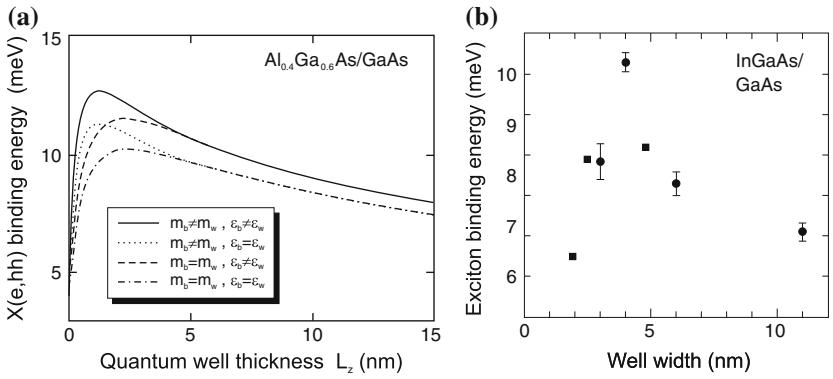


Fig. 12.31 (**a**) Theoretical (variational) calculation (*solid line*) of the heavy-hole exciton binding energy versus QW thickness in a $GaAs/Al_{0.4}Ga_{0.6}As$ quantum well (using also different approximations, *other lines*). Adapted from [1050]. (**b**) Experimental exciton binding energy in $In_xGa_{1-x}As/GaAs$ quantum wells of different thickness. *Circles*: data and error bars from [1051], x unspecified, *squares*: data from [1052], $x = 0.18$

width (Fig. 12.31). In the simple hydrogen-like model with infinite barriers the exciton binding energy is 4 times the bulk binding energy in the limit $L_z \to 0$. In a realistic calculation the effect of different dielectric constants in the well and barrier (image charge effect) need to be considered.

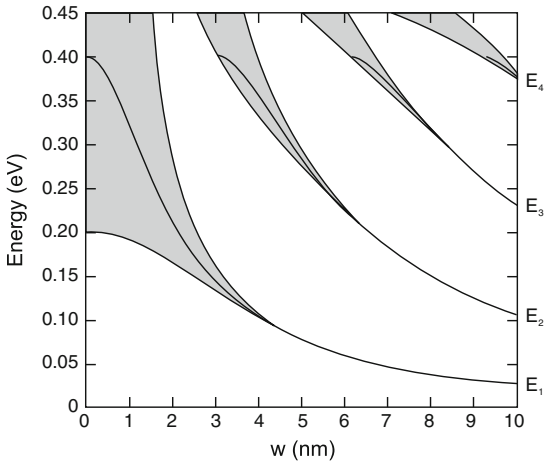


Fig. 12.32 Band structure of a superlattice with a potential depth of 0.4 eV and well and barrier width w ($L_{QW} = L_{barr}$). Adapted from [1053]

12.3.3 Superlattices

In a superlattice, the barrier thickness is so small that carriers can tunnel in neighboring wells or, in other terms, that there exists a significant wavefunction overlap between adjacent wells. This leads to a band structure (Fig. 12.32), similar to the Kronig–Penney model (Appendix F). For the superlattice the bands are called minibands, the gaps are called minigaps. The density of states does not make a step at the subband edge but follows an arccos function. The modification of the density of states, as seen in the absorption spectrum, are shown in Fig. 12.33 for 1, 2, 3 and 10 coupled wells.

12.3.4 Single Heterointerface Between Doped Materials

We consider a single heterointerface between n-doped materials. As an example we take n-AlGaAs/n-GaAs (Fig. 12.34). First, we consider the materials without contact, forming a type-I structure. In thermodynamic equilibrium the system must have a constant Fermi level. Thus, charge is transferred from the region close to the interface from AlGaAs to GaAs. This results in the formation of a triangular potential well in the GaAs close to the interface. A two-dimensional electron gas (2DEG) forms in this potential well (Fig. 12.35). The charge transfer in thermodynamic equilibrium adjusts the band bending and the charge density (quantized levels in the well) in such a way that they are self-consistent. The Poisson equation and the Schrödinger equation are simultaneously fulfilled. Numerically, both equations are iteratively solved and the solution is altered until it is self-consistent, i.e. it fulfills both equations.

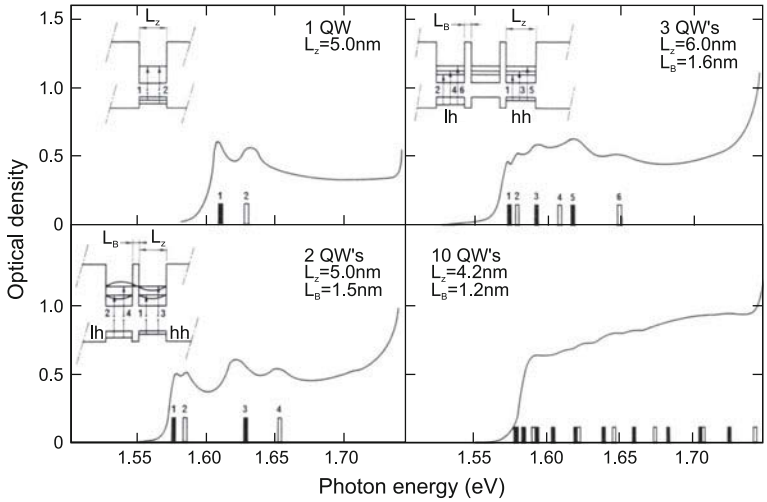


Fig. 12.33 Absorption spectra of a single, double, triple and ten coupled quantum wells. Theoretically predicted transitions with heavy (light) holes are labeled with *filled* (*empty*) *bars* at their respective transition energies. Adapted from [1054]

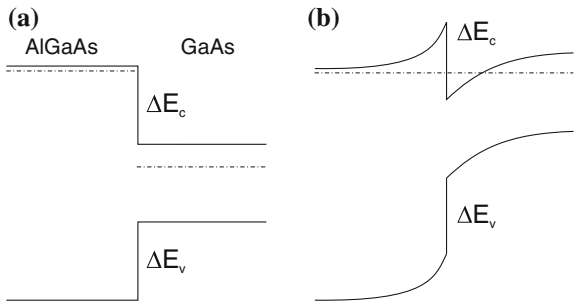


Fig. 12.34 Schematic formation of a triangular potential well in a n-AlGaAs/n-GaAs heterostructure, (**a**) before and (**b**) after equilibration of Fermi levels

If the region of the 2DEG is not doped, the electron gas exists without any dopant atoms and ionized impurity scattering no longer exists. This concept is called *modulation doping*. Mobilities up to 3.1×10^7 cm^2/Vs have been realized (Fig. 12.36). The theoretical limits of mobility in a 2DEG at modulation-doped AlGaAs/GaAs heterointerfaces are discussed in detail in [1056].

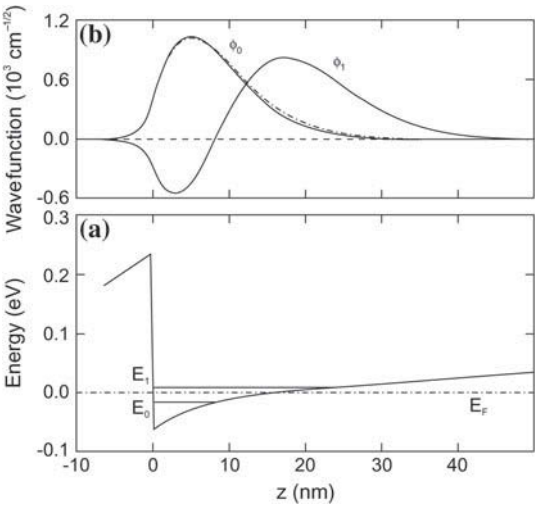


Fig. 12.35 (**a**) Conduction-band edge at a GaAs/$Al_{0.3}Ga_{0.7}As$ heterointerface ($T = 0\,K$) with two confined states at E_0 and E_1 marked with *solid horizontal lines*. In the GaAs channel there are $5 \times 10^{11}\,cm^{-2}$ electrons. The barrier height is 300 meV, $N_D^{GaAs} = 3 \times 10^{14}\,cm^{-3}$. The position of the Fermi level E_F is at $E = 0$ and indicated with a *dash-dotted line*. (**b**) Envelope wavefunctions ϕ_0 and ϕ_1 of the two confined states, *dash-dotted line*: calculation without exchange and correlation for state at E_0. Adapted from [1055]

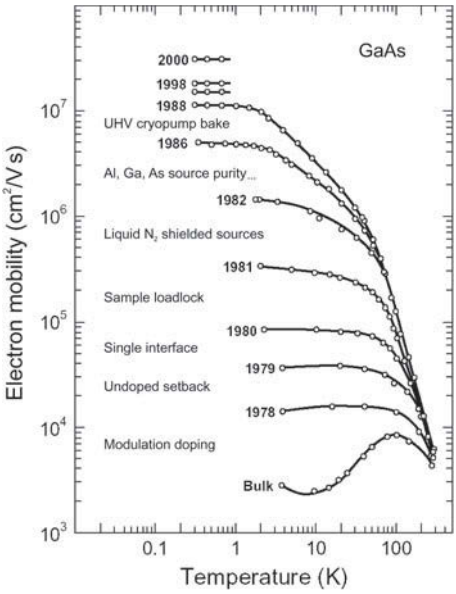


Fig. 12.36 Progress in the achievement of high electron mobility in GaAs, annotated with the technical innovation responsible for the improvement. Adapted from [1057]

12.4 Recombination in Quantum Wells

12.4.1 Thickness Dependence

The energy of exciton recombination in quantum wells is blue-shifted with respect to that in bulk material due to the quantum-confinement energies of electrons and holes (Fig. 12.37). The electron–hole recombination lineshape in quantum wells is given by the product of the joint density of states and the Boltzmann function (when Boltzmann statistics apply). The JDOS is given by a step function (Heavyside function $H(E)$).

$$I(E) \propto H(E - E_{11}) \exp\left(-\frac{E}{kT}\right), \tag{12.10}$$

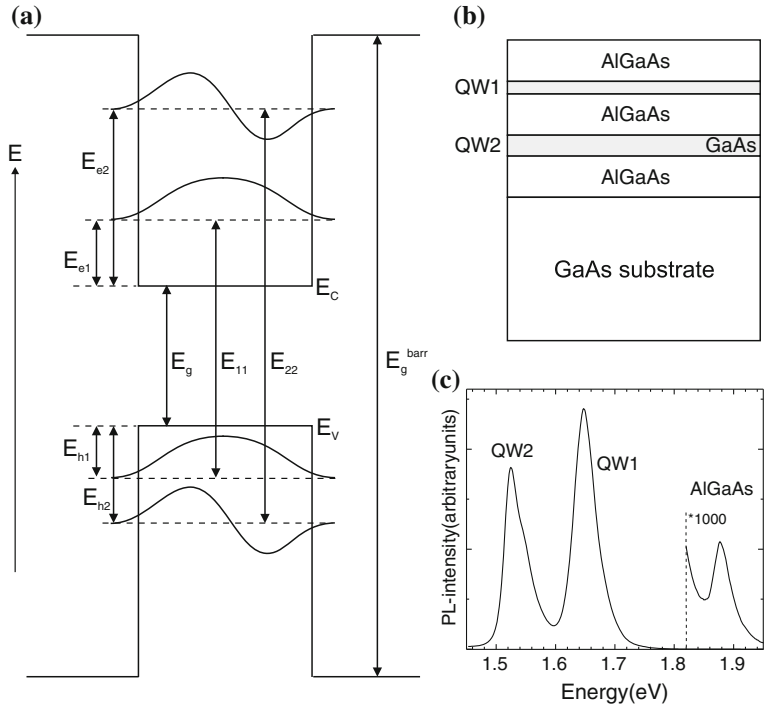


Fig. 12.37 (**a**) Schematic energy diagram of a quantum well with confined electron (e1, e2) and hole (h1, h2) states and recombination between them at energies E_{11} and E_{22}. (**b**) Schematic sample structure with two GaAs/$Al_xGa_{1-x}As$ quantum wells with thicknesses 3 and 6 nm. (**c**) Photoluminescence spectrum ($T = 300$ K) of the structure from part (**b**). A small amount of barrier luminescence appears at 1.88 eV, according to $x = 0.37$ (cf. Fig. 6.19c)

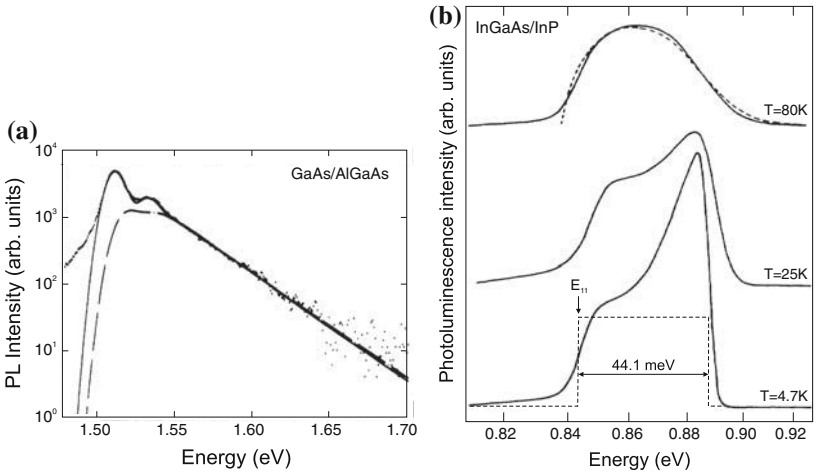


Fig. 12.38 (**a**) Reflectance spectra at various temperatures from a 17 nm thick GaAs/$Al_{0.3}Ga_{0.7}As$ quantum well. The *inset* shows the temperature dependence of the homogeneous linewidth. (**b**) LO phonon broadening parameter (FWHM) for various quantum well widths. Adapted from [1061]s

where $E_{11} = E_g + E_{e1} + E_{h1}$ represents the energy of the E1–H1 subband edge as shown in Fig. 12.37. An experimental spectrum (Fig. 12.39a) shows that excitonic effects influence the recombination lineshape in a GaAs quantum well even at room temperature [1058].

The recombination decay constant of excitons decreases with decreasing well width, partly due to the increase of exciton binding energy as discussed in [1059].

12.4.2 Broadening Effects

Many-Body Effects

At high carrier densities when the electron (quasi-) Fermi level is above the electron subband edge, the spectrum broadens and reflects the Fermi–Dirac distribution (Fig. 12.39b). At low temperatures a many-body effect, multiple electron–hole scattering with electrons at the Fermi edge, leads to an additional peak, termed *Fermi-edge singularity* that is discussed in [1060].

Homogeneous Broadening

The temperature dependence of the homogeneous broadening of quantum well luminescence has been investigated in [1061]. It follows the dependence of the broadening known from bulk material (Sect. 9.6.7) with similar values for the LO broadening parameter. In Fig. 12.38a the reflectance spectra for different temperatures of a 17 nm GaAs/$Al_{0.3}Ga_{0.7}As$ QW are shown. The optical phonon broadening parameter for various well widths is shown in Fig. 12.38b and coincides with the bulk value.

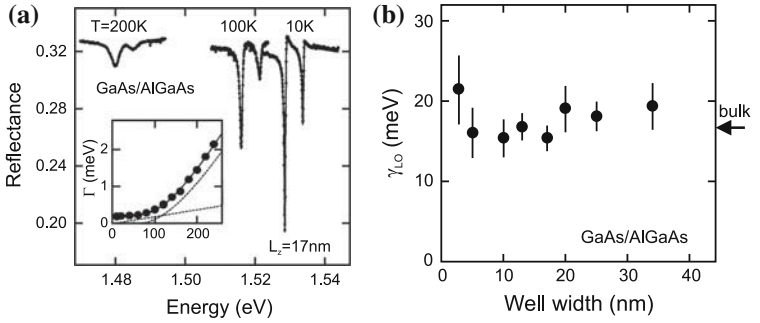


Fig. 12.39 (**a**) Photoluminescence spectrum of a 5-nm GaAs/AlGaAs quantum well at $T = 300\,\mathrm{K}$. The *solid* (*dashed*) line is fit with (without) excitonic effects. The two peaks are due to transitions involving heavy and light holes. Adapted from [1058]. (**b**) Photoluminescence spectra at three different temperatures as labeled of a 10-nm modulation-doped InGaAs/InP quantum well with an electron sheet density $n_s = 9.1 \times 10^{11}\,\mathrm{cm}^{-2}$. The electron quasi-Fermi level is $F_n - (E_C + E_{e1}) = 44.1\,\mathrm{meV}$ from the subband edge. The *dashed line* in the $T = 80\,\mathrm{K}$ spectrum is the lineshape from JDOS and a Fermi–Dirac distribution without enhancement at the Fermi edge. Adapted from [1060]

The homogeneous broadening leads to excitons with in-plane center-of-mass wave-vector $\mathbf{K} \neq 0$ being allowed to recombine radiatively. This leads to a linear increase of exciton lifetime as demonstrated for GaAs quantum wells up to 50 K in [1059]. In [1062] all exciton wave-vectors within the light cone are considered, explaining the linear increase of exciton lifetime in (non-polar) ZnO quantum wells up to room temperature.

Inhomogeneous Broadening

Inhomogeneous broadening affects the recombination lineshape. Since the interfaces of the QW are not ideally flat, the exciton averages over different quantum-well thicknesses within its volume. Also, e.g. for the GaAs/AlGaAs system, the wavefunction in the (binary) quantum well tunnels into the barrier, the amount depending on the QW width, and there 'sees' the alloy broadening (see Sect. 10.3.3). The problem of exciton dynamics in a potential with random fluctuations has been treated in detail [1063, 1064].

A simplified picture is as follows: At low temperatures the excitons populate preferentially the potential minima. A simple lineshape[5] of the QW absorption or joint density of states is given by a step function (cf. Table 9.2) at the QW band edge E_0. The inhomogeneous broadening has a Gaussian probability distribution $p(\delta E) \propto \exp[-(\delta E)^2/2\sigma^2)]$ with δE being the deviation from the QW band edge $\delta E = E - E_0$. The resulting lineshape is given by the convolution of the Gaussian

[5] neglecting excitonic enhancement.

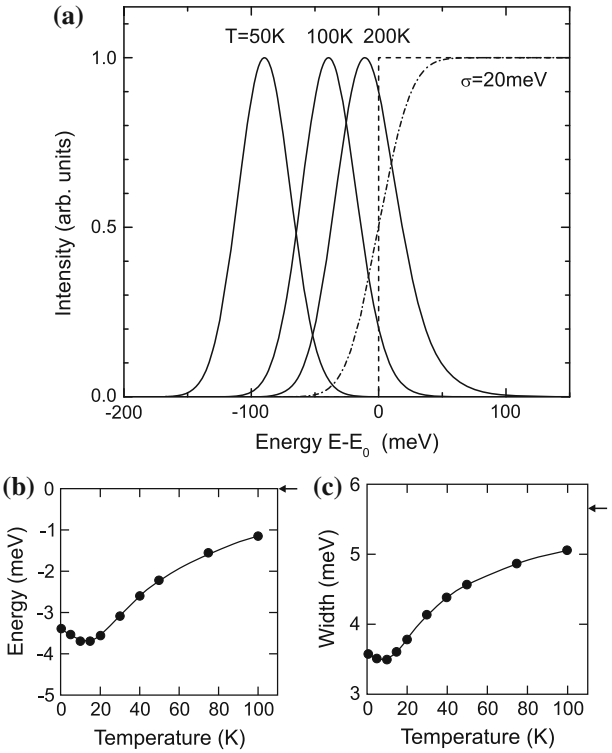


Fig. 12.40 (**a**) Recombination spectra (*solid lines*, scaled to same height) of a model quantum well for different temperatures as labeled and complete thermalization, *dashed* (*dash-dotted*) *line* is unperturbed (inhomogeneously broadened by $\sigma = 20\,\text{meV}$) shape of the QW absorption edge. The energy scale is relative to the energy position of the unperturbed QW absorption edge at E_0. (**b**) Theoretical energy position and (**c**) linewidth of exciton recombination from a model disordered quantum well. The high-temperature limits are marked by *arrows*. Parts (**b**, **c**) adapted from [1063]

with the unperturbed absorption spectrum yielding an error-function-like spectrum[6] as shown in Fig. 12.40a.

For complete thermalization the level population is given by the Boltzmann function. The recombination spectrum is given by the product of the absorption spectrum (or JDOS) and the Boltzmann function. It is (red-) shifted with respect to E_0 by about[7]

$$\Delta E(T) = -\frac{\sigma^2}{kT} = \gamma(T)\, kT. \tag{12.11}$$

This shift between emission and absorption is also called the Stokes shift.

[6]The error function is defined as $\text{erf}(x) = (2/\sqrt{\pi}) \int_0^z \exp -t^2 \,\text{d}t$.

[7]Formula (12.11) is exact for the product of a Gaussian and the Boltzmann function.

Within their lifetime, limited at least by radiative recombination, the excitons are typically unable to reach the energy position required by the Boltzmann function, but only a local minimum. Thus, their thermalization may be incomplete due to insufficient lateral diffusion. This effect is particularly important at low temperatures when thermal emission into adjacent deeper potential minima is suppressed. In this case, the red-shift is smaller than expected from (12.11). A numerical simulation [1063] yields such behavior of the energy position of the recombination line as shown in Fig. 12.40b. Simultaneously, the width of the recombination spectrum also exhibits a minimum (Fig. 12.40c). These findings are in agreement with experiments [1065, 1066]. An analytical model for temperature dependent exciton localization in the presence of disorder has been given in [1067], yielding a value $0 \leq \gamma \leq \gamma_0 = (\sigma/kT)^2$ in (12.11).

A potential fluctuation can localize an exciton laterally at low temperatures [1058] and behave like a quantum dot (cf. Sect. 14.3). Localized and delocalized excitons are separated by a boundary called the mobility edge [1068]. The transition between the two regimes is a Mott transition [1069].

Monolayer Growth Islands

Under certain growth conditions, quantum wells with piecewise very flat interfaces can be fabricated. The thickness difference between such regions (with lateral extension in the μm range) is an integer monolayer. Accordingly, the recombination spectrum yields several, typically two or three, discrete lines (Fig. 12.41).

12.4.3 Quantum Confined Stark Effect

The quantum confined Stark effect (QCSE, Sect. 13.1.2) in quantum wells shifts energy levels when electric fields are present along the width of the quantum well. A strong effect exists for pyro- and piezoelectric material combinations such as *c*-axis oriented (In,Ga)N/(Al,Ga)N [1070, 1071] or (Cd,Zn)O/(Mg,Zn)O [1072, 1073] quantum wells due to the built-in electric field (cf. Sect. 15.2). The QCSE induced red-shift is larger in thicker quantum wells and goes beyond the bulk band gap of the quantum well material (Fig. 12.42b). Also the wavefunction overlap is reduced with increasing well width, leading to an increase of the radiative recombination lifetime as shown in Fig. 12.42. The pyroelectric field and the related modification of lifetime are absent in quantum wells grown on non-polar directions such as [11.0] (Fig. 12.42a).

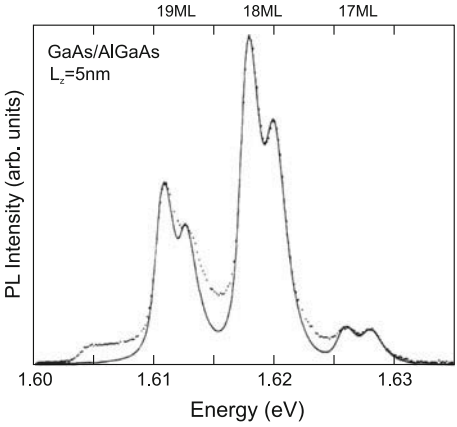


Fig. 12.41 Photoluminescence spectrum ($T = 2\,\mathrm{K}$) (*dots*) of GaAs/AlGaAs quantum well grown by MBE with 120 s growth interruptions. Recombination is due to excitons in islands of 19, 18, and 17 monolayers ($a_0/2$) height. The *solid line* is a lineshape fit including lifetime broadening ($\Gamma = 1.34\,\mathrm{meV}$) and residual inhomogeneous broadening ($\sigma = 0.04\,\mathrm{meV}$) due to AlGaAs barrier alloy fluctuations. Note that the energy separation of the peaks is much larger than kT. The peak doublet structure is discussed in [1058]. Adapted from [1058]

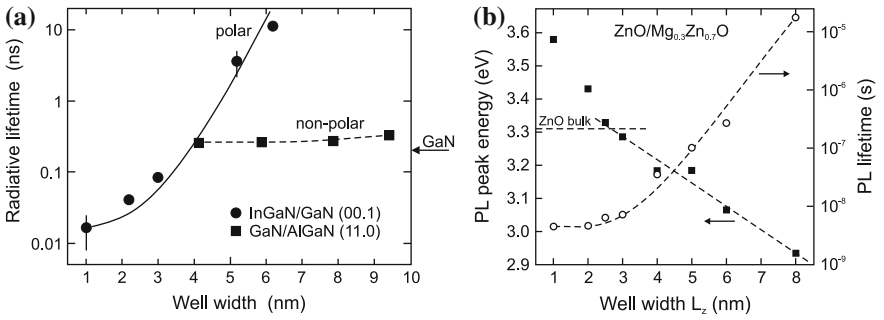


Fig. 12.42 (**a**) Radiative lifetime of electron-hole pairs in polar [00.1]-oriented $In_{0.2}Ga_{0.8}N$/GaN (*circles*) and non-polar [11.0]-oriented GaN/$Al_{0.2}Ga_{0.8}N$ (*squares*) quantum wells of varying thickness. Experimental data are shown in *symbols*. The *solid line* is the (scaled) theoretical dependence of the electron-hole overlap for (In,Ga)N/GaN QWs. The *dashed line* is guide to the eyes. The *arrow* denotes the recombination time constant in bulk GaN. Adapted from [1074, 1075]. (**b**) Low temperature PL peak recombination energy (*solid squares*) for ZnO/$Mg_{0.3}Zn_{0.7}O$ quantum wells for various well widths L_z (barrier width $L_B = 5\,\mathrm{nm}$). *Dashed line* indicates dependence for internal field of 0.9 MV/cm, *horizontal dashed line* indicates recombination energy in ZnO bulk. Carrier lifetime determined from PL (*circles*), *dashed line* is guide to the eyes. Adapted from [1073]

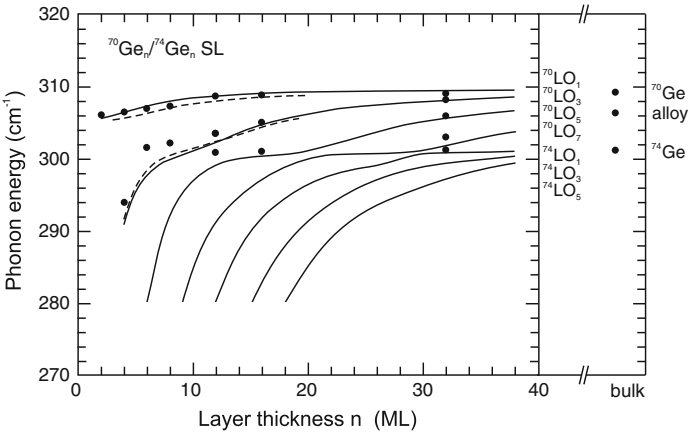


Fig. 12.43 Measured (*full circles*) and theoretical (*solid lines*) confined LO phonon energies in $^{70}Ge_n/^{74}Ge_n$ superlattices versus the layer thickness (number of monolayers) n. The *dashed lines* represent a calculation that considers intermixing at the interfaces. On the right, the energies of bulk modes for isotopically pure ^{70}Ge and ^{74}Ge are shown together with that of an $^{70}Ge_{0.5}{}^{70}Ge_{0.5}$ alloy. Adapted from [1076]

12.5 Isotope Superlattices

A special type of heterostructure is the modulation of the isotope content. The first kind of heterostructures made like this were $^{70}Ge_n/^{74}Ge_n$ symmetric superlattices [1076]. Figure 12.43 shows phonon energies determined from Raman spectroscopy for various layer numbers n. The modes are classified by $^{70}LO_m$ and $^{74}LO_m$ denoting the material in which the amplitude is maximal and m being the number of maxima in that medium.[8] Such modes are visualized in Fig. 12.44a for a $^{69}GaP_{16}/^{71}GaP_{16}$ superlattice. Theoretical mode energies as a function of the superlattice period are shown in Fig. 12.44b.

12.6 Wafer Bonding

Wafer bonding is a fairly recently developed method to join different and dissimilar materials. Two wafers of the respective materials are put together face to face and are adequately fused. The idea is to not only 'glue' the wafers together with a sticky (and compliant) organic material, but to form strong atomic bonds between the two materials with possibly a perfect interface. In some cases, the interface needs to allow charge-carrier transport through it. Less stringent conditions need to be met for photon transport.

[8] Only modes with odd m are Raman-active.

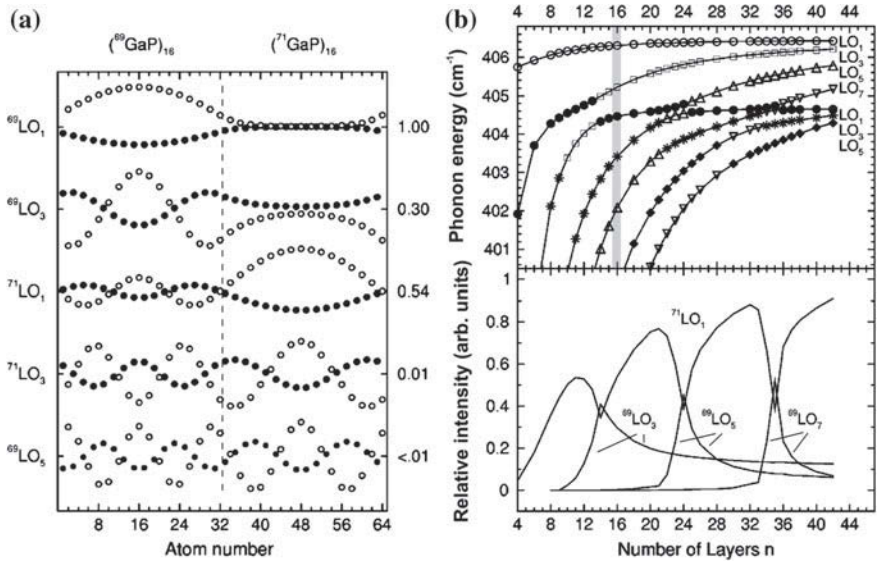


Fig. 12.44 (**a**) Atomic displacements [Ga (*filled dots*) and P (*open circles*)] of odd-index LO modes in a $^{69}Ga_{16}P/^{71}Ga_{16}P$ superlattice unit cell. These modes have even parity with respect to midlayer planes, which are at atom numbers 16 and 48 in this example. The labels on the *left* identify the predominant character of the mode, those on the *right* give the relative Raman intensities with respect to that of the $^{69}LO_1$ mode. The tick marks on the vertical axis indicate zero displacement of the respective mode. (**b**) *Upper panel*: Energies and characters of odd-index LO phonon modes in GaP isotope SLs as calculated within the planar bond charge model for the case of ideal interfaces. $^{69}LO_m$ modes are shown as *open symbols*; $^{71}LO_m$ modes as *full symbols*. The *shaded area* marks $n = 16$ for which the atomic displacements of the modes are shown in part (**a**). *Lower panel*: Calculated intensities of the modes relative to that of the $^{69}LO_1$ phonon mode. Adapted from [335], reprinted with permission, © 1999 APS

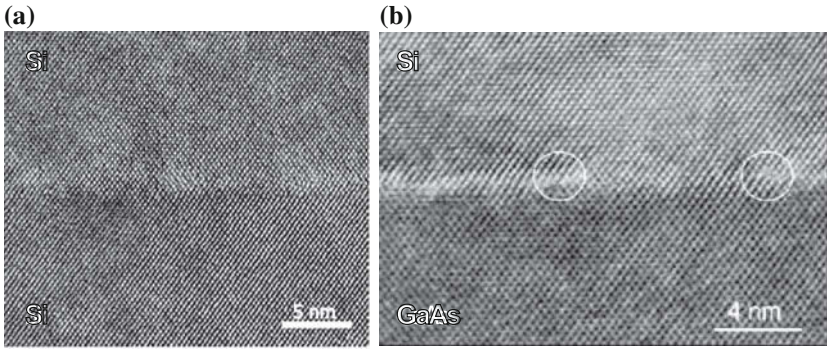


Fig. 12.45 High-resolution TEM images of wafer-bonded (**a**) Si-Si and (**b**) GaAs-Si interfaces. *White circles* indicate the position of misfit dislocations. Part (**a**) reprinted from [1077], © 2003, with permission from Elsevier. Part (**b**) reprinted with permission from [1078], © 1998 AIP

Mechanical deficiencies such as surface roughness, dust particles and the like must be avoided in the wafer-bonding process since they result in voids. Several methods have been developed for bonding various materials [1079, 1080]. Such processes are successful for large substrate sizes. With proper processing, ideal interfaces can be created, as shown in Fig. 12.45. Such structures, if made between a p-doped and a n-doped semiconductor, show diode characteristics.

Chapter 13
External Fields

Abstract The effects of external electric and magnetic fields on the electronic and optical properties of bulk materials and quantum wells are discussed including the Stark effect and quantum-confined Stark effect, the Hall effect and Quantized Hall Effects.

13.1 Electric Fields

13.1.1 Bulk Material

The center-of-mass motion of the exciton is not influenced by an electric field. The Hamilton operator for the relative motion of an electron–hole pair of reduced mass μ along z in the presence of an electric field E along the z direction is

$$\hat{H} = -\frac{\hbar^2}{2\mu}\,\Delta - e\,E\,z. \tag{13.1}$$

Here, the Coulomb interaction, leading to the formation of bound exciton states, is neglected. In the plane perpendicular to the field (here the z direction) the solutions for the relative motion are just plane waves.

In the electric field the bands are tilted (Fig. 13.1), i.e. there is no longer an overall band gap. Accordingly, the wavefunctions are modified and have exponential tails in the energy gap.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_13

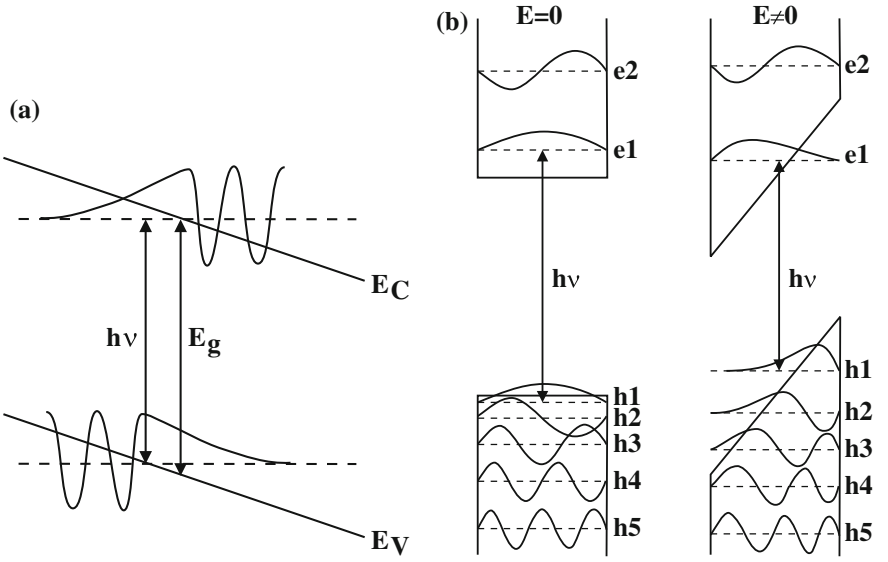


Fig. 13.1 Impact of an electric field on (**a**) bulk material (tilt of bands) and (**b**) a quantum well (quantum confined Stark effect, QCSE)

After separation of the motion in the (x,y) plane the Schrödinger equation for the motion in the z direction is

$$\left(-\frac{\hbar^2}{2\mu}\frac{\mathrm{d}^2}{\mathrm{d}z^2} - e\,E\,z - E_z\right)\phi(z) = 0\,, \tag{13.2}$$

which is of the type

$$\frac{\mathrm{d}^2 f(\xi)}{\mathrm{d}\xi^2} - \xi f(\xi) = 0\,, \tag{13.3}$$

with $\xi = \frac{E_z}{\Theta} - z\left(\frac{2\mu}{\hbar^2}\,e\,E\right)^{1/3}$ and the optoelectronic energy $\Theta = \left(\frac{e^2E^2\hbar^2}{2\mu}\right)^{1/3}$. The solution of (13.3) is given by the Airy function Ai (cf. Fig. 13.2):

$$\phi_{E_z}(\xi) = \frac{\sqrt{e\,E}}{\Theta}\,Ai(\xi). \tag{13.4}$$

The pre-factor guaranties the orthonormality (with regard to the E_z). The absorption spectrum is then given by

$$\alpha(\omega, E) \propto \frac{1}{\omega}\sqrt{\Theta}\,\pi\left[Ai'^2(\eta) - \eta\,Ai^2(\eta)\right]\,, \tag{13.5}$$

with $\eta = (E_\mathrm{g} - E)/\Theta$ and $Ai'(x) = \mathrm{d}Ai(x)/\mathrm{d}x$.

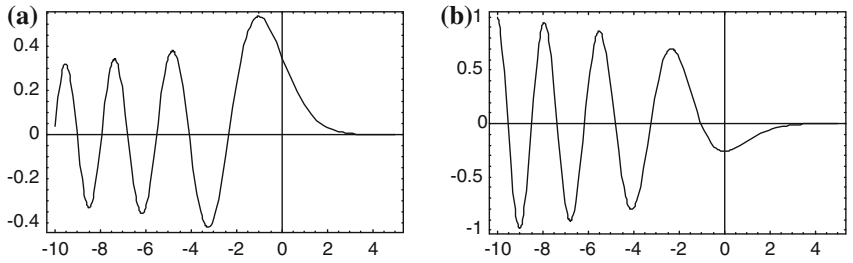


Fig. 13.2 (**a**) Airy function Ai(x), (**b**) Ai'(x)

Optical transitions below the band gap become possible that are photon-assisted tunneling processes. The below-bandgap transitions have the form of an exponential tail. Additionally, oscillations develop above the band gap, the so-called Franz–Keldysh oscillations (FKO) (Fig. 13.3a).

The absorption spectrum scales with the optoelectronic energy Θ. The energy position of the FKO peaks E_n is periodic with ($\nu \sim 0.5$)

$$\left(E_n - E_g\right)^{3/2} \propto (n - \nu)\, E\, \sqrt{\mu}. \tag{13.6}$$

A nonperiodicity can indicate a nonparabolicity of the mass. For a given mass the electric field strength can be determined. Well-pronounced oscillations are only present for homogeneous fields.

Experimental spectra show additionally the peaks due to excitonic correlation (Fig. 13.3b) at low field strength. At higher fields the FKO evolve and the amplitude of the excitonic peaks decreases because the excitons are ionized in the field.

13.1.2 Quantum Wells

In a quantum well an electric field along the confinement direction (z direction) causes electrons and holes to shift their mean position to opposite interfaces (Fig. 13.1b). However, excitons are not ionized due to the electric field. With increasing field (for both field directions) the energy position of the absorption edge and the recombination energy is reduced. This is the quantum confined Stark effect (QCSE). Corresponding experimental data are shown in Fig. 13.4i–v. The shift depends quadratically on the electric field since the exciton has no permanent dipole moment (mirror symmetry of the quantum well). Thus, only the second-order Stark effect is present (as for the hydrogen atom) in which the field first induces a dipole $\mathbf{p} = \alpha\mathbf{E}$. This dipole interacts with the field with an energy $E = -\mathbf{p}\cdot\mathbf{E} = -\alpha\mathbf{E}^2$. The carrier separation in opposite sides of the quantum well (Fig. 13.4b) leads to a reduced overlap of the electron and hole wavefunctions and subsequently to an increased recombination lifetime (see Fig. 12.42).

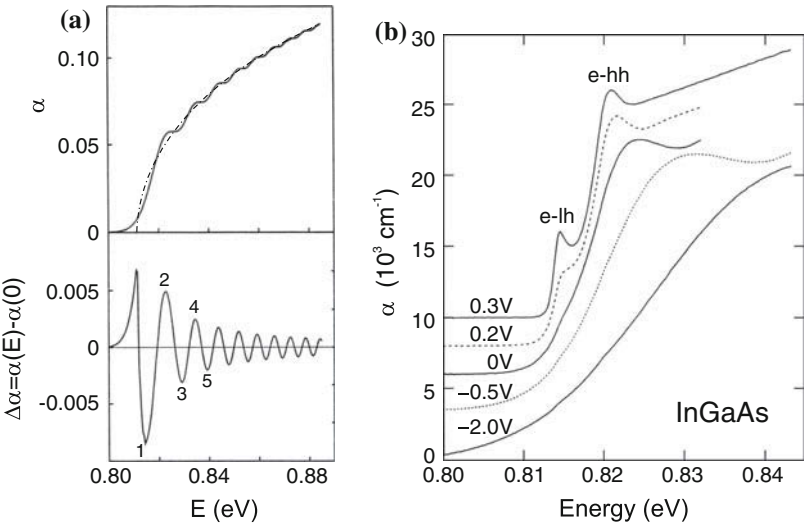


Fig. 13.3 (**a**) Theoretical absorption (*top panel*) with (*solid line*) and without (*dash-dotted line*) electric field for a volume semiconductor (without Coulomb interaction) and theoretical change of absorption (*bottom panel*). (**b**) Experimental absorption spectra of InGaAs on InP at $T = 15\,\mathrm{K}$ for various applied voltages as labeled. Adapted from [1081]

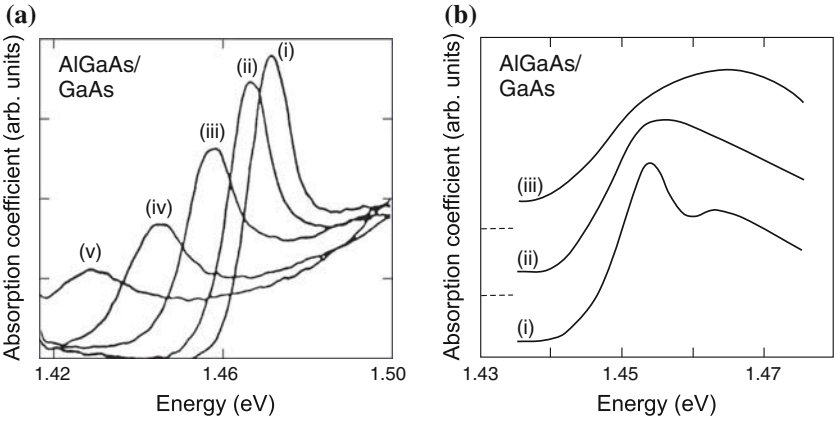


Fig. 13.4 Impact of electric fields on the absorption spectrum of $n \times$ (9.5 nm GaAs/9.8 nm $Al_{0.32}Ga_{0.68}As$) multiple quantum well structures. **a** Electric field along the [001] growth direction ($n = 50$), (**i**)–(**v**): $E = 0$, 0.6, 1.1, 1.5, and 2×10^5 V/cm. **b** Electric field within the interface plane (n = 60), (**i**–**iii**): $E = 0$, 1.1, and 2×10^5 V/cm. Adapted from [1082]

If the field is within the quantum-well interface plane, the field leads to the ionization of excitons without shift of the energy position. The loss of the excitonic peak is visualized in the spectra in Fig. 13.4a–c.

13.2 Magnetic Fields

In magnetic fields, electrons (or holes) perform a cyclotron motion with frequency $\omega_c = eB/m^*$, i.e. a motion perpendicular to the magnetic field on a line of constant energy in **k**-space. This line is the intersection of a plane perpendicular to the magnetic field and the respective isoenergy surface in **k**-space. For semiconductors with anisotropic mass, such as Si and Ge, the quantum theory of cyclotron resonance has been given in [1083]. The physics of semiconductors in magnetic fields is covered in detail in [1084].

The ballistic cyclotron motion can only occur between two scattering events. Thus, a significantly long path along the cyclotron trajectory (classically speaking) and the connected magnetotransport properties are only possible when

- $\omega_c \tau \gg 1$, i.e. when the average scattering time τ is sufficiently large. This requires high mobility.
- the magnetic field is sufficiently strong and the temperature sufficiently low, i.e. $\hbar\omega_c \gg kT$, such that thermal excitations do not scatter electrons between different Landau levels.
- the cyclotron path is free of geometric obstructions.

An external magnetic field also produces a splitting of the spin states. For the electron, the energy splitting ΔE is given by

$$\Delta E = g_e^* \mu_B B \,, \tag{13.7}$$

where B is the magnetic-field amplitude and g_e^* the (effective) electron g-factor. This value differs from the free-electron value in vacuum of $g_e = 2.0023$ due to the presence of spin-orbit interaction (see Sect. 13.2.3). Values for g_e^* at low carrier density and low temperatures are 2 for Si, 1.2 for InP and ZnSe, -1.65 for CdTe, -0.44 for GaAs, -15 in InAs, and -50 for InSb. In [1085] the temperature dependence of g_e^* in GaAs, InP and CdTe is also measured and discussed. The electron g-factor increases in thin GaAs/AlGaAs quantum wells [1086].

13.2.1 Hall Effect

An electrical current along the x (longitudinal) direction in a perpendicular magnetic field $\mathbf{B} = (0, 0, B)$ along z causes an electric field E_y along the transverse (y) direction (Fig. 13.5). The charge accumulation is due to the Lorentz force. The related

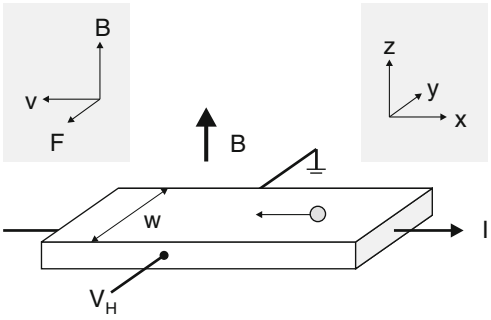


Fig. 13.5 Scheme of the Hall-effect geometry. The movement of one electron in the longitudinal electric current I is shown schematically. The coordinate system (x, y, z) and the directions of the magnetic field $\mathbf{B}$, the drift velocity of an electron $\mathbf{v}$ and the resulting Lorentz force $\mathbf{F}$ are given. The transverse field E_y is given by V_H/w

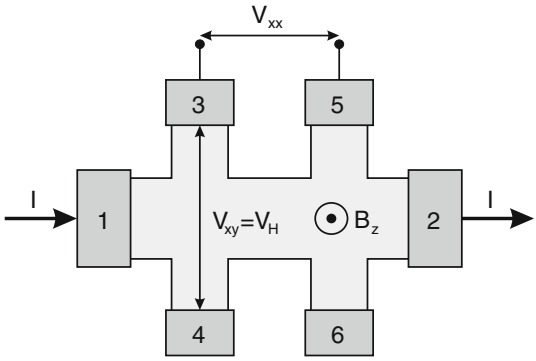


Fig. 13.6 Hallbar geometry. The current I is driven by a source from contact 1 to 2. The Hall voltage is measured between contacts 3 and 4, the longitudinal voltage between 3 and 5

transverse voltage is called the Hall voltage and the resistivity $\rho_{xy} = E_y/j_x$ the Hall resistivity [21, 24, 25]. Many aspects of the Hall effect are discussed in [1087]. For thin-film samples typically Hall bars [1088] (see Fig. 13.6 and also Fig. 13.17, for a reasonable measurement of the Hall voltage the ratio of length and width of the Hall bar should be at least 3) or the van-der-Pauw geometry (Fig. 13.7) and method are used [1089–1091].

For band transport in the relaxation time approximation (Sect. 8.2), the steady-state equation of motion is

$$m^* \frac{\mathbf{v}}{\tau} = q\ (\mathbf{E} + \mathbf{v} \times \mathbf{B})\,. \tag{13.8}$$

We note that this equation of motion is also valid for holes, given the convention of Sect. 6.9.1, i.e. positive effective mass and charge. With the cyclotron frequency $\omega_\mathrm{c} = qB/m^*$ the conductivity tensor is ($\mathbf{j} = qn\mathbf{v} = \boldsymbol{\sigma}\,\mathbf{E}$)

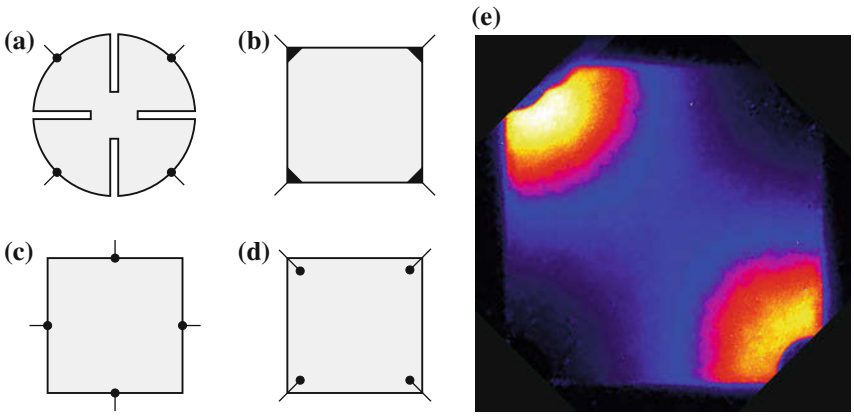


Fig. 13.7 (**a**–**d**) Geometry for van-der-Pauw Hall measurements. (**a**) Best geometry (*cloverleaf*), (**b**) acceptable square geometry with small contacts on the corners, (**c**,**d**) not recommended geometries with contacts on the edge centers or inside the square, respectively. (**e**) Current distribution, as visualized by lock-in thermography [1092], in epitaxial ZnO layer on sapphire with Hall geometry as in part (**b**). *Grey dashed line* indicates the outline of the $10 \times 10\,\mathrm{mm}^2$ substrate, *grey areas* indicate gold ohmic contacts

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & 0 \\ \sigma_{yx} & \sigma_{yy} & 0 \\ 0 & 0 & \sigma_{zz} \end{pmatrix} \tag{13.9a}$$

$$\sigma_{xx} = \sigma_{yy} = \sigma_0 \frac{1}{1 + \omega_c^2 \tau^2} = \sigma_0 \frac{1}{1 + \mu^2 B^2} \tag{13.9b}$$

$$\sigma_{xy} = -\sigma_{yx} = \sigma_0 \frac{\omega_c \tau}{1 + \omega_c^2 \tau^2} = \sigma_0 \frac{\mu B}{1 + \mu^2 B^2} \tag{13.9c}$$

$$\sigma_{zz} = \sigma_0 = \frac{q^2 n \tau}{m^*} = q n \mu. \tag{13.9d}$$

Perpendicular to the magnetic field, the conductivity (σ_{zz}) is given by (8.5). If only one type of carrier (charge q, density n) is considered, the condition $j_y = 0$ leads to $E_y = \mu B E_x$ and $j_x = \sigma_0 E_x$. The Hall coefficient is defined as $R_\mathrm{H} = E_y/(j_x B)$ or more precisely as

$$R_\mathrm{H} = \frac{\rho_{xy}}{B}\,, \tag{13.10}$$

where the resistivity tensor $\boldsymbol{\rho}$ is the inverse of the conductivity tensor $\boldsymbol{\sigma}$,

$$\boldsymbol{\rho} = \boldsymbol{\sigma}^{-1} = \begin{pmatrix} \rho_{xx} & \rho_{xy} & 0 \\ \rho_{yx} & \rho_{yy} & 0 \\ 0 & 0 & \rho_{zz} \end{pmatrix} \tag{13.11a}$$

$$\rho_{xx} = \rho_{yy} = \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2} \tag{13.11b}$$

$$\rho_{xy} = -\rho_{yx} = \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2} \tag{13.11c}$$

$$\rho_{zz} = \frac{1}{\sigma_{zz}} = \frac{1}{\sigma_0}. \tag{13.11d}$$

For a single type of carriers, the Hall coefficient is therefore given by

$$R_\mathrm{H} = \frac{\mu}{\sigma_0} = \frac{1}{q\,n}. \tag{13.12}$$

It is negative (positive) for electron (hole) conduction. We note that electrons and holes are deflected in the *same* y-direction by the magnetic field and collect at the same electrode. Thus the Hall effect allows the determination of the carrier type and the carrier density.[1]

If both types of carriers are present simultaneously, the conductivity (two-band conduction) is given by the sum of electron and hole conductivity (8.11),

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_\mathrm{e} + \boldsymbol{\sigma}_\mathrm{h}. \tag{13.13}$$

The Hall constant (13.10) is then

$$R_\mathrm{H} = \frac{1}{e}\,\frac{-n\,\mu_\mathrm{e}^2\,(1+\mu_\mathrm{h}^2\,B^2) + p\,\mu_\mathrm{h}^2\,(1+\mu_\mathrm{e}^2\,B^2)}{n^2\,\mu_\mathrm{e}^2\,(1+\mu_\mathrm{h}^2\,B^2) - 2\,n\,p\,\mu_\mathrm{e}\,\mu_\mathrm{h}\,(1+\mu_\mathrm{e}\,\mu_\mathrm{h}\,B^2) + p^2\,\mu_\mathrm{h}^2\,(1+\mu_\mathrm{e}^2\,B^2)}. \tag{13.14}$$

Under the assumption of small magnetic fields,[2] i.e. $\mu B \ll 1$, the Hall coefficient is

$$R_\mathrm{H} = \frac{1}{e}\left[\frac{-n\,\mu_\mathrm{e}^2 + p\,\mu_\mathrm{h}^2}{(-n\,\mu_\mathrm{e} + p\,\mu_\mathrm{h})^2} + \frac{n\,p\,(-n+p)\,\mu_\mathrm{e}^2\,\mu_\mathrm{h}^2\,(\mu_\mathrm{e}-\mu_\mathrm{h})^2}{(-n\,\mu_\mathrm{e} + p\,\mu_\mathrm{h})^4}\,B^2 + \cdots\right]. \tag{13.15}$$

For small magnetic field this can be written as

$$R_\mathrm{H} = \frac{1}{e}\,\frac{p - n\,\beta^2}{(p - n\,\beta)^2}\ , \tag{13.16}$$

[1] Using the Hall effect, the net *free* charge carrier concentration is determined. We note that the concentration of *fixed* charges in semiconductors can be investigated by depletion layer spectroscopy (Sect. 21.2.4).

[2] We note that for a mobility of $10^4\,\mathrm{cm^2/Vs}$, μ^{-1} is a field of $B = 1\,\mathrm{T}$.

with $\beta = \mu_e/\mu_h < 0$. For large magnetic fields, i.e. $\mu B \gg 1$, the Hall coefficient is given by

$$R_H = \frac{1}{e}\frac{1}{p-n}. \tag{13.17}$$

In Fig. 13.8, the absolute value of the Hall coefficient for InSb samples with different doping concentrations is shown. The p-doped samples exhibit a reverse of the sign of the Hall coefficient upon increase of temperature when intrinsic electrons contribute to the conductivity. The zero in R_H occurs for $n = p\,\mu_h^2/\mu_e^2 = n_i/|\beta|$. For high temperatures, the Hall coefficient for n- and p-doped samples is dominated by the electrons that have much higher mobility (Table 8.2).

The simultaneous conduction in a band and an impurity band (cmp. Sect. 7.5.7) has been separated with a suitable model assuming two conduction channels for holes [1094] (Fig. 13.9).

In the derivation of the (unipolar) Hall coefficient we had assumed that all carriers involved in the transport have the same properties, in particular that they are subject to the same scattering time. This assumption is generally not the case (cmp. Appendix I) and we need to operate with the ensemble average of the discussed

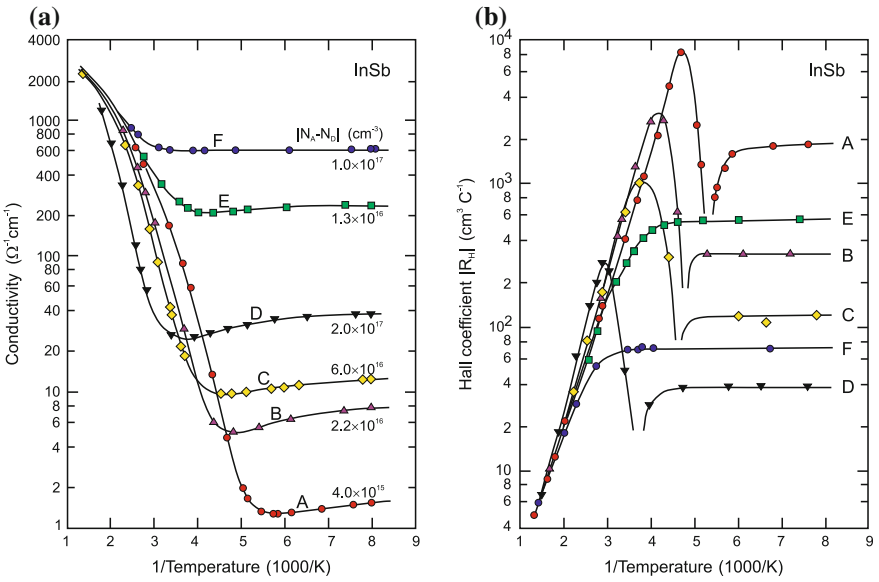


Fig. 13.8 (**a**) Conductivity and (**b**) absolute value of the Hall coefficient versus inverse temperature for four p-doped (A–D) and two n-doped (E, F) InSb samples. The doping levels are given in (**a**). Adapted from [1093]

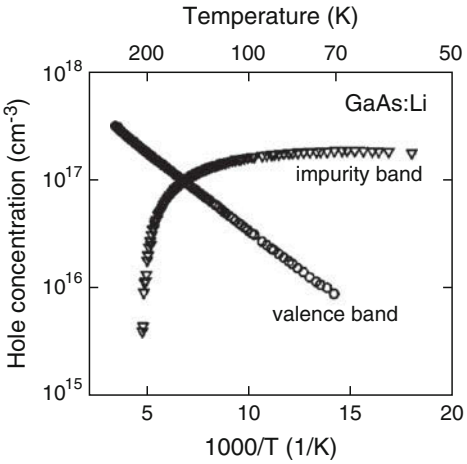


Fig. 13.9 Carrier densities in valence band (*circles*) and impurity band (*triangles*) from evaluating the Hall effect on GaAs doped with lithium (and annealed) taking into account two conduction channels. Adapted from [1094]

quantities. The ensemble average of an energy-dependent quantity $\zeta(E)$ over the (electron) distribution function $f(E)$ is denoted as $\langle\zeta\rangle$ and is given as[3]

$$\langle\zeta\rangle = \frac{\int \zeta(e) f(E)\,\mathrm{d}E}{\int f(E)\,\mathrm{d}E}. \tag{13.18}$$

In particular $\langle\tau\rangle^2$ is now different from $\langle\tau^2\rangle$. Considering the equation $\langle\mathbf{j}\rangle = \langle\boldsymbol{\sigma}\rangle\,\mathbf{E}$ for the ensemble averaged current densities we find (for one type of carrier, cf. (13.12))

$$R_\mathrm{H} = \frac{1}{q\,n}\,r_\mathrm{H}\,, \tag{13.19}$$

with the so-called *Hall factor* r_H given by

$$\begin{aligned} r_\mathrm{H} &= \frac{\gamma}{\alpha^2 + \omega_\mathrm{c}^2\,\gamma^2} \\ \alpha &= \langle\frac{\tau}{1+\omega_\mathrm{c}^2\,\tau^2}\rangle \\ \gamma &= \langle\frac{\tau^2}{1+\omega_\mathrm{c}^2\,\tau^2}\rangle. \end{aligned} \tag{13.20}$$

[3]For this consideration it is assumed that the energy dependence is the decisive one. Generally, averaging may have to be performed over other degrees of freedom as well, such as the spin or, in the case of anisotropic bands, the orbital direction.

The Hall factor depends on the scattering mechanisms and is of the order of 1. For large magnetic fields the Hall factor approaches 1. For small magnetic fields we have

$$R_{\mathrm{H}} = \frac{1}{q\,n}\,\frac{\langle\tau^2\rangle}{\langle\tau\rangle^2}. \tag{13.21}$$

The mobility calculated from (cf. (13.9d)) $\sigma_0 R_{\mathrm{H}}$ is called the *Hall mobility* μ_{H} and is related to the mobility via

$$\mu_{\mathrm{H}} = r_{\mathrm{H}}\,\mu. \tag{13.22}$$

It is assumed so far that the free carrier density and mobility are homogeneous within the volume of current transport. Multi-layer models can be fitted to experimental Hall data in order to account for different conduction channels in different layered parts of the sample [1095]. E.g., in a two-layer model, contributions from bulk and surface/interface conduction can be separated [1096–1098].

The magnetic field dependence of $\boldsymbol{\sigma}$ can be used in a general case to separate contributions of carriers with different density and mobility (including its sign) without assumptions and obtain the mobility spectrum $s(\mu)$ (MSA, mobility spectral analysis),

$$\sigma_{xx} = \int_{-\infty}^{\infty} s(\mu)\,\frac{1}{1+\mu^2B^2}\,\mathrm{d}\mu \tag{13.23a}$$

$$\sigma_{xy} = \int_{-\infty}^{\infty} s(\mu)\,\frac{\mu B}{1+\mu^2B^2}\,\mathrm{d}\mu, \tag{13.23b}$$

as a generalization of (13.13), (13.9b) and (13.9c) [1099-1101]. Examples are the separation of electron conductivity in (GaAs-) Γ- and (InAs-) X-Minima in a GaAs/AlGaAs/InAs double quantum well structures [1099] (Fig. 13.10a), substrate and 2DEG electron conductivity in AlGaN/GaN heterostructures [1102], and electrons and holes in an InAs/GaSb quantum well [1100] (Fig. 13.10b).

In the case of hopping conduction (Sect. 8.7), the theory of Hall effect is more involved [1103, 1104]. If carriers are transported by hopping, generally, they may not be free to move as expected by (13.8) in response to the applied magnetic field and the Lorentz force. A Hall effect occurs only at the junction of *three* (or more) hopping sites [1103]. The sign of the experimentally determined Hall coefficient is often opposite to the one expected from the carrier type according to (13.12), e.g. as studied in a-Si [1105, 1106]. The *sign anomaly* depends on intricacies such as the local site geometry and interference processes taking place among bonding and antibonding orbitals of various numbers as summarized in [1107].

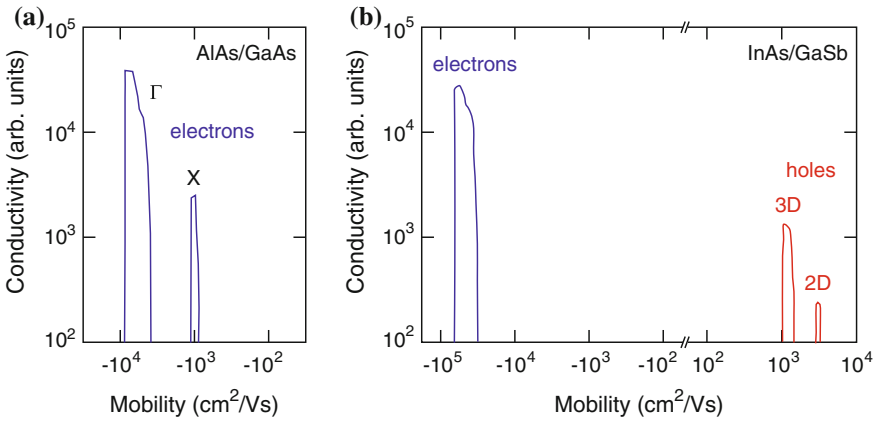


Fig. 13.10 (**a**) Mobility spectrum of GaAs/AlGaAs/InAs double quantum well structure. Adapted from [1099]. (**b**) Mobility spectrum of InAs/GaSb quantum well structure. Adapted from [1100]

13.2.2 Free-Carrier Absorption

The absorption of free carriers was treated in Sect. 9.8.1 without the presence of a static magnetic field. Solving (13.8) for a static magnetic field $\mathbf{B} = \mu_0 \mathbf{H}$ with $\mathbf{H} = H(h_x, h_y, h_z)$ and a harmonic electric field $\mathbf{E} \propto \exp(-\mathrm{i}\omega t)$ yields for the dielectric tensor (cf. (9.60))

$$\boldsymbol{\epsilon} = \frac{\mathrm{i}}{\epsilon_0 \, \omega} \boldsymbol{\sigma}, \tag{13.24}$$

and by comparing to $\mathbf{j} = \boldsymbol{\sigma} \, \mathbf{E} = q \, n \, \mathbf{v}$,

$$\boldsymbol{\epsilon}(\omega) = -\omega_\mathrm{p}^{*2} \left[(\omega^2 + \mathrm{i} \, \omega \, \gamma) \, \mathbf{1} - \mathrm{i} \, \omega_\mathrm{c} \begin{pmatrix} 0 & -h_z & h_y \\ h_z & 0 & -h_x \\ -h_y & h_x & 0 \end{pmatrix} \right]^{-1}, \tag{13.25}$$

where $\mathbf{1}$ denotes the (3 by 3) unity matrix and $\gamma = 1/\tau = q/(m^* \mu)$ is the damping parameter with μ representing the optical carrier mobility (in the non-isotropic case a tensor $\boldsymbol{\gamma}$ needs to be used). The (unscreened) plasma frequency is given by (cmp. (9.65))

$$\omega_\mathrm{p}^* = \sqrt{n \, \frac{e^2}{\epsilon_0 \, m^*}}. \tag{13.26}$$

The free-carrier cyclotron frequency is

$$\omega_\mathrm{c} = e\,\frac{\mu_0 H}{m^*}. \tag{13.27}$$

If the effective mass is treated as a tensor, $1/m^*$ is replaced by $\mathbf{m}^{*-1}$ in (13.26) and (13.27). For zero magnetic field the classical Drude theory for one carrier species is recovered (cf. (9.62a))

$$\boldsymbol{\epsilon}(\omega) = -\frac{\omega_\mathrm{p}^{*2}}{\omega\,(\omega + \mathrm{i}\gamma)}. \tag{13.28}$$

With the magnetic field perpendicular to the sample surface, i.e. $\mathbf{B} = \mu_0\,(0, 0, H)$ the magneto-optic dielectric tensor simplifies to (cf. (13.9d))

$$\boldsymbol{\epsilon}(\omega) = -\frac{\omega_\mathrm{p}^{*2}}{\omega^2}\begin{pmatrix} \tilde{\epsilon}_{xx} & \mathrm{i}\tilde{\epsilon}_{xy} & 0 \\ -\mathrm{i}\tilde{\epsilon}_{xy} & \tilde{\epsilon}_{xx} & 0 \\ 0 & 0 & \tilde{\epsilon}_{zz} \end{pmatrix} \tag{13.29a}$$

$$\tilde{\epsilon}_{xx} = \frac{1 + \mathrm{i}\gamma/\omega}{(1 + \mathrm{i}\gamma/\omega)^2 - (\omega_\mathrm{c}/\omega)^2} \tag{13.29b}$$

$$\tilde{\epsilon}_{zz} = \frac{1}{(1 + \mathrm{i}\gamma/\omega)} \tag{13.29c}$$

$$\tilde{\epsilon}_{xy} = \frac{\omega_\mathrm{c}/\omega}{(1 + \mathrm{i}\gamma/\omega)^2 - (\omega_\mathrm{c}/\omega)^2}. \tag{13.29d}$$

The in-plane component ϵ_{xx} provides information about ω_p^* and γ, i.e. two of the three parameters n, μ and m^* are known. Additionally, the antisymmetric tensor component ϵ_{xy} is linear in the cyclotron frequency and provides q/m^*. This subtle but finite birefringence depends on the strength (and orientation) of the magnetic field and can be experimentally determined in the infrared using magneto-ellipsometry [1108, 1109]. Such 'optical Hall effect' experiment allows the determination of the carrier density n, the mobility μ, the carrier mass[4] m^* and the sign of the carrier charge $\mathrm{sgn}(q)$ with optical means. The electrical Hall effect (Sect. 13.2.1) can reveal n, μ and $\mathrm{sgn}(q)$ but cannot reveal the carrier mass.

13.2.3 Energy Levels in Bulk Crystals

In a 3D electron gas (the magnetic field is along z, i.e. $\mathbf{B} = B[0, 0, 1]$) the motion in the (x, y) plane is described by Landau levels. Quantum mechanically they correspond to levels of a harmonic oscillator. The magnetic field has no impact on

[4] We note that mobility and effective mass defined and measured in this way may be referred to as 'optical'. Other definitions and approaches to the mobility or effective mass may give different results.

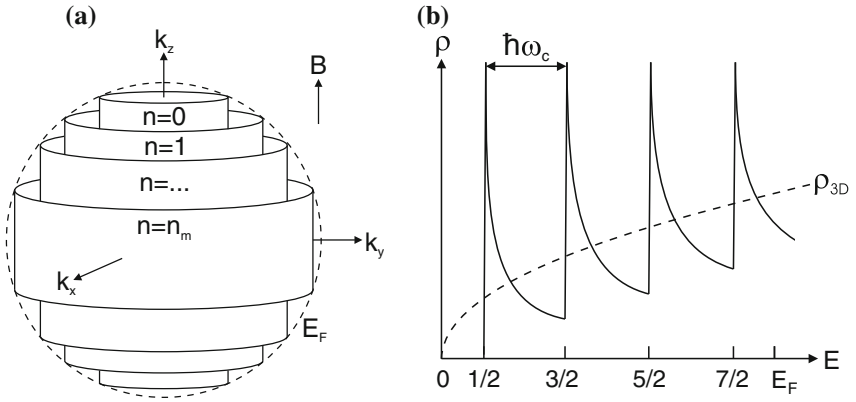


Fig. 13.11 3D electron gas in an external magnetic field. (**a**) Allowed states in **k**-space for magnetic field along the z direction. (**b**) Density of states (DOS) ρ versus energy (in units of $\hbar\omega_c$). *Dashed line* is three-dimensional DOS without magnetic field. Based on [991]

the motion of electrons along z, such that in this direction a free dispersion relation $\propto k_z^2$ is present. The energy levels are given as

$$E_{nk_z} = \left(n + \frac{1}{2}\right)\hbar\omega_c + \frac{\hbar^2}{2m}k_z^2. \tag{13.30}$$

Thus, the states are on concentric cylinders in **k**-space (Fig. 13.11a). The populated states of the 3D electron gas (at 0 K) lie within the Fermi vector of length k_F. For the 3D system the density of states at the Fermi energy is a square root function of the Fermi energy (6.67). In the presence of a magnetic field the density of states diverges every time that a new cylinder (with a one-dimensional density of states, (6.75)) touches the Fermi surface at E_F. In real systems, the divergence will be smoothed, however, a pronounced peak or the periodic nature of the density of states is often preserved.

The period is given by the number n_m of cyclotron orbits (Landau levels) within the Fermi surface.

$$\left(n_m + \frac{1}{2}\right)\hbar\omega_c = E_F. \tag{13.31}$$

If the number of carriers is constant, the density of states at the Fermi energy at varying magnetic field varies periodically with $1/B$. From the conditions $\left(n_m + \frac{1}{2}\right)\hbar\frac{eB_1}{m} = E_F$ and $\left(n_m + 1 + \frac{1}{2}\right)\hbar\frac{eB_2}{m} = E_F$ with $\frac{1}{B_2} = \frac{1}{B_1} + \frac{1}{\Delta B}$ we find

$$\frac{1}{\Delta B} = \frac{e\hbar}{m^* E_F}. \tag{13.32}$$

This periodicity is used to determine experimentally, e.g., the properties of the Fermi surface in metals using the Shubnikov–de Haas oscillations (of the magnetoresistance) or the de Haas–van Alphén effect (oscillation of the magnetic susceptibility).

Equation (13.30) needs to be extended to account for the splitting (13.7) of the Landau level due to the electron spin. According to [1110] the electron Landau level energy can be written as

$$E_n = \left(n + \frac{1}{2}\right) \frac{\hbar e B}{m^*(E)} \pm g_e^*(E)\, \mu_B\, B\ , \tag{13.33}$$

with energy dependent effective mass and g-factor

$$\frac{1}{m^*(E)} = \frac{1}{m^*(0)} \frac{E_g\,(E_g + \Delta_0)}{3E_g + 2\Delta_0} \left(\frac{2}{E + E_g} + \frac{1}{E + E_g + \Delta_0}\right) \tag{13.34a}$$

$$g_e^*(E) = g_e^*(0) \frac{E_g(E_g + \Delta_0)}{\Delta_0} \left(\frac{1}{E + E_g} - \frac{1}{E + E_g + \Delta_0}\right) \tag{13.34b}$$

The band edge value $m^*(0)$ of the effective mass is given by (6.39) and that of the g-factor by

$$g_e^*(0) = 2\left[1 - \frac{2\,\Delta_0}{3\,E_g\,(E_g + \Delta_0)} E_P\right]. \tag{13.35}$$

For large spin-orbit splitting, the value of the g-factor deviates strongly from 2 and becomes negative.

13.2.4 Energy Levels in a 2DEG

In a 2D electron gas (2DEG), e.g. in a quantum well or a potential well at a modulation-doped heterointerface, a free motion in z is not possible and k_z is quantized. The energy levels (for each 2D subband) are only given by the cyclotron energy (Fig. 13.12a). The density of states is a sequence of δ-like peaks (Fig. 13.12b). Each peak contributes (degeneracy $\hat{g}$ of a Landau level) a total number of

$$\hat{g} = \frac{e\,B}{h} \tag{13.36}$$

states (per unit area without spin degeneracy and without the degeneracy of the band extremum). In reality, disorder effects lead to an inhomogeneous broadening of these peaks. The states in the tails of the peaks correspond to states that are localized in real space.

Also, in a 2D system several physical properties exhibit an oscillatory behavior as a function of Fermi level, i.e. with varying electron number, and as a function of the magnetic field at fixed Fermi energy, i.e. at fixed electron number (Fig. 13.13).

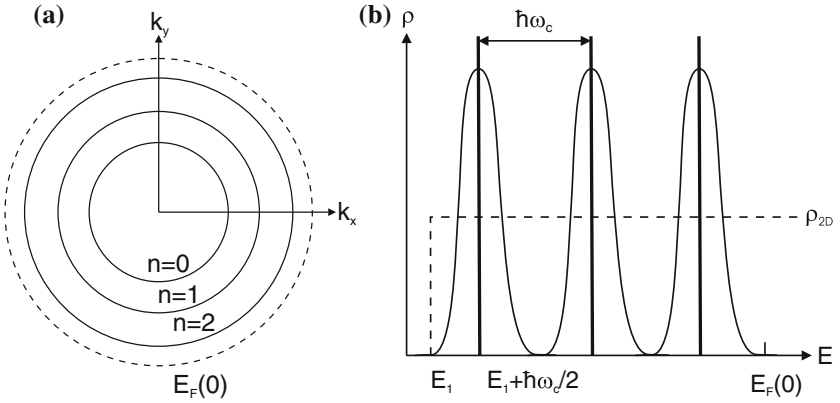


Fig. 13.12 2D electron gas in an external magnetic field. (**a**) Allowed states in **k**-space. (**b**) Density of states (DOS) ρ versus energy. *Dashed line* is two-dimensional DOS without magnetic field. *Thick vertical lines*: δ-like DOS without broadening, *curves*: broadened DOS. Based on [991]

13.2.5 Shubnikov–de Haas Oscillations

From the 2D density of states (per unit area including spin degeneracy) $D^{2D}(E) = m^*/\pi\hbar^2$ the sheet density of electrons n_s can be expressed as a function of the Fermi level (at $T = 0\,\mathrm{K}$ without spin degeneracy)

$$n_s = \frac{m^*}{2\pi\,\hbar^2}\,E_F. \tag{13.37}$$

Using (13.32) we thus find (without spin degeneracy, without valley degeneracy), that the period of $1/B$ is $\propto n_s$:

$$\frac{1}{\Delta B} = \frac{e}{h}\,\frac{1}{n_s}. \tag{13.38}$$

The carrier density of a 2DEG can therefore be determined from the oscillations of magnetoresistance, and is proportional to the density of states at the Fermi level (Shubnikov–de Haas effect). A corresponding measurement with varying field and fixed electron density is shown in Fig. 13.14. The periodicity with $1/B$ is obvious. Since only the component of the magnetic field perpendicular to the layer affects the (x, y) motion of the carriers, no effect is observed for the magnetic field parallel to the layer.

In another experiment the carrier density was varied at constant field (Fig. 13.15). The electron density in an inversion layer in p-type silicon is (linearly) varied with the gate voltage of a MOS (metal–oxide–semiconductor) structure (inset in Fig. 13.15, for MOS diodes cf. Sect. 24.5). In this experiment, the Fermi level was shifted through the Landau levels. The equidistant peaks show that indeed each Landau level contributes the same number of states.

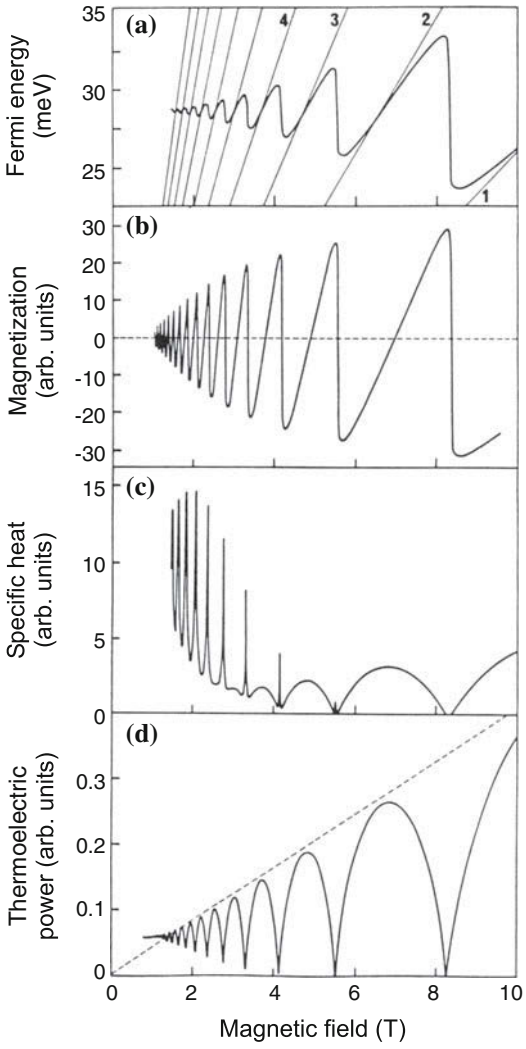


Fig. 13.13 Oscillatory (theory, $T = 6\,\mathrm{K}$) behavior of a 2DEG (GaAs/AlGaAs) in a magnetic field: (**a**) Fermi level, (**b**) magnetization, (**c**) specific heat, (**d**) thermoelectric power. A Gaussian broadening of 0.5 meV was assumed. Adapted from [991, 1111]

13.3 Quantum Hall Effect

In high magnetic fields, at low temperatures and for high-mobility, 2D electron gases exhibit a deviation from the classical behavior. We recall that the classical Hall effect (i.e. considering the Lorentz force, classical Drude theory), the generation of a field E_y perpendicular to a current flow j_x (cf. Sect. 13.2.1), was described with the conductivity tensor $\boldsymbol{\sigma}$ (here, for the (x, y)-plane only)

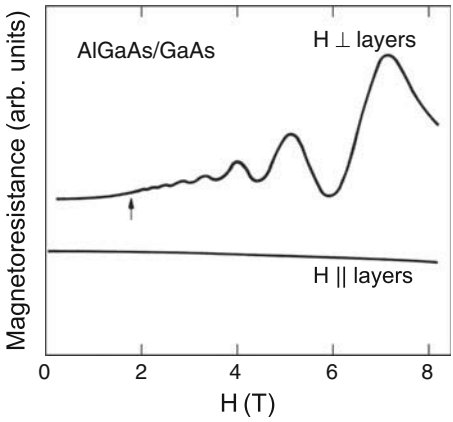


Fig. 13.14 Shubnikov–de Haas oscillations at a modulation-doped AlGaAs/GaAs heterostructure with a 2DEG, $n = 1.7 \times 10^{17}\,\mathrm{cm}^{-2}$ and $\mu = 11\,400\,\mathrm{cm}^2/\mathrm{Vs}$. Data from [1112]

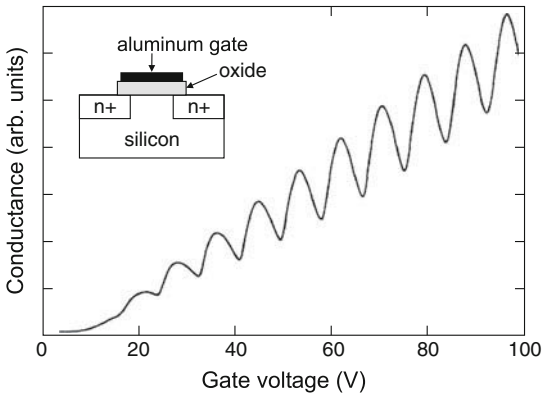


Fig. 13.15 Shubnikov–de Haas oscillations of a 2DEG at the (100) surface of p-type silicon (100 Ω cm) at a magnetic field of 33 kOe and $T = 1.34\,\mathrm{K}$. The *inset* shows schematically the contact geometry. Data from [1113]

$$\boldsymbol{\sigma} = \frac{\sigma_0}{1+\omega_\mathrm{c}^2\tau^2}\begin{pmatrix} 1 & \omega_\mathrm{c}\tau \\ -\omega_\mathrm{c}\tau & 1 \end{pmatrix} \tag{13.39a}$$

$$\sigma_{xx} = \sigma_0 \frac{1}{1+\omega_\mathrm{c}^2\tau^2} \to 0 \tag{13.39b}$$

$$\sigma_{xy} = \sigma_0 \frac{\omega_\mathrm{c}\tau}{1+\omega_\mathrm{c}^2\tau^2} \to \frac{n\,e}{B}\,, \tag{13.39c}$$

where σ_0 is the zero-field conductivity $\sigma_0 = ne^2\tau/m^*$ (8.5). The arrows denote the limit for $\omega_\mathrm{c}\tau \to \infty$, i.e. large fields. The resistivity tensor $\boldsymbol{\rho} = \boldsymbol{\sigma}^{-1}$ is given by

$$\boldsymbol{\rho} = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ -\rho_{xy} & \rho_{xx} \end{pmatrix} \tag{13.40a}$$

$$\rho_{xx} = \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2} \to 0 \tag{13.40b}$$

$$\rho_{xy} = \frac{-\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2} \rightarrow -\frac{B}{n\,e}. \tag{13.40c}$$

13.3.1 Integral QHE

Experiments yield strong deviations from the linear behavior of the transverse resistivity $\rho_{xy} = E_y/j_x = BR_H$ with the Hall coefficient $R_H = -1/(ne)$ with increasing magnetic field is observed at low temperatures for samples with high carrier mobility, i.e. $\omega_c\tau \gg 1$ (Fig. 13.16). In Fig. 13.17a,b, Hall bars are shown for 2DEGs in silicon metal–oxide–semiconductor field-effect transistor (Si-MOSFET) electron inversion layers and at GaAs/AlGaAs heterostructures, respectively.

The Hall resistivity exhibits extended Hall plateaus with resistivity values that are given by

$$\rho_{xy} = \frac{1}{i}\,\frac{h}{e^2}\,, \tag{13.41}$$

i.e. integer fractions of the quantized resistance $\rho_0 = h/e^2$ =25812.807… Ω, which is also called the von-Klitzing constant. In Fig. 13.16, a spin splitting is seen for the $n = 1$ Landau level (and a small one for the $n = 2$). We note that the topmost Hall

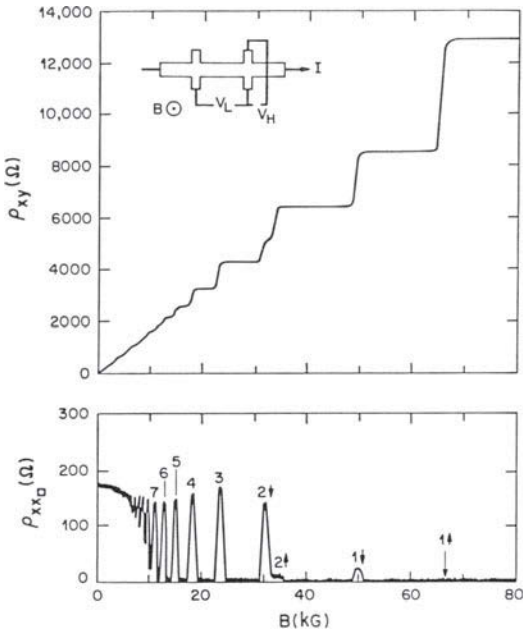


Fig. 13.16 Hall resistivity ρ_{xy} and longitudinal resistivity ρ_{xx} for a modulation-doped GaAs/AlGaAs heterostructure ($n = 4 \times 10^{11}\,\mathrm{cm}^{-2}$, $\mu = 8.6 \times 10^4\,\mathrm{cm}^2/\mathrm{Vs}$) at 50 mK as a function of magnetic field (10 kG=1 T). The numbers refer to the quantum number and spin polarization of the Landau level involved. The inset shows schematically the Hall bar geometry, V_L (V_H) denotes the longitudinal (Hall) voltage drop. Reprinted with permission from [1114], ©1982 APS

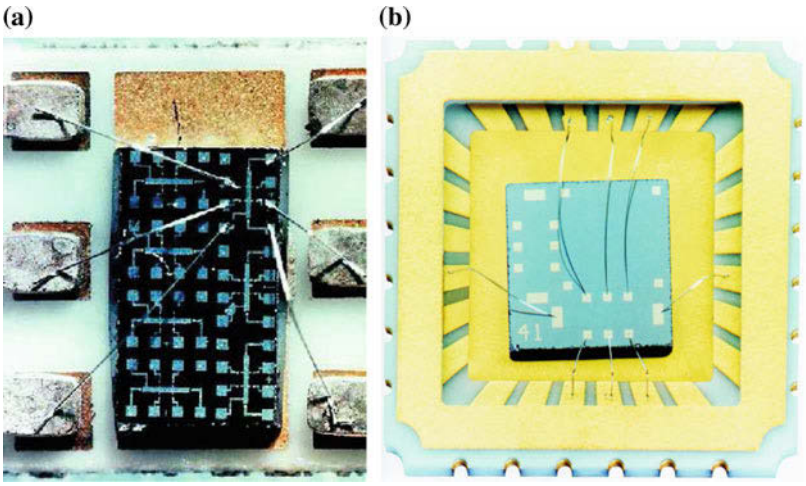


Fig. 13.17 (**a**) Silicon MOS (metal–oxide–semiconductor) structure of K. v. Klitzing's et al. original experiments. (**b**) GaAs/AlGaAs heterostructure sample grown with molecular beam epitaxy for QHE measurements, chip carrier and bond wires. Reprinted with permission from [1115]

plateau is due to the completely filled $n = 0$ Landau level; the resistance is $\rho_0/2$ due to the spin degeneracy of 2.

The integral quantum Hall effect, first reported in [1116, 1117], and the value for ρ_0 are found for a wide variety of samples and conditions regarding sample temperature, electron density or mobility of the 2DEG and the materials of the heterostructure.

Within the plateau the resistivity is well defined within 10^{-7} or better up to 4×10^{-9}. A precise determination allows for a new normal for the unit Ohm [1118, 1119], being two orders of magnitude more precise than the realization in the SI system, and an independent value for the fine-structure constant $\alpha = \frac{e^2}{\hbar c}\frac{1}{4\pi\epsilon_0}$. At the same time, the longitudinal resistivity, starting from the classical value for small magnetic fields, exhibits oscillations and eventually it is zero for the plateaus in ρ_{xy}. For ρ_{xx} values of $10^{-10}\,\Omega/\square$ have been measured, which corresponds to $10^{-16}\,\Omega$/cm for bulk material, a value three orders of magnitude smaller than for any nonsuperconductor.

The interpretation of the quantum Hall effect(s) is discussed in [1120] among many other treatises. The simplest explanation is that the conductivity is zero when a Landau level is completely filled and the next is completely empty, i.e. the Fermi level lies between them. The temperature is small, i.e. $kT \ll \hbar\omega_c$, such that no scattering between Landau levels can occur. Thus no current, similar to a completely filled valence band, can flow. The sheet carrier density n_s is given by counting the i filled Landau levels (degeneracy according to (13.36)) as

$$n_s = i\,\frac{eB}{h}. \tag{13.42}$$

In the transverse direction energy dissipation takes place and the Hall resistivity $\rho_{xy} = B/(n_s e)$ takes the (scattering-free) values given in (13.41).

However, this argument is too simple as it will not explain the extension of the plateaus. As soon as the system has one electron more or less, the Fermi energy will (for a system with δ-like density of states) be located in the upper or lower Landau level, respectively. Then, the longitudinal conductivity should no longer be zero and the Hall resistivity deviates from the integer fraction of ρ_0.

Therefore (in the localization model) an inhomogeneous (Gaussian) broadening of the density of states is assumed. Additionally, the states in the tails of the distribution are considered nonconducting, i.e. localized, while those around the peak are considered conductive. This mobility edge is schematically shown in Fig. 13.18a–c.

When the Fermi level crosses the density of states of the broadened Landau level (upon increase of the magnetic field), it first populates localized states and ρ_{xx} remains at zero. When the Fermi level crosses the delocalized, conducting states, the longitudinal conductivity shows a peak and the Hall conductivity increases from the plateau value by e^2/h to the next plateau. When the Fermi level lies within the upper nonconducting states, the Hall resistivity remains constant in the next plateau. For the sample in Fig. 13.16 it has been estimated that 95 % of the Landau level states are localized [1114].

In the localization model it is still astonishing that the samples exhibit the step of e^2/h in the conductivity as if *all* electrons on the Landau level contribute to conduction, cf. (13.42). According to the calculations in [1123] the Hall current lost by the localized states is compensated by an extra current by the extended states. The behavior of the electrons in the quantum Hall regime ('quantum Hall liquid') can be considered similar to that of an incompressible fluid where obstructions lead the fluid to move with increased velocity. Nevertheless, the single particle picture seems to be insufficient to model the IQHE.

Another model for the explanation of the QHE, supported now with plenty of experimental evidence, is the edge state model where quantized one-dimensional conductivity of edge channels, i.e. the presence of conductive channels along the sample boundaries, is evoked [1124]. Due to depletion at the boundary of the sample, the density of the 2DEG varies at the edge of the sample and 'incompressible' stripes develop for which $\partial\mu/\partial n_s \to \infty$. When the filling factor is far from an integer, the Hall voltage is found to vary linearly across the conductive channel and the current is thus homogeneous over the sample (Fig. 13.18d). In the Hall plateau, the Hall voltage is flat in the center of the channel and exhibits drops at the edges, indicating that the current flows along the boundary of the sample (edge current) [1122] in agreement with predictions from [1125]. Although the current pattern changes with varying magnetic field, the Hall resistivity remains at its quantized value.

The most fundamental arguments for the explanation of the IQHE come from gauge invariance and the presence of a macroscopic quantum state of electrons and magnetic flux quanta [1126]. This model holds as long as there are any extended states at all.

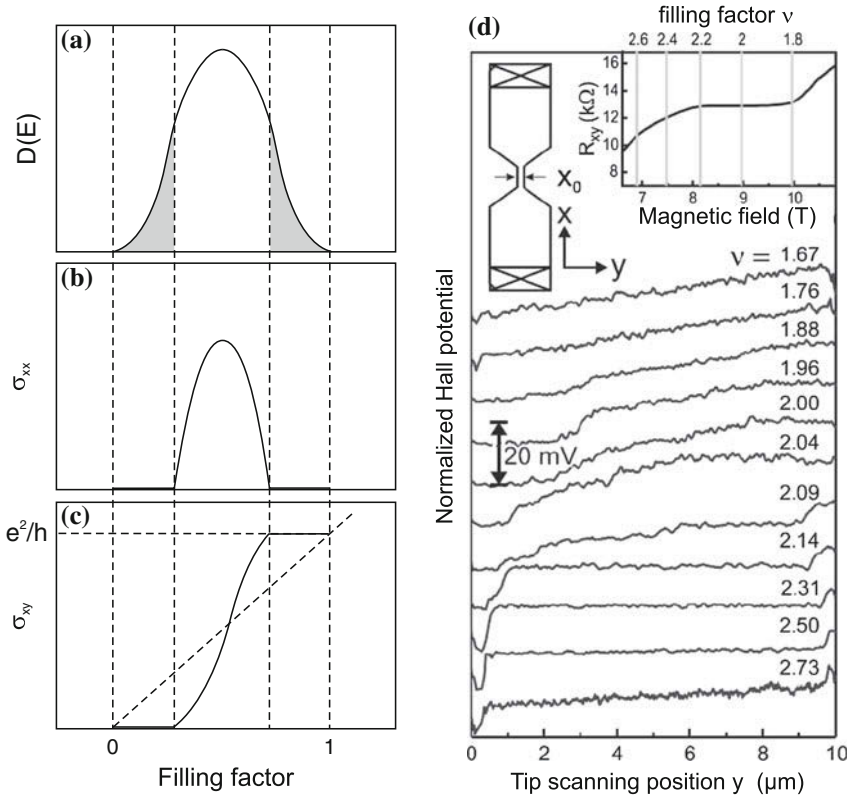


Fig. 13.18 (**a**) Density of states $D(E)$, (**b**) longitudinal conductivity σ_{xx} and (**c**) Hall conductivity σ_{xy} for a Landau level as a function of the filling factor $\mu = n/d$, where n is the electron density and d the degeneracy of the level. In (**a**) the filling factor relates via the position of the Fermi level to an energy scale. The *grey areas* in (**a**) denote localized states. The *slanted dashed line* in (**c**) has a slope of ne/B. Adapted from [1121]. (**d**) Normalized Hall potential profile for different magnetic fields around filling factor $\nu = 2$. The overall voltage drop corresponds to 20 mV. The *insets* show the sample geometry and transport data. The 2DEG is from a GaAs/$Al_{0.33}Ga_{0.67}$As modulation-doped heterostructure, $n_s = 4.3 \times 10^{11}\,\mathrm{cm}^{-2}$, $\mu = 5 \times 10^5\,\mathrm{cm}^2/\mathrm{Vs}$, $T = 1.4\,\mathrm{K}$. Adapted from [1122]

13.3.2 Fractional QHE

For very low temperatures and in the extreme quantum limit, novel effects are observed when the kinetic energy of the electrons is smaller than their Coulomb interaction. New quantum Hall plateaus are observed at various fractional filling factors $\nu = p/q$. We note that the effects of the fractional quantum Hall effect (FQHE) in Fig. 13.19 mostly arise for magnetic fields beyond the $n = 1$ IQHE plateau. The filling factor $\nu = n/(eB/h)$ is now interpreted as the number of electrons per magnetic flux quantum $\phi_0 = h/e$.

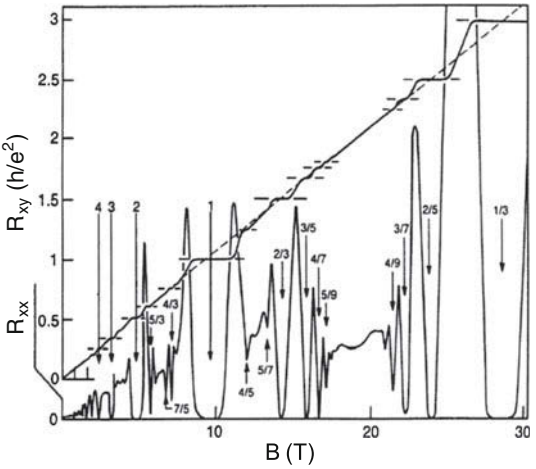


Fig. 13.19 Hall resistance R_{xy} and magnetoresistance R_{xx} of a two-dimensional electron system (GaAs/AlGaAs heterostructure) of density $n = 2.33 \times 10^{11}\,\mathrm{cm}^{-2}$ at a temperature of 85 mK, versus magnetic field B. Numbers identify the filling factor ν, which indicates the degree to which the sequence of Landau levels is filled with electrons. Plateaus are due to the integral ($\nu = i$) quantum Hall effect (IQHE) and fractional ($\nu = p/q$) quantum Hall effect (FQHE). Adapted from [1127], reprinted with permission, ©1990 AAAS

The effects of the FQHE cannot be explained by single-electron physics. The plateaus at fractional fillings ν occur when the Fermi energy lies within a highly degenerate Landau (or spin) level and imply the presence of energy gaps due to many-particle interaction and the result of correlated 2D electron motion in the magnetic field.

A decisive role is played by the magnetic flux quanta. The presence of the magnetic field requires the many-electron wavefunction to assume as many zeros per unit area as there are flux quanta penetrating it. The decay of the wavefunction has a length scale of the magnetic length $l_0 = \sqrt{\hbar/(eB)}$. Since the magnetic field implies a 2π phase shift around the zero, such an object is also termed a *vortex*, being the embodiment of the magnetic flux quanta in the electron system. Such a vortex represents a charge deficit (compared to a homogeneous charge distribution) and thus electrons and vortices attract each other. If a vortex and an electron are placed onto each other, considerable Coulomb energy is gained. At $\nu = 1/3$, there are three times more vortices than there are electrons, each vortex representing a charge deficit of $1/3\,e$. Such a system is described with many-particle wavefunctions, such as the Laughlin theory for $\nu = 1/q$ [1126] and novel quasi-particles called *composite fermions*[1128, 1129] for other fractional fillings. For further reading we refer readers to [1130]and references therein.

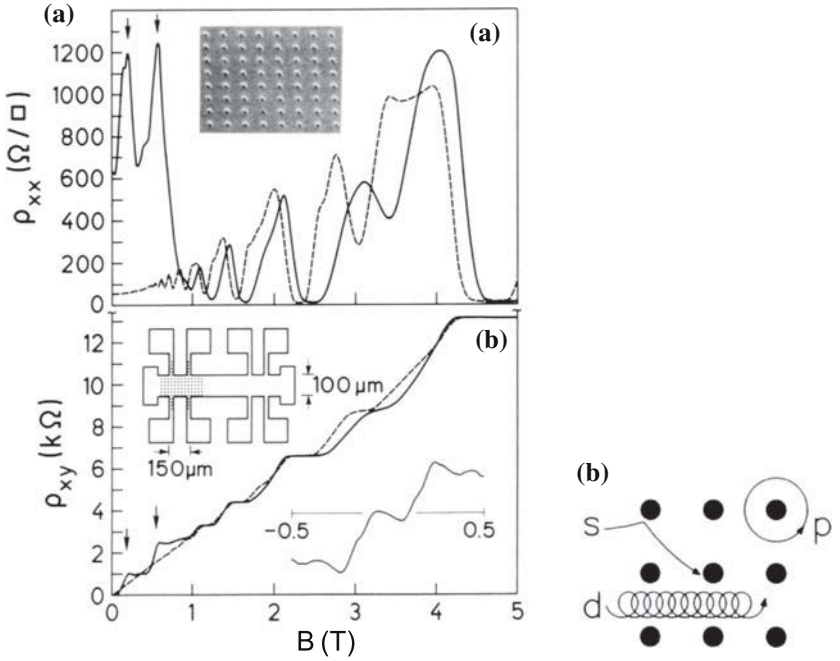


Fig. 13.20 Weiss oscillations: (**a**) magnetoresistance and (**b**) Hall resistance of an antidot lattice (*inset* in (**a**)) with pattern (*solid lines*) and without pattern (*dashed lines*) at $T = 1.5$ K. (**b**) Schematic of the different orbits: ('p': pinned, 'd': drifting, 's': scattered). Reprinted with permission from [1131], ©1991 APS

13.3.3 Weiss Oscillations

In Fig. 13.20, measurements are shown for a Hall bar in which an array of antidots (in which no conduction is possible) has been introduced by dry etching. The antidot size is 50 nm (plus depletion layer) and the period is 300 nm. These obstructions for the cyclotron motion lead to a modification of the magnetotransport properties.

Before etching of the antidot array the 2DEG has a mean free path length of 5–10 μm at 4 K for the mobility of $\approx 10^6\,\mathrm{cm}^2/\mathrm{Vs}$. At low magnetic fields there is a strong deviation of the Hall resistivity from the straight line to which the QHE levels converge. Similarly, ρ_{xx} shows a strong effect as well.

These effects are related to commensurability effects between the antidot lattice and the cyclotron resonance path. When the cyclotron orbit is equal to the lattice period, electrons can fulfill a circular motion around one antidot (pinned orbit, Fig. 13.20b) that leads to a reduction of conductivity. At high fields, drifting orbits for which the cyclotron orbit is much smaller than the lattice period occur. At small fields, scattering orbits also contribute for which the cyclotron radius is large and the electron has antidots from time to time. Resonances in the Hall resistivity have been found due to pinned orbits enclosing 1, 2, 4, 9 or 21 antidots.

Chapter 14
Nanostructures

The principles of physics, as far as I can see,
do not speak against the possibility
of maneuvering things atom by atom.
R.P. Feynman, 1959 [1132]

Abstract One-dimensional nanostructures (quantum wires) and zero-dimensional ones (quantum dots) are discussed with regard to their various fabrication methods and the tunable physical properties in such systems. Main effects covered are the modified density of states, confined energy levels, (envelope) wavefunction symmetry and the resulting novel electrical and optical properties.

14.1 Introduction

When the structural size of functional elements enters the size range of the de Broglie matter wavelength, the electronic and optical properties are dominated by quantum-mechanical effects. The most drastic impact can be seen from the density of states (Fig. 14.1).

The quantization in a potential is ruled by the Schrödinger equation with appropriate boundary conditions. These are simplest if an infinite potential is assumed. For finite potentials, the wavefunction leaks out into the barrier. Besides making the calculation more complicated (and more realistic), this allows electronic coupling of nanostructures. Via the Coulomb interaction, a coupling is even given if there is no wavefunction overlap. In the following, we will discuss some of the fabrication techniques and properties of quantum wires (QWR) and quantum dots (QD). In particular for the latter, several textbooks can also be consulted [1133, 1134].

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_14

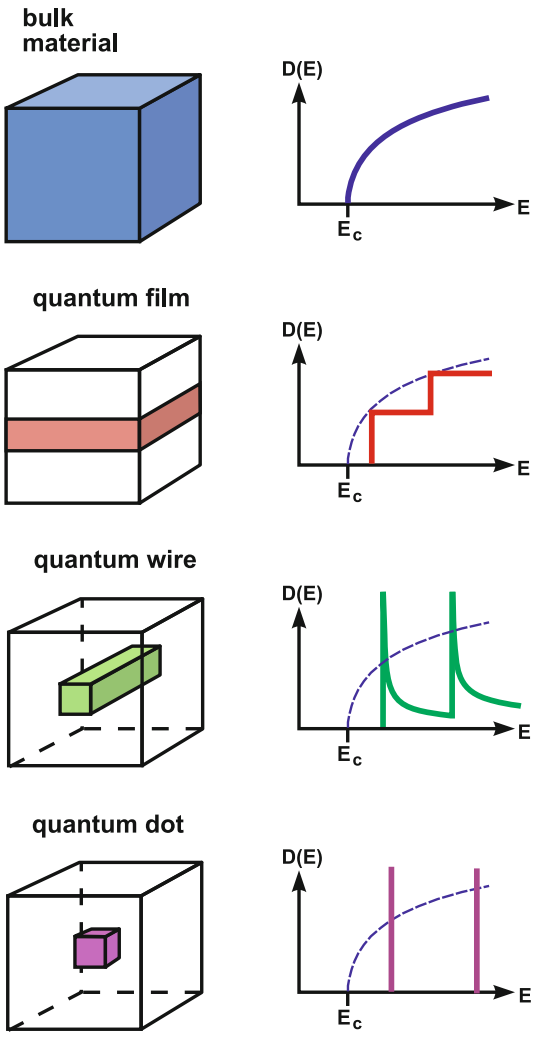


Fig. 14.1 Schematic geometry and density of states for 3D, 2D, 1D and 0D electronic systems

14.2 Quantum Wires

14.2.1 V-Groove Quantum Wires

Quantum wires with high optical quality, i.e. high recombination efficiency and well-defined spectra, can be obtained by employing epitaxial growth on corrugated substrates. The technique is shown schematically in Fig. 14.2. A V-groove is etched, using, e.g., an anisotropic wet chemical etch, into a GaAs substrate. The groove direction is along $[1\bar{1}0]$. Even when the etched pattern is not very sharp on the

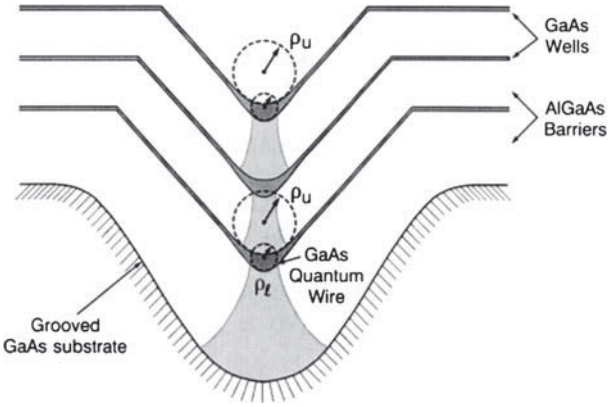


Fig. 14.2 Schematic cross section of a GaAs/AlGaAs heterostructure grown on a channeled substrate, illustrating the concept of self-ordered quantum-wire fabrication. Adapted from [1135]

bottom, subsequent growth of AlGaAs sharpens the apex to a self-limited radius ρ_l of the order of 10 nm. The side facets of the groove are {111}A. Subsequent deposition of GaAs leads to a larger upper radius $\rho_u > \rho_l$ of the heterostructure. The GaAs QWR formed in the bottom of the groove is thus crescent-shaped as shown in Fig. 14.3. A thin GaAs layer also forms on the side facets (sidewall quantum well) and on the top of the ridges. Subsequent growth of AlGaAs leads to a resharpening of the V-groove to the initial, self-limited value ρ_l. The complete resharpening after a sufficiently thick AlGaAs layer allows vertical stacking of crescent-shaped QWRs of virtually identical size and shape, as shown in Fig. 14.4. In this sense, the self-limiting reduction of the radius of curvature and its recovery during barrier-layer growth leads to self-ordering of QWR arrays whose structural parameters are determined solely by growth parameters. The lateral pitch of such wires can be down to 240 nm.

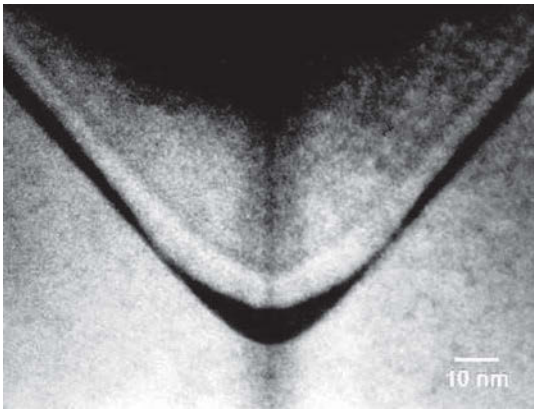


Fig. 14.3 Transmission electron microscopy cross-sectional image of a crescent-shaped single GaAs/AlGaAs quantum wire. From [1136]

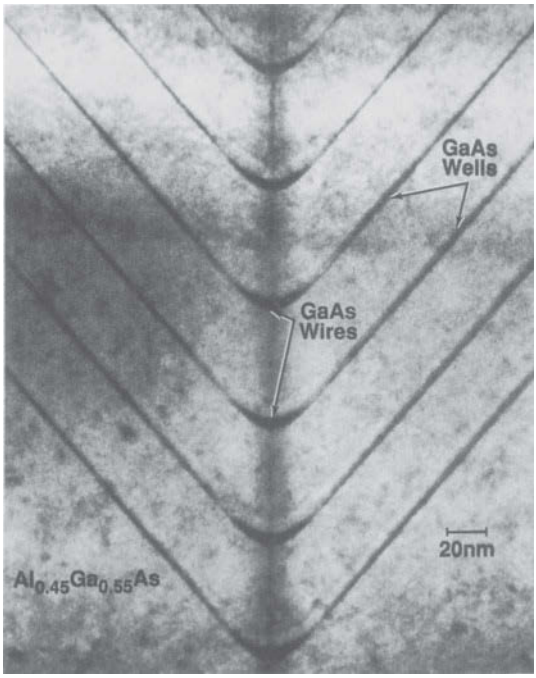


Fig. 14.4 TEM cross-sectional image of a vertical stack of identical GaAs/AlGaAs crescent-shaped QWRs. From [1135]

To directly visualize the lateral modulation of the band gap, a lateral cathodoluminescence (CL) linescan perpendicular across the wire is displayed in Fig. 14.5. In Fig. 14.5a, the secondary electron (SE) image of the sample from Fig. 14.3 is shown in plan view. The top ridge is visible in the upper and lower parts of the figure, while in the middle the sidewalls with the QWR in the center are apparent. In Fig. 14.5b, the CL spectrum along a linescan perpendicular to the wire (as indicated by the white line in Fig. 14.5a) is displayed. The x-axis is now the emission wavelength, while the y-axis is the lateral position along the linescan. The CL intensity is given on a logarithmic scale to display the full dynamic range. The top QW shows almost no variation in band gap energy ($\lambda = 725\,\text{nm}$); only directly at the edge close to the sidewall does a second peak at lower energy ($\lambda = 745\,\text{nm}$) appear, indicating a thicker region there. The sidewall QW exhibits a recombination wavelength of 700 nm at the edge to the top QW, which gradually increases to about 730 nm at the center of the V-groove. This directly visualizes a linear tapering of the sidewall QW from about 2.1 nm thickness at the edge to 3 nm in the center. The QWR luminescence itself appears at about 800 nm.

After fast capture from the barrier into the QWs and, to a much smaller extent corresponding to its smaller volume, into the QWR, excess carrierswill diffuse into the QWR via the adjacent sidewall QW and the vertical QW. The tapering of the sidewall QW induces an additional drift current.

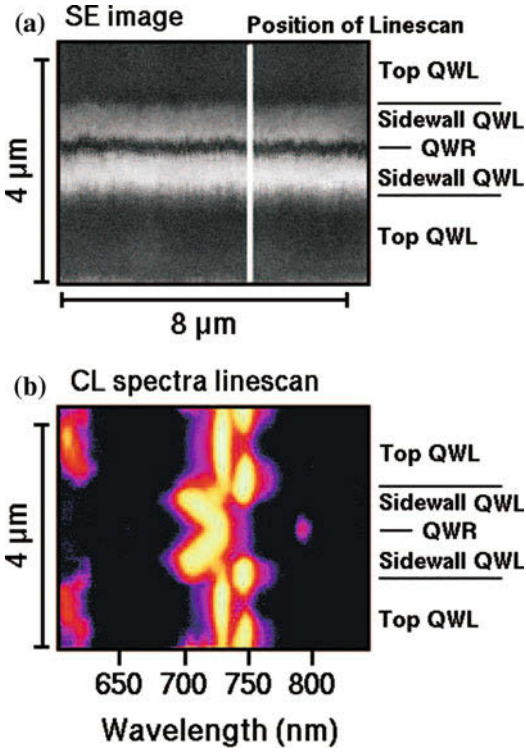


Fig. 14.5 (**a**) Plan-view SE image of single QWR (sample A), showing top and sidewall with QWR in the center. The *white line* indicates the position of the linescan on which the CL spectra linescan (**b**) has been taken at $T = 5\,\mathrm{K}$. The CL intensity is given on a logarithmic *grey scale* to display the full dynamic range as a function of wavelength and position along the *white line* in (**a**). From [1136]

14.2.2 Cleaved-Edge Overgrowth Quantum Wires

Another method to create quantum wires of high structural perfection is cleaved-edge overgrowth (CEO) [1137], shown schematically in Fig. 14.6. First, a layered structure is grown (single or multiple quantum wells or superlattice). Then, a {110} facet is fabricated by cleaving (in vacuum) and epitaxy is continued on the cleaved facet. At the junctures of the {110} layer and the original quantum wells QWRs form. Due to their cross-sectional form they are also called *T-shaped* QWRs. A second cleave and another growth step allow fabrication of CEO quantum dots [1138, 1139] (Fig. 14.6c).

14.2.3 Nanowhiskers

Whiskers are primarily known as thin metal spikes and have been investigated in detail [1140]. Semiconductor whiskers can be considered as (fairly short) quantum wires. They have been reported for a number of materials, such as Si, GaAs, InP and ZnO [1141]. A field of ZnO whiskers is shown in Fig. 14.7. If heterostructures are

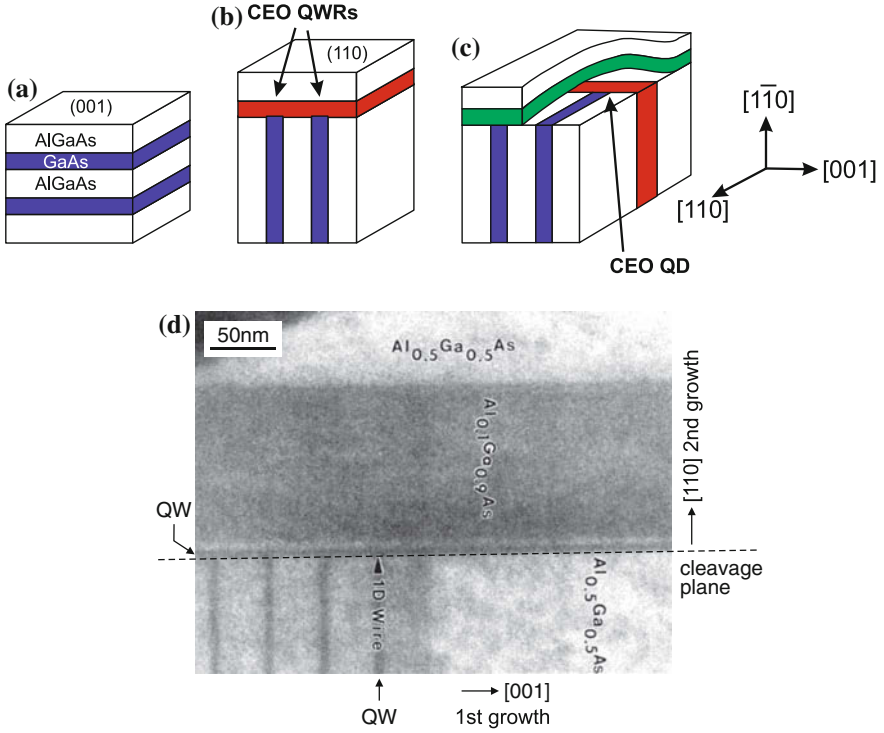


Fig. 14.6 Principle of CEO quantum wires and 2-fold CEO quantum dots. Part (**a**) depicts a layered structure (quantum wells or superlattice, *blue*), (**b**) describes the growth on the cleaved facet used for fabrication of quantum wires. In (**c**) a second cleave and growth on top of the plane allows the fabrication of quantum dots. From [1138]. (**d**) Cross-sectional TEM image of CEO GaAs/AlGaAs quantum wires. Two quantum wells (QW) and the QWR at their junction are labeled. The first epitaxy was from *left* to *right*. The second epitaxy step was on top of the cleavage plane (*dashed line*) in the upward direction. Adapted from [1139], reprinted with permission, ©1997 APS

incorporated along the whisker axis [1142], quantum dots or tunneling barriers can be created (Fig. 14.8). The nanocrystal can also act as a nanolaser [1143, 1144]. In ZnO nanowhiskers the conversion of mechanical energy into electrical energy has been demonstrated [1145] based on the piezoelectric effect (Sect. 15.4).

The critical thickness h_c in nanowire heterostructure is strongly modified from the 2D situation (Sect. 5.4.1). Based on the strain distribution of a misfitted slab in a cylindrical wire [1148] the dependence of critical thickness on the nanowhisker radius r was developed [1149, 1150]. For given misfit ϵ there is a critical radius r_c for which h_c is infinite for $r < r_c$ (Fig. 14.9).

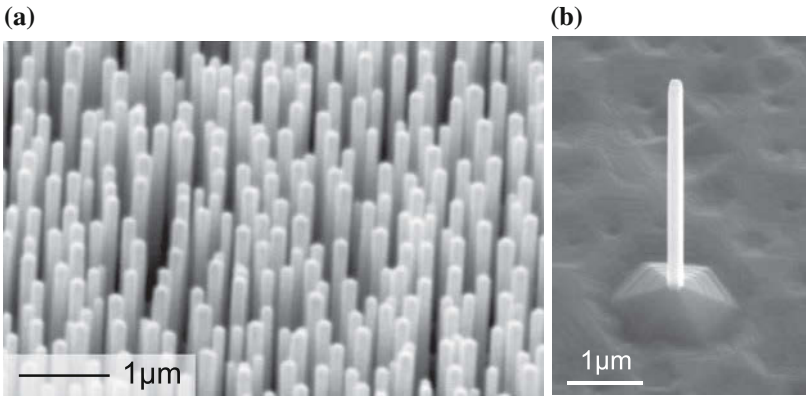


Fig. 14.7 (**a**) Array of ZnO nanowhiskers on sapphire, fabricated using thermal evaporation. Adapted from [1146]. (**b**) Single, free-standing ZnO nanowire fabricated using PLD. Adapted from [1147]

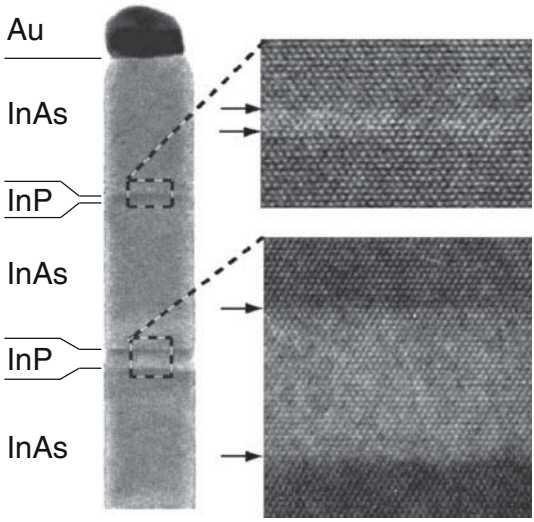


Fig. 14.8 TEM image of a part of an InAs whisker 40 nm in diameter that contains InP barriers. The zooms show sharp interfaces. On top of the whisker is a gold droplet from the so-called vapor–liquid–solid growth mechanism. The whisker axis is [001], the viewing direction is [110]. Adapted from [1142], reprinted with permission, ©2002 AIP

14.2.4 Nanobelts

A number of belt-shaped nanostructures has been reported [1141]. These are wire-like, i.e. very long in one dimension. The cross-section is rectangular with a high aspect ratio. In Fig. 14.10a ZnO nanobelts are shown. The wire direction is $[2\bar{1}.0]$. The large surface is (00.1), the thickness of the belt extends in [01.0]-direction. High resolution transmission microscopy (Fig. 14.10b) shows that these structures are defect-free. The pyroelectric charges on the ZnO (0001) surfaces (Sect. 15.2) lead to the formation of open (Fig. 14.11c) spirals. Closed spirals (Fig. 14.11a) occur if

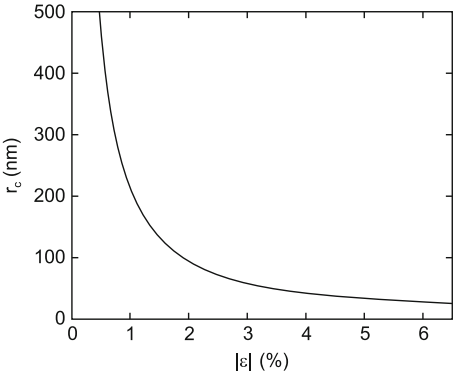


Fig. 14.9 Critical radius r_c above which an infinitely thick layer with misfit ϵ grows coherently on a cylindrical nanowire (relaxation by 60° dislocations, $b = 0.4\,\text{nm}$, $\nu = 1/3$). Adapted from [1150]

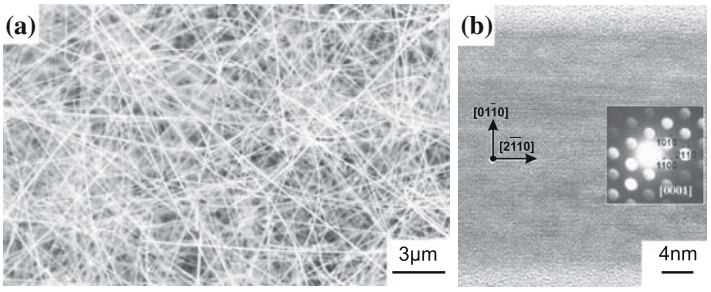


Fig. 14.10 (**a**) SEM image of an ensemble of ZnO nanobelts. (**b**) HRTEM image of a single ZnO nanobelt, viewing direction is [00.1]. The *inset* shows the diffraction pattern. Adapted from [1151]

the short dimension is along [00.1] and alternating charges become compensated in a 'slinky'-like ring (Fig. 14.11b).

14.2.5 Quantization in Two-Dimensional Potential Wells

The motion of carriers along the quantum wire is free. In the cross-sectional plane the wavefunction is confined in two dimensions. The simplest case is for constant cross section along the wire. However, generally the cross section along the wire may change and therefore induce a potential variation along the wire. Such potential variation will impact the carrier motion along the longitudinal direction. Also, a twist of the wire along its axis is possible.

In Fig. 14.12, the electron wavefunctions in a V-groove GaAs/AlGaAs QWR are shown. Further properties of V-groove QWRs have been reviewed in [1153].

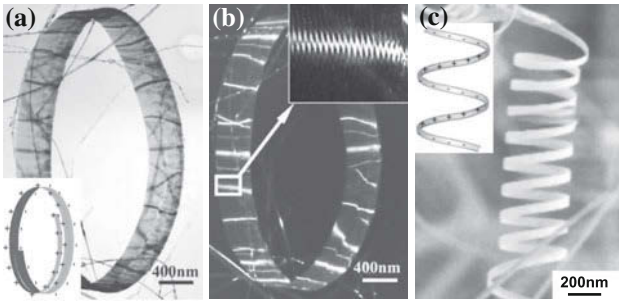


Fig. 14.11 (**a**) Bright field and (**b**) dark field TEM image of a ZnO nanoring formed by the 'slinky'-like growth of a nanobelt. (**c**) SEM image of an open ZnO nanospiral. The *insets* in (**a**, **c**) show schematically the surface charge distribution. Adapted from [1152]

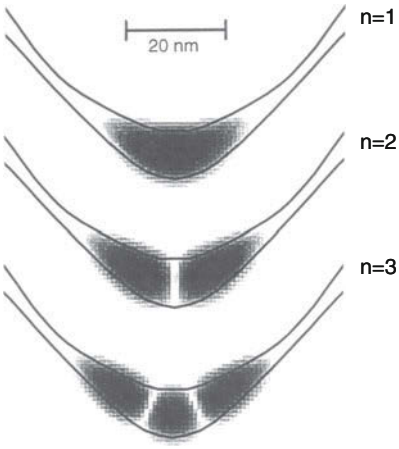


Fig. 14.12 Electron wavefunctions ($|\Psi|^2$ on logarithmic grey scale) for the first three confined levels for the QWR of Fig. 14.3. From [1136]

In Fig. 14.13, the excitonic electron and hole wavefunctions are shown for a (strained) T-shaped QWR.

In Fig. 14.14a the atomic structure of a very thin ZnO nanowhisker with a cross-section consisting of seven hexagonal unit cells is shown. The theoretical one-dimensional band structure [1155] is shown in Fig. 14.14b together with the charge density of the lowest conduction band state (LUMO) and the highest valence band state (HOMO). The band gap is generally too small because of the LDA method used.[1] In [1155] also the properties of nanowires with various diameters are compared. The HOMO at Γ lies only 80 meV above the top of valence band of bulk ZnO, and its position changes little with the wire diameter. It is mainly composed by surface oxygen 2p like dangling bonds (Fig. 14.14d). The LUMO (Fig. 14.14c) is delocalized in the whole nanowire, indicating that it is a bulk state. The delocalized

[1]The LDA in [1155] yields $E_g = 0.63\,\text{eV}$ for the bulk ZnO band gap; its experimental value is 3.4 eV.

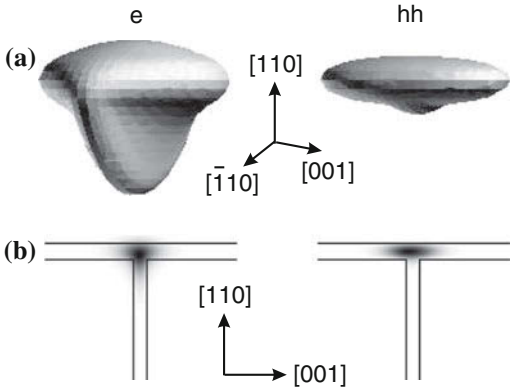


Fig. 14.13 (**a**) Three-dimensional view of the electron and (heavy) hole part of the excitonic wavefunction in a 4 × 5 nm T-shaped $In_{0.2}Ga_{0.8}As/GaAs$ QWR. The orbitals correspond to 70 % probability inside. (**b**) Cross section through the electron and hole orbitals in their center along the wire direction. Reprinted with permission from [1154], ©1998 APS

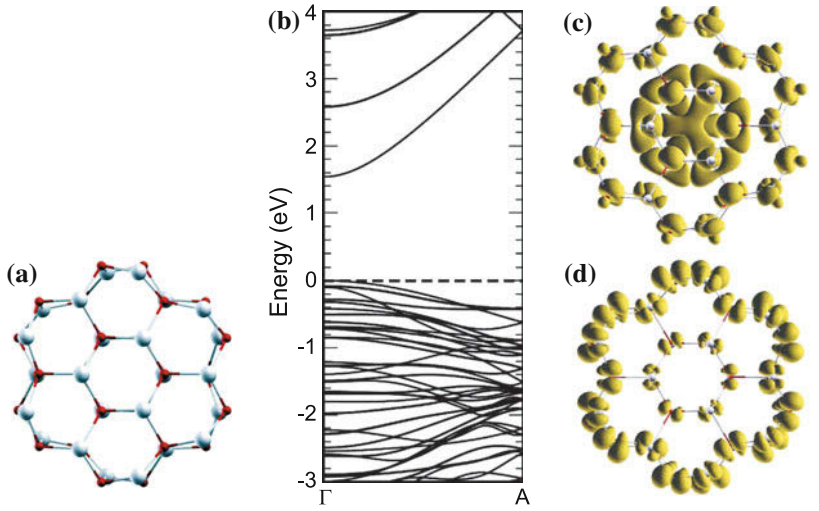


Fig. 14.14 (**a**) Atomic arrangement of a 1 nm wide ZnO nanowire. (**b**) Theoretical band structure and charge density of the (**c**) lowest conduction band and (**d**) highest valence band state. Adapted from [1155]

distribution is also responsible for the large dispersion of the LUMO from Γ to A. The energy of the LUMO increases substantially with decreasing diameter due to the radial confinement.

14.3 Quantum Dots

14.3.1 Quantization in Three-Dimensional Potential Wells

The solutions for the d-dimensional ($d = 1$, 2, or 3) harmonic oscillator, i.e. the eigenenergies for the Hamiltonian

$$\hat{H} = \frac{\mathbf{p}^2}{2m} + \sum_{i=1}^{d} \frac{1}{2}\, m\, \omega_0^2\, x_i^2 \tag{14.1}$$

are given by

$$E_n = \left(n + \frac{d}{2}\right) \hbar\omega_0, \tag{14.2}$$

with $n = 0, 1, 2, \ldots$. More detailed treatments can be found in quantum-mechanics textbooks.

Next, we discuss the problem of a particle in a centrosymmetric finite potential well with different masses m_1 in the dot and m_2 in the barrier. The Hamiltonian and the potential are given by

$$\hat{H} = \nabla \frac{\hbar^2}{2m} \nabla + V(r) \tag{14.3}$$

$$V(r) = \begin{cases} -V_0 & , r \le R_0 \\ 0 & , r > R_0 \end{cases}. \tag{14.4}$$

The wavefunction can be separated into radial and angular components $\Psi(\mathbf{r}) = R_{nlm}(r)\; Y_{lm}(\theta, \phi)$, where Y_{lm} are the spherical harmonic functions. For the ground state ($n = 1$) the angular momentum l is zero and the solution for the wavefunction (being regular at $r = 0$) is given by

$$R(r) = \begin{cases} \frac{\sin(k\, r)}{k\, r} & , r \le R_0 \\ \frac{\sin(k\, R_0)}{k\, R_0}\, \exp\left(-\kappa\, (r - R_0)\right) & , r > R_0 \end{cases} \tag{14.5a}$$

$$k^2 = \frac{2\, m_1\, (V_0 + E)}{\hbar^2} \tag{14.5b}$$

$$\kappa^2 = -\frac{2\, m_2\, E}{\hbar^2}. \tag{14.5c}$$

From the boundary conditions that both $R(r)$ and $\frac{1}{m}\frac{\partial R(r)}{\partial r}$ are continuous across the interface at $r = R_0$, the transcendental equation

$$k\, R_0\, \cot\left(k\, R_0\right) = 1 - \frac{m_1}{m_2}\, (1 + \kappa\, R_0) \tag{14.6}$$

is obtained. From this formula the energy of the single particle ground state in a spherical quantum dot can be determined. For a given radius, the potential needs a certain strength $V_{0,\min}$ to confine at least one bound state; this condition can be written as

$$V_{0,\min} = \frac{\pi^2\,\hbar^2}{8\,m^*\,R_0^2} \tag{14.7}$$

for $m_1 = m_2 = m^*$. For a general angular momentum l, the wavefunctions are given by spherical Bessel functions j_l in the dot and spherical Hankel functions h_l in the barrier. Also, the transcendental equation for the energy of the first excited level can be given:

$$k\,R_0\,\cot(k\,R_0) = 1 + \frac{k^2\,R_0^2}{\frac{m_1}{m_2}\,\frac{2+2\kappa\,R_0+\kappa^2\,R_0^2}{1+\kappa\,R_0} - 2}. \tag{14.8}$$

In the case of infinite barriers ($V_0 \to \infty$), the wavefunction vanishes outside the dot and is given by (normalized)

$$R_{nml}(r) = \sqrt{\frac{2}{R_0^3}}\,\frac{j_l(k_{nl}\,r)}{j_{l+1}(k_{nl}\,R_0)}, \tag{14.9}$$

where k_{nl} is the nth zero of the Bessel function j_l, e.g. $k_{n0} = n\pi$. With two-digit precision the lowest levels are determined by

k_{nl}	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
$n=0$	3.14	4.49	5.76	6.99	8.18	9.36
$n=1$	6.28	7.73	9.10	10.42		
$n=2$	9.42					

The $(2l+1)$ degenerate energy levels E_{nl} are ($V_0 = \infty$, $m = m_1$):

$$E_{nl} = \frac{\hbar^2}{2m}\,\frac{k_{nl}^2}{R_0^2}. \tag{14.10}$$

The 1s, 1p, and 1d states have smaller eigenenergies than the 2 s state.

A particularly simple solution is given for a cubic quantum dot of side length a_0 and infinite potential barriers. One finds the levels $E_{n_x n_y n_z}$:

$$E_{n_x n_y n_z} = \frac{\hbar^2}{2m}\,\pi^2\,\frac{n_x^2 + n_y^2 + n_z^2}{a_0^2}, \tag{14.11}$$

with n_x, n_y, $n_z = 1, 2, \ldots$. For a sphere, the separation between the ground and first excited state is $E_1 - E_0 \approx E_0$, for a cube and a two-dimensional harmonic

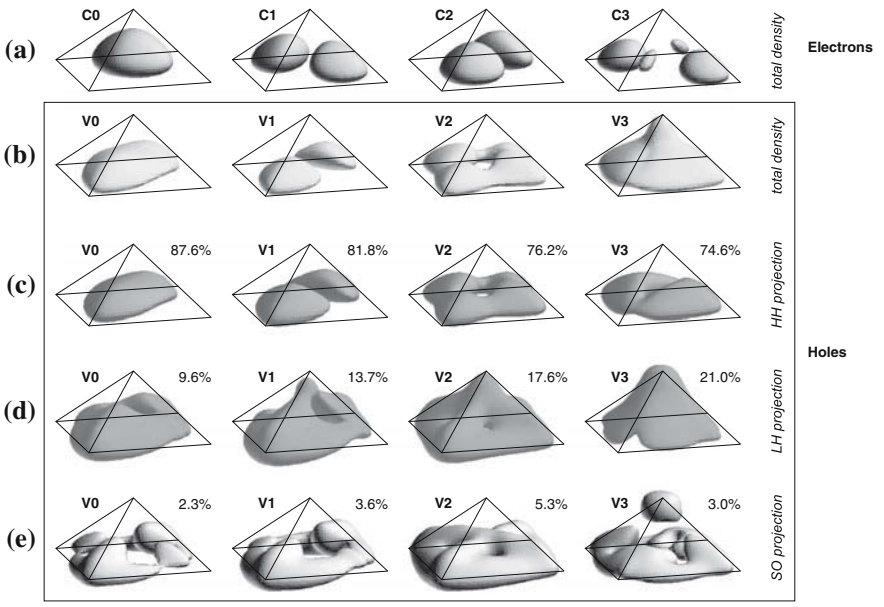


Fig. 14.15 Isosurface plots (25 % of maximum value) of the total probability densities (**a**, **b**) and valence-band projections (**c**)–(**e**) of bound electron (**a**) and hole (**b**)–(**e**) states in a model pyramidal InAs/GaAs quantum dot with base length $b = 11.3$ nm. The percentages are the integrals of the projections to the bulk heavy, light and split-off hole bands, respectively, and the isosurfaces show the corresponding projection shapes. For each valence-band state the difference from 100 % is the integral $\int_{-\infty}^{\infty} |\psi_{s\uparrow}|^2 + |\psi_{s\downarrow}|^2 \mathrm{d}^3\mathbf{r}$ of the s-type (conduction band) Bloch function projection (not shown). Reprinted with permission from [1159]

oscillator it is exactly E_0. For a three-dimensional harmonic oscillator this quantity is $E_1 - E_0 = 2E_0/3$.

For realistic quantum dots a full three-dimensional simulation of strain, piezoelectric fields and the quantum-mechanical confinement must be performed [1156, 1157]. In Fig. 14.15, the lowest four electron and hole wavefunctions in a pyramidal InAs/GaAs quantum dot (for the strain distribution see Fig. 5.27 and for the piezoelectric fields see Fig. 15.16) are shown. The figure shows that the lowest hole states have dominantly heavy-hole character and contain admixtures of the other hole bands. The wavefunction in such quantum dots can be imaged using scanning tunneling microscopy [1158].

14.3.2 Electrical and Transport Properties

The classical electrostatic energy of a quantum dot with capacitance C that is capacitively coupled to a gate (Fig. 14.16) at a bias voltage V_g is given by

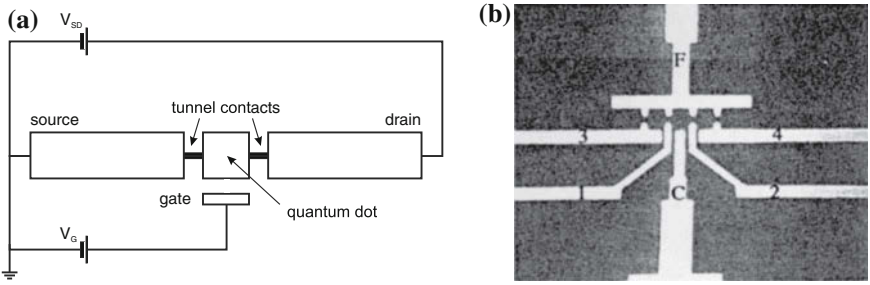


Fig. 14.16 (**a**) Schematic drawing of a quantum dot with tunnel contacts and gate electrode. (**b**) Realization with an in-plane gate structure. The distance between 'F' and 'C' (gate electrode) is 1 μm. Electron transport occurs from a 2DEG between 3/F to 4/F through the quantum points contacts 1/3 and 2/4. Part (**b**) reprinted with permission from [1160]

$$E = \frac{Q^2}{2C} - Q\,\alpha\,V_\mathrm{g}, \tag{14.12}$$

where α is a dimensionless factor relating the gate voltage to the potential of the island and Q is the charge of the island.

Mathematically, minimum energy is reached for a charge $Q_\mathrm{min} = \alpha\,C\,V_\mathrm{g}$. However, the charge has to be an integer multiple of e, i.e. $Q = Ne$. If V_g has a value, such that $Q_\mathrm{min}/e = N_\mathrm{min}$ is an integer, the charge cannot fluctuate as long as the temperature is low enough, i.e.

$$kT \ll \frac{e^2}{2C}. \tag{14.13}$$

Tunneling into or out of the dot is suppressed by the Coulomb barrier $e^2/2C$, and the conductance is very low. Analogously, the differential capacitance is small. This effect is called *Coulomb blockade*. Peaks in the tunneling current (Fig. 14.17b), conductivity (Fig. 14.17a) and the capacitance occur, when the gate voltage is such that the energies for N and $N+1$ electrons are degenerate, i.e. $N_\mathrm{min} = N + \frac{1}{2}$. The expected level spacing is

$$e\,\alpha\,\Delta V_\mathrm{g} = \frac{e^2}{C} + \Delta\epsilon_N, \tag{14.14}$$

where $\Delta\epsilon_N$ denotes the change in lateral (kinetic) quantization energy for the added electron. e^2/C will be called the charging energy in the following. A variation of the source–drain voltage (for a given gate voltage) leads to a so-called Coulomb staircase since more and more channels of conductivity contribute to the current through the device (Fig. 14.18). Single electron tunneling (SET) is investigated with respect to metrology for a novel ampere standard [1118].

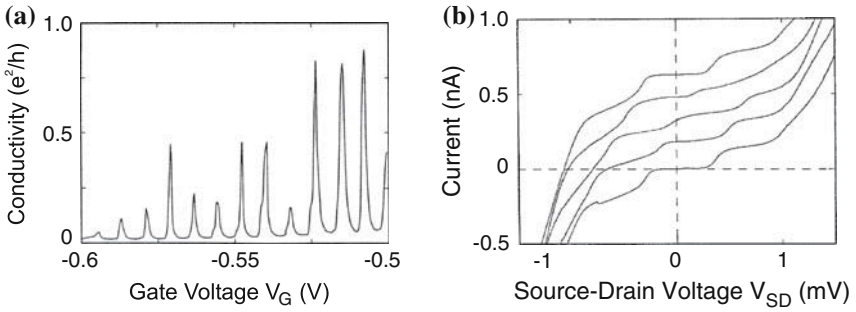


Fig. 14.17 (**a**) Conductivity (Coulomb oscillations) and (**b**) current–voltage diagram at different gate voltages (Coulomb staircase, shifted vertically for better readability) of a tunnel junction with a quantum dot as in Fig. 14.16. Adapted from [1160]

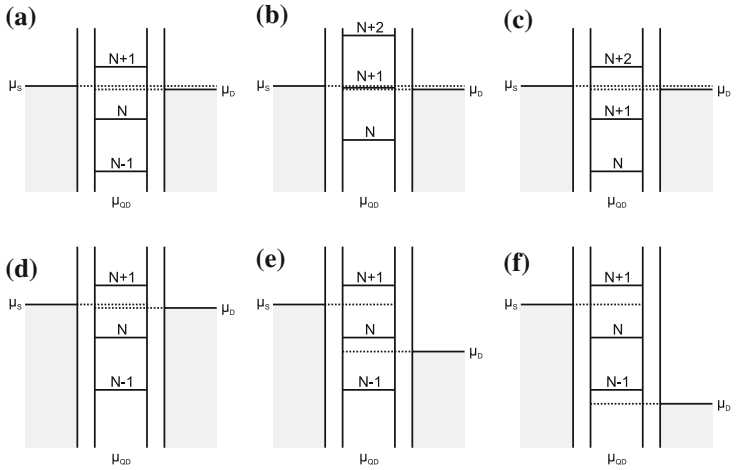


Fig. 14.18 Chemical potentials of source and drain and of a quantum dot in between them. (**a**), (**b**), and (**c**) show the sequence for a variation of the gate voltage and visualize the origin of the Coulomb oscillations (see Fig. 14.17a). (**d**), (**e**) and (**f**) visualize a variation of the source–drain voltage and the origin of the Coulomb staircase (cf. Fig. 14.17b)

A lot of research so far has been done on lithographically defined systems where the lateral quantization energies are small and smaller than the Coulomb charging energy. In this case, periodic oscillations are observed, especially for large N. A deviation from periodic oscillations for small N and a characteristic shell structure (at $N = 2, 6, 12$) consistent with a harmonic oscillator model ($\hbar\omega_0 \approx 3\,\text{meV}$) has been reported for ≈500-nm diameter mesas (Fig. 14.19b, c). In this structure, a small mesa has been etched and contacted (top contact, substrate back contact and side gate). The quantum dot consists of a 12-nm $In_{0.05}Ga_{0.95}As$ quantum well that is laterally constricted by the 500-nm mesa and vertically confined due to 9- and 7.5-nm thick $Al_{0.22}Ga_{0.68}As$ barriers (Fig. 14.19a). By tuning the gate voltage, the number

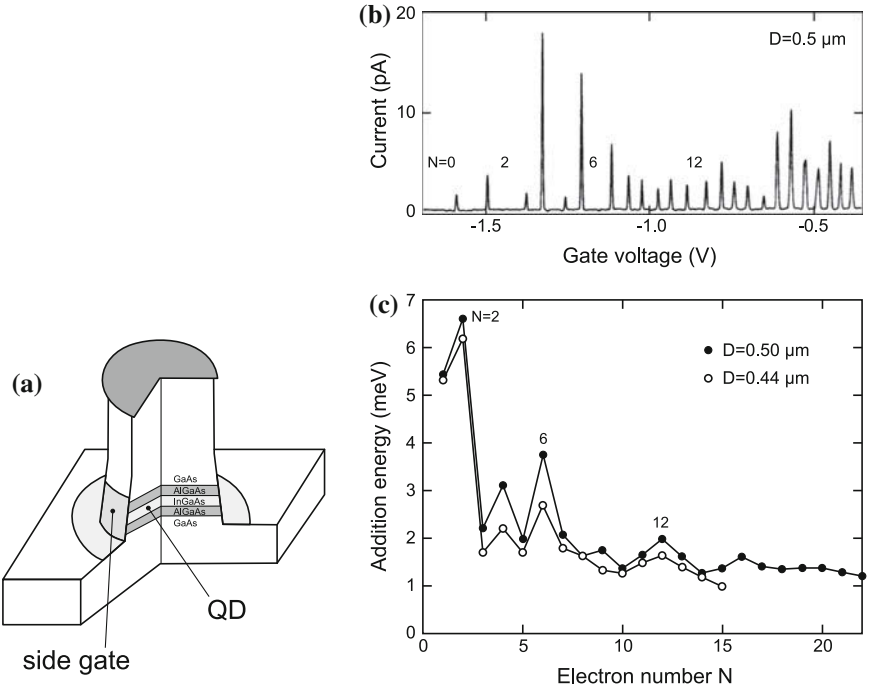


Fig. 14.19 (**a**) Schematic sample geometry for side-gated $In_{0.05}Ga_{0.95}As/Al_{0.22}Ga_{0.68}As$ disk-shaped quantum dot. (**b**) Coulomb oscillations in the current versus gate voltage at $B = 0\,\mathrm{T}$ observed for a $D = 0.5\,\mu\mathrm{m}$ disk-shaped dot. (**c**) Addition energy versus electron number for two different dots with $D = 0.50$ and $0.44\,\mu\mathrm{m}$. Adapted from [1161]

of electrons can be varied within 0 and 40. Measurements are typically carried out at a sample temperature of 50 mK.

In the sample shown in Fig. 14.20, self-assembled QDs are positioned in the channel under a split-gate structure. In a suitable structure, tunneling through a single QD is resolved.

In small self-assembled quantum dots single-particle level separations can be larger than or similar to the Coulomb charging energy. Classically, the capacitance for a metal sphere of radius R_0 is given as

$$C_0 = 4\pi\,\epsilon_0\,\epsilon_\mathrm{r}\,R_0, \tag{14.15}$$

e.g. $C_0 \approx 6\,\mathrm{aF}$ for $R_0 = 4\,\mathrm{nm}$ in GaAs, resulting in a charging energy of 26 meV. Quantum mechanically, the charging energy is given in first-order perturbation theory by

$$E_{21} = \langle 00|W_\mathrm{ee}|00\rangle = \iint \Psi_0^2(\mathbf{r}_\mathrm{e}^1)\,W_\mathrm{ee}(\mathbf{r}_\mathrm{e}^1,\mathbf{r}_\mathrm{e}^2)\,\Psi_0^2(\mathbf{r}_\mathrm{e}^2)\,\mathrm{d}^3\mathbf{r}_\mathrm{e}^1\,\mathrm{d}^3\mathbf{r}_\mathrm{e}^2, \tag{14.16}$$

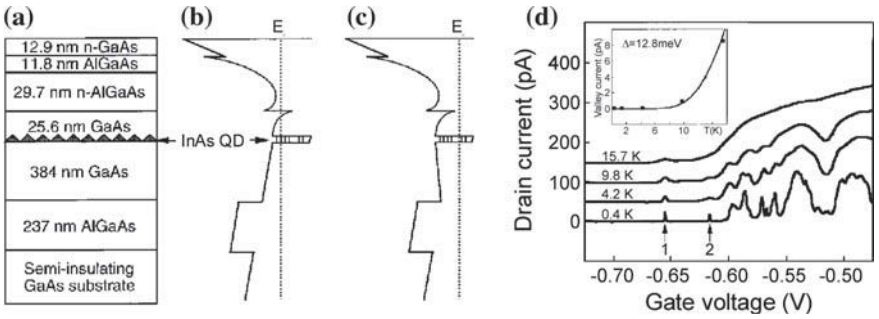


Fig. 14.20 (**a**) Schematic layer sequence of epitaxial structure comprising a n-AlGaAs/GaAs heterointerface with a two-dimensional electron gas and a layer of InAs/GaAs quantum dots. (**b**) and (**c**) are corresponding band diagrams with no gate bias and gate voltage below the critical value, respectively. (**d**) Experimental dependence of drain current on gate voltage in a split-gate structure at a drain source voltage of 10 μV. *Inset*: Dependence of valley current on temperature (*squares*) with theoretical fit. Reprinted with permission from [1162], ©1997 AIP

where W_{ee} denotes the Coulomb interaction of the two electrons and Ψ_0 the ground state (single particle) electron wavefunction. The matrix element gives an upper bound for the charging energy since the wavefunctions will rearrange to lower their overlap and the repulsive Coulomb interaction. For lens-shaped InAs/GaAs quantum dots with radius 25 nm a charging energy of about 30 meV has been predicted.

14.3.3 Self-Assembled Preparation

The preparation methods for QDs split into top-down (lithography and etching) and bottom-up (self-assembly) methods. The latter achieve typically smaller sizes and require less effort (at least concerning the machinery).

Artificial Patterning

Using artificial patterning, based on lithography and etching (Fig. 14.21), quantum dots of arbitrary shape can be made (Fig. 14.22). Due to defects introduced by high-energy ions during reactive ion etching the quantum efficiency of such structures is very low when they are very small. Using wet-chemical etching techniques the damage can be significantly lowered but not completely avoided. Since the QDs have to compete with other structures that can be made structurally perfect, this is not acceptable.

Template Growth

Template growth is another technique for the formation of nanostructures. Here, a mesoscopic structure is fabricated by conventional means. The nanostructure is created using size-reduction mechanisms, e.g. faceting, (Fig. 14.23). This method

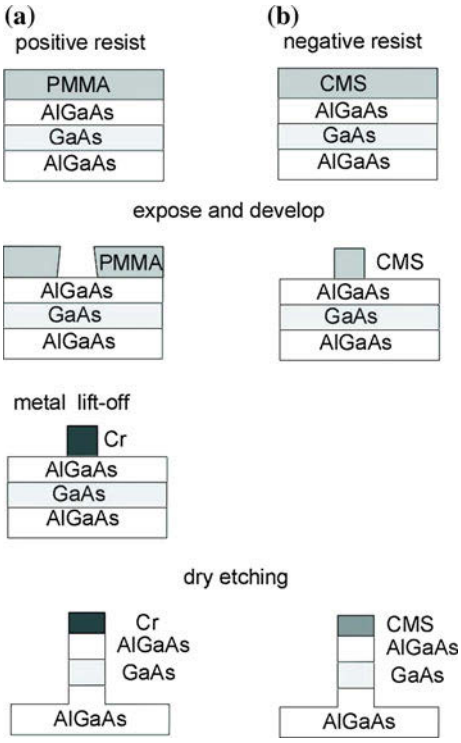


Fig. 14.21 Lithography and etching techniques for the fabrication of semiconductor structures

can potentially suffer from low template density, irregularities of the template, and problems of reproducibility.

Colloids

Another successful route to nanocrystals is the doping of glasses with subsequent annealing (color filters). When nanocrystals are prepared in a sol-gel process, the nanoparticles are present as a colloid in wet solution (Fig. 14.24). With the help of suitable stabilizing agents they are prevented from sticking to each other and can be handled in ensembles and also individually.

Mismatched Epitaxy

The self-assembly (or self-organization) relies on strained heterostructures that achieve energy minimization by island growth on a wetting layer (Stranski–Krastanow growth mode, see Sect. 12.2.3 and [1133]). Additional ordering mechanisms [1166, 1167] lead to ensembles that are homogeneous in size[2] [1168] and shape [1169] (Fig. 14.25).

[2] The ordering in size is remarkable. Typically Ostwald ripening (due to the Gibbs–Thomson effect; smaller droplets have larger vapor pressure and dissolve, larger droplets accordingly grow) occurs in an ensemble of droplets or nuclei. In the case of strained QDs, surface energy terms stabilize a certain QD size.

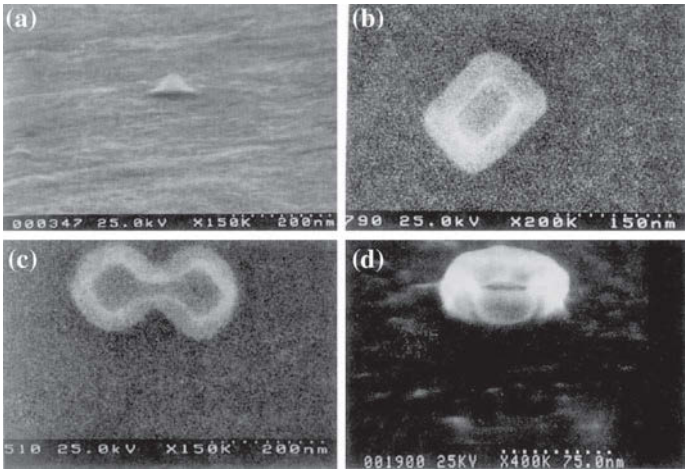


Fig. 14.22 Quantum dots of various shapes created by lithography and etching techniques. From [1163]

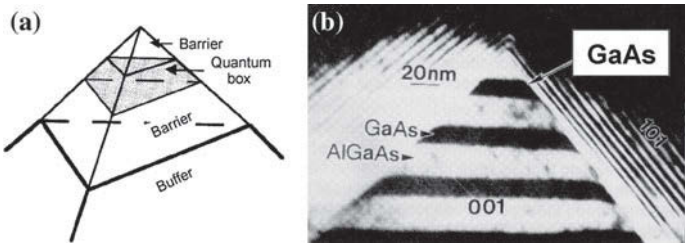


Fig. 14.23 (**a**) Schematic representation of growth on top of a predefined template, (**b**) cross-sectional TEM of quantum dot formation at the apex. Reprinted with permission from [1164], ©1992 MRS

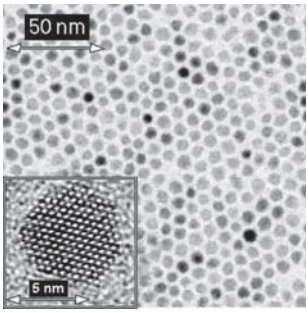


Fig. 14.24 CdSe colloidal nanoparticles. From [1165]

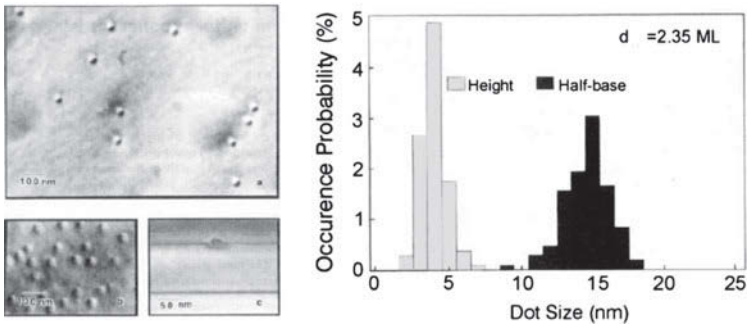


Fig. 14.25 Self-organized formation of InGaAs/GaAs quantum dots during epitaxy. *Left* Plan-view and cross-sectional transmission electron micrographs. *Right* Histogram of vertical and lateral size of the quantum dots. Reprinted with permission from [1170], ©1993 AIP

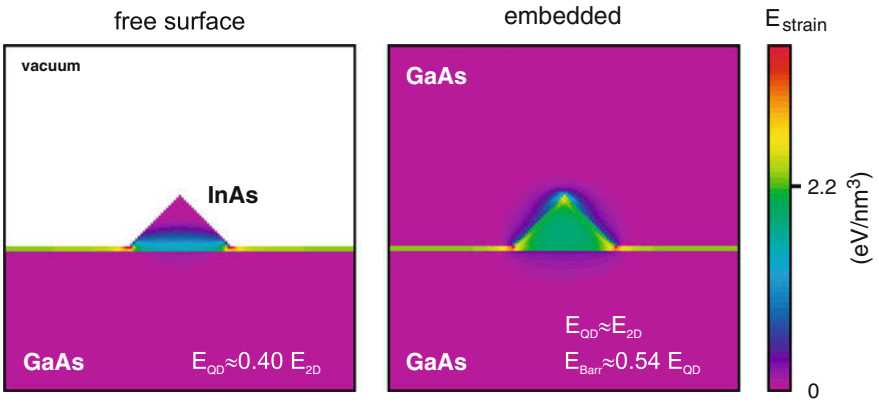


Fig. 14.26 Distribution of strain energy for (*left*) uncapped island and (*right*) island embedded in host matrix. Numerical values are for InAs/GaAs

When a thin layer of a semiconductor is grown on top of a flat substrate with different lattice constant, the layer suffers a tetragonal distortion (Sect. 5.3.3). Strain can only relax along the growth direction (Fig. 14.26). If the strain energy is too large (highly strained layer or large thickness), plastic relaxation via dislocation formation occurs. If there is island geometry, strain can relax in all three directions and about 50 % more strain energy can relax, thus making this type of relaxation energetically favorable. When the island is embedded in the host matrix, the strain energy is similar to the 2D case and the matrix becomes strained (metastable state).

When such QD layers are vertically stacked, the individual quantum dots grow on top of each other (Fig. 14.27) if the separation is not too large (Fig. 14.29). This effect is due to the effect of the underlying QD. In the case of InAs/GaAs (compressive strain), the buried QD stretches the surface above it (tensile surface strain). Thus, atoms impinging in the next QD layer find a smaller strain right on top of the buried

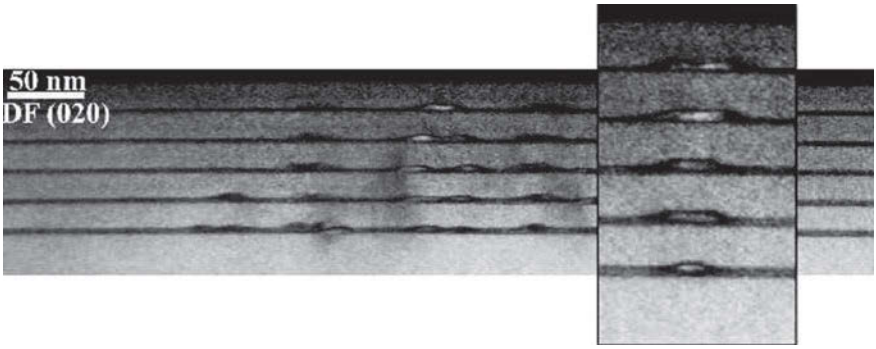


Fig. 14.27 Cross-sectional TEM image of a stack of five layers of quantum dots. Due to strain effects, vertical arrangement is achieved

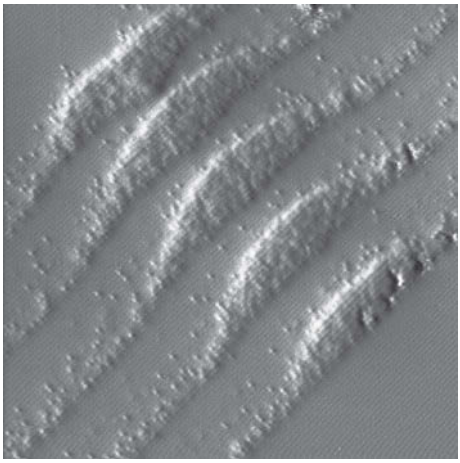


Fig. 14.28 Cross-sectional STM image of a stack of five InAs quantum dots in a GaAs matrix. Individual In atoms can be observed in-between the wetting layers and the quantum dots. Each quantum dot layer was formed by growing 2.4 ML of InAs. The intended distance between the quantum dot layers was 10 nm. Image size is $55 \times 55\,\text{nm}^2$. Reprinted with permission from [1171], ©2003 AIP

QDs. In STM images of the cross section through (XSTM) such a stack (Fig. 14.28) individual indium atoms are visible and the shape can be analyzed in detail [1171].

The vertical arrangement can lead to further ordering since a homogenization in lateral position takes place. If two QDs in the first layers are very close, their strain fields overlap and the second layer 'sees' only one QD.

The lateral (in-plane) ordering of the QDs with respect to each other occurs in square or hexagonal patterns and is mediated via strain interaction through the substrate. The interaction energy is fairly small, leading only to short-range in-plane order [1166] as shown in Fig. 14.30. The in-plane ordering can be improved up to the point that regular one- or two-dimensional arrays form or individual quantum dots are placed on designated positions using directed self-assembly [1133]. Among others, dislocation networks buried under the growth surface of the nanostructure, surface patterning and modification have been used to direct the QD positioning.

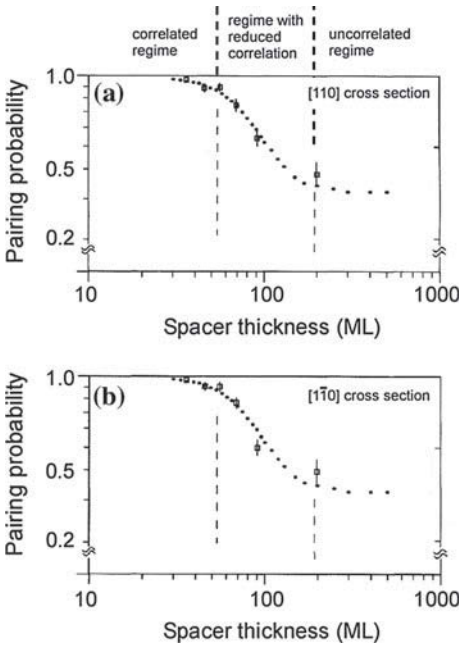


Fig. 14.29 Experimentally observed pairing probability in MBE-grown stacks of InAs/GaAs quantum dots as a function of the spacer-layer thickness. Data are taken from (**a**) (110) and (**b**) (1–10) cross-sectional TEM images. The *filled circles* are fit to data from theory of correlated island formation under strain fields. Reprinted with permission from [1172], ©1995 APS

Ion-Beam Erosion

During the erosion of a surface with low-energy ion beam sputtering ordered patterns of dots appear [1173–1176]. Isotropic [1177] and hexagonal [1173, 1175] (Fig. 14.31) near-range ordering has been observed. The pattern formation mechanism is based on the morphology-dependent sputter yield and further mechanisms of mass redistribution [1178]. Also linear patterns have been reported [1179].

14.3.4 Optical Properties

The optical properties of QDs are related to their electronic density of states. In particular, optical transitions are allowed only at discrete energies due to the zero-dimensional density of states.

Photoluminescence from a single QD is shown in Fig. 14.32. The δ-like sharp transition is strictly true only in the limit of small carrier numbers ($\ll$1 exciton per dot on average) since otherwise many-body effects come into play that can encompass recombination from charged excitons or multiexcitons. At very low excitation

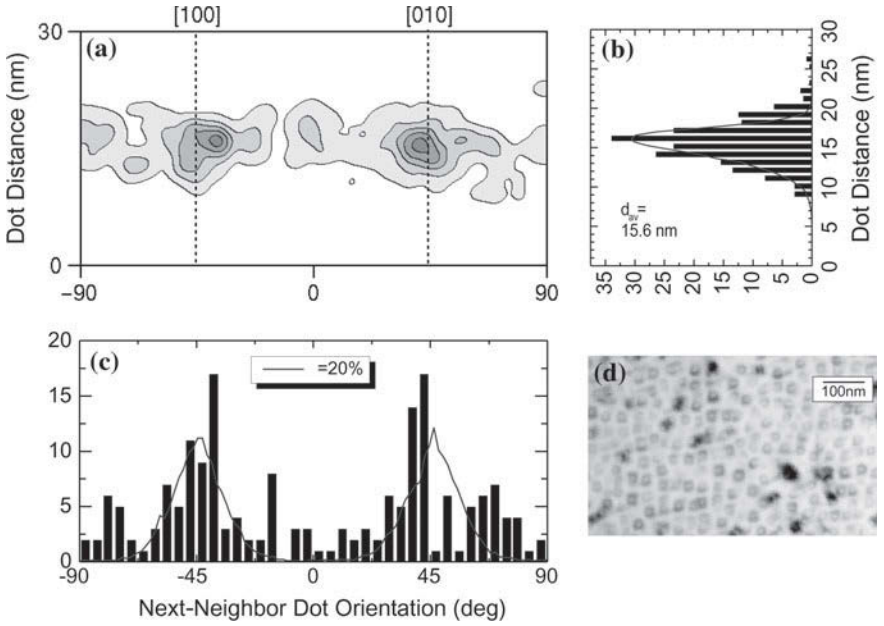


Fig. 14.30 Lateral ordering of QD array. (**d**) Plan-view TEM of QD array on which the statistical evaluation is based. (**a**) Two-dimensional histogram of QDs as a function of the nearest-neighbor distance and direction, (**b**,**c**) projections of part (**a**). *Solid lines* in (**b**) and (**c**) are theory for square array with $\sigma = 20\,\%$ deviation from ideal position. Adapted from [1133] and [1166]

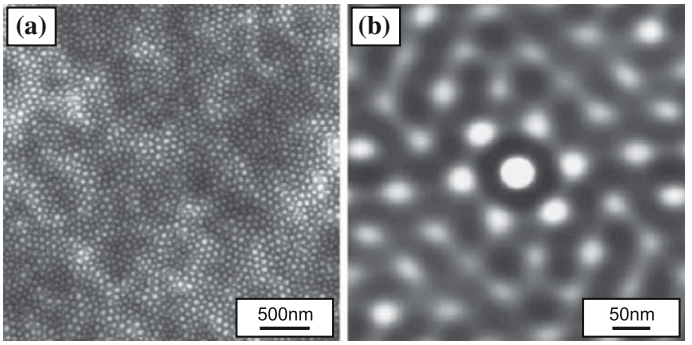


Fig. 14.31 (**a**) AFM image of a Si (001) substrate after 960 min of ion sputtering (1.2 keV Ar^+, normal incidence). (**b**) Two-dimensional autocorrelation function from a $400 \times 400\,nm^2$ area of image in part (**a**). Adapted from [1175]

density the recombination spectrum consists only of the one-exciton (X) line. With increasing excitation density small satellites on either side of the X-line develop that are attributed to charged excitons (trions) X^+ and X^-. On the low-energy side, the biexciton (XX) appears. Eventually, the excited states are populated and

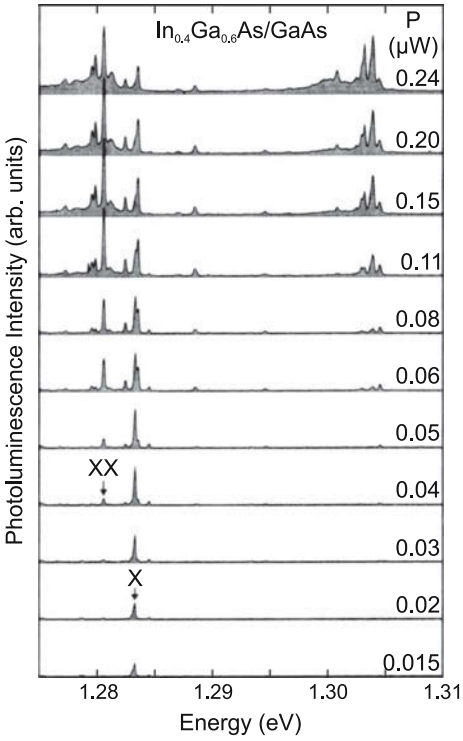


Fig. 14.32 Optical emission spectra ($T = 2.3$ K) of a single InGaAs/GaAs quantum dot at different laser excitation levels P as labeled. The single exciton (X) and biexciton (XX) lines are indicated. Adapted from [1180]

a multitude of states contribute with rich fine structure. In bulk material the biexciton (Sect. 9.6.10) is typically a bound state, i.e. its recombination energy E_{XX} is lower than that of the exciton E_X. A similar situation is present in Fig. 14.32. It was pointed out in [1181] that in QDs the biexciton recombination energy can also be *larger* than the exciton recombination energy. In [1182] the modification of the QD confinement potential of InAs/GaAs QDs by annealing was reported. The exciton binding energy (E_X-E_{XX}) is tuned from positive ('normal') to negative values upon annealing (Fig. 14.33).

The charging state of the exciton can be controlled in a field-effect structure. The recombination energy is modified due to Coulomb and exchange effects with the additional carriers. In charge-tunable quantum dots [1183] and rings [1184] exciton emission has been observed in dependence of the number of additional electrons. The electron population can be controlled in a Schottky-diode-like structure through

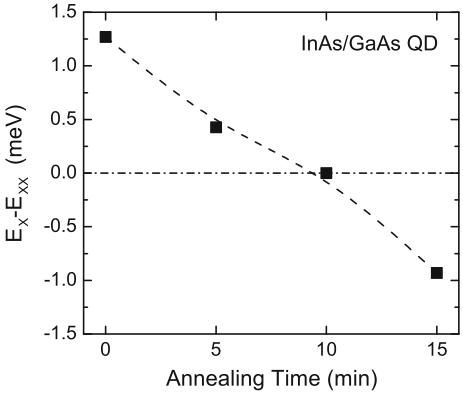


Fig. 14.33 Biexciton binding energy determined for a single InAs/GaAs quantum dot for various annealing times. Data from [1182]

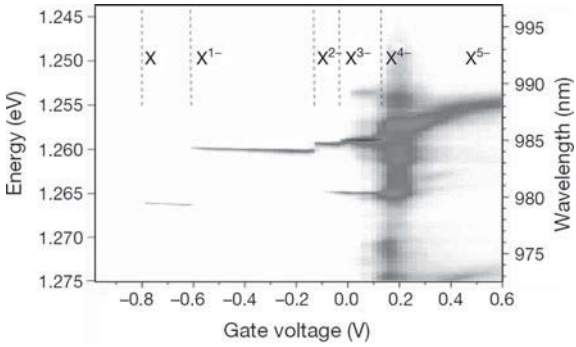


Fig. 14.34 Luminescence of charged excitons from a single quantum ring at $T = 4.2\,\mathrm{K}$ versus the bias voltage with which the number of electrons in the quantum dot N is tuned from zero to $N > 3$. Adapted from [1184], reprinted with permission from Nature, ©2000 Macmillan Magazines Limited

the manipulation of the Fermi level with the bias voltage. At high negative bias all charge carriers tunnel out of the ring and no exciton emission is observed. A variation of the bias then leads to an average population with $N = 1, 2, 3, \ldots$ electrons. The recombination of additional laser-excited excitons depends (due to the Coulomb interaction) on the number of the electrons present (Fig. 14.34). The singly negatively charged exciton X^- is also called a trion.

The interaction of a spin with an exciton in a CdTe quantum dot has been observed in [1185]. If the CdTe quantum dot is pure, a single line arises. If the dot contains a single Mn atom, the exchange interaction of the exciton with the Mn $S = 5/2$ spin leads to a six-fold splitting of the exciton line (Fig. 14.35). In an external magnetic field a splitting into a total of twelve lines due to Zeeman effect at the Mn spin is observed.

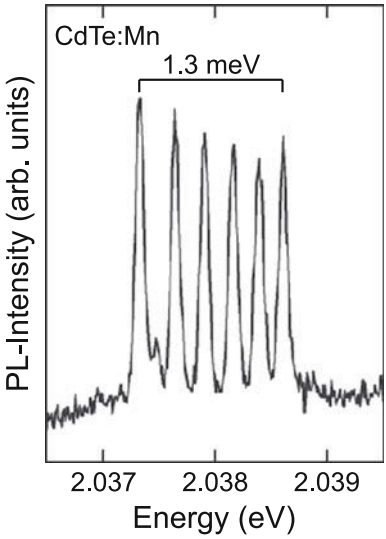


Fig. 14.35 Photoluminescence spectrum of a single CdTe/ZnSe quantum dot containing a single Mn atom ($T = 5\,\mathrm{K}$)

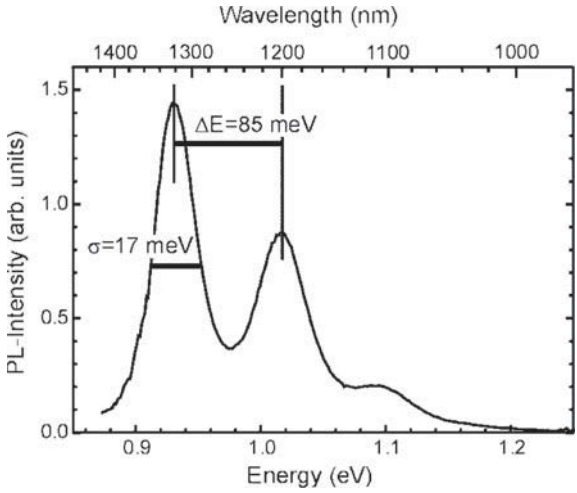


Fig. 14.36 Ensemble photoluminescence spectrum ($T = 293\,\mathrm{K}$, excitation density $500\,\mathrm{W/cm^2}$) of InAs/GaAs QDs

In a QD ensemble, optical transitions are inhomogeneously broadened due to fluctuations in the QD size and the size dependence of the confinement energies (Fig. 14.36). Interband transitions involving electrons and holes suffer from the variation of the electron and hole energies:

$$\sigma_E \propto \left(\left| \frac{\partial E_\mathrm{e}}{\partial L} \right| + \left| \frac{\partial E_\mathrm{h}}{\partial L} \right| \right) \delta L. \tag{14.17}$$

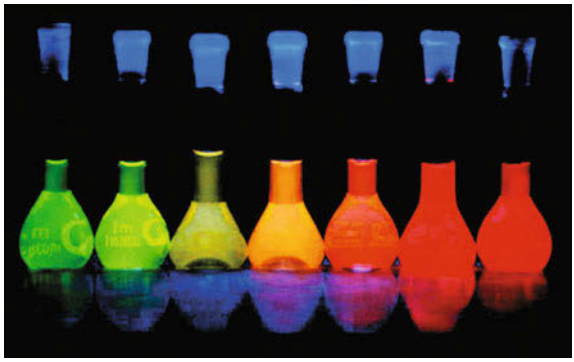


Fig. 14.37 Luminescence (under UV excitation) from flasks of colloidal CdTe quantum dots with increasing size (from *left* to *right*). From [1165]

A typical relative size inhomogeneity of σ_L/L of 7 % leads to several tens of meV broadening. Additional to broadening due to different sizes fluctuations of the quantum dot shape can also play a role. The confinement effect leads to an increase of the recombination energy with decreasing quantum-dot size. This effect is nicely demonstrated with colloidal quantum dots of different size as shown in Fig. 14.37.

Chapter 15
Polarized Semiconductors

Abstract In this chapter semiconductors with a spontaneous polarization and ferroelectric semiconductors are discussed. Also the effect of piezoelectricity is treated in some detail for zincblende and wurtzite materials.

15.1 Introduction

Semiconductors can have an electric polarization. Such polarization can be induced by an external electric field (Fig. 15.1a). This phenomenon, i.e. that the semiconductor is dielectric, has been discussed already in Chap. 9. In this chapter, we discuss pyroelectricity, i.e. a spontaneous polarization without an external field (Fig. 15.1b), ferroelectricity,[1] i.e. pyroelectricity with a hysteresis (Fig. 15.1c) and piezoelectricity, i.e. a polarization due to external stress.

15.2 Spontaneous Polarization

The reason for a spontaneous polarization $P_{\rm sp}$ (without external electric field) is the static, relative shift of positive and negative charges in the unit cell. For a slab of semiconductor material (thus ignoring depolarization effects present in other geometries), the polarization causes polarization charges located at the upper and lower surfaces (Fig. 15.2a). The polarization vector $\mathbf{P}$ points from the negative to the positive charge. The electric field due to the polarization charges has the opposite direction. In the absence of free charges, the Maxwell equation $\nabla \cdot \mathbf{D} = 0$ yields for piecewise constant fields at a planar interface (Fig. 15.2b) $(\mathbf{D}_2 - \mathbf{D}_1) \cdot \mathbf{n}_{12} = 0$ where $\mathbf{n}_{12}$ is the surface normal pointing from medium 1 to medium 2. Therefore, the polarization charge $\sigma_{\rm P} = \epsilon_0 \nabla \cdot \mathbf{E}$ is given by

[1] Ferromagnetic semiconductors are discussed in Chap. 16. We note that there exist so called *multiferroic* materials that possess more than one ferroic property [1186, 1187].

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_15

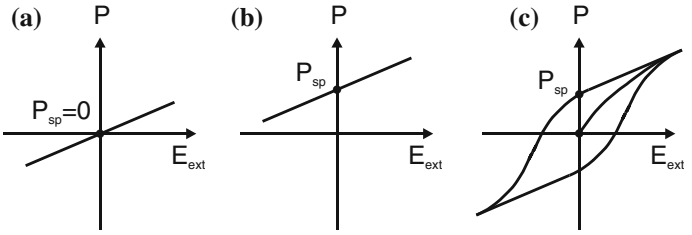


Fig. 15.1 Schematic representation of the polarization versus external electric field dependence for (**a**) dielectric, (**b**) pyroelectric and (**c**) ferroelectric semiconductors

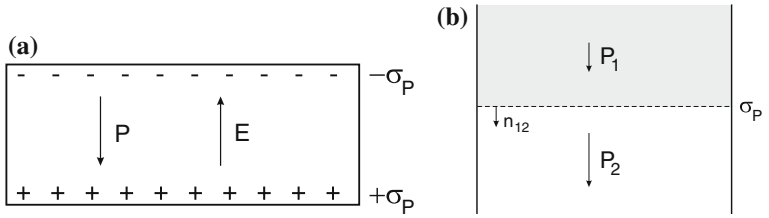


Fig. 15.2 (**a**) Surface polarization charges σ_P on a slab of semiconductor material with polarization. The electric field is given by $\mathbf{E} = -\mathbf{P}/\epsilon_0$. (**b**) Polarization charge σ_P at an interface between two semiconductors with different polarization. In the depicted situation σ is negative

$$\sigma_P = -(\mathbf{P}_2 - \mathbf{P}_1) \cdot \mathbf{n}_{12}. \tag{15.1}$$

Polarization charges develop at interfaces where the polarization is discontinuous, e.g. an interface between two semiconductors with different spontaneous polarization. Vacuum (at a surface) represents a special case with $\mathbf{P} = 0$.

For cubic zincblende structure semiconductors, P_{sp} is typically fairly small. The anisotropy of the wurtzite structure allows for sizeable effects (Fig. 15.3). The main cause is the nonideality of the cell-internal parameter u (cf. Sect. 3.4.5).

15.3 Ferroelectricity

Ferroelectric semiconductors exhibit a spontaneous polarization in the ferroelectric phase and zero spontaneous polarization in the paraelectric phase. As a function of temperature, the ferroelectric material undergoes a phase transition from the high-temperature paraelectric phase into the ferroelectric phase. There can be further phase transitions between different ferroelectric phases that differ in the direction of the polarization. The literature until 1980 is summarized in [1189]. A more recent treatment can be found in [1190,1191].

$PbTiO_3$ has perovskite structure (cf. Sect. 3.4.10). It exhibits a phase transition at $T_C = 490\,°C$ from the cubic into the (ferroelectric) tetragonal phase as shown in Fig. 15.4a. Mostly the cell symmetry changes, while the cell volume remains almost constant. A more complicated situation arises for $BaTiO_3$. At 120 °C the transition

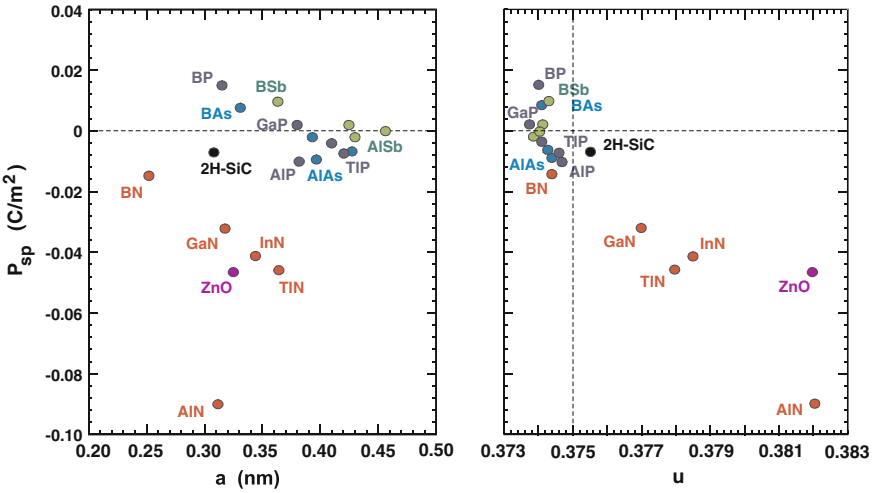


Fig. 15.3 Spontaneous polarization P_{sp} for various semiconductors as a function of the lattice constant a (*left*) and the cell-internal parameter u (*right*). Based on [1188]

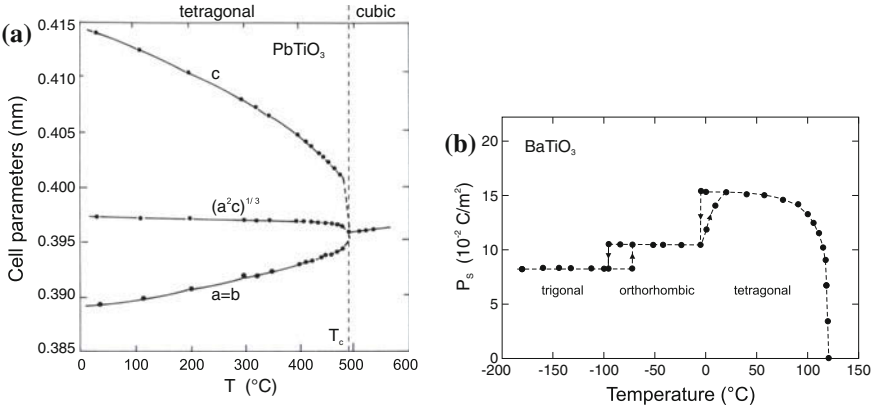


Fig. 15.4 (**a**) Cell parameters of $PbTiO_3$ as a function of temperature. Adapted from [1194]. (**b**) Phase transitions of $BaTiO_3$ as a function of temperature. The spontaneous polarization P_S points along ⟨100⟩, ⟨110⟩ and ⟨111⟩ in the tetragonal (C_{4v}), orthorhombic (C_{2v}) and trigonal (C_{3v}, rhombohedral) phase, respectively. Adapted from [1195]

into the ferroelectric phase occurs (Fig. 15.4b) that is tetragonal with the polarization in the [100] direction. At −5 °C and −90 °C transitions occur into an orthorhombic and a rhombohedral (trigonal) phase, respectively. The largest polarization is caused by a displacement of the negatively (O) and positively (Ba, Ti) charged ions of the

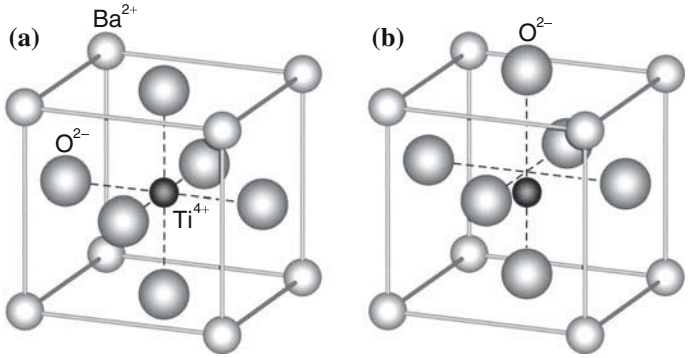


Fig. 15.5 (**a**) Crystal structure of $BaTiO_3$ (see Fig. 3.26). (**b**) Schematic tetragonal deformation below the Curie temperature, generating a dipole moment

unit cell by $\delta \approx 0.02\,\text{nm}$ (Fig. 15.5). Such an origin of the spontaneous polarization is called a *displacement* transition.[2]

15.3.1 Materials

A large class of ferroelectric semiconductors are of the type ABO_3, where A stands for a cation with larger ionic radius and B for an anion with smaller ionic radius. Many ferroelectrics have perovskite ($CaTiO_3$) structure. They are $A^{2+}B^{4+}O_3^{2-}$, e.g. (Ba,Ca,Sr) (Zi,Zr)O_3 or $A^{1+}B^{5+}O_3^{2-}$, e.g. (Li,Na,K) (Nb,Ta)O_3. Ferroelectrics can also be alloyed. Alloying in the B component yields, e.g. $PbTi_xZr_{1-x}O_3$ also called PZT. PZT is widely used for piezoelectric actuators. Also, alloying in the A component is possible, e.g. $Ba_xSr_{1-x}TiO_3$.

Another class of ferroelectrics are $A^{V}B^{VI}C^{VII}$ compounds, such as SbSI, SbSBr, SbSeI, BiSBr. These materials have a width of the forbidden band in the $\sim$2 eV range. A further class of ferroelectric semiconductors are $A_2^{V}B_3^{VI}$ compounds, such as Sb_2S_3.

15.3.2 Soft Phonon Mode

The finite displacement of the sublattices in the ferroelectric means that the related lattice vibration has no restoring force. The displacement is, however, finite due to higher-order terms (anharmonicity). Thus, for $T \to T_C$ $\omega_{TO} \to 0$. Such a mode

[2]The widely accepted model for such ferroelectric is that the basic displacement occurs into the ⟨111⟩-direction at low temperature. The three higher symmetry phases at higher temperature are the result of 2 (orthorhombic), 4 (tetragonal) or 8 (cubic) allowed ⟨111⟩ orientations which make the macroscopically averaged polarization appear in ⟨110⟩- or ⟨100⟩-direction or vanishing altogether, respectively [1192, 1193].

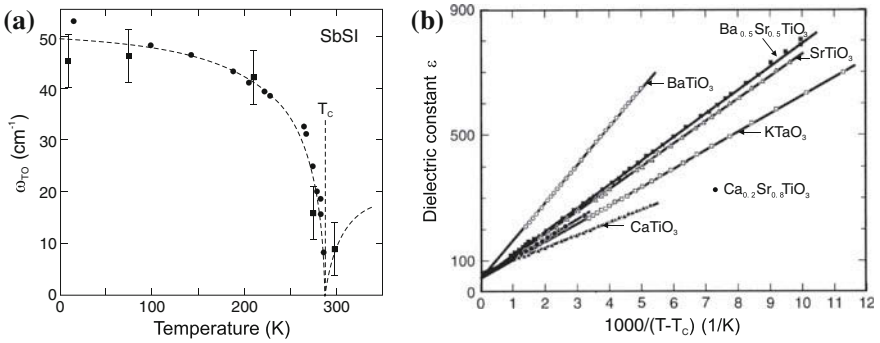


Fig. 15.6 (**a**) Decrease of the transverse phonon mode of SbSI close to the Curie temperature of $T_C = 288\,\mathrm{K}$. The *dashed curved line* represents a $|T - T_C|^{1/2}$ dependence. Adapted from [1196]. (**b**) Dielectric constant of various perovskites versus $1/(T - T_C)$ in the paraelectric phase ($T > T_C$). Adapted from [1197]

is called a *soft phonon mode*. The decrease of the phonon frequency is shown in Fig. 15.6a for SbSI.

From the LST relation (5.49), it follows that the static dielectric function must increase strongly. The increase is $\propto (T - T_C)^{-1}$ (Fig. 15.6b).

15.3.3 Phase Transition

In the case of ferroelectrics, the order parameter for the Ginzburg–Landau theory of phase transitions is the spontaneous polarization P. The free energy F of the ferroelectric crystal is written in terms of the free energy of the paraelectric phase F_0 and is expanded in powers of P (here up to P^6) as

$$F = F_0 + \frac{1}{2}\,\alpha\,P^2 + \frac{1}{4}\,\beta\,P^4 + \frac{1}{6}\,\gamma\,P^6. \tag{15.2}$$

In this equation, we have neglected effects due to charge carriers, an external electric field or external stresses and we assume homogeneous polarization. In order to obtain a phase transition, it has to be assumed that α has a zero at a certain temperature T_C and we assume (expanding only to the linear term)

$$\alpha = \alpha_0\,(T - T_C). \tag{15.3}$$

Second-Order Phase Transition

For modeling a second-order phase transition, we set $\gamma = 0$. Thus, the free energy has the form (Fig. 15.7a)

$$F = F_0 + \frac{1}{2}\,\alpha\,P^2 + \frac{1}{4}\,\beta\,P^4. \tag{15.4}$$

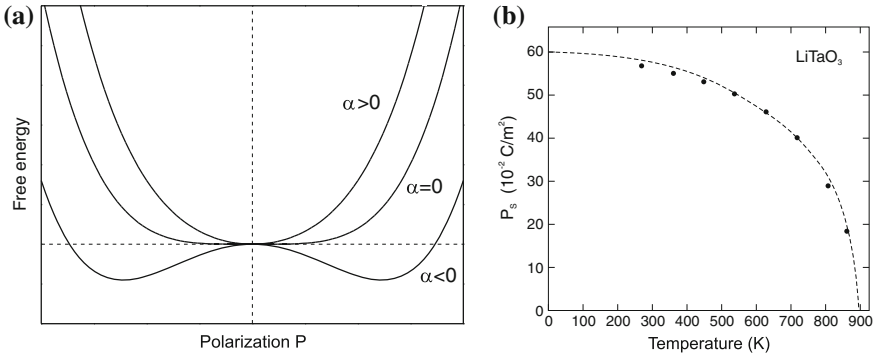


Fig. 15.7 (**a**) Schematic plot of the free energy versus spontaneous polarization for a second-order phase transition. $\alpha > 0$ ($\alpha < 0$) corresponds to the paraelectric (ferroelectric) phase. (**b**) Spontaneous polarization of $LiTaO_3$ as a function of temperature exhibiting a second-order phase transition. The *dashed line* is theory with suitable parameters. Adapted from [1198]

The equilibrium condition with regard to the free energy yields a minimum for

$$\frac{\partial F}{\partial P} = \alpha\, P + \beta\, P^3 = 0 \tag{15.5a}$$

$$\frac{\partial^2 F}{\partial P^2} = \alpha + 3\,\beta\, P^2 > 0\,. \tag{15.5b}$$

Equation (15.5a) yields two solutions. $P = 0$ corresponds to the paraelectric phase. $P^2 = -\alpha/\beta$ is the spontaneous polarization in the ferroelectric phase. The condition from (15.5b) yields that $\alpha > 0$ in the paraelectric phase, while $\alpha < 0$ in the ferroelectric phase. Also, $\beta > 0$ below the Curie temperature (β is assumed to be temperature independent in the following). Using (15.3), the polarization is given as (Fig. 15.7b)

$$P^2 = \frac{\alpha_0}{\beta}\,(T - T_\mathrm{C})\,. \tag{15.6}$$

Therefore, the entropy $S = -\frac{\partial F}{\partial T}$ and the discontinuity ΔC_p of the heat capacity $C_\mathrm{p} = T\left(\frac{\partial S}{\partial T}\right)_\mathrm{p}$ at the Curie point T_C are given by

$$S = S_0 + \frac{\alpha_0^2}{\beta}\,(T - T_\mathrm{C}) \tag{15.7a}$$

$$\Delta C_\mathrm{p} = \frac{\alpha_0^2}{\beta}\,T_\mathrm{C}, \tag{15.7b}$$

with $S_0 = -\frac{\partial F_0}{\partial T}$ being the entropy of the paraelectric phase. This behavior is in accordance with a second-order phase transition with vanishing latent heat (continuous entropy) and a discontinuity of the heat capacity. The dielectric function in

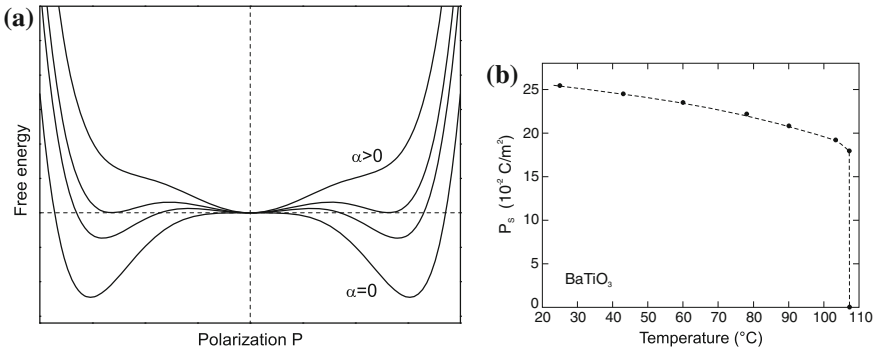


Fig. 15.8 (**a**) Schematic plot of the free energy versus spontaneous polarization for a first-order phase transition. The lowest curve is for $\alpha = 0$, the others are for $\alpha > 0$. (**b**) Spontaneous polarization in $BaTiO_3$ as a function of temperature exhibiting a first-order phase transition. The *dashed line* is a guide to the eye. Adapted from [1199]

the paraelectric phase is $\propto 1/\alpha$ and in the ferroelectric phase $\propto -1/\alpha$. The latter relation is usually written as the Curie–Weiss law

$$\epsilon = \frac{C}{T - T_C}. \tag{15.8}$$

First-Order Phase Transition

When the P^6 term is included in (15.2) ($\gamma \neq 0$), a first-order phase transition is modeled. However, in order to obtain something new, compared to the previous consideration, now $\beta < 0$ (and $\gamma > 0$) is necessary. The dependence of the free energy on P is schematically shown in Fig. 15.8a for various values of α. The condition $\frac{\partial F}{\partial P} = 0$ yields

$$\alpha P + \beta P^3 + \gamma P^5 = 0, \tag{15.9}$$

with the solutions $P = 0$ and

$$P^2 = -\frac{\beta}{2\gamma}\left(1 + \sqrt{1 - \frac{4\alpha\gamma}{\beta^2}}\right). \tag{15.10}$$

For a certain value of α, i.e. at a certain temperature $T = T_1$, the free energy is zero for $P = 0$ *and* also for another value $P = P_0$ (second curve from the top in Fig. 15.8a). From the condition

$$\frac{1}{2}\alpha(T_1)\, P_0^2 + \frac{1}{4}\beta P_0^4 + \frac{1}{6}\gamma P_0^6 = 0, \tag{15.11}$$

the values for P_0 and α at the transition temperature $T = T_1$ are given by

$$P_0^2 = -\frac{3}{4}\frac{\beta}{\gamma} \tag{15.12a}$$

$$\alpha(T_1) = \frac{3}{16}\frac{\beta^2}{\gamma} > 0. \tag{15.12b}$$

The schematic dependence of P at the phase transition temperature T_1 is depicted in Fig. 15.8b.

For $T \leq T_1$ the absolute minimum of the free energy is reached for finite polarization $P > P_0$. However, between $F(P = 0)$ and the minimum of the free energy an energy barrier (second lowest curve in Fig. 15.8) is present for T close to T_1. The energy barrier disappears at the Curie–Weiss temperature T_0. At the phase transition temperature, the entropy has a discontinuity

$$\Delta S = \alpha_0 P_0^2, \tag{15.13}$$

that corresponds to a latent heat $\Delta Q = T\Delta S$. Another property of the first-order phase transition is the occurrence of hysteresis in the temperature interval between T_1 and T_0

$$\Delta T \approx T_1 - T_0 = \frac{1}{4\alpha_0}\frac{\beta^2}{\gamma}, \tag{15.14}$$

in which an energy barrier is present to hinder the phase transition. For decreasing temperature, the system tends to remain in the paraelectric phase. For increasing temperature, the system tends to remain in the ferroelectric phase. Such behavior is observed for $BaTiO_3$, as shown in Fig. 15.4b.

15.3.4 Domains

Similar to ferromagnets, ferroelectrics form domains with different polarization directions in order to minimize the total energy by minimizing the field energy outside the crystal. The polarization can have different orientations, 6 directions for P along $\langle < 100 \rangle >$ (tetragonal phase), 12 directions for P along $\langle < 110 \rangle >$ (orthorhombic phase) and 8 directions for P along $\langle < 111 \rangle >$ (rhombohedral phase). In Fig. 15.9, such domains are visualized for $BaTiO_3$. Due to the restricted geometry, domain formation in thin films is different from that in bulk material.

Domains can also be artificially created by so-called poling. The ferroelectric semiconductor is heated to the paraelectric phase. With electrodes, appropriate electric fields are applied and the material is cooled. The polarization is then frozen in the ferroelectric phase. The domains of a periodically poled structure in $LiNbO_3$ (PPLN) are shown in Fig. 15.10b. The nonlinear optical properties in such structures can be used for efficient second harmonic generation (SHG).

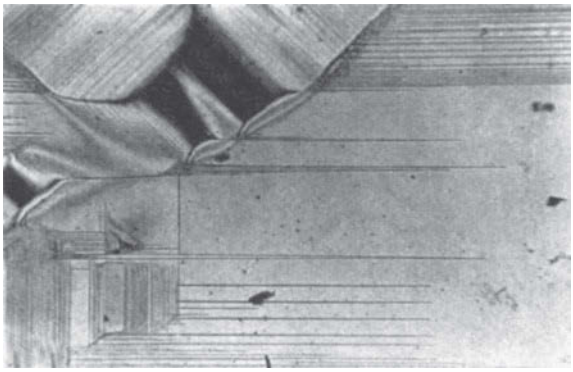


Fig. 15.9 Ferroelectric domains in a $BaTiO_3$ single crystal visualized by birefringence contrast. Reprinted with permission from [1200], © 1949 APS

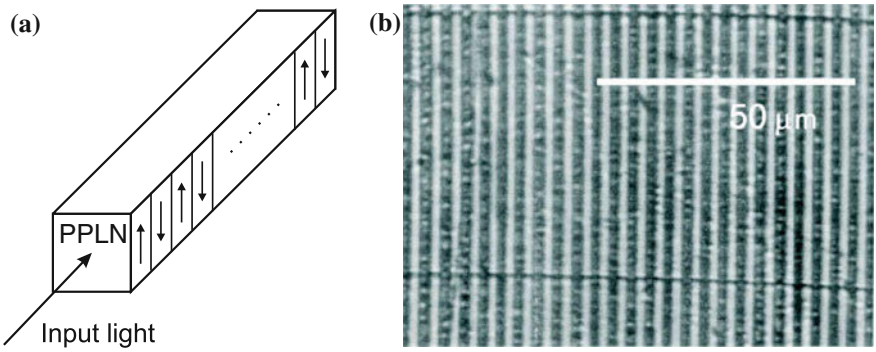


Fig. 15.10 (**a**) Scheme of PPLN (perodically poled lithium niobate), *arrows* denote the direction of spontaneous polarization. (**b**) Polarization microscopy image (vertical stripes are domains, horizontal dark lines are scratches)

15.3.5 Optical Properties

The first-order phase transition of $BaTiO_3$ manifests itself also in a discontinuity of the band gap (Fig. 15.11). The coefficient $\partial E_g/\partial T$ for the temperature dependence of the band gap is also different in the para- and ferroelectric phases.

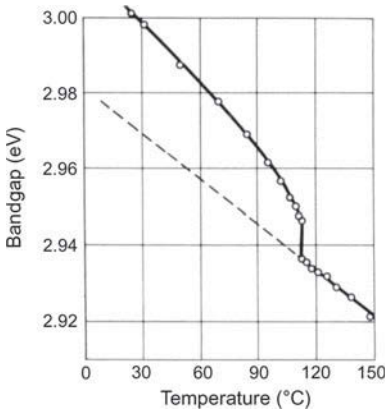


Fig. 15.11 Temperature dependence of the band gap in $BaTiO_3$ (for polarized light with $\mathbf{E} \perp c$). Experimental data from [1201]

15.4 Piezoelectricity

15.4.1 Piezoelectric Effect

External stress causes atoms in the unit cell to shift with respect to each other. In certain directions, such a shift can lead to a polarization. Generally, all ferroelectric materials are piezoelectric. However, there are piezoelectric materials that are not ferroelectric, e.g. quartz, GaAs and GaN. Piezoelectricity can occur only when no center of inversion is present. Thus, e.g., GaAs is piezoelectric along $\langle < 111 \rangle >$, but Si is not. Also, the cubic perovskite structure (in the paraelectric phase) is not piezoelectric. Generally, the piezoelectric polarization is related to the strains via the tensor e_{ijk} of the piezoelectric modules[3]

$$P_i = e_{ijk}\, \epsilon_{jk}. \tag{15.15}$$

15.4.2 Zincblende Crystals

In zincblende semiconductors, the polarization (with respect to $x = [100]$, $y = [010]$, $z = [001]$) is due to shear strains only and is given as

$$\mathbf{P}_{\text{pe}} = 2\, e_{14} \begin{pmatrix} \epsilon_{yz} \\ \epsilon_{xz} \\ \epsilon_{xy} \end{pmatrix}. \tag{15.16}$$

[3] Another formulation is used with parameters d_{ijk} and $P_i = d_{ijk}\, \sigma_{jk}$.

Table 15.1 Piezoelectric coefficients (in $\mathrm{C\,m^{-2}}$) for various zincblende and wurtzite semiconductors

Zincblende				Wurtzite			
III–V	e_{14}	II–VI	e_{14}		e_{33}	e_{31}	e_{15}
InSb	−0.123	CdTe	0.054	CdSe	0.347	−0.16	−0.138
InAs	−0.078	ZnSe	0.049	CdS	0.385	−0.262	−0.183
GaSb	−0.218	ZnS	0.254	ZnS	0.265	−0.238	−0.118
GaAs	−0.277	ZnO	0.69	ZnO	0.89	−0.51	−0.45
AlSb	−0.118	GaN	0.375	GaN	0.73	−0.49	−0.3
				AlN	1.46	−0.60	−0.48
				BeO	0.02	−0.02	

Data from [690,1202,1205], for zb-ZnO from [1206], for zb-GaN from [1207]

The values of e_{14} for various zincblende compound semiconductors are listed in Table 15.1. We note that the sign of e_{14} reverses from negative in cubic III–V to positive in cubic II–VI semiconductors. This non-trivial behavior involves strain effects on ionic and electronic polarization and ionicity as discussed in [163,1202]. The coefficient e_{33}, the equivalent to e_{14} in wurtzite semiconductors (see below), is positive for III–V and II-VI semiconductors.

The strain in pseudomorphic heterostructures (cf. Sect. 5.3.3) can cause piezoelectric polarization in a piezoelectric semiconductor. In zincblende, the main effect is expected when the growth direction is along [111] and the strain has a purely shear character. In this case, the polarization is in the [111] direction, i.e. perpendicular to the interface ($P_\perp$). For the [001] growth direction, no piezoelectric polarization is expected. For the [110] growth direction, the polarization is found to be parallel to the interface ($P_\parallel$). The situation is shown for various orientations of the growth direction in Fig. 15.12.

15.4.3 Wurtzite Crystals

In wurtzite crystals, the piezoelectric polarization (with respect to x = [2–1.0], y = [01.0], z = [00.1]) is given by[4]

$$\mathbf{P}_{\mathrm{pe}} = \begin{pmatrix} 2\,e_{15}\,\epsilon_{xz} \\ 2\,e_{15}\,\epsilon_{yz} \\ e_{31}\,(\epsilon_{xx}+\epsilon_{yy}) + e_{33}\,\epsilon_{zz} \end{pmatrix}. \tag{15.17}$$

[4] In Voigt notation, $P_x = e_{15}\,e_5$, $P_y = e_{15}\,e_4$.

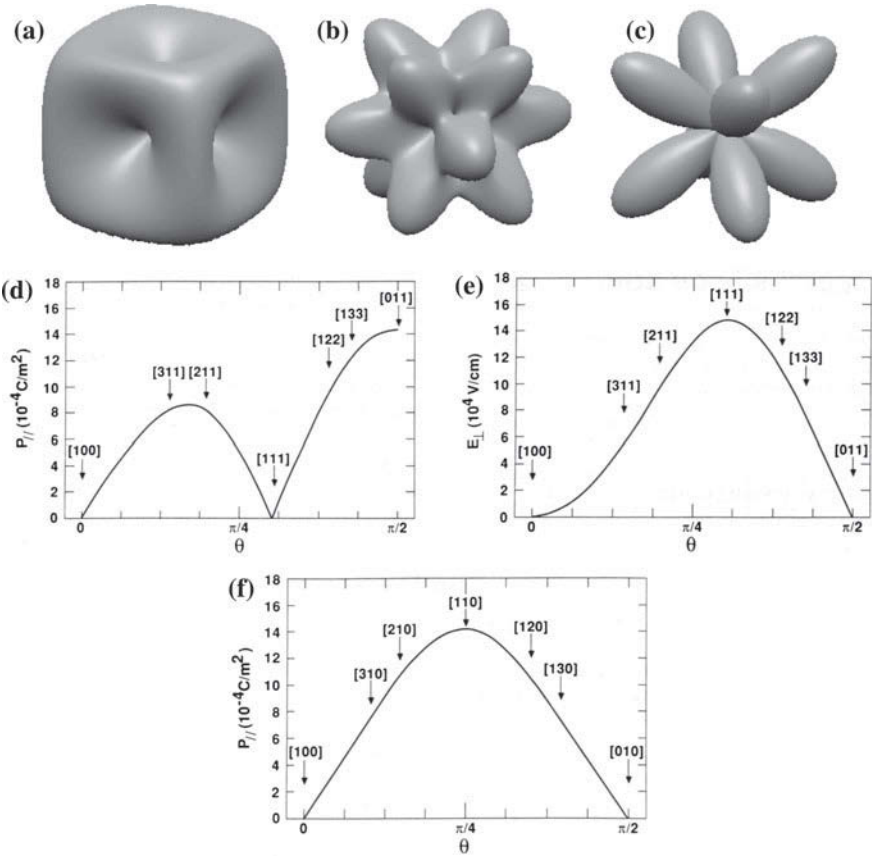


Fig. 15.12 Three-dimensional view of the (**a**) total, (**b**) longitudinal and (**c**) transverse polarization in uniaxially compressed GaAs. Adapted from [1203]. (**d**) Transverse polarization $P_\parallel$ (parallel to the interface) and (**e**) longitudinal electric field $E_\perp$ (perpendicular to the interface) in the InGaAs layer of a GaAs/$In_{0.2}Ga_{0.8}As$ superlattice with joint in-plane lattice constant (obtained from energy minimization, 1.4 % lattice mismatch, the InGaAs is under compressive and the GaAs under tensile strain). The layer thicknesses of the GaAs and InGaAs layers are identical. The quantities are shown for various orientations of the growth direction. The vector of the growth direction varies in the (01–1) plane ($\phi = \pi/4$) with polar angle θ reaching from [100] (0°) over [111] to [011] (90°). Image (**f**) depicts the transverse polarization $P_\parallel$ ($P_\perp = 0$ in this geometry) for growth directions in the (001) plane ($\phi = 0$). Parts (**d**, **e**, **f**) reprinted with permission from [1204], © 1988 AIP

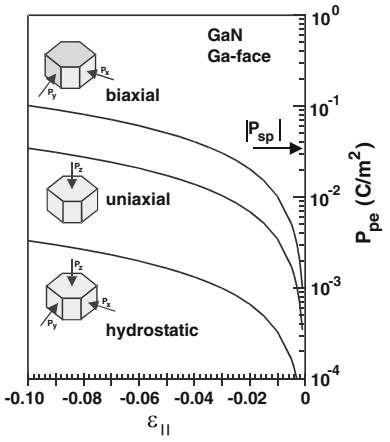


Fig. 15.13 Piezoelectric polarization P_{pe} in GaN (Ga-face) versus in-plane strain $\epsilon_{\|} = (a - a_0)/a_0$ for biaxial, uniaxial and hydrostatic strain. The value of the spontaneous polarization P_{sp} is indicated by an *arrow*. From [1209]

Values for the piezoelectric coefficients of several wurtzite semiconductors[5] are listed in Table 15.1.

The polarization (along c) for biaxial strain in heteroepitaxy (5.73) on the [00.1] surface is

$$P_{pe} = 2\,\epsilon_{\|} \left(e_{31} - \frac{C_{13}}{C_{33}}\, e_{33} \right), \tag{15.18}$$

where $\epsilon_{\|} = (a - a_0)/a_0$ is the in-plane strain. The dependence of the magnitude for GaN on the in-plane strain is shown in Fig. 15.13 together with the polarization for uniaxial stress along [00.1] and hydrostatic strain. In the latter two cases, the polarization is smaller.

The difference of spontaneous polarization of the constituent materials in heterostructures and piezoelectric effects in strained quantum wells lead to quantum confined Stark effect (QCSE, Sect. 13.1.2). The spatial separation of electrons and holes leads to a larger radiative lifetime (Fig. 12.42) and thus reduced radiative recombination rate in the presence of nonradiative channels, an effect unwanted in light emitting diodes. In particular nitride based LEDs grown along the c-direction are prone to this effect. Therefore growth on nonpolar surfaces such as $(1\bar{1}00)$ (m-plane) and $(11\bar{2}0)$ (a-plane) has been investigated [1210,1211]. However, crystal quality for these growth planes seems limited.

An alternative route are 'semipolar' planes, e.g. $(10\bar{1}\bar{1})$ or $(11\bar{2}2)$ [1212], with at least reduced polarization effects. The $(11\bar{2}2)$-plane is tilted about 58° (for GaN) from the c-axis, close to the theoretically predicted zero of the internal electric field for InGaN/GaN quantum wells [1213] (Fig. 15.14). Promising experimental results have been published on semipolar growth, especially for InGaN/GaN on $(11\bar{2}2)$-oriented GaN substrates [1214].

[5] If e_{14} for a zincblende material is transformed to a coordinate system along [111], the wurtzite-like piezoelectric constants are $e_{33} = 2\,e_{14}/\sqrt{3}$ and $e_{31} = -e_{14}/\sqrt{3}$ [1208].

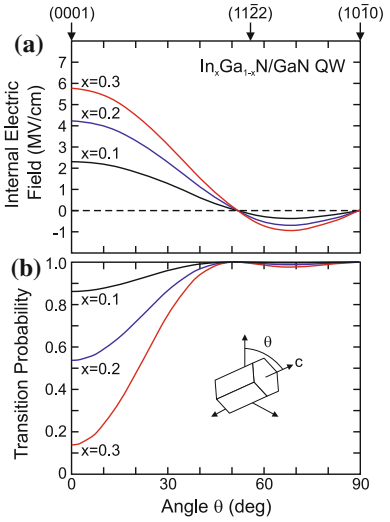


Fig. 15.14 (**a**) Internal electric field and (**b**) transition probability for electron-hole pairs in pseudomorphic $In_xGa_{1-x}N$/GaN quantum well ($L_w = 3\,nm$) for three different indium contents x as labeled. θ denotes the angle between the c-axis and the normal of the interface plane (see *inset*). Based on [1213]

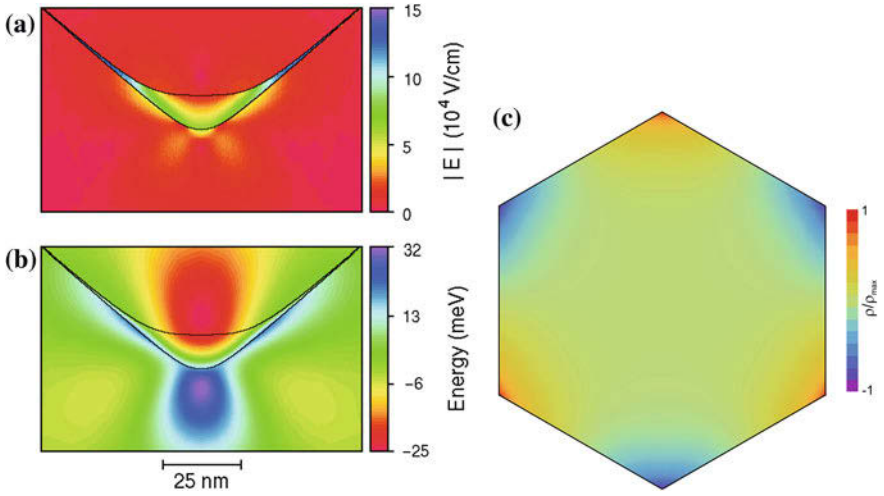


Fig. 15.15 (**a**) Electric field and (**b**) additional confinement potential for electrons due to piezoelectric charges for a strained $In_{0.2}Ga_{0.8}As$/GaAs quantum wire. Adapted from [1203]. (**c**) Piezoelectric charge density in a (111)-oriented GaAs wire with hexagonal cross section. Adapted from [1215]

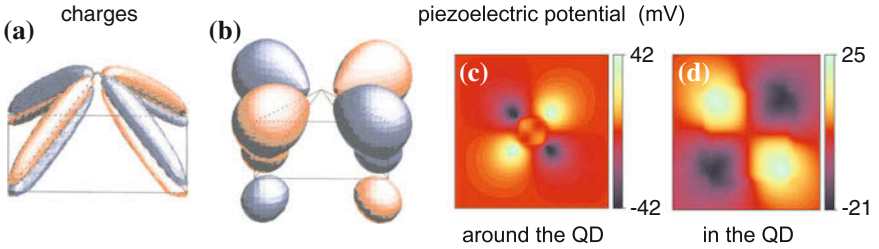


Fig. 15.16 (**a**) Piezoelectric charges and (**b**)–(**d**) resulting Coulomb potential for InAs/GaAs quantum dot with base length $b = 12\,\text{nm}$. (**a**) Isosurfaces corresponding to volume charges $\pm 0.3\,\text{e}\,\text{nm}^{-3}$. (**b**) Isosurfaces for the Coulomb potential at $\pm 30\,\text{meV}$. (**c**, **d**) Cross section through the Coulomb potential somewhat above the wetting layer in two different magnifications, (**d**) is a zoom into (**c**). The InAs/GaAs interface is visible in (**d**) due to the image charge effect. Parts (**a**) and (**b**) reprinted with permission from [367], © 1995 APS

15.4.4 Piezoelectric Effects in Nanostructures

The strain distribution around zincblende strained quantum wires [1203], (111)-oriented wires under torsion [1215] and epitaxial (embedded) quantum dots [367] contains shear components and thus generates piezoelectric fields. In Fig. 15.15a, b, the electric field and potential due to the piezoelectric charges are shown for a strained $In_{0.2}Ga_{0.8}As/GaAs$ quantum wire. In Fig. 15.15c the three-fold symmetric piezoelectric charge density within the cross-section of a (111)-oriented GaAs wire under torsion (due to second order piezoelectricity) is depicted.

In Fig. 15.16, the piezoelectric charges and potential are shown for the quantum dot from Fig. 5.27. The piezoelectric potential has quadrupole character and thus reduces the symmetry of the QD (to C_{2v}) [367].[6] Piezoelectric effects are particularly important in wurtzite nanostructures [1216].

[6]The strain distribution has C_{2v} symmetry for a square-based pyramid for zincblende materials. The energy levels and wavefunctions are more strongly impacted by the piezoelectric effects than by the strain asymmetry [1156, 1157].

Chapter 16
Magnetic Semiconductors

Abstract Materials and properties of two types of semiconductors with spontaneous magnetization or ferromagnetic and paramagnetic properties are explained: Compound materials and diluted magnetic semiconductors. A short introduction to semiconductor spintronics covers the concepts of spin transistor and spin LED.

16.1 Introduction

Magnetic semiconductors exhibit spontaneous magnetic order. Even ferromagnetism, important for spin polarization, as needed in spinelectronics (also called *spintronics*), can occur below the Curie temperature that is characteristic of the material. Magnetic semiconductors can be binary compounds such as EuTe (antiferromagnetic) or EuS (ferromagnetic). Another class of magnetic semiconductors contains paramagnetic ions in doping concentration (typically $< 10^{21}\,\mathrm{cm}^{-1}$) or alloy concentration x (typically $x \geq 0.1\%$). Such materials are termed *diluted magnetic semiconductors* (DMS). The incorporation of the magnetic atoms leads first to conventional alloy effects, such as the modification of the lattice constant, the carrier concentration or the band gap. The status of the field up to the mid-1980s can be found in [1217, 1218], mostly focused on II–VI DMS. A review of work on III–V based materials for spintronics, mostly GaAs:Mn, can be found in [1219]. A 2003 review of wide band gap ferromagnetic semiconductors is given in [1220], a 2014 review of Mn-containing DMS in [1221].

16.2 Magnetic Semiconductors

In a magnetic semiconductor, one sublattice is populated with paramagnetic ions. The first two ferromagnetic semiconductors discovered were $CrBr_3$ [1222] in 1960 and EuO [1223] one year later. Europium monoxide has an ionic $Eu^{2+}O^{2-}$ character, such that the electronic configuration of europium is $[Xe]4f^7 5d^0 6s^0$ and that of oxygen is $1s^2 2s^2 2p^6$. Some properties of europium chalcogenides [1224] are summarized in Table 16.1.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_16

Table 16.1 Material properties of Eu chalcogenides

Material	E_g (eV)	Magnetic order	T_C, T_N (K)
EuO	1.12	FM	69.3
EuS	1.65	FM	16.6
EuSe	1.8	AF	4.6
		FM	2.8
EuTe	2.00	AF	9.6

'FM' ('AF') denotes ferromagnetic (antiferromagnetic) order. T_C (T_N) denotes the Curie (Néel) temperature. Data collected in [1226]

EuO can be modeled as a Heisenberg ferromagnet with dominant nearest- and next-nearest Eu–Eu interactions [1225]. The Heisenberg exchange parameters J_1 and J_2 for these four compounds are shown in Fig. 16.1. In the nearest-neighbor interaction J_1 a 4f electron is excited to the 5d band, experiences an exchange interaction with the 4f spin on a nearest neighbor and returns to the initial state. This mechanism generally leads to ferromagnetic exchange. The next-nearest-neighbor interaction J_2 is weakly ferromagnetic (EuO) or antiferromagnetic (EuS, EuSe, EuTe). In the superexchange process, electrons are transferred from the anionic p states to the 5d states of the Eu^{2+} cations, resulting in an antiferromagnetic coupling.

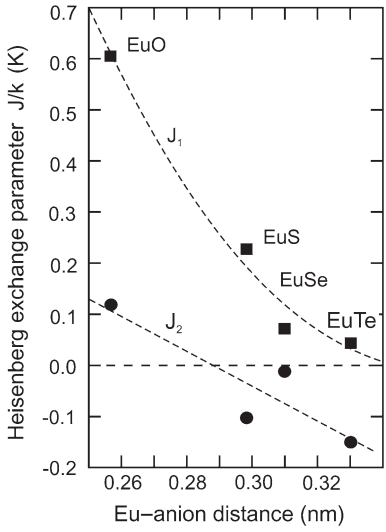


Fig. 16.1 Heisenberg nearest (J_1, *squares*) and next-nearest (J_2, *circles*) exchange parameters (in units of $J_{1,2}/k_B$) for the Eu chalcogenides versus the Eu–anion distance. *Dashed lines* are guides to the eye. Experimental data from [1226]

16.3 Diluted Magnetic Semiconductors

In Table 16.2, the transition metals and their electron configurations are summarized. The 3d transition metals are typically used for magnetic impurities in DMS due to their partially filled 3d shell. Due to Hund's rule, the spins on the 3d shell are filled in parallel for the first five electrons up to half filling (in order to allow the electrons to get out of their way in real space). Thus, the atoms have a sizeable spin and a magnetic moment. The spin of Mn is $S = 5/2$. Most transition metals have a $4s^2$ configuration that makes them isovalent in II–VI compounds. We note that Zn has a complete 3d shell and thus no net spin. In Fig. 16.2, an overview of the crystallographic properties is given for Mn-alloyed II–(Se, S, Te, O) based DMS [1227] (DMS with Se, S, and Te have been discussed in [1228]).

As an example, the properties of $Hg_{1-x}Mn_xTe$ are discussed. This alloy is semiconducting (positive band gap ϵ_0) for $x > 0.075$ and a zero-gap material (negative interaction gap ϵ_0) for smaller Mn concentration (cf. Fig. 6.40). The transitions between the Γ_6 and Γ_8 bands can be determined with magnetoabsorption spectra in the infrared [1229]. In Fig. 16.3a, the magnetic field dependence of transition

Table 16.2 3d, 4d and 5d transition metals and their electron configurations

Sc^{21}	Ti^{22}	V^{23}	Cr^{24}	Mn^{25}	Fe^{26}	Co^{27}	Ni^{28}	Cu^{29}	Zn^{30}
$3d$	$3d^2$	$3d^3$	$3d^5$	$3d^5$	$3d^6$	$3d^7$	$3d^8$	$3d^{10}$	$3d^{10}$
$4s^2$	$4s^2$	$4s^2$	$4s$	$4s^2$	$4s^2$	$4s^2$	$4s^2$	$4s$	$4s^2$
Y^{39}	Zr^{40}	Nb^{41}	Mo^{42}	Tc^{43}	Ru^{44}	Rh^{45}	Pd^{46}	Ag^{47}	Cd^{48}
$4d$	$4d^2$	$4d^4$	$4d^5$	$4d^6$	$4d^7$	$4d^8$	$4d^{10}$	$4d^{10}$	$4d^{10}$
$5s^2$	$5s^2$	$5s$	$5s$	$5s$	$5s$	$5s$	–	$5s$	$5s^2$
La^{57}	Hf^{72}	Ta^{73}	W^{74}	Re^{75}	Os^{76}	Ir^{77}	Pt^{78}	Au^{79}	Hg^{80}
$5d$	$5d^2$	$5d^3$	$5d^4$	$5d^5$	$5d^6$	$5d^7$	$5d^9$	$5d^{10}$	$5d^{10}$
$6s^2$	$6s^2$	$6s^2$	$6s^2$	$6s^2$	$6s^2$	$6s^2$	$6s$	$6s$	$6s^2$

Note that Hf^{72} has an incompletely filled 4f-shell with $4f^{14}$

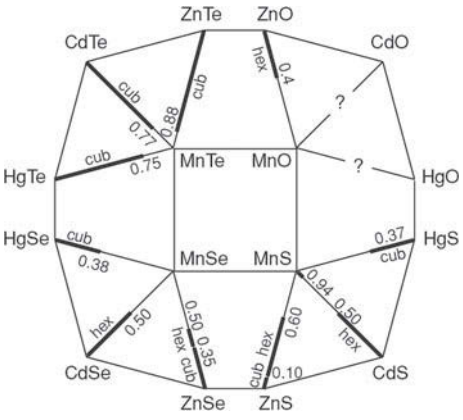


Fig. 16.2 Diagrammatic overview of $A^{II}_{1-x}Mn_xB^{VI}$ alloys and their crystal structures. The *bold lines* indicate ranges of the molar fraction x for which homogeneous crystal phases form. 'Hex' and 'Cub' indicate wurtzite and zincblende, respectively. From [1227]

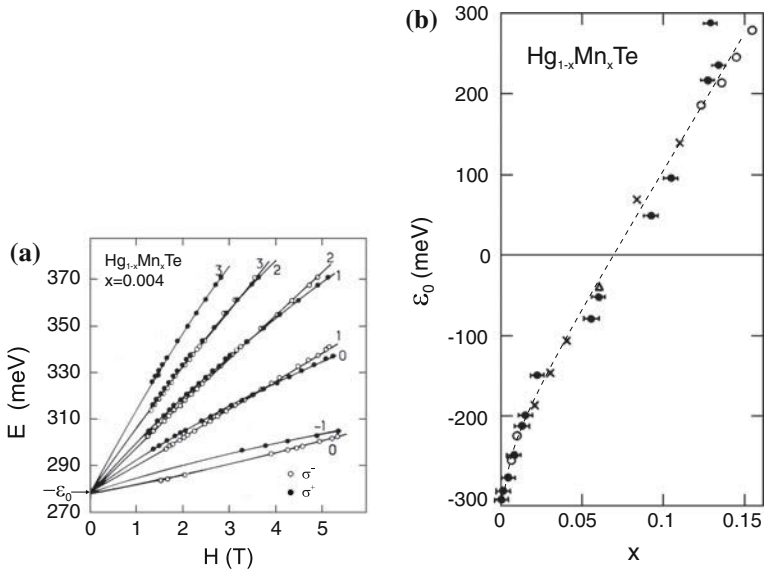


Fig. 16.3 (**a**) Energies of $\Gamma_6 \rightarrow \Gamma_8$ transitions versus magnetic field for $Hg_{0.996}Mn_{0.004}Te$ at $T = 2\,K$. *Symbols* are experimental values for two polarization directions as indicated. Numbers denote quantum numbers of transitions. *Solid lines* are theoretical fits. (**b**) Interaction gap versus Mn concentration for $Hg_{1-x}Mn_xTe$ at $T = 4.2\,K$. Various *symbols* represent data from different authors and methods. *Dashed line* is a guide to the eye. Adapted from [1229]

energies between different Landau levels is shown that can be extrapolated to yield the interaction gap. The interaction gap is shown in Fig. 16.3b as a function of the Mn concentration.

For small Mn concentrations, the DMS behaves like a paramagnetic material. For larger concentrations, the Mn atoms have increasing probability to be directly neighbored by another Mn atom and suffer superexchange interaction (cf. (3.23b)). At a certain critical concentration x_c, the cluster size becomes comparable with the size of the sample. If interaction up to the first, second or third neighbor are taken into account for a fcc lattice, the critical concentrations are given by $x_c = 0.195$, 0.136, and 0.061, respectively [1230]. The nearest-neighbor interaction between Mn atoms in such DMS as (Zn,Cd,Hg)Mn(S,Se,Te) was found to be antiferromagnetic,[1] i.e. neighboring spins are aligned antiparallel. Due to frustration of antiferromagnetic long-range order on a fcc lattice, an antiferromagnetic spin glass forms. The transition temperature T_C between the paramagnetic and spin-glass phases of $Hg_{1-x}Mn_xTe$ is shown in Fig. 16.4.

In III–V compounds, the 3d transition metals represent an acceptor if incorporated on the site of the group-III element as, e.g., in the much investigated compound $Ga_{1-x}Mn_xAs$. This material will be used in the following to discuss some prop-

[1] Such superexchange leads to antiferromagnetic interaction if the bond angle is 'close' to 180°.

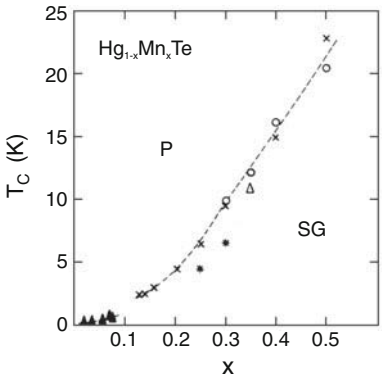


Fig. 16.4 Magnetic phase diagram of $Hg_{1-x}Mn_xTe$, 'P' ('SG') denotes the paramagnetic (spin glass) phase. Various *symbols* represent data from different authors and methods. *Dashed line* is a guide to the eye. Adapted from [1229]

erties of magnetic semiconductors. It seems currently well understood and has a fairly high Curie temperature of $T_C \approx 160\,K$. Ferromagnetism in a diluted magnetic semiconductor is believed to be caused by indirect exchange through itinerant charge carriers. The ferromagnetic coupling can be invoked by the Ruderman–Kittel–Kasuya–Yoshida (RKKY) interaction, i.e. the spins of the paramagnetic ions are aligned via interaction with the free carriers in the semiconductor. A related concept is the double exchange[2] [1231–1233] in which carriers move in a narrow Mn-derived d-band (for d-wave character see Fig. 7.16c). Such a mechanism was first invoked for PbSnMnTe [1234]. Later, ferromagnetism was discovered in InMnAs [1235] and GaMnAs [1236]. In (In,Ga)MnAs a Mn ion (spin up) spin polarizes the surrounding hole gas (spin down), which has been supplied from the Mn acceptors. This mechanism lowers the energy of the coupled system. The interaction

$$H = -\beta\, N_0\, x\, S\, s \tag{16.1}$$

between the Mn d-shell electrons ($S = 5/2$) and the p-like free holes ($s = 1/2$) is facilitated by p–d hybridization of the Mn states. N_0 denotes the concentration of cation sites in the $A_{1-x}Mn_xB$ alloys. The coupling via electrons is much weaker (coupling coefficient α). The holes interact with the next Mn ion and polarize it (spin up), thus leading to ferromagnetic order. The ferromagnetic properties are evident from the hysteresis shown in Fig. 16.6a. Without the carrier gas such interaction is not present and the material is only paramagnetic. Theoretical results for the Curie temperature of various p-type semiconductors are shown in Fig. 16.5. Generally, the quest for higher Curie temperatures (well above room temperature) is underway and wide band gap materials such as GaN or ZnO doped with transition metals have shown some encouraging results. Mn-substituted chalcopyrite semiconductors are analyzed theoretically in [1237] and are predicted to exhibit less-stable ferromagnetism than III–V semiconductors of comparable band gap.

[2] This model is also called the Zener model.

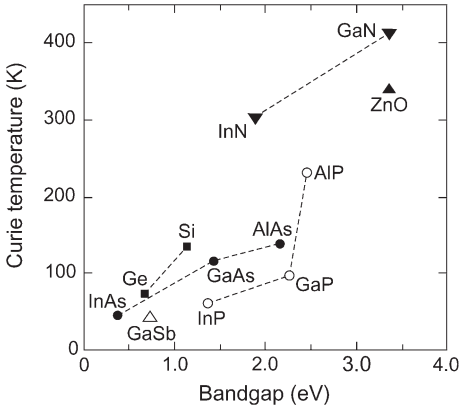


Fig. 16.5 Computed values of the Curie temperature T_C for various p-type semiconductors plotted versus the band gap (*dashed lines* are guides to the eye). All materials contain 5 % Mn on the cation sublattice and a hole concentration of $p = 3.5 \times 10^{20}\,\mathrm{cm}^{-3}$. Values for T_C taken from [1233]

The carrier density and thus magnetic properties in a DMS can be controlled in a space-charge region (cf. Sect. 21.2.2) as demonstrated in [1238]. In Fig. 16.6, results are shown for hydrogen- (deuterium-) passivated GaMnAs that exhibits ferromagnetism as 'as-grown' thin film. The deuterium is incorporated in similar concentration as the Mn, assumes a back-bond position (forming a H–As–Mn complex) and compensates the hole gas from the Mn (cf. Sect. 7.8). The low-temperature conductivity drops nine orders of magnitude [1239]. Such material displays only paramagnetic behavior. An optimal Mn concentration for ferromagnetic $Ga_{1-x}Mn_xAs$ is around $x = 0.05$. For smaller Mn concentrations, the hole density is too small and the Curie

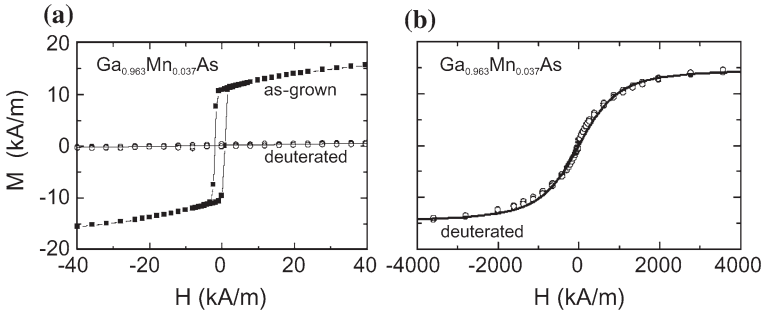


Fig. 16.6 Magnetization M versus magnetic field H for $Ga_{0.963}Mn_{0.037}As$ at low temperature. (**a**) Comparison of as-grown (*full squares*) and deuterated (*open circles*) thin film with magnetic field in the layer plane at $T = 20\,\mathrm{K}$. (**b**) Magnetization of the deuterated sample at $T = 2\,\mathrm{K}$ for larger magnetic fields. *Solid line* is Brillouin function for $g = 2$ and $S = 5/2$. Adapted from [1239]

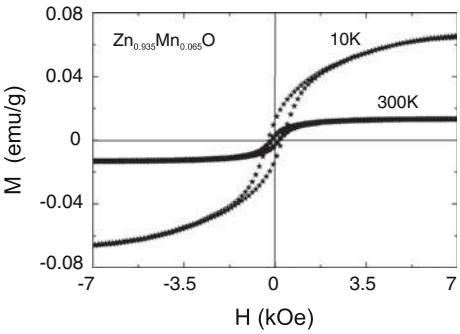


Fig. 16.7 Magnetization M versus magnetic field H for $Zn_{0.935}Mn_{0.065}O$ thin film at $T = 10$ and 300 K. A hysteresis is obvious for both temperatures

temperature drops; for larger Mn concentrations, the structural properties of the alloy degrade (phase separation into GaAs and MnAs).[3]

Magnetic hysteresis has been found in nearly compensated Mn-doped ZnO [1240, 1241] (Fig. 16.7). Such material is interesting due to its small spin-orbit coupling. The exchange mechanism is under debate.

16.4 Spintronics

Spintronics (as opposed to electronics) is an emerging field that uses the electron *spin* rather than its *charge* for transport, processing and storage of information. Prototype devices are the spin transistor and the spin LED. A crucial point is spin injection, i.e. the creation of (highly) spin-polarized currents. It remains to be seen whether spintronics can be developed to its theoretically envisioned potential and will play a commercially important role in the course of microelectronics. The spin degree of freedom also promises potential for quantum information processing due to its weak coupling to charge and phonons and the resulting long dephasing time.

It shall be mentioned here that the magnetization alters also the 'classical' transport properties of the semiconductor. The Hall effect is modified strongly and reflects the magnetic hysteresis, also termed *anomalous* Hall effect (AHE).[4] This effect was already discovered by Hall in 1881 for Ni and Co [1242]; a review of the AHE can be found in [1243]. Also the optical excitation of spin-polarized carriers in non-magnetic materials causes AHE [1244].

[3] MnAs is a ferromagnetic metal. MnAs clusters can be a problem since they create ferromagnetic properties but not in the way the DMS is supposed to work.

[4] This use of the term 'anomalous' should be distinguished from the historical use of 'anomalous' Hall effect for the sign reversal of the Hall voltage for hole conductors.

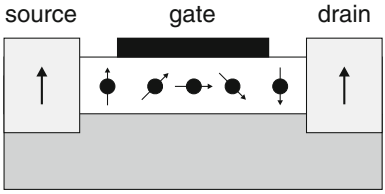


Fig. 16.8 Scheme of spin transistor after the proposal of [1247]. Source and drain are ferromagnets with their magnetization shown schematically as arrows. The channel under the gate transports electrons whose spin rotates in the electric field under the gate

16.4.1 Spin Transistor

In this device (for regular transistors cf. Chap. 24), spin-polarized electrons are injected from contact 1, transported through a channel and detected in contact 2. During the transport, the spin rotates (optimally by π) such that the electrons cannot enter contact 2 that has the same magnetization as contact 1 (Fig. 16.8). The spin rotation is caused by spin-orbit interaction due to the electric field under the gate contact. This effect is called the Rashba effect and is purely relativistic [1245]. As channel material, a semiconductor with strong spin-orbit coupling such as InAs or (In,Ga)Sb is preferable. However, the use of narrow-gap semiconductors and the increase of spin scattering at elevated temperatures [1246] make the realization of such a transistor at room temperature difficult.

16.4.2 Spin LED

In a spin LED (for LEDs see Sect. 23.3), the injection of spin-polarized carriers into the active layer leads to circularly polarized luminescence. The spin alignment can be achieved with semimagnetic semiconductors grown on top of the active layer or via spin injection from a ferromagnetic metal into the semiconductor (for metal–semiconductor junctions cf. Sect. 21.2). In Fig. 16.9a, a Fe/AlGaAs interface is shown.

Ideally, the spin-polarized electrons from the ferromagnetic metal tunnel into the semiconductor and transfer to the recombination region. Subsequently, the emission is circularly polarized (Fig. 12.29b). The degree of circular polarization is

$$P_\sigma = \frac{I_{\sigma+} - I_{\sigma-}}{I_{\sigma+} + I_{\sigma-}}, \tag{16.2}$$

with $I_{\sigma\pm}$ being the intensity of the respective polarization. The degree of polarization depends on the magnetization of the metal. For the saturation magnetization of Fe, the maximum polarization is about 30 % at $T = 4.5\,\mathrm{K}$ (Fig. 16.9b) [1248]. The interface

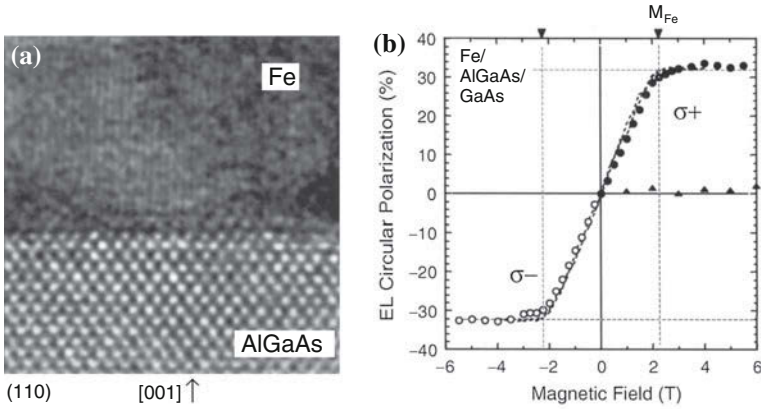


Fig. 16.9 (**a**) Transmission electron microscopy image of the (110) cross section of the Fe/AlGaAs interface of a spin LED. The *vertical lines* in Fe are the (110) planes with 0.203 nm distance. (**b**) Magnetic-field dependence of the circular polarization ratio P_σ at $T = 4.5\,\mathrm{K}$ (16.2) (*filled* and *empty* circles) and the out-of-plane component of the Fe-film magnetization (*dashed line*, scaled to the maximum of P_σ). Reproduced from [1249] by permission of the MRS Bulletin

and its structural nonideality of the interface presumably prevent the spin injection from being 100 % efficient [1249].

Chapter 17
Organic Semiconductors

Abstract An introduction to organic semiconductors, based on the sp^2 bonding scheme is given, covering small organic molecules and polymers. Their electronic structure, doping schemes, as well as their particular transport and optical properties are covered.

17.1 Introduction

Organic semiconductors are based on carbon compounds. The main structural difference from inorganic semiconductors is the bond based on sp^2 hybridization (cf. Sect. 2.2.3) as present in benzene (and graphite). Diamond, although consisting of 100 % carbon, is not considered an organic semiconductor. We note that carbon can form further interesting structures based on sp^2 bonds, such as carbon nanotubes (Sect. 18.2), (single or few layer) graphene sheets (Sect. 18.1) rolled up to form cylinders, or fullerenes, e.g. soccer-ball-like molecules such as C_{60}.

In the 1980 Handbook on Semiconductors only a good ten pages were devoted to organic semiconductors [1250]. Now several textbooks are available [1251, 1252] for a much more detailed treatment than given here.

17.2 Materials

17.2.1 Small Organic Molecules, Polymers

The prototype organic molecule is the benzene molecule with its ring-like structure (Fig. 2.8).

There is a large number of organic, semiconducting molecules that differs by the number of benzene rings (Fig. 17.1), the substitution of carbon atoms by nitrogen or sulfur (Fig. 17.2a, b), the polymerization (Fig. 17.2c) or the substitution of hydrogen atoms by side groups (Fig. 17.2d). Since PPV is insoluble, typically derivatives

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_17

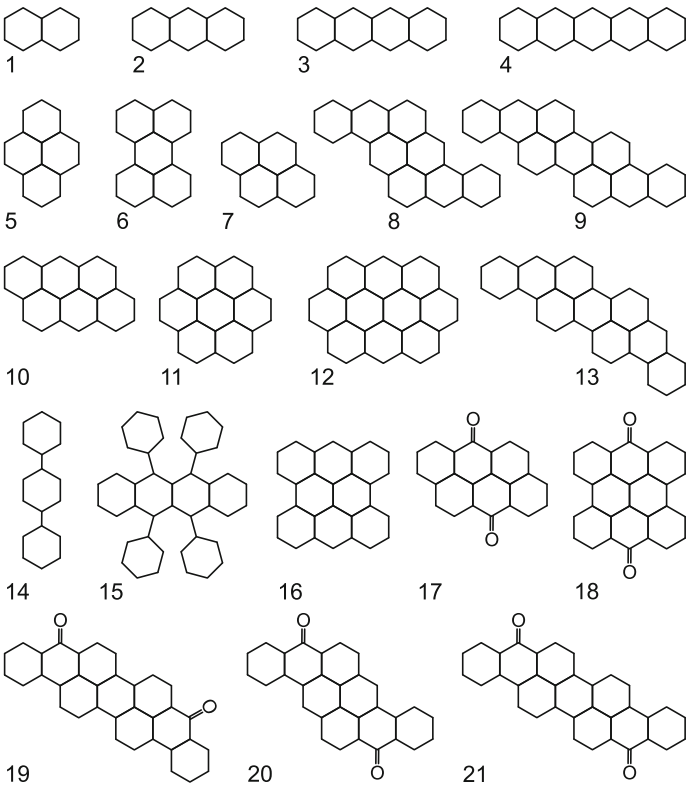


Fig. 17.1 Various organic compounds: 1: naphtalene, 2: anthracene, 3: tetracene, 4: pentacene, 5: pyrene, 6: perylene, 7: chrysene, 8: pyranthrene, 9: isoviolanthrene, 10: anthanthrene, 11: coronene, 12: ovalene, 13: violanthrene, 14: p-terphenyl, 15: rubrene, 16: m-dinaphthanthrene, 17: anthanthrone, 18: m-dinaphthanthrone, 19: violanthrone, 20: pyranthrone, 21: isoviolanthrone

such as MEH-PPV[1] [1254] that are soluble in organic solvents are used. Compared to benzene, the substitution of one carbon atom by nitrogen (pyridine) represents doping with one electron. In Fig. 17.3, the most important building blocks of organic molecules are shown.

17.2.2 Organic Semiconductor Crystals

Small organic molecules can crystallize into solids, so-called *organic molecular crystals* (OMC), due to van-der-Waals interaction. In Fig. 17.4a, the monoclinic unit cell of an anthracene crystal [1255] is shown as an example. Also tetracene and

[1] 2-ethoxy,5-(2'-ethyl-hexyloxy)-1,4-phenylene vinylene.

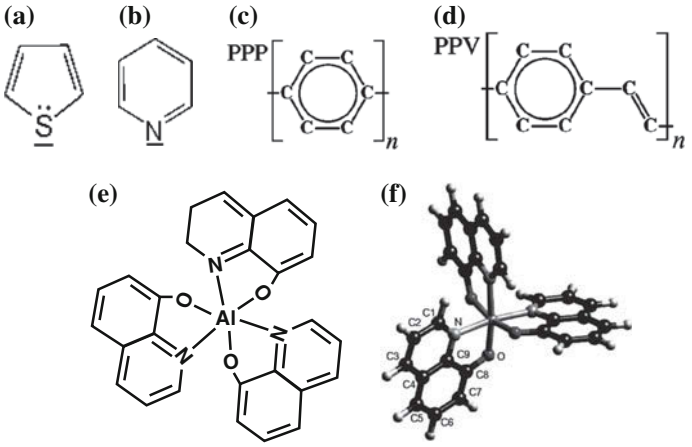


Fig. 17.2 Organic compounds: (**a**) thiophene, (**b**) pyridine, (**c**) poly-(p-phenyl), (**d**) poly-(p-phenylvinyl), (**e**) Alq_3 (tris-(8-hydroxyquinolate)-aluminum) and (**f**) a three-dimensional view of the Alq_3 molecule. Part (**f**) reprinted with permission from [1253], © 1998 AIP

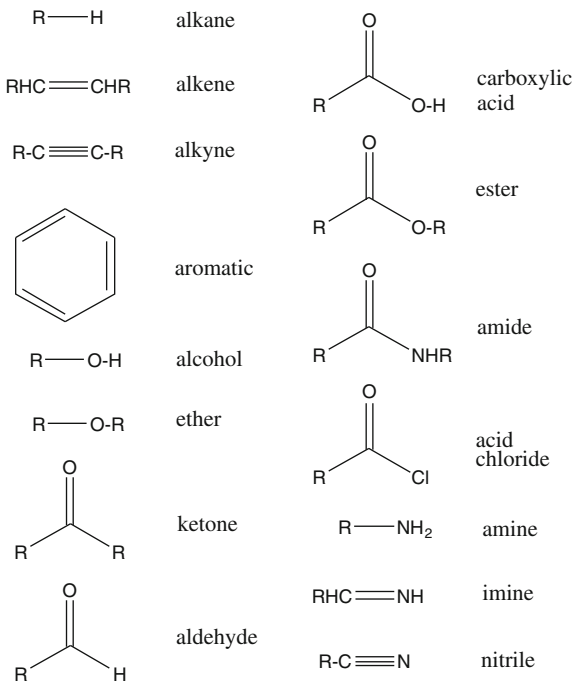


Fig. 17.3 Building blocks of organic molecules, 'R'=alkyl group, i.e. CH_3 (methyl-), CH_3CH_2 (buthyl-), . . .

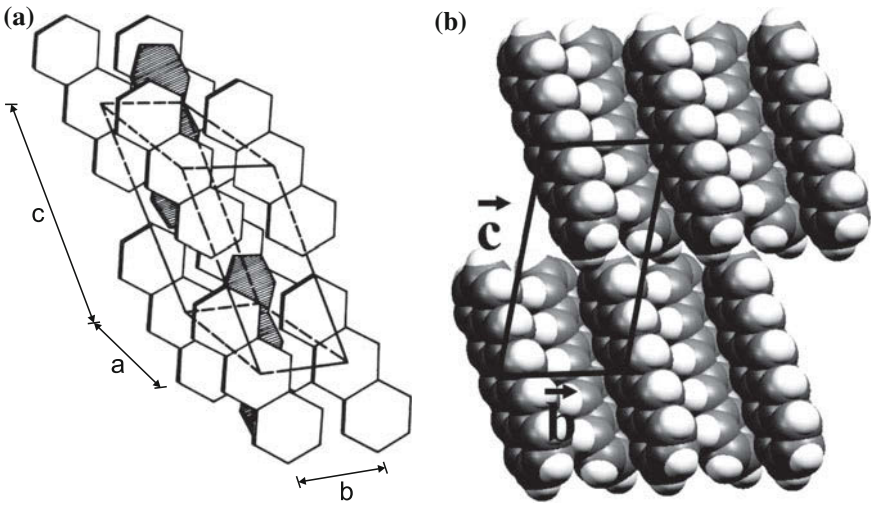


Fig. 17.4 (**a**) Monoclinic unit cell (for size see Table 17.1) of anthracene crystal. (**b**) Two herringbone layers of pentacene in a projection onto the bc plane of the triclinic unit cell. Adapted from [1263]

Table 17.1 Properties of oligoacene crystals

Property	Naphtalene	Anthracene	Tetracene	Pentacene
Melting point (°C)	80	217	357	>300 °C
Crystal system	Monoclinic	Monoclinic	Triclinic	Triclinic
a (nm)	0.824	0.856	0.798	0.793
b (nm)	0.600	0.604	0.614	0.614
c (nm)	0.866	1.116	1.357	1.603
α (°)	90	90	101.3	101.9
β (°)	122.9	124.7	113.2	112.6
γ (°)	90	90	87.5	85.8

Melting point and unit cell parameters. Data from [1256]

pentacene (Fig. 17.4b) have this layered 'herringbone' structure. A comparison of the unit cells of oligoacene crystals is given in Table 17.1.

Crystal growth of single crystal OMC is achieved with a variety of methods, among them sublimation, Bridgman- and Czochralski-type methods [1257, 1258], vapor phase growth [1259, 1260] or from solution [1261, 1262]. Single organic molecular crystals exhibit intrinsic material properties. The practical use of organic semiconductors involves thin films, e.g. in LEDs (OLED, Sect. 23.3.7) and transistors (OFET, Sect. 24.6.3). Thin films of organic molecules are typically disordered and their performance parameters are inferior to that of OMCs.

17.3 Electronic Structure

The p_z orbitals in benzene are partially filled and there is an energy gap between HOMO and LUMO (Fig. 2.11). A similar consideration is valid for polymers. The coupling of orbitals along the polymer chain leads to broadening of the π and π^* states into a (filled) valence and an (empty) conduction band, respectively (Fig. 17.5). The HOMO and LUMO positions of various organic semiconductors are shown in Fig. 17.6 relative to the vacuum level (cmp. Fig. 12.21 for inorganic semiconductors). The HOMO is also known as *ionization energy* (IE), the LUMO as *electron affinity* (EA). With layered organic semiconductors heterostructures can be built, e.g. to design recombination pathways (recombination layer, electron blocking layers (EBL) and hole blocking layers (HBL)). For electron injection and electron extraction (hole injection) contacts, metals with appropriate work function (in connection with a possible interface dipole layer) have to be used. More details on injection and extraction contacts to organic semiconductors are discussed in Sect. 21.2.7.

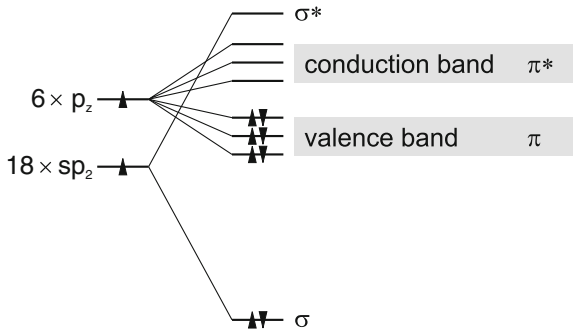


Fig. 17.5 Schematic band structure of a polymer originating from the states of the benzene molecule (see Fig. 2.11)

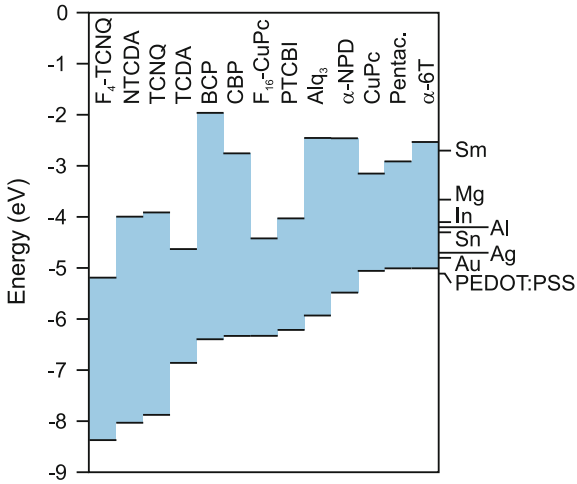


Fig. 17.6 Position of HOMO and LUMO for a variety of organic semiconductors (relative to a common vacuum level at $E = 0\,\text{eV}$). Based on values from [1264]. On the right hand side the work functions of several metals are shown for comparison

17.4 Doping

The doping of organic semiconductors can be achieved by

- partial oxidation or reduction of the organic molecule,
- substitution of atoms in the organic molecule,
- mixing of the matrix with 'dopant' molecules.

The systematic shift of the Fermi level with dopant concentration has been reported in [1265]. Typically, the conductivity increases superlinearly with the doping concentration (Fig. 17.7a), an effect discussed in detail in [1266]. While the mobility remains constant, the thermal activation energy E_a for carriers decreases with increasing doping (Fig. 17.7b) due to electrostatic interaction [515, 1267], an effect already discussed in Sect. 7.5.7. The activation energy in the dilute limit $E_{a,0}$ is modified to (cmp. (7.53))

$$E_a = E_{a,0} - \beta N_D^{1/3} \,. \tag{17.1}$$

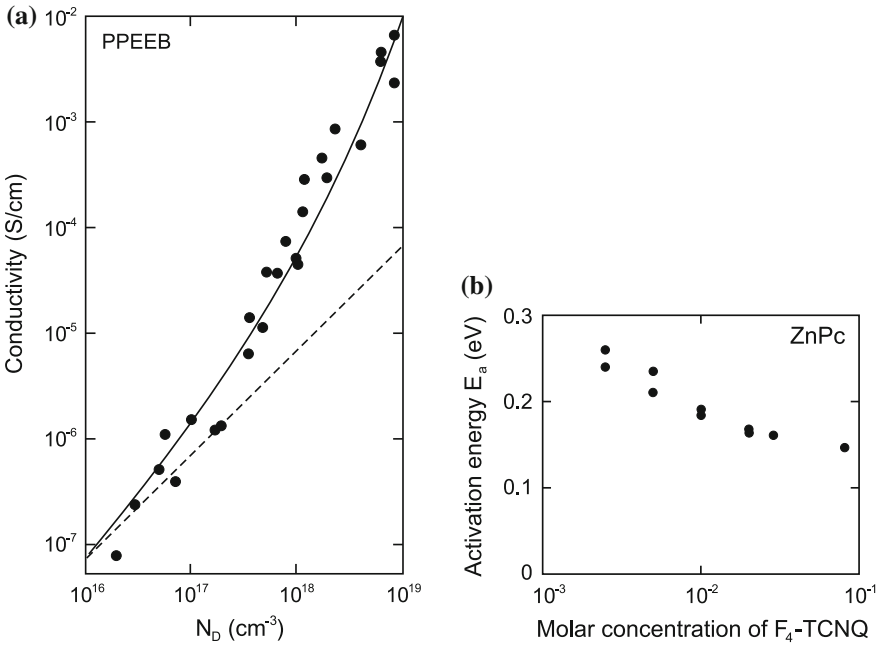


Fig. 17.7 (**a**) Conductivity (at 0.9 V/μm) versus dopant concentration N_D for PPEEB films. Experimental data (*circles*) and fit (*solid line*) according to (17.1) with activation energy $E_{a,0} = 0.23\,\text{eV}$ and $\beta = 6.5 \times 10^{-8}\,\text{eV cm}$ ($\mu = 0.2\,\text{cm}^2/\text{Vs}$). *Dashed line* denotes linear relation between conductivity and N_D. Adapted from [1268]. (**b**) Thermal activation energy E_a of carriers (holes) in ZnPc:F_4-TCNQ as a function of the molar dopant concentration. Adapted from [1265]

17.5 Transport Properties

Transport in organic semiconductors is characterized by

- strong polaronic effects,
- hopping conduction,
- low mobility, low drift saturation velocity.

The interaction of charges with lattice deformations leads to the formation of polarons [1269]. In organic materials these are often 'small', i.e. the extension of the deformation is on atomic scale. Such self-trapping of charges reduces their mobility. Two charges can share the same deformation (bipolaron) or oppositely charged polarons can attract (similar to an exciton). If these charges are on the same (neighboring) polymer chain, the polaron is called intrachain (interchain).

The conductivity within a molecule, e.g. a long polymer chain, and the conductivity between different molecules have to be distinguished. The conduction between different molecules occurs via *hopping*. Typically, the conductivity is thermally activated according to

$$\sigma = \sigma_0 \exp\left(-\frac{E_a}{kT}\right), \tag{17.2}$$

where E_a is an energy of the order of 1 eV. Such activation also pertains to the mobility alone, e.g. $E_a = 0.48$ eV for PPV [1270].

The maximum low-field mobility of many crystalline organic semiconductors at room temperature is around 1 cm^2/Vs with a weak temperature dependence [1271]. Such mobility is much smaller than that of crystalline silicon and rather comparable to that of amorphous silicon. Improved purity and handling of organic semiconductors has allowed to achieve intrinsic material properties (Fig. 17.8). The mobility increases

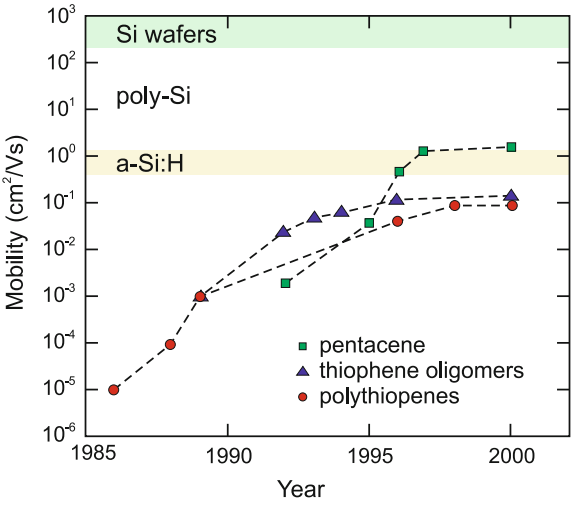


Fig. 17.8 Historic development of the experimentally achieved mobility of organic semiconductors at room temperature

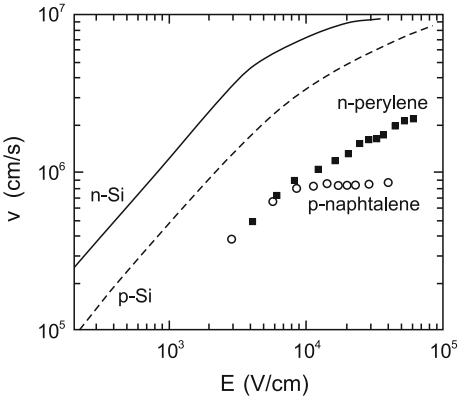


Fig. 17.9 Carrier velocity in ultrapure and highly perfect single crystals of (n-conducting) perylene at $T = 30\,\mathrm{K}$ and (p-conducting) naphtalene at $T = 4.3\,\mathrm{K}$. For comparison electron (hole) velocity in silicon at room temperature is shown as *solid* (*dashed*) line. Adapted from [1258]

at low temperatures, e.g. below 100 K in naphtalene [1272]. This has been attributed to the freeze-out of phonons and the transition from hopping to band transport. The drift velocity at higher fields shows saturation but the values, even at low temperature, are much smaller than in silicon (Fig. 17.9). An analytical model for the described main features of transport in organic semiconductors has been given in [1273].

17.6 Optical Properties

Organic molecules can emit light efficiently and are thus useful for light emitters. For the photo-physics of organic materials it is essential to recall the molecular physics of singlet and triplet states. In the singlet (triplet) state, the total spin quantum number of the unpaired electrons is $S = 0$ ($S = 1$). A simple energy scheme includes a ground state S_0 and excited singlet (S_1) and triplet (T_1) states. The recombination transition $S_1 \rightarrow S_0$ is allowed and its lifetime short. Such luminescence is termend 'fluorescence'. Recombination from the triplet state is forbidden or at least very slow ('phosphorescence').

As an example for a small organic molecule, the fluorescence lifetime of Alq_3 is about 12 ns [1274]. The triplet lifetime is in the 10 µs range [1275]. Luminescence and absorption spectra of Alq_3 are shown in Fig. 17.10. The luminescence peak is redshifted with respect to the absorption edge due to the Frank-Condon principle (Fig. 10.21). The density of excited (empty) states of the Alq_3 molecule is shown in Fig. 17.11 together with the orbitals associated with the four prominent states. The lowest orbital is the LUMO and leads to the visible luminescence of the Alq_3 in the red.

In Fig. 17.12a, the photoluminescence (PL) and absorption of a polymer, polythiophene are shown. The recombination is below the band gap of 2.1–2.3 eV on an excitonic level at 1.95 eV. There are several phonon replica whose separation of 180 meV corresponds to the C–C stretching mode. The PL excitation (PLE) spectrum

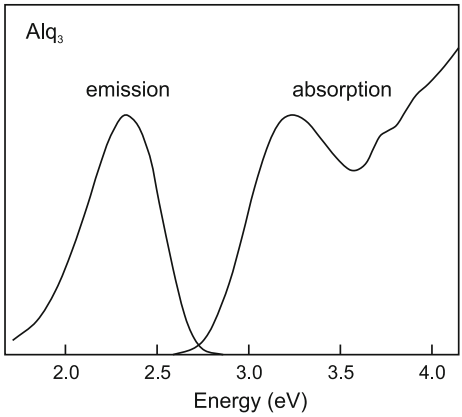


Fig. 17.10 Luminescence and absorption spectra of Alq_3 (vapor-deposited 150 nm thin film on a quartz substrate) at room temperature. Adapted from [1276]

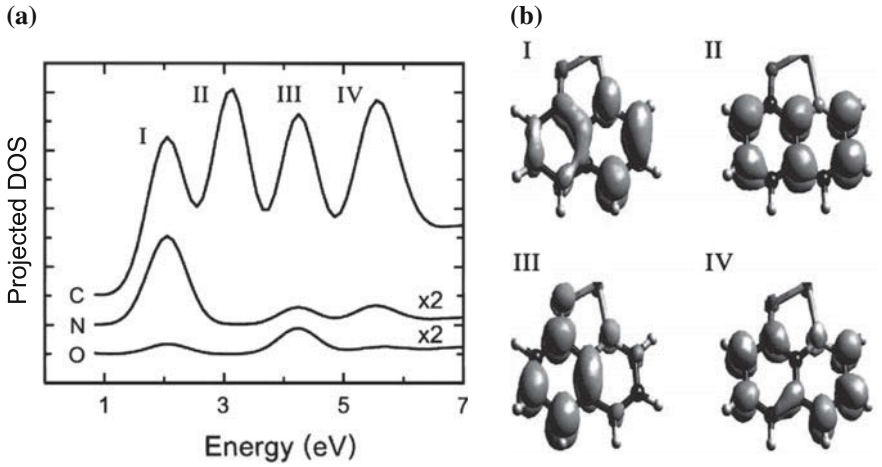


Fig. 17.11 (**a**) Projected density of states (for C, N, and O) of excited states in an Alq_3 molecule. The origin of the energy axis is the HOMO level. (**b**) Orbitals for the four states labeled I–IV in (**a**). Reprinted with permission from [1253], © 1998 AIP

of poly-thiophene demonstrates that the PL at 1.83 eV can be excited via the exciton level. The theoretical band structure of poly-thiophene is shown in Fig. 17.13a. The Brillouin zone is one-dimensional. The situation I corresponds to a single molecular chain, the situation II pertains to the chain embedded in a medium with a dielectric constant $\epsilon = 3$. The predicted band gaps are 3.6 and 2.5 eV, respectively. The exciton binding energy is about 0.5 eV. The exciton is a Frenkel exciton that has a small extension and is localized. The high binding energy is favorable for radiative recombination since the exciton is stable at room temperature. For photovoltaic applications,

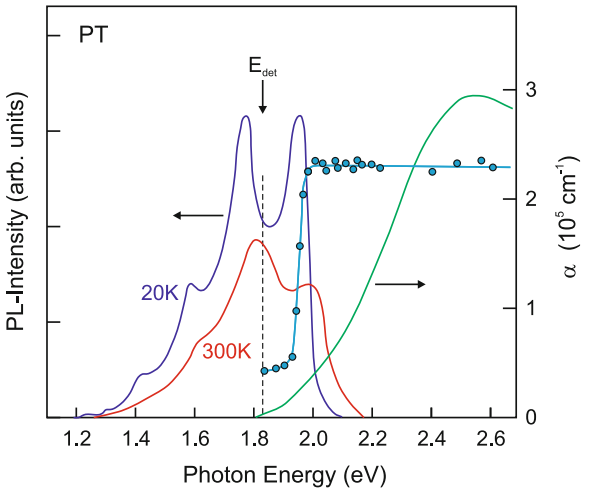


Fig. 17.12 Photoluminescence (PL) spectra at $T = 20$ and 300 K and absorption spectrum (*green line*) of poly-thiophene. The *vertical dashed line* denotes the detection energy ($E_{\mathrm{det}} = 1.83$ eV) of the PL excitation (PLE) spectrum (*blue circles*) ($T = 20$ K). Adapted from [1277]

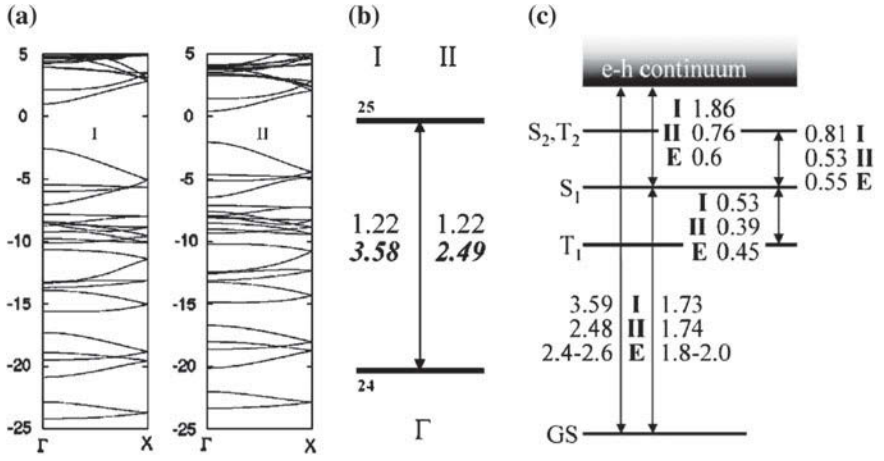


Fig. 17.13 (**a**) Band structure of poly-thiophene ('I': naked chain, 'II': chain in a dielectric medium ($\epsilon = 3$)), (**b**) single-particle energies and band gap, (**c**) exciton levels ('E': experimental values). Reprinted with permission from [1278], © 2002 APS

it is unfavorable since it has to be overcome in order to separate electrons and holes (after absorption). Generally, intrachain excitons (as here) and interchain excitons, where electron and the hole are localized on different chains, are distinguished.

Collection of carriers in the 'dark' triplet state poses a problem limiting the quantum efficiency to 25 % in a simple model [1279]. Harvesting luminescence from *all* exciton states would yield significantly higher efficiencies than is possible in purely fluorescent materials (or devices).

A successful route is the use of a phosphorescent guest material. Radiative transitions from triplet states become partially allowed when the excited singlet and triplet states are mixed. This is typically achieved in metalorganic molecules with heavy metal atoms, providing large spin-orbit interaction effects [1280, 1281]. Most prominently Ru-, Pt- and Ir-containing compounds are used, e.g. *fac* tris(2-phenylpyridine) iridium [$Ir(ppy)_3$] in 4,4'-N,N'-dicarbazole-biphenyl [CBP] (Fig. 17.14a, b) [1282]. The luminescence spectrum of $Ir(ppy)_3$ is shown in Fig. 17.14d. The radiative decay constant of the $Ir(ppy)_3$ triplet state is about 800 ns and observable if energy transfer from the host triplet state is exotherm ($\Delta G = G_G - G_H < 0$ [1275], see Fig. 17.15a) and fast. This is the case for CBP:$Ir(ppy)_3$ (Fig. 17.15b), $\Delta G \approx -0.2$ eV. Actually, reverse transfer from $Ir(ppy)_3$ to CBP seems responsible for some loss in luminescence efficiency and the decrease in phosphorescent lifetime from 800 to 400 ns. In the case of N,N'-diphenyl-N,N'-bis(3-methylphenyl)-[1,1'-biphenyl]-4,4'-diamine

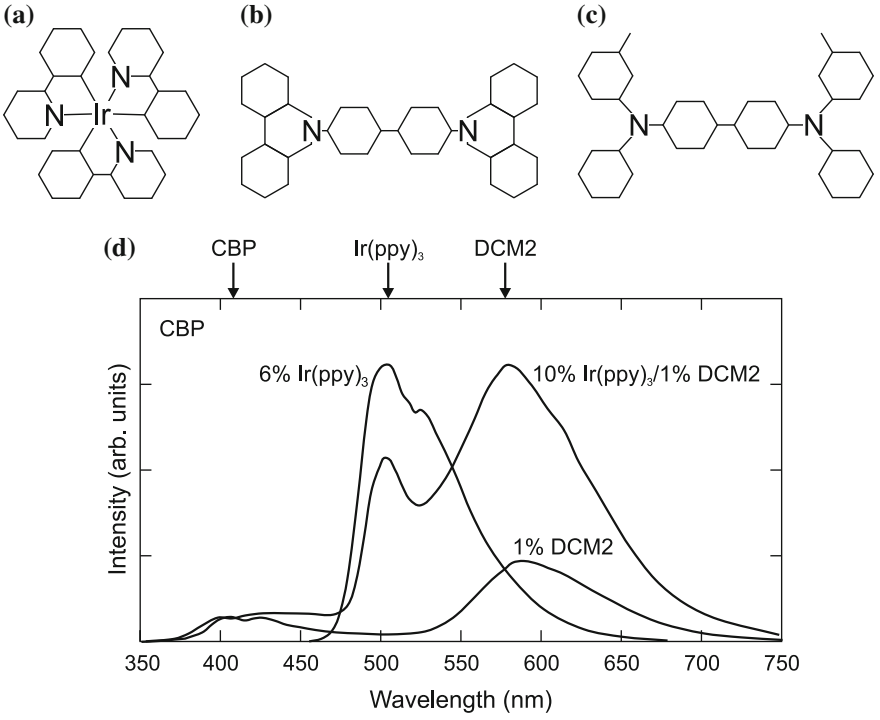


Fig. 17.14 Molecular structure of (**a**) $Ir(ppy)_3$, (**b**) CBP and (**c**) TDP (see text). (**d**) Electroluminescence spectra (at room temperature) of CBP:6 % $Ir(ppy)_3$, CBP:10 % $Ir(ppy)_3$/1 %DCM2 and CBP:2 %DCM2. Based on data from [1282, 1283]

[TDP] (Fig. 17.14c) host, the triplet energy transfer to the phosphorent $Ir(ppy)_3$ guest is endotherm ($\Delta G \approx +0.1\,eV$) and represents the rate limiting step [1275]. In this case the recombination of $Ir(ppy)_3$ has a decay constant of about 15 µs (Fig. 17.15b). The thermal activation character is confirmed by even longer decay times at low

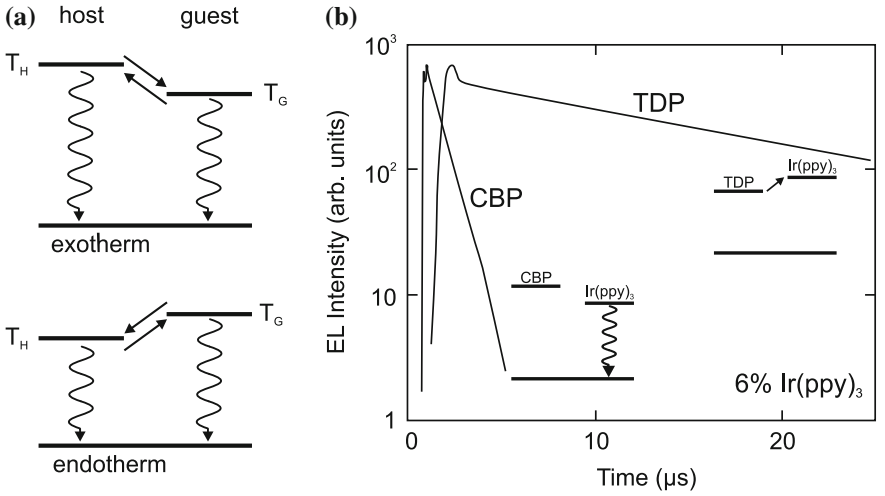


Fig. 17.15 (**a**) Schematic term diagram of host (T_H) and guest (T_G) triplet states. *Straight arrows* denote energy transfer between triplet states, *wiggly arrows* denote radiative transitions to the (singlet) ground state. (**b**) Electroluminescence transients (at room temperature, detected in the range 500–560 nm [cmp. Fig. 17.14b]) of CBP:6 % $Ir(ppy)_3$ ($\tau \approx 1\,\mu s$) and TDP:6 % $Ir(ppy)_3$ ($\tau \approx 15\,\mu s$). *Insets*: term schemes with arrow denoting the rate limiting step. Based on data from [1275]

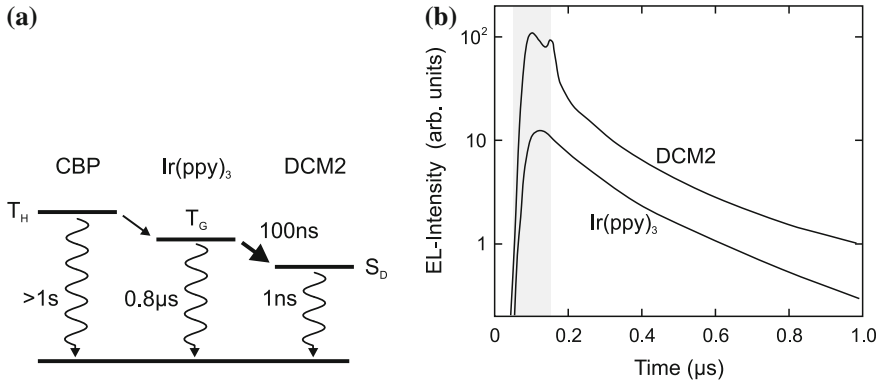


Fig. 17.16 (**a**) Schematic term scheme of CBP:$Ir(ppy)_3$/DCM2 and energy transfer and recombination paths (cmp. spectrum in Fig. 17.14). The rate constants are shown for various processes, the rate limiting step is shown with a *bold arrow*. (**b**) Electroluminescence transients after 100 ns excitation pulse (*grey area*) of $Ir(ppy)_3$ and DCM2 luminescence from CBP:10 % $Ir(ppy)_3$/1 % DCM2. Based on data from [1283]

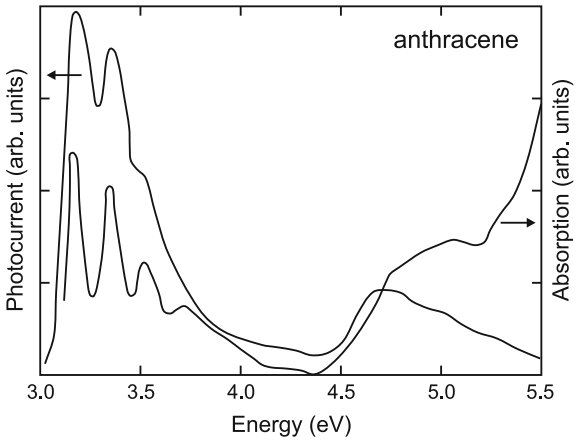


Fig. 17.17 Photoconductivity and absorption spectrum of anthracene

temperatures ($\tau \approx 80\,\mu s$ at $T = 200\,K$) [1275]. Endothermic transfer allows to pump a blue guest phosphor without a blue host material.

Further, subsequent Förster energy transfer [1284] from the guest triplet state to a fast and efficient singlet state (S_D) of a fluorescent dye is possible, e.g. from CBP:Ir(ppy)$_3$ to DCM2 [1283]. The transient lifetime of pure DCM2 is about 1 ns. In a mixture of CBP:10 % Ir(ppy)$_3$/1 % DCM2 the luminescence of DCM2 appears with the same 100 ns decay constant as that of Ir(ppy)$_3$ (Fig. 17.16b). This decay constant (rate limiting step, see Fig. 17.16a) corresponds to the energy transfer depleting the triplet state of Ir(ppy)$_3$ to DCM2 and is much faster than the pure Ir(ppy)$_3$ radiative lifetime.

The photoconductivity of organic semiconductors is typically related to their absorption spectrum as shown for anthracene in Fig. 17.17.

Chapter 18
Graphene and Carbon Nanotubes

Abstract The phononic properties and the band structure of graphene and carbon nanotubes of different chirality are given and derived within elementary models. Also inorganic atomic sheets and nanotubes such as BN are briefly discussed.

18.1 Graphene

18.1.1 Structure

Graphene is a single sheet of carbon atoms in hexagonal arrangement (Fig. 18.1). The carbon bond length is $d_{C-C} = 0.142\,\text{nm}$. It can be prepared from graphite via micromechanical cleavage, i.e. mechanical exfoliation (repeated peeling) of small mesas of highly oriented pyrolytic graphite [1285, 1286]. Graphite is a stacked arrangement of such graphene sheets (Fig. 18.2a) held together by van-der-Waals forces as shown in Fig. 18.2b. The carbon atoms bond in-plane via sp^2 hybridization. Organic molecules such as, e.g., anthracene or coronene (Fig. 17.1) can be understood as molecular-size pieces of such two-dimensional graphene sheet with hydrogen saturating the outside, broken bonds. In literature single layer graphene sheets (SLG) and few-layer graphene (FLG) sheets are distinguished. Ideally such two-dimensional crystal is infinitely extended, e.g. for band structure calculations. Real crystals have a boundary (surface) which is topologically a line or very thin sidewall. The mechanical properties of graphene are discussed in [1287].

The phonon dispersion of graphene is shown in Fig. 18.3. For the ZA and ZO modes the displacement is perpendicular to the graphene plane (out-of plane modes).

18.1.2 Band Structure

A single layer of graphene is a zero-gap semiconductor (cf. Fig. 6.40) which shows a linear photon-like spectrum

$$E = \hbar k\, c_* \tag{18.1}$$

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_18

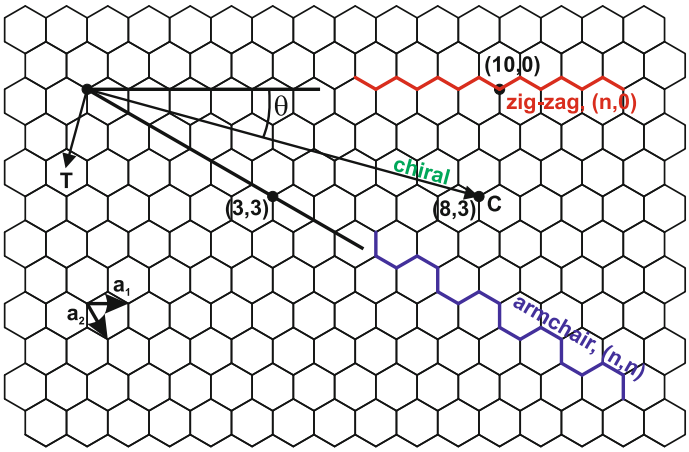


Fig. 18.1 Schematic atomic arrangement in graphene; the C–C bond length is $d_{C-C} = 0.142\,\text{nm}$. Several vectors for making carbon nanotubes (cf. Sect. 18.2) are shown

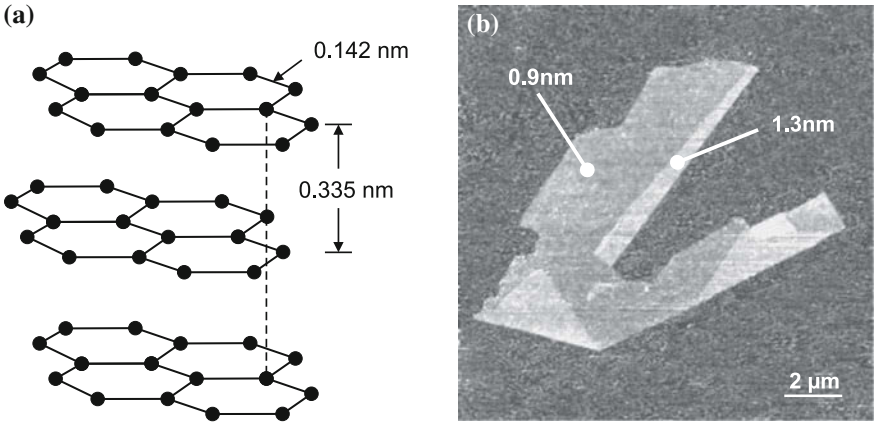


Fig. 18.2 (**a**) Schematic layer structure of graphite with bond length and layer distance labelled. (**b**) AFM image of graphene on oxidized silicon. The height of two areas relative to the background is labelled. Adapted from [1288]

around the Fermi energy at the K point (Fig. 18.4a). This point is also called Dirac point. The important bands close to the Fermi level stem from the π-orbitals. The linear dispersion around the K point is similar to that of relativistic particles without rest mass. The electrons in graphene are of course not really massless, their velocity (6.31) being $v = c_* \approx 10^6$ m/s, about 300 times smaller than the speed of light [1290, 1291].

In the simplest tight-binding approximation, the band structure is given as

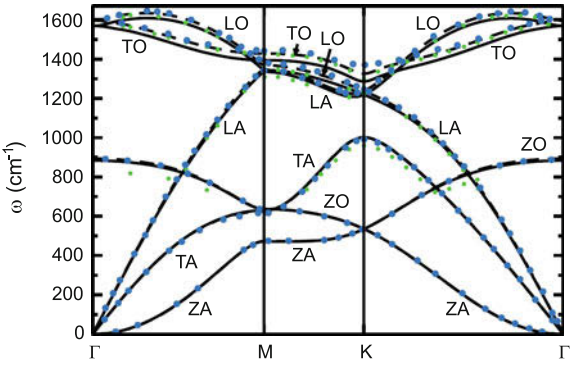


Fig. 18.3 Phonon dispersion in graphene. *Symbols* are experimental data from various methods. *Dashed lines* is DFT-LDA theory, *solid lines* GGA. Adapted from [1289]

$$E(\mathbf{k}) = \pm T \sqrt{1 + 4\cos\left(\frac{\sqrt{3}a\,k_y}{2}\right)\cos\left(\frac{a\,k_x}{2}\right) + 4\cos^2\left(\frac{a\,k_x}{2}\right)}, \qquad (18.2)$$

where a is the lattice constant and $T \approx 3\,\mathrm{eV}$ is the next-neighbor hopping energy [1292]. In the meantime more elaborate tight-binding schemes have been reported [1293, 1294] as shown in Fig. 18.4b. The two-dimensional band structure is visualized in Fig. 18.4c. Such band structure has been directly confirmed experimentally [1295] as shown in Fig. 18.5a–e. The distortion of the band structure with increasing electron concentration (Fig. 18.5e–h) from the conical bands is due to strong electron–electron, electron–phonon, and electron–plasmon coupling effects [1295].

The band structure of FLG has been theoretically analysed in [1296]. For bilayers experimental data on the band structure can be found in [1297]. Subtle differences exist for different stacking orders of the graphene sheets. Bulk graphite shows a semimetallic behavior with a band overlap of about 41 meV. For more than ten graphene layers the difference with the band overlap in bulk graphite is less than 10 %.

18.1.3 Electrical Properties

The Shubnikov-de Haas (SdH) oscillations from a graphene sheet exhibit a behavior [1291]

$$\frac{1}{\Delta B} = \frac{4e}{h}\frac{1}{n_S}, \qquad (18.3)$$

which corresponds to (13.38) for a two-dimensional electron system and a spin- and valley-degeneracy[1] of two. The cyclotron mass has been determined from the

[1] Each of the six valleys at the K points is shared by three Brillouin zones.

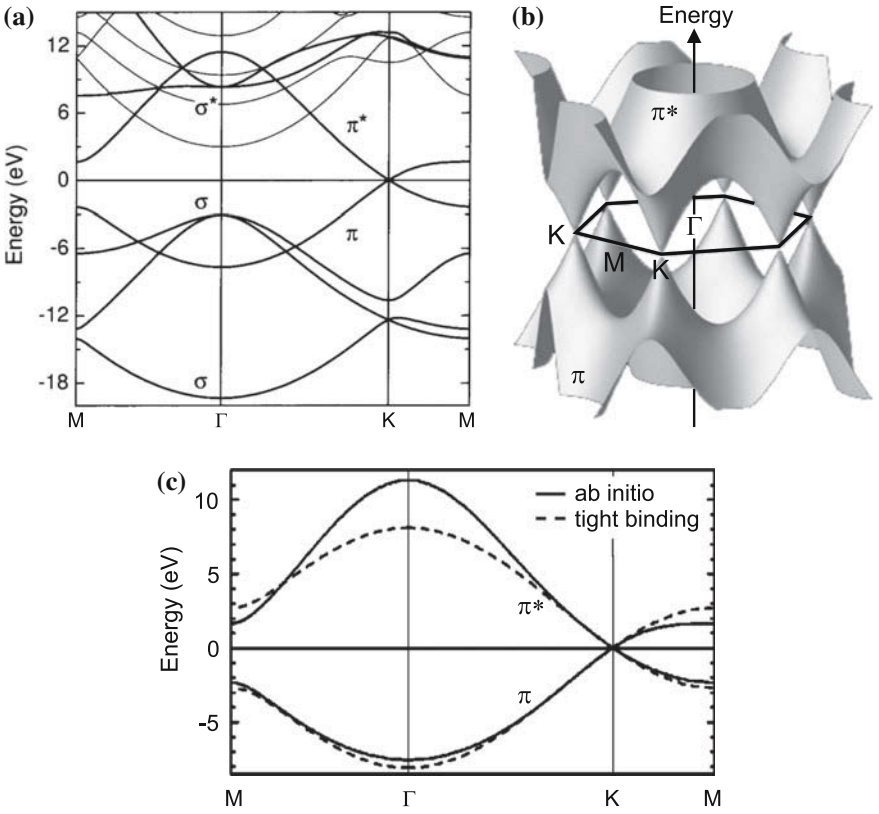


Fig. 18.4 (**a**) Band structure of graphene from first principles. (**b**) Three-dimensional representation $E(k_x, k_y)$ of the π-bands of graphene. (**c**) Band structure of graphene with only π-bands shown, *solid lines* are ab-initio calculation, *dashed lines* are calculated with tight-binding approximation. Adapted from [1294]

temperature dependence of the SdH oscillations to be proportional[2] to $\sqrt{n}$ (Fig. 18.6). The cyclotron mass is generally related [1298] to the area $S(E) = \pi k^2$ in k-space of the orbits at the Fermi energy via

$$m_c = \frac{\hbar^2}{2\pi} \frac{\partial S(E)}{\partial E} \tag{18.4}$$

With the linear dispersion (18.1) we can write (18.4) as

$$m_c = \frac{\hbar^2}{2\pi} \frac{2\pi E}{\hbar^2 c_*^2} = \frac{E}{c_*^2}. \tag{18.5}$$

[2] In a parabolic dispersion as in (13.37), the cyclotron mass is independent of n.

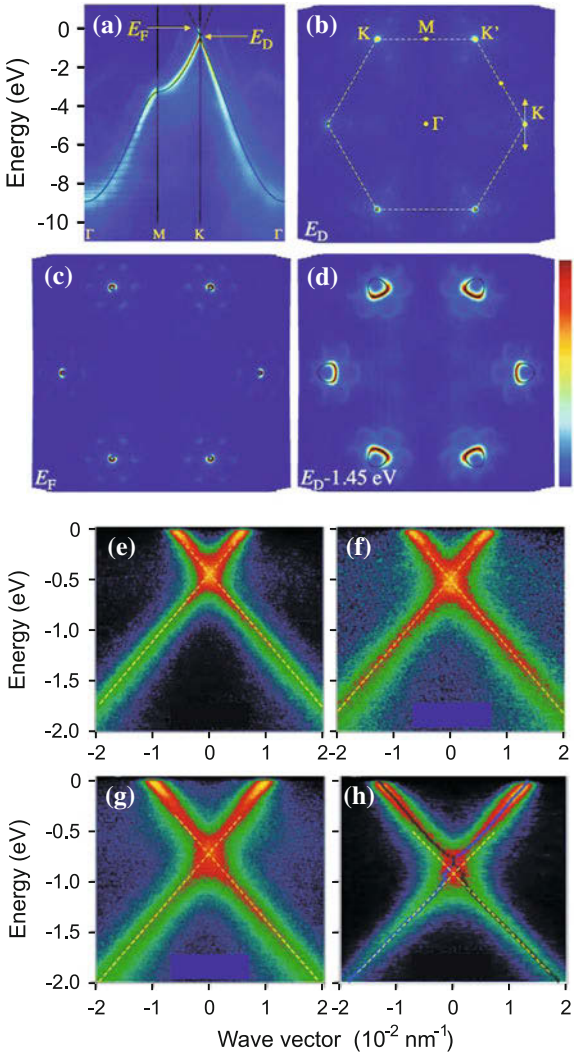


Fig. 18.5 Experimental bandstructure of graphene (on (0001) 6H-SiC) as determined from ARPES. (**a**) Energy distribution of states as a function of momentum along principal directions in the Brillouin zone. The single-orbital tight-binding model (18.2) with $T = 2.82\,\text{eV}$ is shown as *solid lines*. The Fermi level is shifted by 0.435 eV due to doping. (**b**) Constant energy map of the states at binding energy corresponding to the Dirac energy E_D; the boundary of the Brillouin Zone boundary is superimposed as *dashed line*. The arrow at the K point indicates the directions over which the data in (**e**–**h**) were acquired. (**c**, **d**) Constant energy maps at the Fermi energy ($E_F = E_D + 0.45\,\text{eV}$) and $E_D - 1.5\,\text{eV}$, respectively. (**e**–**h**) Experimental energy bands along the line through the K point parallel to Γ–M direction as indicated in (**b**). The *dashed lines* are an extrapolation of the lower bands below the Dirac crossing energy, which are observed not to pass through the upper bands (above E_D), suggesting the kinked shape of the bands around E_D. The sheet electron density is $n_S = 1.1$, 1.5, 3.7, and $5.6 \times 10^{13}\,\text{cm}^{-2}$ for (**e**)–(**h**), respectively, due to increased doping upon potassium adsorption. Adapted from [1295]

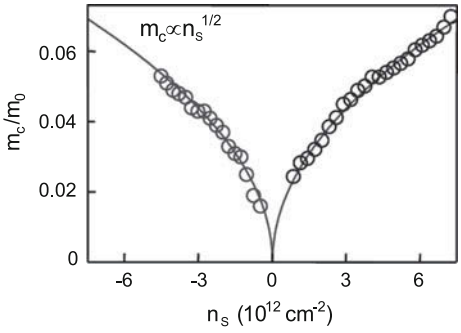


Fig. 18.6 Cyclotron mass in graphene as a function of the sheet electron concentration n_S (negative values relate to hole concentration, $E_F < E_D$). Adapted from [1291]

For the linear energy dispersion (18.1) the number of states up to energy E_F is (with a degeneracy of 4)

$$N(E_F) = 4\,\frac{\pi\,k_F^2}{(2\pi/L)^2} = A\,\frac{4\pi\,E_F^2}{h^2\,c_*^2}, \tag{18.6}$$

where A is the system area. Therefore we have (at low temperature) with (18.5)

$$n_S = \frac{4\pi}{h^2}\,\frac{E_F^2}{c_*^2} \propto m_c^2, \tag{18.7}$$

as determined experimentally. Therefore the behavior of the SdH oscillations confirms the linear dispersion relation. The experimental value for the velocity is $c_* \approx 10^6$ m/s. From (18.7) the density of states (per area and energy) around the Dirac point increases linearly with energy,

$$D(E) = \frac{8\pi}{h^2\,c_*^2}\,E. \tag{18.8}$$

The carrier density in a graphene sheet can be controlled via the field effect. The graphene is positioned on an insulator/semiconductor structure, typically SiO_2/Si (cf. Sect. 21.3). The carrier density is then related to the applied (gate) voltage V_g via (21.89) and (21.91), i.e.

$$n_S = \frac{\epsilon_i\,V_g}{e\,d}, \tag{18.9}$$

where d is the thickness of the insulator and ϵ_i its dielectric constant. By applying positive (negative) bias electron (holes) can be induced in the sheet. The electron and hole densities depend on the Fermi energies[3]

[3]Here we assume the linear dispersion for all thermally populated states.

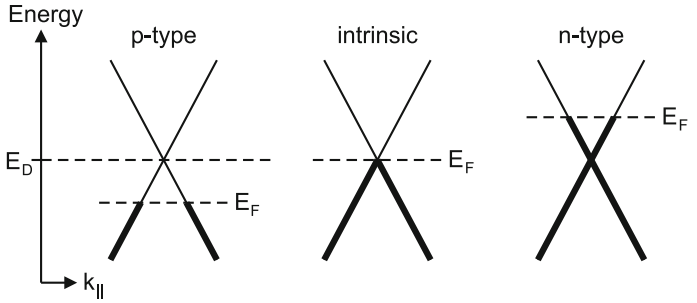


Fig. 18.7 Band structure of graphene with various positions of the Fermi energy E_F in relation to the Dirac energy E_D. States occupied with electrons are shown with in *bold*

$$n_S = \frac{8\pi}{h^2\, c_*^2} \int_{E_D}^{\infty} \frac{E - E_D}{1 + \exp\left[(E - E_F)/kT\right]}\, dE \tag{18.10a}$$

$$p_S = \frac{8\pi}{h^2\, c_*^2} \int_{-\infty}^{E_D} \frac{-(E - E_D)}{1 + \exp\left[-(E - E_F)/kT\right]}\, dE \tag{18.10b}$$

as visualized in Fig. 18.7. These relations cannot be inverted to obtain $E_F(n, p)$ analytically. The total charge carrier density is $\rho_S = e(p_S - n_S)$.

The Hall effect (Fig. 18.8) shows the expected ambipolar dependence according to (13.16) which takes the form

$$R_H = \frac{1}{e}\, \frac{p_S - n_S}{(n_S + p_S)^2} \tag{18.11}$$

for equal electron and hole mobilities.[4]

From Hall effect measurements the mobility has been determined to be about $10^4\,\mathrm{cm^2/Vs}$, independent of temperature between 10 and 100 K, and the same for electrons and holes. This value, however, is much smaller than the in-plane mobility of about $10^6\,\mathrm{cm^2/Vs}$ at 4.2 K in high quality samples of highly ordered pyrolithic graphite (HOPG) [1299]. In suspended graphene a mobility of $2.3 \times 10^5\,\mathrm{cm^2/Vs}$ has been found, limited by finite sample size [1300]. Thus in graphene layers on a solid surface extrinsic effects such as charge traps, interfacial phonons, ripples or fabrication residue seem to limit the carrier mobility.

The QHE has been observed in graphene [1291], reportedly up at room temperature [1301]. The plateaus $(4e^2/h)(n + 1/2)$ correspond to unusual half-integer filling, the first plateau occurring at $2e^2/h$, as has been also suggested from theory, and related to the 'pseudospin'. Another consequence of the Dirac-like behavior of the fermions in graphene is the presence of finite maximum resistiv-

[4]Including sign, $\mu_h = -\mu_e$. For $T = 0$ and $E_F = E_D$, $n_S = p_S = 0$ and thus $1/R_H$ should be zero. For finite temperatures there is always $n_S > 0$ and $p_S > 0$, even for $E_F = E_D$. Thus $1/R_H \propto 1/(p_S - n_S)$ diverges at $\rho_S = e(p_S - n_S) = 0$ with a change of sign.

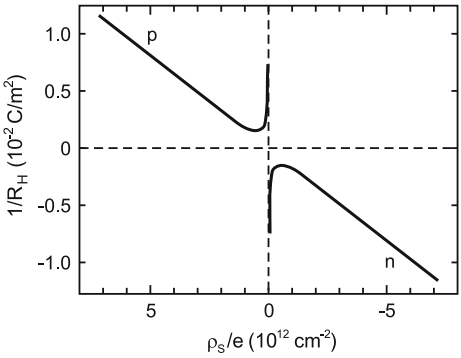


Fig. 18.8 Hall coefficient ($T = 10\,\mathrm{K}$) for a graphene sheet as a function of the free carrier sheet density $\rho_S/e = p_S - n_S$ (positive values indicate p-type). Data shown as *solid lines* from [1291]

ity $\rho_{max} = h/4e^2 = 6.45\,\mathrm{k\Omega}$ even at low temperature and $E_F = E_D$. It is due to the fact that localization effects, leading to insulating behavior, are strongly suppressed. Then each carrier keeps a mean free path in the order of its Fermi wavelength.

The Klein paradox [1302, 1303], the efficient tunneling of Dirac particles through high and thick barriers seems accessible in transport experiments with graphene [1304].

18.1.4 Other Two-Dimensional Crystals

Two-dimensional crystals have been reported also for other materials such as BN, MoS_2, $NbSe_2$, $Bi_2Sr_2CaCu_2O_x$ [1286] or ZnO [1305]. The phonon spectra of graphene, BN- and BC_2N-sheets are compared in Fig. 18.9. The spectrum of BC_2N is similar to the superposition of the C and BN spectra [1306].

18.2 Carbon Nanotubes

18.2.1 Structure

A carbon nanotube (CNT) is a part of a graphene sheet rolled up to form a cylinder. They were first described as multi-walled nanotubes by Iijima [1307] in 1991 (Fig. 18.10b) and in their single-walled form (Fig. 18.10a) in 1993 [1308]. Reviews can be found in [1293, 1309].

The chirality and diameter of a nanotube are uniquely described by the chiral vector

$$\mathbf{c}_h = n_1\,\mathbf{a}_1 + n_2\,\mathbf{a}_2 \equiv (n_1, n_2), \tag{18.12}$$

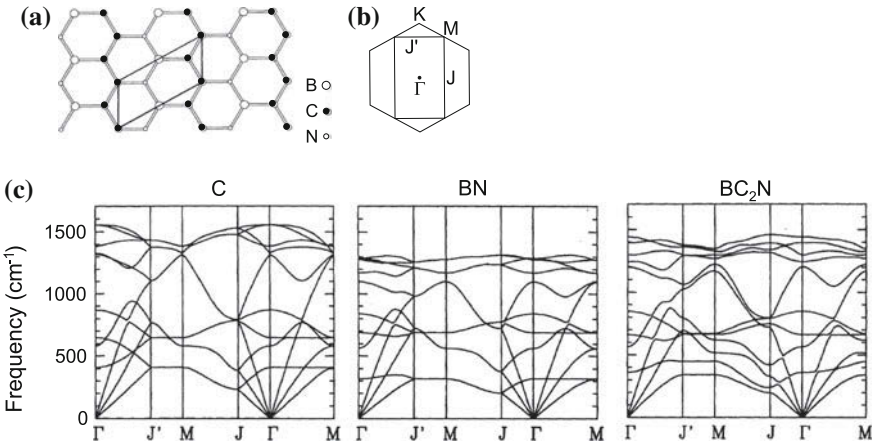


Fig. 18.9 (**a**) Crystal structure of BC_2N. (**b**) Brillouin zone of BC_2N (*rectangle*) and C, BN (*hexagon*). (**c**) Phonon dispersion of graphene (C) and BN- and BC_2N-sheets. Adapted from [1306]

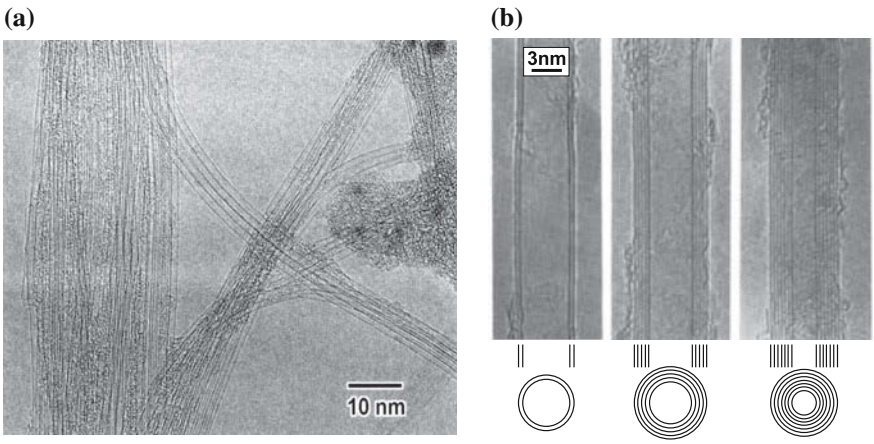


Fig. 18.10 (**a**) TEM image of single-walled carbon nanotubes (SWNT). (**b**) TEM images of various multi-walled carbon nanotubes (MWNT). Adapted from [1307]

where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the unit vectors of the graphene sheet. The chiral vector denotes two crystallographic equivalent sites which are brought together along the circumference of the nanotube. The possible vectors are visualized in Fig. 18.1 for $-30° \leq \theta \leq 0°$. The fiber diameter is given by

$$d = \frac{|\mathbf{c}_\mathrm{h}|}{\pi} = \frac{a}{\pi}\left(n_1^2 + n_1 n_2 + n_2^2\right), \tag{18.13}$$

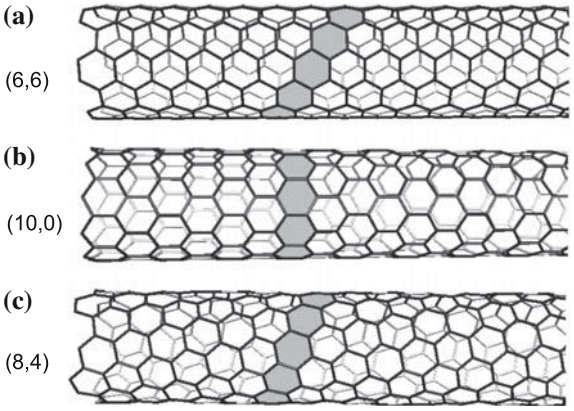


Fig. 18.11 Structure of different types of carbon nanotubes that have similar diameter of 0.8 nm. (**a**) Armchair (6,6), (**b**) zigzag (8,0) and (**c**) chiral symmetry. Adapted from [1309]

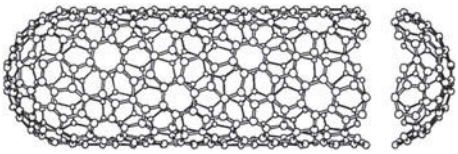


Fig. 18.12 A chiral nanotube (chiral vector is (10,5), $\theta = -19.11°$) with hemispherical caps at both ends based on an icosahedral C_{140} fullerene. The tube diameter is 1.036 nm. Adapted from [1311]

with the graphene lattice constant $a = \sqrt{3}\, d_{C-C} = 0.246\,\text{nm}$. Ab-initio calculations show that the diameter becomes a function of the chiral angle below 0.8 nm; deviations from (18.13) are below 2 % for tube diameters $d > 0.5\,\text{nm}$ [1310]. The $(n, 0)$ tubes ($\theta = 0$) are termed 'zig-zag' and an example is depicted in Fig. 18.11b. Nanotubes with $\theta = \pm\pi/6$, i.e. of the (n, n) (and $(2n, -n)$) type, are called 'armchair'. All others are termed 'chiral'.

The extension along the wire axis is large compared to the diameter. The tip of a nanotube is part of a buckminster-fullerene type molecule (Fig. 18.12). When the nanotube if formed by rolling a single sheet of graphene (SLG), a single-walled nanotube (SWNT) is formed. A FLG sheet creates a multi-walled nanotube (MWNT). For small number of layers they are called double-walled, triple-walled and so forth.

The mechanical strength of carbon nanotubes is very large. For SWNT Young's moduli of 10^3 GPa have been found experimentally [1312] in agreement with theoretical predictions [1313].

18.2.2 Band Structure

In carbon nanotubes there is some mixing of the $\pi(2p_z)$ and $\sigma(2s$ and $2p_z)$ carbon orbitals due to the radial curvature. This mixing is, however, small and can

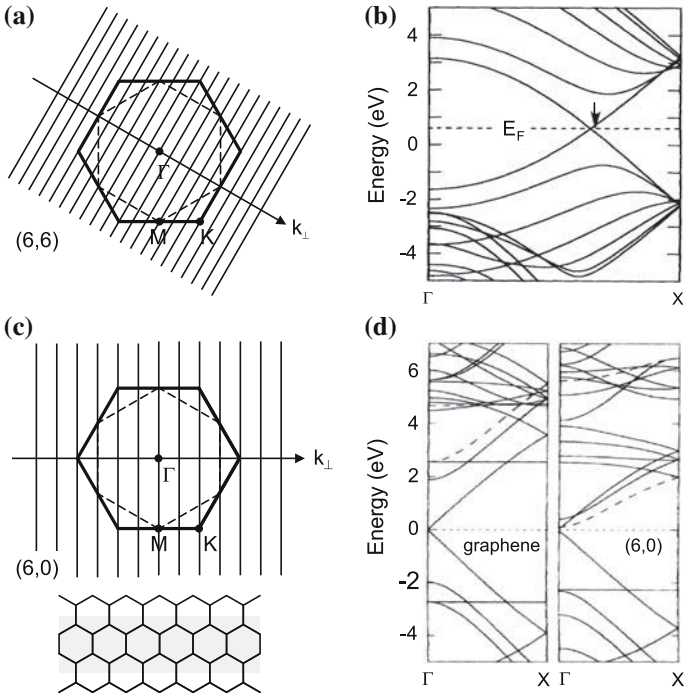


Fig. 18.13 (**a**) Brillouin zone of the graphene lattice (*bold line*) and allowed k-values for a (6,6) armchair nanotube. (**b**) Band structure of a (6,6) carbon nanotube. Adapted from [1315]. (**c**) Brillouin zone of the graphene lattice (*bold line*) and allowed k-values for a (6,0) zig-zag carbon nanotube. In the *lower part* the real space structure is visualized. (**d**) Band structure of graphene (*left*) and a (6,6) nanotube (*right*). Adapted from [1316]

be neglected near the Fermi level [1314]. The band structure of a nanotube is mainly determined by zone-folding of the graphene band structure. The vector along the (infinitely extended) wire k_z is continuous. The vector $k_\perp$ around the nanotube is discrete with the periodic boundary condition

$$\mathbf{c}_\mathrm{h} \cdot \mathbf{g}_\perp = 2\pi\, m, \tag{18.14}$$

where m is an integer. The distance of allowed $k_\perp$-values is (5.5)

$$\Delta k_\perp = \frac{2\pi}{\pi\, d} = \frac{2}{d}. \tag{18.15}$$

The character of the nanotube band structure depends on how the allowed k-values lie relative to the graphene Brillouin zone and its band structure. This is visualized in Fig. 18.13. For the case of an armchair tube (n, n), as shown in Fig. 18.13a, the K-point of the graphene band structure always lies on an allowed k-point. Therefore,

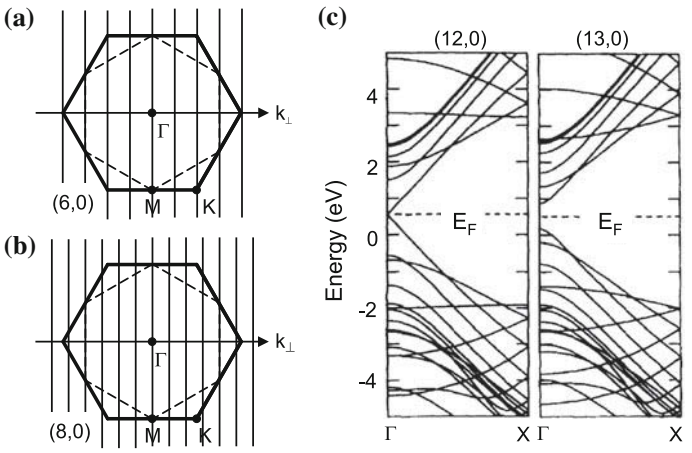


Fig. 18.14 (**a**, **b**) Brillouin zone of the graphene lattice (*bold line*) and allowed k-values for a (**a**) (6,0) and a (**b**) (8,0) zig-zag nanotube. (**c**) Band structures of a (12,0) metallic and (13,0) semiconducting armchair carbon nanotube. Adapted from [1315]

the nanotube is metallic, i.e. zero-gap, as seen in the bandstructure in Fig. 18.13b. The Dirac point is between Γ and X. For a zig-zag nanotube, the k-space is shown in Fig. 18.13c for a $(6, 0)$ nanotube. The corresponding band structure for a $(6, 0)$ nanotube is also metallic (Fig. 18.13d) with the Dirac point at the Γ point.

In Fig. 18.14c the band structure of another metallic $(12, 0)$ zig-zag nanotube is shown. However, only for $(3m, 0)$ the K-point is on an allowed state and thus the tube metallic. For the other cases, as shown for the k-space of a $(8, 0)$ nanotube in Fig. 18.14b, this is not the case. The corresponding band structure (Fig. 18.14c for $(13, 0)$) has a gap and thus the nanotube is a semiconductor. Generally, the condition for a nanotube to be metallic is with an integer m

$$n_1 - n_2 = 3\,m. \tag{18.16}$$

There are two semiconducting 'branches' with $\nu = (n_1 - n_2) \bmod 3 = \pm 1$. The tubes with $\nu = +1$ have a small band gap, those with $\nu = -1$ have a larger band gap.

The density of states is a series of one-dimensional DOS, proportional to $\sqrt{E}$ (6.75). It is compared in Fig. 18.15 for a metallic and a semiconducting nanotube. Within 1 eV from the Fermi energy the DOS can be expressed in an universal term [1317].

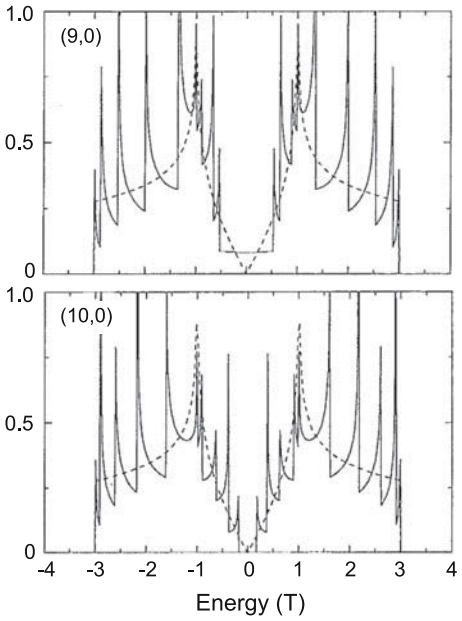


Fig. 18.15 Density of states for a (9,0) metallic and (10,0) semiconducting zig-zag carbon nanotube within the tight-binding approximation (18.2). The energy scale is given in units of the tight-binding parameter $T \approx 3\,\text{eV}$. The *dashed lines* are the DOS of graphene. Adapted from [1311]

18.2.3 Optical Properties

Optical transitions occur with high probability between the van-Hove singularities of the DOS. The theoretical absorption spectrum of a (10,0) nanotube is shown in Fig. 18.16.

In an ensemble of nanotubes various types and sizes occur. The transition energies of all possible nanotubes sorted by diameter are assembled in the Kataura plot (Fig. 18.17a). Experimental data are shown in Fig. 18.17b. The two branches of semiconducting nanotubes $\nu = \pm 1$ yield different transition energies. The overall dependence of the transition energy follows a $1/d$-law.

18.2.4 Other Anorganic Nanotubes

Structures similar to carbon nanotubes have been reported for BN [1322, 1323]. A boron nitride nanotube is a cylindrically rolled part of a BN sheet. BN tubes are always semiconducting (Fig. 18.18) and have a band gap beyond 5 eV similar to hexagonal BN which is mostly independent on chirality and diameter [1324].

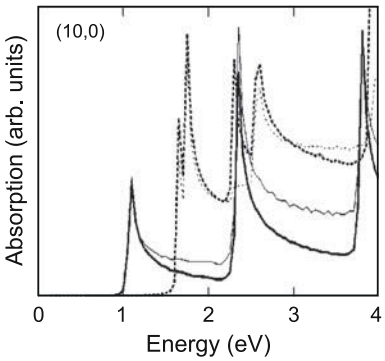


Fig. 18.16 Calculated absorption spectra for a (semiconducting) (10,0) carbon nanotube for parallel (*solid line*) and perpendicular (*dotted line*) polarization. The *thick* (*thin*) *lines* are calculated with (without) the matrix element included. Adapted from [1318]

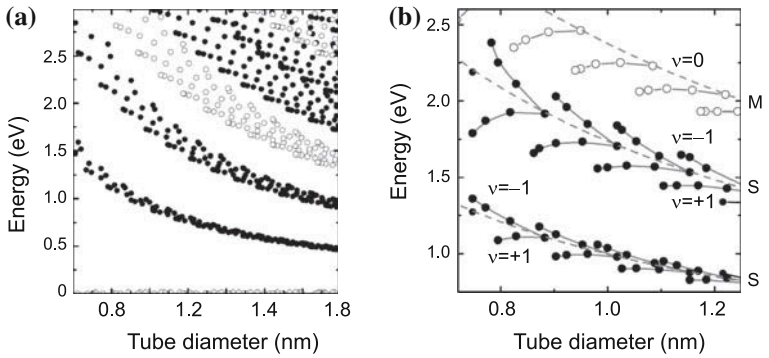


Fig. 18.17 (**a**) Theoretical transition energies of semiconducting (*filled symbols*) and metallic (*open symbols*) carbon nanotubes as a function of tube diameter (Kataura plot). Energies are calculated from van-Hove singularities in the JDOS within the third-order tight-binding approximation [1294]. (**b**) Experimental Kataura plot for the first two semiconducting (S, *closed symbols*) and the first metallic (M, *open symbols*) transition. *Dashed lines* connect the (near-to) armchair tubes; *full lines* connect tubes in a branch, $\nu = (n_1 - n_2)$ mod 3. Data from photoluminescence [1319] and resonant Raman scattering [1320]. Adapted from [1321]

Thus, while carbon nanotubes appear black since they absorb within 0–4 eV, BN is transparent (or white if scattering). For high energies larger than 10 eV C and BN tubes are quite similar since they are isoelectronic and the high-lying unoccupied states are less sensitive to the difference in the nuclear charges than the states at and below the Fermi energy [1325].

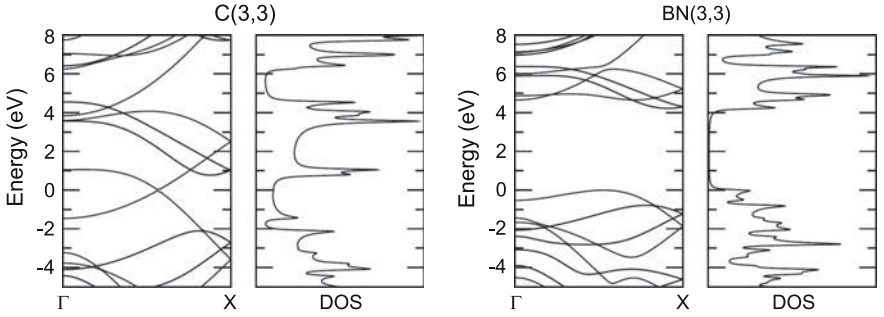


Fig. 18.18 Band structure and density of states (DOS) of C(3,3) and BN(3,3) nanotubes, calculated with DFT-LDA. Adapted from [1325]

Chapter 19
Dielectric Structures

Abstract Dielectric structures, in particular periodic dielectric structures are treated. A general one-dimensional model is developed describing Bragg mirrors. Examples for photonic band gap materials in one, two and three dimensions are given. Different types of dielectric cavities and microscopic resonators including Fabry–Pérot and whispering gallery resonators are treated. Quantum electrodynamic physical effects from light matter coupling such as Purcell effect and strong coupling are treated.

19.1 Photonic Band Gap Materials

Layered structures of dielectric materials with different index of refraction are used as optical elements such as filters or reflection and anti-reflection coatings [1326]. In this section we discuss the use of such concepts in one-, two- and three-dimensional photonic band gap materials.

19.1.1 Introduction

A structure with a so-called *photonic* band gap (PBG) exhibits an energy range (color range) in which photons cannot propagate in any direction. In the photonic band gap, there are no optical modes, no spontaneous emission and no vacuum (zero-field) fluctuations. We recollect that spontaneous emission is not a necessary occurrence: Looking at Fermi's golden rule (19.18) for the transition probability integrated over all final states

$$w(E) = \frac{2\pi}{\hbar} |M|^2 \rho_{\mathrm{f}}(E), \tag{19.1}$$

we see that the decay rate depends on the density ρ_{f} of final states at energy E. In the case of spontaneous emission, this is the (vacuum) density D_{em} of electromagnetic

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_19

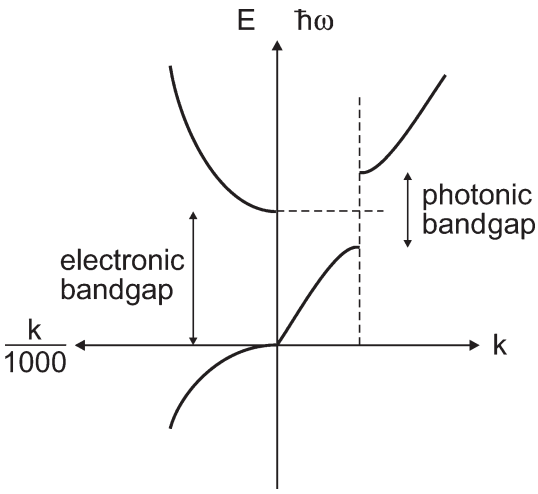


Fig. 19.1 *Right* electromagnetic dispersion with a forbidden gap at the wavevector of the periodicity. *Left* Electron-wave dispersion typical of a direct-gap semiconductor. When the photonic band gap straddles the electronic band gap, electron–hole recombination into photons is inhibited since the photons have no place to go (zero final density of states)

modes (per energy per volume) that varies $\propto \omega^2$:

$$D_{\mathrm{em}}(E) = \frac{8\pi}{(hc)^3} E^2. \tag{19.2}$$

In a homogeneous optical medium c must be replaced with c/n (cmp. Sect. 10.2.3).

If the band gap of a PBG is tuned to the electronic gap of a semiconductor, the spontaneous emission, and also induced emission, can be suppressed. Thus, one mode has to be left by 'doping' the structure. In this mode all emission will disappear and an efficient single-mode (monochromatic) LED or 'zero-threshold' laser could be built. A schematic comparison of the band structure of electrons and photons is given in Fig. 19.1.

19.1.2 General 1D Scattering Theory

The formation of a photonic band gap in a one-dimensional dielectric can be calculated to a large extent analytically and thus with direct insight. Let $n(x)$ be the spatially varying index of refraction (no losses or nonlinear optical effects). The one-dimensional wave equation (Helmholtz equation) reads for the electric field E

$$\frac{\partial^2 E(x)}{\partial x^2} + n^2(x)\,\frac{\omega^2}{c^2}\,E(x) = 0. \tag{19.3}$$

A comparison with a one-dimensional Schrödinger equation

$$\frac{\partial^2 \Psi(x)}{\partial x^2} - \frac{2m}{\hbar^2}\,[V(x) - E]\,\Psi(x) = 0 \tag{19.4}$$

shows that the Helmholtz equation corresponds to the quantum-mechanical wave equation of zero external potential V and a spatially modulated mass, i.e. a case that is usually not considered.

Let us consider now the amplitude a_k of the k eigenvector. The eigenvalue is then ω_k. The one-dimensional mode density $\rho(\omega)$ (per energy and per unit length) is

$$\rho(\omega) = \frac{\mathrm{d}k}{\mathrm{d}\omega}, \tag{19.5}$$

which is the inverse of the group velocity.

We follow one-dimensional scattering theory as presented in [1327]. At this point we do not rely on any specific form of $n(x)$ (Fig. 19.2a). The (complex) transmission coefficient t for any index structure is

$$t = x + \mathrm{i}y = \sqrt{T}\,\exp(\mathrm{i}\phi), \tag{19.6}$$

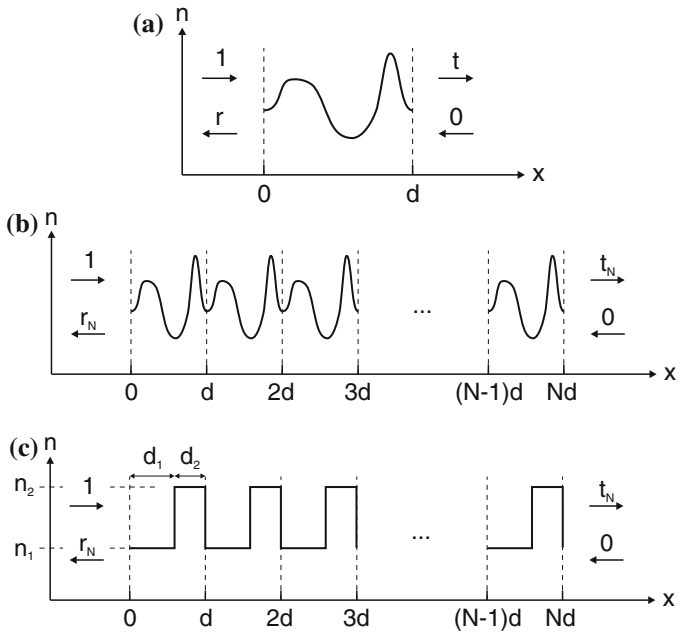


Fig. 19.2 1D scattering problem: (**a**) General scattering of an index of refraction distribution, (**b**) N-period stack, (**c**) two-layer (quarter-wave) stack

where $\tan\phi = y/x$. ϕ is the total phase accumulated during propagation through the structure. It can be written as the product of the physical thickness of the structure d and the effective wave number k. Hence we have the dispersion relation

$$\frac{\mathrm{d}}{\mathrm{d}\omega}\tan(kd) = \frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{y}{x}\right). \tag{19.7}$$

Evaluating the derivative we find

$$\frac{d}{\cos^2(kd)}\frac{\mathrm{d}k}{\mathrm{d}\omega} = \frac{y'x - x'y}{x^2}, \tag{19.8}$$

where the prime denotes derivation with respect to ω. Using the relation $\cos^2\theta = (1+\tan^2\theta)^{-1}$, we obtain the general expression

$$\rho(\omega) = \frac{\mathrm{d}k}{\mathrm{d}\omega} = \frac{1}{d}\frac{y'x - x'y}{x^2+y^2}. \tag{19.9}$$

19.1.3 Transmission of an N-Period Potential

Now, the behavior of N periods of a given index distribution $n(x)$ within a thickness d of one period (Fig. 19.2b) is investigated. The scattering matrix $\mathbf{M}$ connects the intensity at $x = 0$ with that at $x = d$. We use the column vector $\mathbf{u} = \begin{pmatrix} u^+ \\ u^- \end{pmatrix}$ containing the right- and left-going waves (labeled '+' and '−', respectively), $u^\pm = f^\pm \exp(\pm \mathrm{i}kx)$,

$$\mathbf{u}(0) = \mathbf{M}\,\mathbf{u}(d). \tag{19.10}$$

Using the boundary conditions $\mathbf{u}(0) = (1, r)$ and $\mathbf{u}(d) = (t, 0)$, we find that $\mathbf{M}$ has the structure

$$\mathbf{M} = \begin{pmatrix} 1/t & r^*/t^* \\ r/t & 1/t^* \end{pmatrix}. \tag{19.11}$$

The conservation of energy requires that $\det\mathbf{M} = (1-R)/T = 1$. The eigenvalue equation for $\mathbf{M}$ is

$$\mu^2 - 2\mu\,\mathrm{Re}(1/t) + 1 = 0. \tag{19.12}$$

The two eigenvalues $\mu^\pm$ are related by $\mu^+\mu^- = \det\mathbf{M} = 1$. If we consider an infinite, periodic structure, we know from Bloch's theorem that the eigenvector varies between unit cells only via a phase factor, i.e. $|\mu| = 1$. Therefore, the eigenvalues can be written as

$$\mu^\pm = \exp(\pm \mathrm{i}\beta), \tag{19.13}$$

where β corresponds to the Bloch phase of a hypothetical infinite periodic structure. This phase β should not be confused with ϕ defined earlier, which is associated with the unit cell transmission. We find the condition

$$\mathrm{Re}(1/t) = \cos\beta \tag{19.14}$$

for the Bloch phase. Since every matrix obeys its own eigenvalue equation, we have also (**1** being the unity matrix)

$$\mathbf{M}^2 - 2\,\mathbf{M}\cos\beta + \mathbf{1} = 0. \tag{19.15}$$

By induction one can show that the N-period case has the scattering matrix

$$\mathbf{M}^N = \mathbf{M}\,\frac{\sin(N\beta)}{\sin\beta} - \mathbf{1}\,\frac{\sin((N-1)\beta)}{\sin\beta}. \tag{19.16}$$

The solution for the finite period case can be written in terms of the Bloch phase of the infinite potential. The transmission and reflection of the N-period system are given by

$$\frac{1}{t_n} = \frac{1}{t}\,\frac{\sin(N\beta)}{\sin\beta} - \frac{\sin((N-1)\beta)}{\sin\beta} \tag{19.17a}$$

$$\frac{r_n}{t_n} = \frac{r}{t}\,\frac{\sin(N\beta)}{\sin\beta}\,. \tag{19.17b}$$

The transmission of intensity can be written as ($T = t^*t$)

$$\frac{1}{T_N} = 1 + \frac{\sin^2(N\beta)}{\sin^2\beta}\left(\frac{1}{T} - 1\right). \tag{19.18}$$

Again, up to this point no specific distribution of the index of refraction within the unit cell has been specified.

From (19.17a), a general formula for the mode density $\rho_N(\omega)$ of the N-stack can be obtained as [1327]

$$\rho_N = \frac{1}{N\,d}\,\frac{\frac{\sin(2N\beta)}{2\sin\beta}\left(\eta' + \frac{\eta\xi\xi'}{1-\xi^2}\right) - \frac{N\eta\xi'}{1-\xi^2}}{\cos^2(N\beta) + \eta^2\left(\frac{\sin(N\beta)}{\sin\beta}\right)^2}, \tag{19.19}$$

where $\xi = x/T = \cos\beta$ and $\eta = y/T$.

19.1.4 The Quarter-Wave Stack

A quarter-wave stack, also known as a Bragg mirror, exhibits a one-dimensional photonic band gap. One period consists of two regions with thickness and index of refraction (d_1, n_1) and (d_2, n_2), respectively (Fig. 19.2c). In the quarter-wave stack each region has an optical thickness of $\lambda/4$ (the wave accumulates in each region a phase of $\pi/2$) for a particular wavelength λ_0 or (midgap) frequency ω_0. Thus, the condition reads

$$n_1 d_1 = n_2 d_2 = \frac{\lambda_0}{4} = \frac{\pi}{2}\frac{c}{\omega_0}. \tag{19.20}$$

Using the Fresnel formulas, the transmission of an arbitrary two-layer cell is

$$t = \frac{T_{12}\,\exp(\mathrm{i}(p+q))}{1 + R_{12}\,\exp(2\mathrm{i}q)}, \tag{19.21}$$

where $p = n_1 d_1 \omega/c$ and $q = n_2 d_2 \omega/c$ are the phases accumulated in the two layers, respectively. The values of T_{12} and R_{12} are given as

$$T_{12} = \frac{4\,n_1\,n_2}{(n_1+n_2)^2} \tag{19.22a}$$

$$R_{12} = \frac{(n_1-n_2)^2}{(n_1+n_2)^2}\,. \tag{19.22b}$$

For the quarter-wave stack ($p = q = \pi/2$), we obtain for (19.21)

$$t = \frac{T_{12}\,\exp(\mathrm{i}\pi\tilde{\omega})}{1 + R_{12}\,\exp(\mathrm{i}\pi\tilde{\omega})}, \tag{19.23}$$

where $\tilde{\omega} = \omega/\omega_0$ is the frequency scaled to the midgap value.

The transmission of a single two-layer cell is

$$T = \frac{T_{12}^2}{1 - 2\,R_{12}\,\cos(\pi\tilde{\omega}) + R_{12}^2}, \tag{19.24}$$

and the Bloch phase is given by

$$\cos\beta = \xi = \frac{\cos(\pi\tilde{\omega}) - R_{12}}{T_{12}} \tag{19.25a}$$

$$\eta = \frac{\sin(\pi\tilde{\omega})}{T_{12}}\,. \tag{19.25b}$$

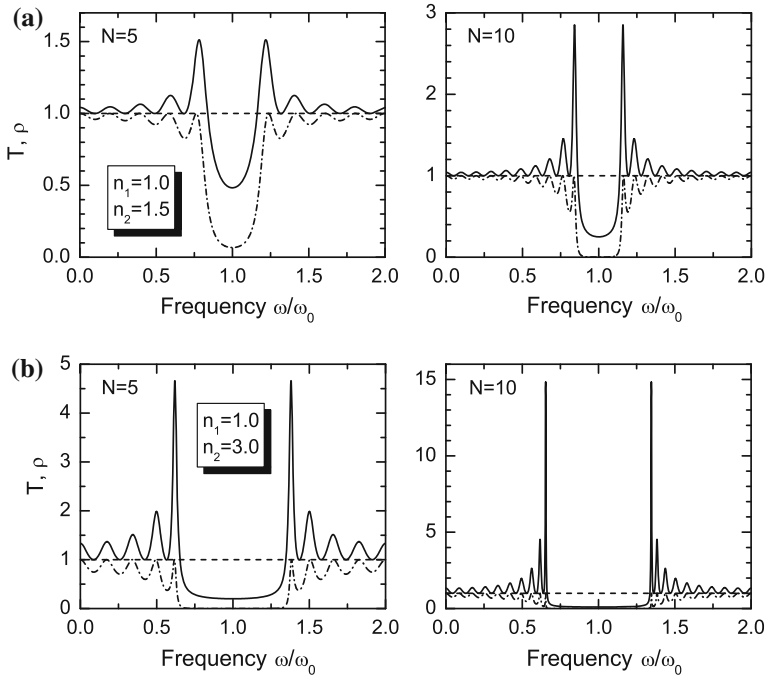


Fig. 19.3 Quarter-wave stack with indices of refraction (**a**) n_1, n_2 = 1.0, 1.5 and (**b**) 1.0, 3.0. *Solid lines* dimensionless density of modes ρ_N (19.19), *dashed lines* transmission T_N (19.26) for two different numbers of pairs $N = 5$ (*left panels*) and 10 (*right panels*) versus the dimensionless frequency $\tilde{\omega}$

For the N-period quarter-wave stack the transmission is given by

$$T_N = \frac{1 + \cos\beta}{1 + \cos\beta + 2\,(R_{12}/T_{12})\,\sin^2(N\,\beta)}. \tag{19.26}$$

A band gap forms. Within the band gap, the density of modes is lowered, at the edges it is enhanced (Figs. 19.3 and 19.4). The transmission at midgap decreases $\propto \left(n_i/n_j\right)^{2N}$, where $n_i < n_j$.

In the limit of large N the complete width $\Delta\tilde{\omega}$ of the band gap is implicitly given by

$$\cos\left(\frac{\pi}{2}\,\Delta\tilde{\omega}\right) = 1 - 2\left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2. \tag{19.27}$$

If $|n_1 - n_2| \ll n_1 + n_2$, we find

$$\Delta\tilde{\omega} \approx \frac{4}{\pi}\,\frac{|n_1 - n_2|}{n_1 + n_2}. \tag{19.28}$$

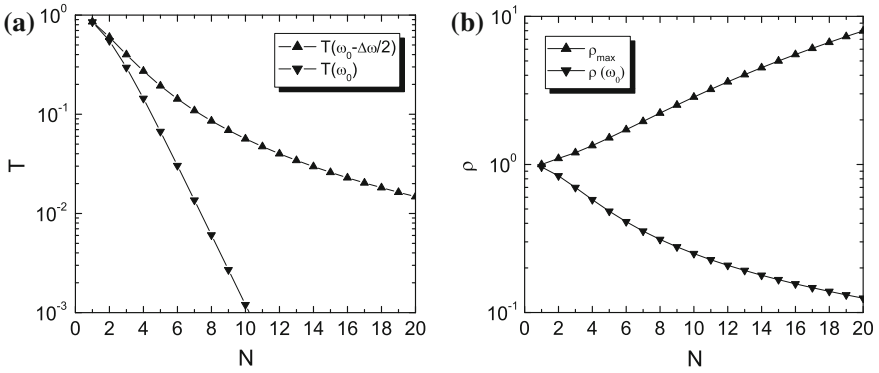


Fig. 19.4 Quarter-wave stack with indices of refraction n_1, n_2 = 1.0, 1.5: (**a**) Transmission T_N at midgap ($\tilde{\omega} = 1$, *down triangles*) and at the band edge ($\tilde{\omega} = 1 - \Delta\tilde{\omega}/2$, *up triangles*) versus number of pairs N. (**b**) Dimensionless density of modes ρ_N at maximum near the band edge and at midgap versus number of pairs N

The principle of the quarter-wave stack is scalable to frequencies other than visible light.[1] In Fig. 19.5 the reflectance of various quarter-wave stacks from yttria-stabilized zirconia (YSZ [1328], high index material, Fig. 19.6a) and Al_2O_3 are shown [1329]. The different design wavelengths have been achieved solely by varying the layer thicknesses.

In Fig. 19.6b the three Bragg mirrors from Fig. 19.5 with $N = 15.5$ pairs are replotted in relative frequency units $\tilde{\omega}$. The spectra look very similar; subtle differences in the width of the reflectance band are due to slightly larger index contrast at higher design energy (cmp. Fig. 19.6a). The width of the gap is approximately $\Delta\tilde{\omega} \approx 0.18$ in agreement with (19.28).

As further example, a Mo/Si Bragg mirror with a period of 6.7 nm is shown in Fig. 19.7. Such a mirror works in the extreme UV and is used for soft X-ray optics, possibly in advanced lithography systems. Dielectric thin films can also be designed for anti-reflection coatings, edge filters or pass and stop band filters as detailed in [1326].

19.1.5 Formation of a 3D Band Structure

For other applications, e.g. waveguides with minimized footprint, 3D (or at least 2D) photonic band gap structures are needed. Details can be found in dedicated textbooks [1332–1334]. In [1335] planar, cylindrical and spherical Bragg mirrors are discussed.

[1] This is a general property of Maxwell's equations which do not contain a specific length scale.

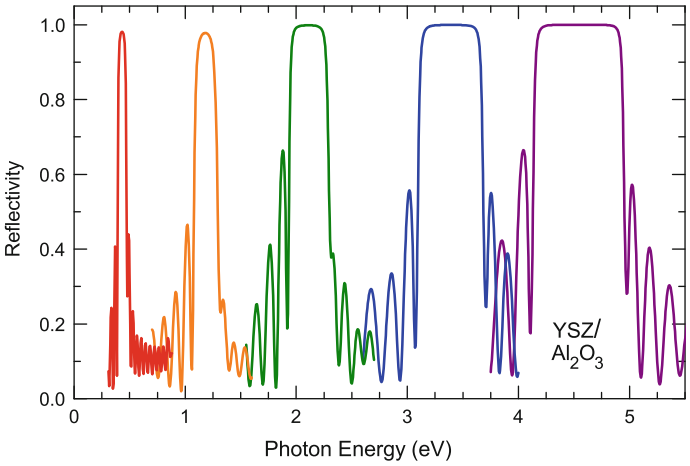


Fig. 19.5 Reflectance of various Bragg mirrors from YSZ/Al_2O_3 grown by pulsed laser deposition on sapphire. The different layer thicknesses result in the design energies 0.43 eV ($N = 10.5$, $R_{max} = 0.9812$, *red*), 1.19 eV ($N = 10.5$, $R_{max} = 0.9779$, *orange*), 2.11 eV ($N = 15.5$, $R_{max} = 0.99953$, *green*), 3.39 eV ($N = 15.5$, $R_{max} = 0.99946$, *blue*) and 4.54 eV ($N = 15.5$, $R_{max} = 0.99989$, *purple*)

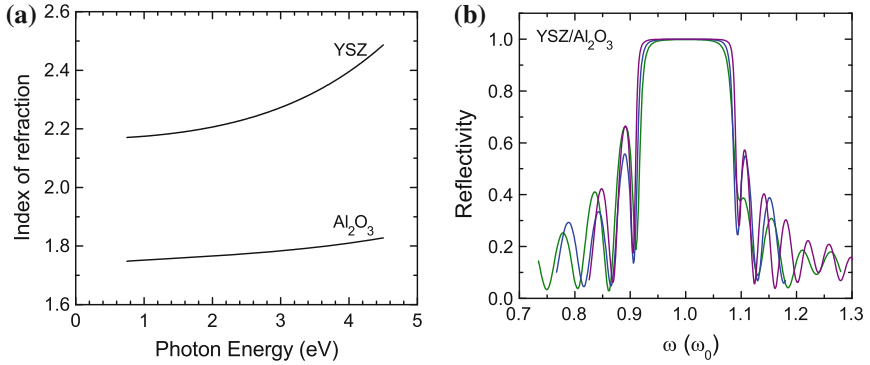


Fig. 19.6 (**a**) Index of refraction of YSZ and Al_2O_3 as a function of photon energy. (**b**) Reflectance spectra of the 2.11 eV, 3.39 eV and 4.54 eV Bragg mirrors of Fig. 19.5 replotted as a function of the scaled frequency $\tilde{\omega} = \omega/\omega_0$

Since we want a photonic band gap that is present for *all* directions of propagation, a Brillouin zone with a shape close to a sphere is preferable. Then, the main directions are at similar k-values (Fig. 19.8). One of the best suited is the fcc lattice. Since the L point is centered at ≈14 % lower frequency than the X point, the forbidden gaps for different directions must be, however, sufficiently wide to create a forbidden frequency band overlapping at all points along the surface of the Brillouin zone. For example, the bcc lattice has a Brillouin zone that is less symmetric than that of the fcc lattice (see Fig. 3.34) and thus is less suited for the creation of an omnidirectional

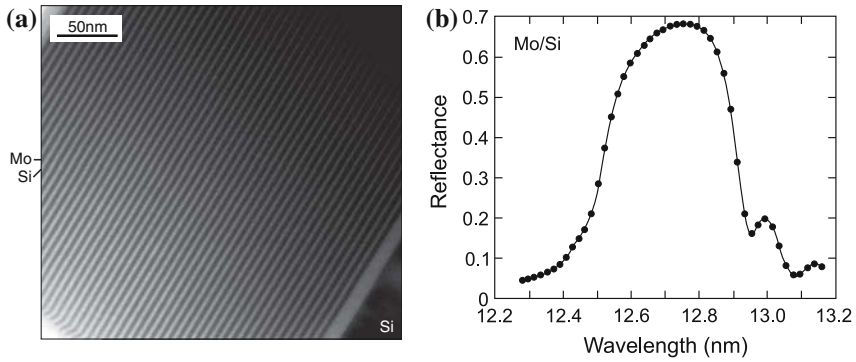


Fig. 19.7 (**a**) Cross-sectional TEM of Mo/Si superlattice with 2.7 nm Mo and 4.0 nm Si (period: 6.7 nm) on Si(001) substrate. From [1330]. (**b**) Reflection spectrum for a SL with period of 6.5 nm and 88.5° angle of incidence. Data points are shown as *circles*, the *solid line* is a fit with a period of 6.45 nm. Adapted from [1331]

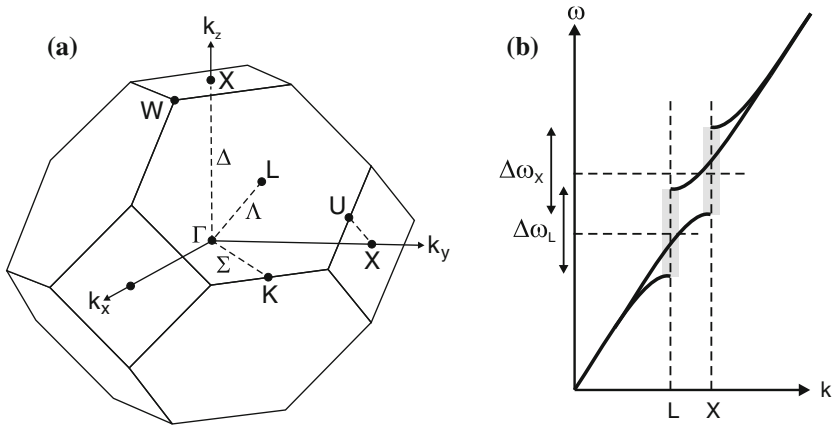


Fig. 19.8 (**a**) The Brillouin zone of the fcc lattice. (**b**) Schematic forbidden gaps at the L and X points

photonic band gap. However, the photonic band gap must not arise above the first band, relaxing problems due to asymmetry of the Brillouin zone (cf. Table 19.1).

Maxwell's equations (zero charge density) for monochromatic waves $\propto \exp(\mathrm{i}\omega t)$

$$\nabla \cdot \mathbf{D} = 0 \tag{19.29a}$$

$$\nabla \times \mathbf{E} = \mathrm{i}\,\frac{\mu\,\omega}{c}\mathbf{H} \tag{19.29b}$$

$$\nabla \times \mathbf{H} = \mathrm{i}\,\frac{\omega}{c}\mathbf{D} \tag{19.29c}$$

$$\nabla(\mu\mathbf{H}) = 0\,, \tag{19.29d}$$

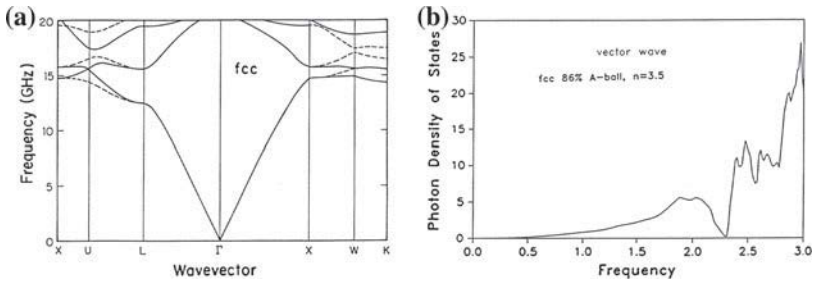


Fig. 19.9 (**a**) Calculated photonic band gap structure of a fcc dielectric structure composed of air spheres in a dielectric background of refractive index $n = 3.5$. The filling ratio is 86 % air and 14 % dielectric material. *Dotted* and *solid lines* represent coupling to s- and p-polarized light, respectively. Reprinted with permission from [1336], ©1990 APS. (**b**) Density of states for the band structure of part (**a**)

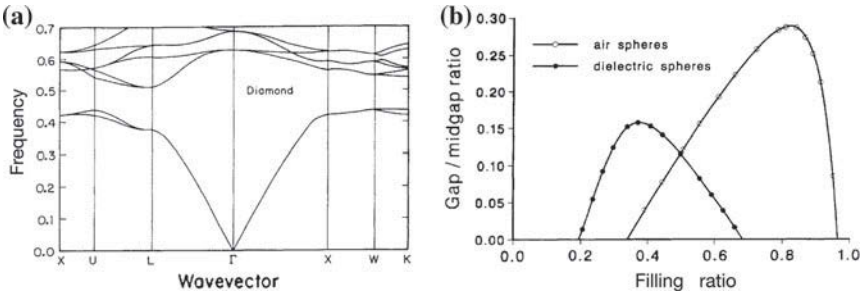


Fig. 19.10 (**a**) Calculated photonic band structure for a diamond dielectric structure consisting of overlapping air spheres in a dielectric material with $n = 3.6$. Filling ratio of air is 81 %. The frequency is given in units of c/a, a being the cubic lattice constant of the diamond lattice and c being the velocity of light. (**b**) Gap-to-midgap frequency ratio for the diamond structure as a function of filling ratio for dielectric spheres $n = 3.6$ in air (*solid circles*) and air spheres in dielectric n. Optimal case: air spheres with 82 % filling ratio. Reprinted with permission from [1336], ©1990 APS

together with $\mathbf{D}(\mathbf{r}) = \epsilon(\mathbf{r})\mathbf{E}(\mathbf{r})$ and $\mu = 1$ they are combined into the wave equation

$$\nabla \times \left[\epsilon^{-1}(\mathbf{r})\, \nabla \times \mathbf{H}(\mathbf{r})\right] + \frac{\omega^2}{c^2}\, \mathbf{H}(\mathbf{r}) = 0. \tag{19.30}$$

This equation is numerically solved for planar waves with wavevector $\mathbf{k}$.

In the following, results are shown for various structures. In a fcc lattice of air spheres in a dielectric medium with $n = 3.6$ (a typical semiconductor), no band gap can be achieved (Fig. 19.9a), only a pseudogap (Fig. 19.9b) appears.

In a diamond lattice (two fcc lattices shifted by $1/4\,\langle 111\rangle$), a complete photonic band gap is possible [1336] (Fig. 19.10). Recently, a periodic array of spirals (Fig. 19.11) has been predicted to exhibit a large photonic band gap [1337]. Glancing-angle deposition [1338] (GLAD) is a way to realize such structures. Another method

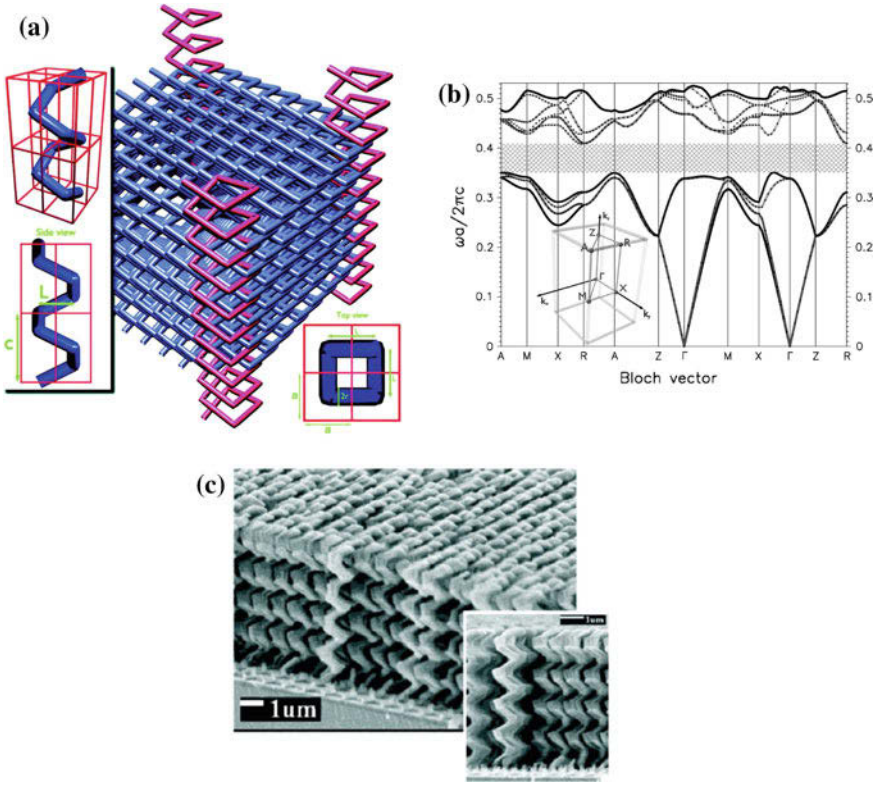


Fig. 19.11 (**a**) Tetragonal square-spiral photonic crystal. The crystal shown here has a solid filling fraction of 30 %. For clarity, spirals at the corners of the crystal are highlighted with a different shade and height. The tetragonal lattice is characterized by lattice constants a and b. The geometry of the square spiral is illustrated in the insets and is characterized by its width, L, cylinder radius, r, and pitch, c. The *top left inset* shows a single spiral coiling around four unit cells. (**b**) Band structure for the direct structure crystal characterized by [L, C, r]=[1.6, 1.2, 0.14] and a spiral filling factor f_{spiral} = 30 %. The lengths are given in units of a, the lattice constant. The width of the PBG is 15.2 % relative to the center frequency for background dielectric constant $\epsilon_b = 1$ and spiral dielectric constant $\epsilon_s = 11.9$. The positions of high-symmetry points are illustrated in the *inset*. Reprinted with permission from [1337], ©2001 AAAS. (**c**) Oblique and edge views of a tetragonal square spiral structure grown using the GLAD (glancing-angle deposition) process. Reprinted with permission from [1338], ©2002 ACS

to fabricate structures with arbitrary geometry within a material is two-photon lithography or two-photon holography. Another path to PBG structures are so-called inverted opals. First, a close-packed structure of spheres, e.g. monodisperse silica spheres, is fabricated by sedimentation or self-assembly. The gaps are filled with a high-index medium and the template is subsequently removed, e.g. by etching or dissolving. The resulting structure is shown in Fig. 19.12a. Such a structure has a photonic band gap (Fig. 19.12b) if the refractive index is sufficiently high (>2.85) [1339].

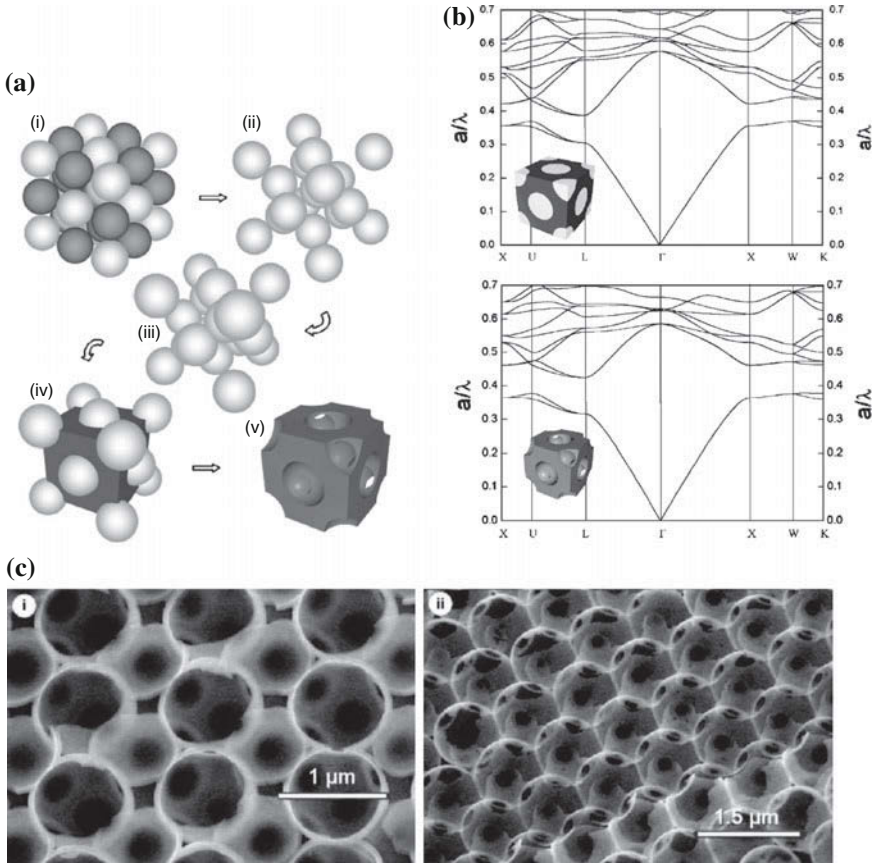


Fig. 19.12 (**a**) Cartoon showing, in five steps, the fabrication of an inverse diamond structure with a full photonic band gap. First, (i) a mixed body-centered cubic lattice is assembled (ii) after which the latex sublattice is removed; (iii) then the structure is sintered to a filling fraction of ∼50 % after that (iv) silicon or germanium infiltration takes place and finally (v) silica elimination. (**b**) Photonic band diagrams of (upper panel) a silicon/silica composite diamond opal and (lower panel) made of air spheres in silicon resulting from the removal of the silica spheres from the former. The filling fraction for silicon is 50 %. The inset shows the corresponding real space structures. Reprinted with permission from [1347], ©2001 AIP. (**c**) SEM images of internal facets of silicon inverse opal: (i) (110) facet, (ii) (111) facet. Adapted from [1348], reprinted with permission from Nature, ©2000 Macmillan Magazines Limited

The band gap in this case is between the 4th and 5th band. Table 19.1 offers a compilation of various PBG structures and their properties.

Table 19.1 Various photonic band gap structures and some of their properties

Name	Crystal type	n	$\Delta\tilde{\omega}$ (%)	Reference
Diamond	Diamond	2	29	[1336]
Yablonovite	fcc	2	19	[1340]
Woodpile	fc tetragonal	2	20	[1341]
Spirals	sc	4	17	[1342]
Square-spirals	Tetragonal	4	24	[1337]
Layered 3D	bc orthorhombic	4	23	[1343]
Inverted scaffold	sc	5	7	[1344]
Inverse opal	fcc	8	4.25	[1345]
Inverse hcp	hcp	16	2.8	[1346]

The band gap is between the nth and $(n+1)$th band, $\Delta\tilde{\omega}$ is given for air/silicon ($\epsilon \approx 12$)

19.1.6 Disorder

A real photonic band gap structure deviates from the ideal, perfectly periodic system by slight variations of the position and possibly also the size of the dielectric 'atoms'. This is schematically shown in Fig. 19.13a. The difference between the real and ideal structure is a (bipolar) spatial distribution of $\Delta\epsilon(\mathbf{r})$ which acts as a source of scattering and hence exponential attenuation of coherent beams propagating through photonic crystals over lengths l, named the '(extinction) mean free path'. After propagating over such distance l, a light beam is converted to a diffuse glow that corrupts the functionality of any photonic integrated circuit. Experimentally for opals a mean free path consistent with 5 % fabrication accuracy has been found (Fig. 19.13b). For such disorder and a lattice constant $a \approx \lambda$, the mean free path is about only 10 wavelengths, $l \approx 10\lambda$.

19.1.7 Defect Modes

Similar to a perfect periodic atomic arrangement leading to the formation of the electronic band structure, a perfectly periodic dielectric structure leads to the photonic band structure. As we know from semiconductor physics, much of the interesting physics and numerous applications lie in defect modes, i.e. localized electronic states due to doping and recombination at such centers. The equivalent in PBG structures are point defects (one unit missing) or line defects (a line of units, straight, bend or with sharp angles, missing). Such defects create localized states, i.e. regions for light localization. In the case of line defects we deal with waveguides that can be conveniently designed and could help to reduce the size of photonic and optoelectronic integrated circuits.

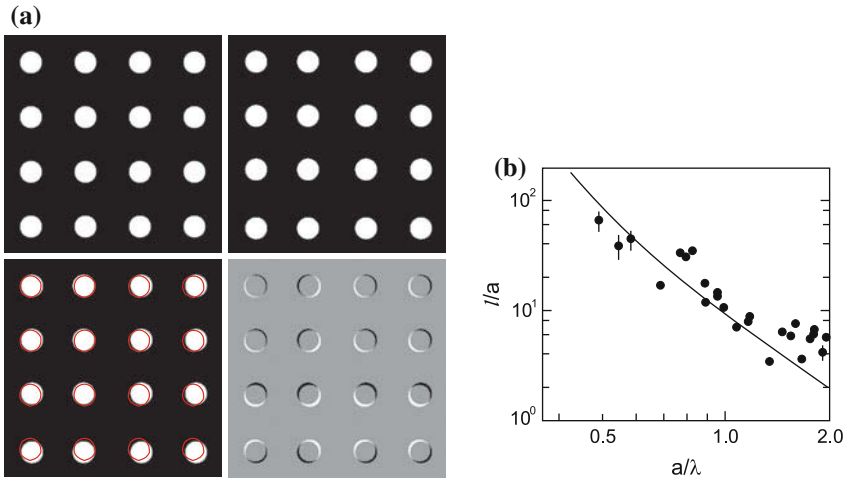


Fig. 19.13 (**a**) Schematic photonic band gap structure with perfect (*upper left*) and disordered (*upper right*) periodicity. In the *lower left* panel the disordered structure is overlayed with the ideal structure (*red circles*). In the *lower right* panel, the difference between ideal and disordered structure is shown. (**b**) Optical mean free path in an opal photonic band gap structure for various lattice constants. *Solid line* is theory for 5 % fabrication accuracy. Adapted from [1349]

1D Model

We revisit our 1D scattering theory and create now a 'defect'. A simple defect is the change of the width of the center n_2-region in a quarter-wave stack. For the numerical example, we choose $N = 11$, $n_1 = 1$, $n_2 = 2$.

In Fig. 19.14, the transmission curves are shown for the undisturbed quarter-wave stack and the microcavity with $n_2\, d_2^{\text{center}} = 2\lambda_0/4 = \lambda_0/2$. A highly transmissive mode at $\omega = \omega_0$ arises that is quite sharp with $\Delta\omega = 3 \times 10^{-4}$. Thus, the quality factor Q, also called the Q-factor or finesse,

$$Q = \frac{\omega_0}{\Delta\omega}, \tag{19.31}$$

with ω_0 being the resonance frequency and $\Delta\omega$ being the linewidth, is 3.3×10^3 in this case.

If the thickness is varied (Fig. 19.15), the mode shifts away from the center. A similar scenario arises for higher-order $n_l/2$-cavities, e.g. $n_2 d_2^{\text{center}} = 4\lambda_0/4 = \lambda_0$ (Fig. 19.16).

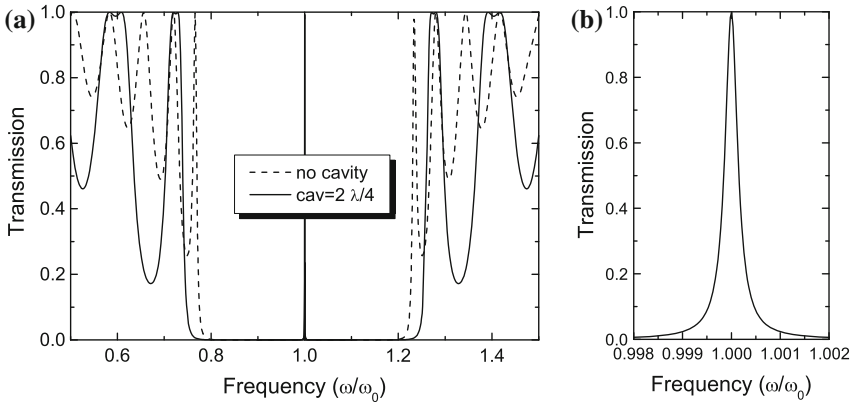


Fig. 19.14 Defect mode in 1D photonic band gap: (**a**) Transmission of $N = 11$ quarter-wave stack exhibiting a photonic band gap ($n_1 = 1, n_2 = 2$) (*dashed line*) and of microcavity (*solid line*) with center n_2-region of width $\lambda_0/2$ (instead of $\lambda_0/4$). (**b**) Relative width of mode is about 3×10^{-4}

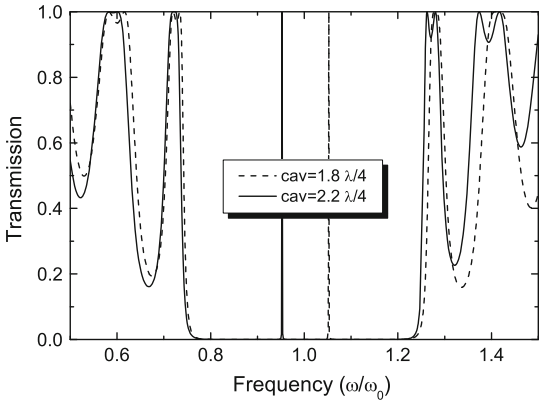


Fig. 19.15 Transmission of $N = 11$ quarter-wave stack ($n_1 = 1, n_2 = 2$) with center n_2-region of widths $1.8\,\lambda_0/4$ (*dashed line*) and $2.2\,\lambda_0/4$ (*solid line*)

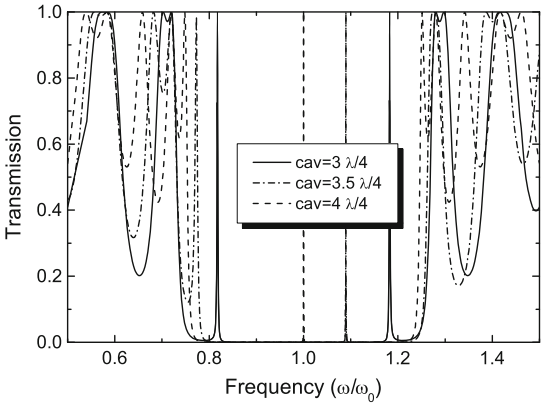


Fig. 19.16 Transmission of $N = 11$ quarter-wave stack ($n_1 = 1, n_2 = 2$) with center n_2-regions of widths $3\,\lambda_0/4$ (*solid line*), $3.5\,\lambda_0/4$ (*dash-dotted line*) and $4\,\lambda_0/4$ (*dashed line*)

2D or 3D Defect Modes

An example of 2D waveguiding is shown in Fig. 19.17. Point defects can be used for high-finesse wavelength filtering. Emitters surrounded by a photonic band gap material with a defect mode can emit into the defect mode only, leading to spectrally filtered, highly directional emission.

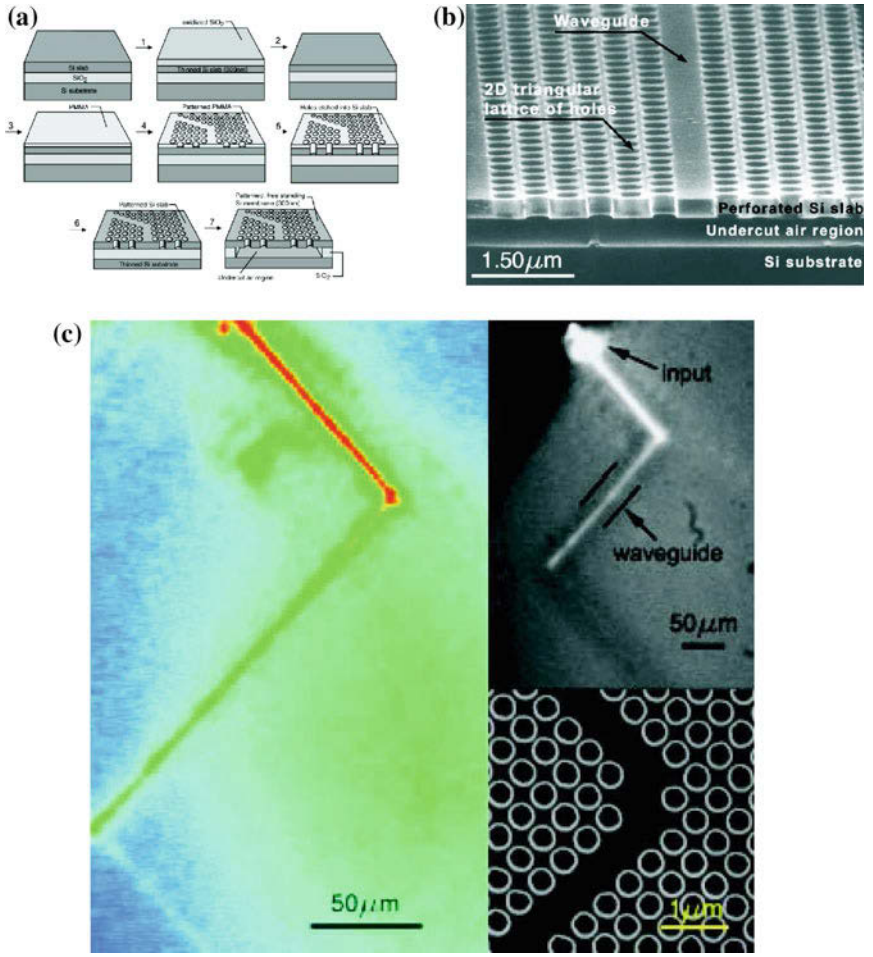


Fig. 19.17 2D photonic band gap waveguide structure. (**a**) Fabrication principle, (**b**) SEM image of the structure, (**c**) light guiding at a 90° bend. Reprinted with permission from [1350], ©2000 AIP

19.1.8 Coupling to an Electronic Resonance

In a vertical-cavity surface-emitting laser (cf. Sect. 23.4.14), an optical defect mode in a 1D dielectric structure is coupled to an electronic excitation, such as an exciton in a quantum well or dot. In the simplest picture, the oscillator must emit its radiation into the cavity mode since other modes do not exist in the Bragg band. Thus, the emission energy is given and fixed by the cavity mode. However, the photon mode (field oscillator) and the electronic oscillator form a coupled system that generally must be described using quantum electrodynamics. Energy is periodically exchanged between the two oscillators with the Rabi frequency. An analogous phenomenon is investigated in the field of atom–cavity interactions. A necessary condition for the observation of such an oscillation is that the radiation energy remains long enough in the cavity that can be expressed as [1351, 1352] (cf. (19.37))

$$\alpha\, d \gg 1 - R \approx \pi/Q, \tag{19.32}$$

where α is the absorption coefficient of the electronic transition, d is the length of the absorbing medium, R is the reflectance of the cavity mirror and Q is the finesse of the cavity given in (19.31). This situation is called the *strong coupling* regime since it leads to anticrossing behavior of the cavity mode and electronic resonance. In the *weak coupling* regime for small absorption, the resonances cross (within their linewidth). For resonance, the emission intensity of the oscillator into the cavity mode is enhanced and its lifetime is reduced (Purcell effect), which is discussed in Sect. 19.2.2.

The transmission T of a Fabry–Perot cavity with two (equal and lossless) mirrors of transmission $T_\mathrm{m} = 1 - R_\mathrm{m}$ is given by

$$T(\omega) = \frac{T_\mathrm{m}^2 \, \exp\left(-2\, L\, \alpha(\omega)\right)}{\left|1 - R_\mathrm{m} \, \exp\left(\mathrm{i}\, 2\, n^*\, L\, \omega/c\right)\right|^2}, \tag{19.33}$$

with the complex index of refraction $n^* = n_\mathrm{r} + \mathrm{i}\kappa = \sqrt{\epsilon}$ and $\alpha = 2\omega\kappa/c$ (cf. (9.9)). For an empty cavity, i.e. a (small) background absorption α_B and a background index of refraction $n_\mathrm{r} = n_\mathrm{B}$, the resonances occur when the phase shift $2n_\mathrm{B}L\omega/c$ is an integer multiple of 2π, i.e. for

$$\omega_\mathrm{m} = m\, \frac{\pi\, c}{n_\mathrm{B}\, L}, \tag{19.34}$$

with $m \geq 1$ being a natural number. In the vicinity of the resonance, i.e. for $\omega = \omega_\mathrm{m} + \delta\omega$, we can expand $\exp\left(2n_\mathrm{B}L\omega/c\right) \approx 1 + \mathrm{i}2n_\mathrm{B}L\delta\omega/c$ and obtain from (19.33) a Lorentzian for the transmission

$$T(\omega) \approx \frac{T_\mathrm{m}^2 \exp\left(-2L\alpha(\omega)\right)}{\left|1 - R_\mathrm{m}(1 + \mathrm{i}2n_\mathrm{B}L\, \delta\omega/c)\right|^2} = \frac{(T_\mathrm{m}/R_\mathrm{m})^2 \exp\left(2L\alpha(\omega)\right)}{(\delta\omega)^2 + \gamma_\mathrm{c}^2}. \tag{19.35}$$

The frequency width (HWHM) γ_c of the empty-cavity resonance is given by

$$\gamma_c = \frac{1 - R'}{R'} \frac{c}{2\, n_B\, L}, \tag{19.36}$$

where $R' = R_m \exp(-2L\alpha)$. Thus, the decay rate (photon loss from the cavity) is proportional to $T_m + \alpha_B L$ if both terms are small. The quality factor of the cavity resonance m is given by

$$Q = \frac{\omega_m}{2\,\gamma_c} \approx \frac{m\,\pi}{1 - R}. \tag{19.37}$$

Now, the electronic resonance is put into the cavity leading to a change in the dielectric function to (cf. (D.11))

$$\epsilon = n_B^2 \left[1 + \frac{f}{1 - (\omega^2 + \mathrm{i}\omega\Gamma)/\omega_0^2}\right], \tag{19.38}$$

where the index of refraction due to the electronic resonance is given by $n(\omega) = \sqrt{\epsilon}$ and (D.13a, b). For resonance of the cavity mode and the electronic oscillator, i.e. $\omega_m = \omega_0$, the solution for the cavity resonance condition $2n_r\omega L/c = m2\pi$ is obtained, using (19.34), from

$$n_r(\omega) = m\,\frac{\pi\, c}{\omega\, L} = n_B\,\frac{\omega_m}{\omega}. \tag{19.39}$$

A graphical solution (Fig. 19.18a) yields three intersections of the left and right hands of (19.39). The very high absorption at the central solution ($\omega = \omega_0$) results in very low transmission. The other two solutions[2] yield the frequencies of the coupled normal mode peaks. For $f \ll 1$, we use (D.13a) in (19.39) and find for the splitting $\pm\Omega_0/2$ of the two modes

$$\Omega_0^2 = f\,\omega_0^2 - \Gamma^2. \tag{19.40}$$

This frequency is called the Rabi frequency. If the dielectric function of the oscillator is put into (19.33), the splitting is found to be

$$\Omega_0^2 = f\,\omega_0^2 - (\Gamma - \gamma_c)^2. \tag{19.41}$$

A splitting will only be observable if $\Omega_0 \gg \Gamma, \gamma_c$. If the two resonances ω_c and ω_0 are detuned by $\Delta = \omega_c - \omega_0$, the splitting Ω of the transmission peaks shows the typical anticrossing behavior of two coupled oscillators

$$\Omega^2 = \Omega_0^2 + \Delta^2. \tag{19.42}$$

[2]These solutions only occur for sufficient oscillator strength $f > (\Gamma/\omega_0)^2$, i.e. in the strong coupling regime where $\Omega_0^2 > 0$. The absorption coefficient at ω_0 must be larger than $\Gamma n_\infty/c$.

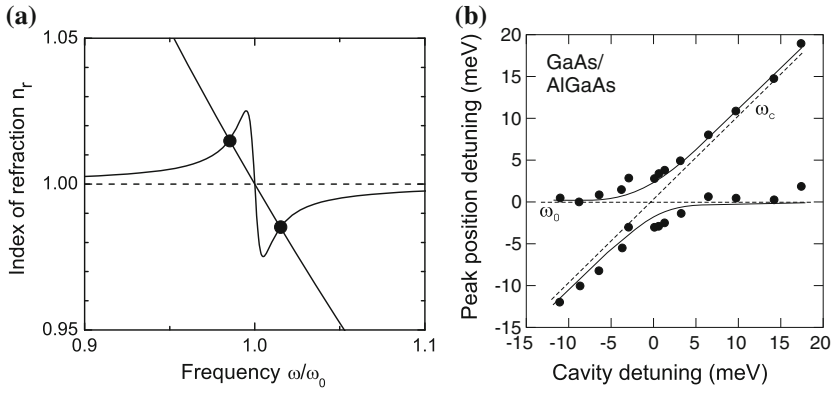


Fig. 19.18 (**a**) Graphical representation of (19.39) with the two solutions marked with *circles* for $n_\infty = 1$ (*dashed line*), $f = 10^{-3}$, $\Gamma/\omega_0 = 10^{-2}$ and $\omega_0 = \omega_m$. (**b**) Reflectance peak positions (experimental data (*circles*) at $T = 5\,\mathrm{K}$) versus cavity detuning $\omega_c - \omega_0$ for a cavity with two GaAs/AlGaAs Bragg mirrors (24/33 pairs for the front/bottom mirror) and five embedded quantum wells whose resonances are closely matched. *Solid lines* are a theoretical fit according to (19.42) with $\Omega_0 = 4.3\,\mathrm{meV}$. The *dashed lines* show the electronic resonance ω_0 and the cavity resonance ω_c. Part (**b**) adapted from [1352]

In the experiment, typically the electronic resonance remains fixed at ω_0 and the cavity resonance is detuned by variation of the cavity length across the wafer (Fig. 19.18b).

A detailed theory of cavity polaritons is given in [1353]. The nonlinear optics of normal mode coupling in semiconductor microcavities is reviewed in [1354].

The in-plane dispersion of the cavity polaritons depends on the coupling strength. First, the photon dispersion is given by

$$E_{\mathrm{ph}}(\mathbf{k}) = \hbar\omega = \hbar c\, k = \hbar c\, \left(k_\parallel^2 + k_z^2\right)^{1/2}, \tag{19.43}$$

where $k_\parallel$ is the in-plane k-vector and k_z is given by the resonance condition, $k_z = \omega_m/c$ with (19.34),

$$k_z = m\,\frac{\pi}{n_\mathrm{B}\,L}. \tag{19.44}$$

Thus the dispersion relation is no longer linear as for freely propagating light.

For small $k_\parallel$ this leads to an (in-plane) effective photon 'rest mass', applying (6.34),

$$\frac{1}{m^*_{\mathrm{ph}}} = \frac{1}{\hbar^2}\,\frac{\partial^2 E_{\mathrm{ph}}}{\partial k^2}. \tag{19.45}$$

We find

$$m^*_{\mathrm{ph}} = \frac{\hbar k_z}{c} = \frac{\hbar\omega(k_\parallel = 0)}{c^2}. \tag{19.46}$$

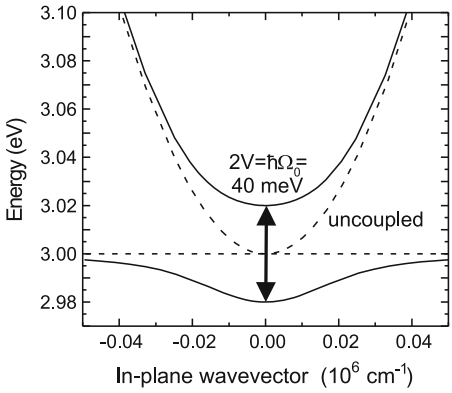


Fig. 19.19 Dispersion of cavity photon mode and electronic resonance at $E_{\mathrm{el}} = 3.0\,\mathrm{eV}$ (*dashed lines*) and coupled modes (*solid lines*) for $2V = 40\,\mathrm{meV}$

Now we assume the electronic oscillator to be in resonance with the photon dispersion at $k_\parallel = 0$, i.e. $E_{\mathrm{el}} = \hbar\omega(k_\parallel = 0)$. The electronic resonance shall have vanishing dispersion for simplicity since the exciton mass is much larger than (19.46). The eigenwert equation of the coupled system, resembling (6.57), is

$$\begin{vmatrix} E - E_{\mathrm{ph}} & V \\ V & E - E_{\mathrm{el}} \end{vmatrix} = 0, \tag{19.47}$$

with two solutions, called the upper and lower cavity polariton branch, visualized in Fig. 19.19. Their splitting at $k_\parallel = 0$ is $2V$. Thus the coupling parameter $V = \hbar\Omega_0/2$ corresponds [1353] to the Rabi frequency (19.40). Experimental values for the splitting of 3–15 meV in InAlGaAs based [1352, 1355–1358], 17–44 meV in CdZnTeSe based [1359], 6–60 meV in (Al,In,Ga)N based [1360–1364] and 78 meV in ZnO based [1365] microcavities were found. It is possible to condensate cavity polaritons in the minimum of the dispersion around $k_\parallel = 0$ (Bose–Einstein condensation). In [1366] stimulated scattering and gain from cavity polaritons have been reported. Further details on cavity polaritons can be found in [1367, 1368].

19.2 Microscopic Resonators

19.2.1 Microdiscs

A microdisc is a cylindrical resonator with a thickness d that is small compared to the radius R. It can be fabricated from semiconductors and semiconductor heterostructures using patterning and material-selective etching. With underetching a mostly free-standing disc can be made that resides on a post (Fig. 19.20).

The coordinate system is (ρ, ϕ, z) with the z direction being perpendicular to the disc area. Typically, the disc is so thin that there is only one node along z. Solving the

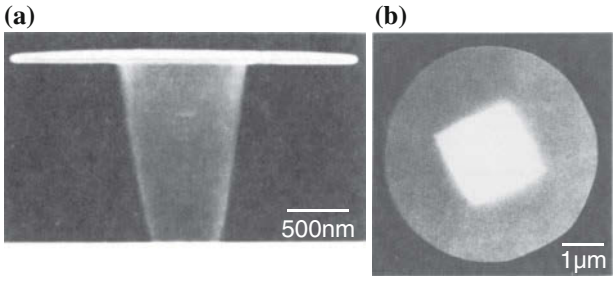


Fig. 19.20 (**a**) Side view of a 3-μm diameter disc containing one 10-nm InGaAs quantum well between 20-nm InGaAsP barriers standing on an InP pillar that has been selectively underetched using HCl. (**b**) Top view SEM image of a 5-μm diameter InGaAsP microdisc. The pedestal shape is a rhombus due to anisotropic etching of the HCl. Adapted from [1369], reprinted with permission, ©1992 AIP

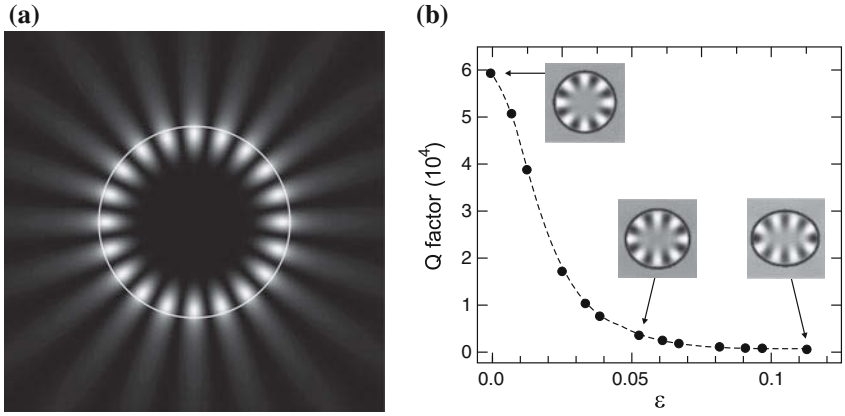


Fig. 19.21 (**a**) Field intensity for whispering gallery mode (10, 0) (TM-polarized) for a circle with 1 μm radius (shown as *white line*) and $n = 1.5$. The image size is $4 \times 4\,\mu\text{m}^2$. (**b**) Theoretical quality factor of a 2-μm InP microdisc as a function of the deformation parameter (19.49). The *insets* show (8,0) whispering gallery modes at a wavelength of 1.55 μm for $n = 3.4$. Part (**b**) adapted from [1372]

wave equation in this geometry [1370], the modes are characterized by two numbers (m, l). m describes the number of zeros along the azimuthal direction ϕ with the field amplitude being proportional to $\exp(\pm i m\phi)$. Thus, except for $m = 0$, the modes are simply degenerate. Modes with $E_z = 0$ are called TE modes. This is the preferred polarization of emission. The number l denotes the number of zeros in the radial direction. Only for modes with $|m| = 1$, is the intensity nonzero on the axis, i.e. for $\rho = 0$. All other modes have vanishing intensity in the disc center.

The light intensity in whispering gallery modes is preferentially concentrated along the circumference of the disc as shown in Fig. 19.21a. Since the light can only escape via evanescent waves, the light is well 'captured' in such a mode.

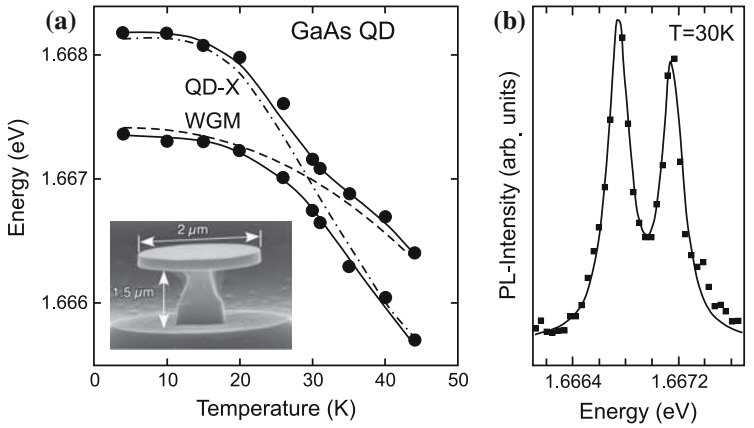


Fig. 19.22 Strong coupling of a single QD exciton (due to monolayer fluctuation in a 13 ML thick GaAs/$Al_{0.33}Ga_{0.67}As$ QW) with a WGM in a microdisk of 2 μm diameter (*inset*). (**a**) Anti-crossing of upper and lower peak for various temperatures. *Symbols* are data points, *solid lines* are theory considering coupling. The *dashed* (*dash-dotted line*) is the expected temperature shift of the WGM mode (exciton energy). (**b**) Photoluminescence spectrum at the anti-crossing point ($T = 30$ K). Experimental data (*squares*) and fit with two peaks (*solid line*). Adapted from [1373]

The Q-factor (19.31) is extremely high and takes values of several 10^4. In order to couple light out of such a disc, deformed resonators, e.g. with a defect in the form of protrusions [1371], were devised. Deformed resonators are discussed in more detail in the next section.

The strong coupling of a QD exciton to a whispering gallery mode is shown in Fig. 19.22 where anti-crossing behavior is observed at low temperatures. Tuning is achieved by temperature variation. Behavior of a similar system in the weak coupling regime is shown in Fig. 19.24.

19.2.2 Purcell Effect

According to Fermi's golden rule (19.1), the probability of an optical transition depends on the density of available optical modes (final states). If the density of modes is enhanced compared to its vacuum value (19.2) at a resonance of an optical cavity, the lifetime of the electronic state decreases by the Purcell factor [1374]

$$F_{\mathrm{P}} = \frac{3}{4\pi^2}\, Q\, \frac{(\lambda/n)^3}{V}, \tag{19.48}$$

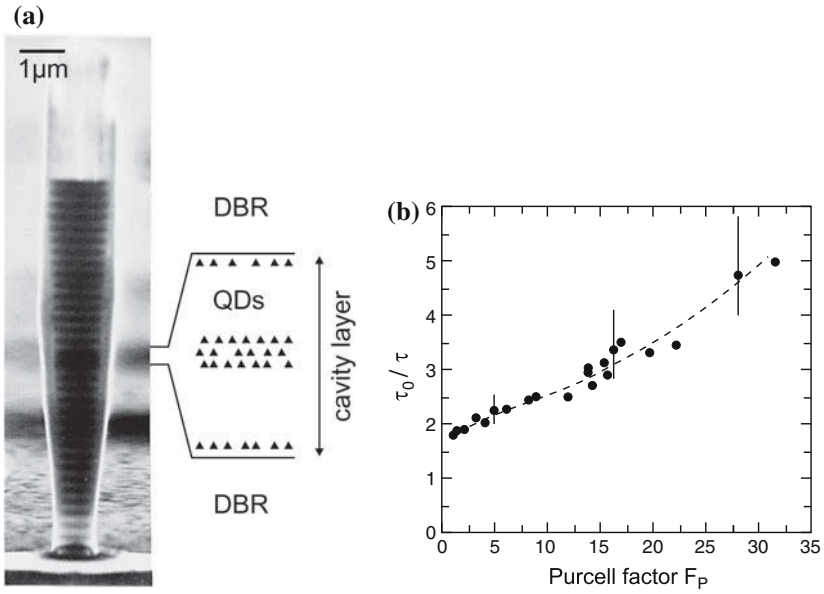


Fig. 19.23 (**a**) Micropillar with MBE-grown GaAs/AlAs DBRs and a cavity containing five layers of InAs quantum dots as indicated. The pillar has been prepared by reactive ion etching. Reprinted with permission from [1375], ©1998 APS. (**b**) Experimental decay time τ of on-resonance quantum dot luminescence scaled by off-resonance lifetime $\tau_0 = 1.1\,\text{ns}$ (close to lifetime in a QD in bulk) for a variety of micropillars with different Purcell factors F_P. The *error bars* correspond to the measurement accuracy of the decay time ($\pm 70\,\text{ps}$), the *dashed line* is a guide to the eye. Adapted from [1375]

where n is the refractive index of the medium, Q is the quality factor of the cavity resonance and V is the effective mode volume.[3] Experiments on the emission of quantum dots (that generally provide small absorption and thus allow for the weak coupling regime) in etched micropillars containing a microcavity (Fig. 19.23a) have shown that indeed the luminescence decay is faster for cavities with large Purcell factor (Fig. 19.23b) [1375]. The resonance of cavity mode and emitter leads to an enhanced emission intensity as shown in Fig. 19.24 for the exciton emission of a single quantum dot in a microdisc [1376].

19.2.3 Deformed Resonators

The whispering gallery modes in circular (or spherical) cavities are long-lived and emission goes into all angles. Light escape is based only on the exponentially slow

[3] V is given by the spatial integral of the vacuum field intensity for the cavity mode, divided by its maximum value.

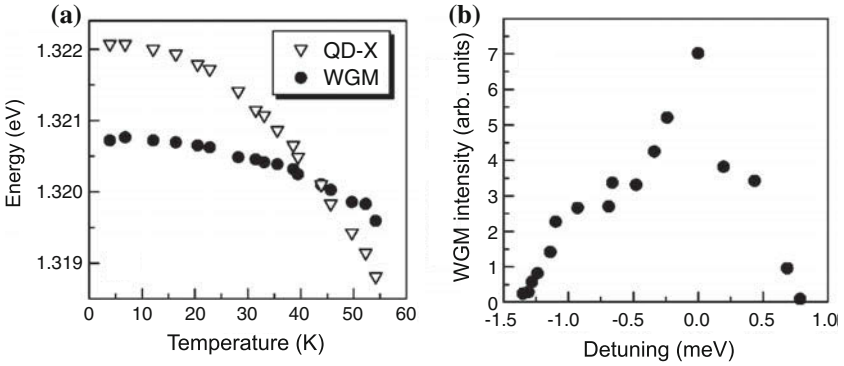


Fig. 19.24 (**a**) Temperature dependence of the energy positions of the whispering gallery mode (WGM) of a 5-μm diameter AlGaAs/GaAs microdisc ($Q = 6500$) and the single-exciton resonance of a single InAs quantum dot contained within the disc. (**b**) Intensity of WGM mode as a function of the detuning $E_{\mathrm{WGM}} - E_{\mathrm{QD-X}}$ from the QD single exciton resonance. The excitation density was $15\,\mathrm{Wcm}^{-2}$ for all data. Adapted from [1376]

process of evanescent leakage (neglecting disorder effects such as surface roughness). In order to overcome the isotropic light emission, the resonator needs to be deformed. This can be accomplished with an ellipsoidal shape, i.e.

$$r(\phi) = R\,[1 + \epsilon\,\cos\phi]\,, \tag{19.49}$$

where $1 + 2\epsilon$ is the aspect ratio of the ellipse. The increased radiation leads to a decrease of the Q-factor as shown in Fig. 19.21b. Also, a new decay process, refractive escape, becomes possible. A ray that is initially in a whispering gallery trajectory diffuses in phase space until finally an angle smaller than the critical angle for total reflection (9.11) is reached. The ray dynamics becomes partially chaotic [1377].

One other possible deformation of the circular disc geometry is a 'flattened quadrupole' as shown in Fig. 19.25a. This shape can be parameterized by a deformation parameter ϵ and the angle-dependent radius $r(\phi)$ given by

$$r(\phi) = R\,\left[1 + 2\,\epsilon\,\cos^2(2\phi)\right]^{1/2}. \tag{19.50}$$

For small deformation, the whispering gallery modes become chaotic and exhibit preferred emission along the long axis of the resonator (Fig. 19.25b). For larger deformations ($\epsilon \geq 0.14$), a stronger and qualitatively different directionality occurs in the shape of a bow-tie [1378] as shown in Fig. 19.25c. The optical laser power extracted from deformed resonators was found to increase exponentially with ϵ; for $\epsilon = 0.2$ it was 50 times larger than for the circular resonator.

Another modification that can be applied to the microdisc in order to increase outcoupling of light, is the spiral resonator [1379] as shown in Fig. 19.26a. The radius is parameterized by

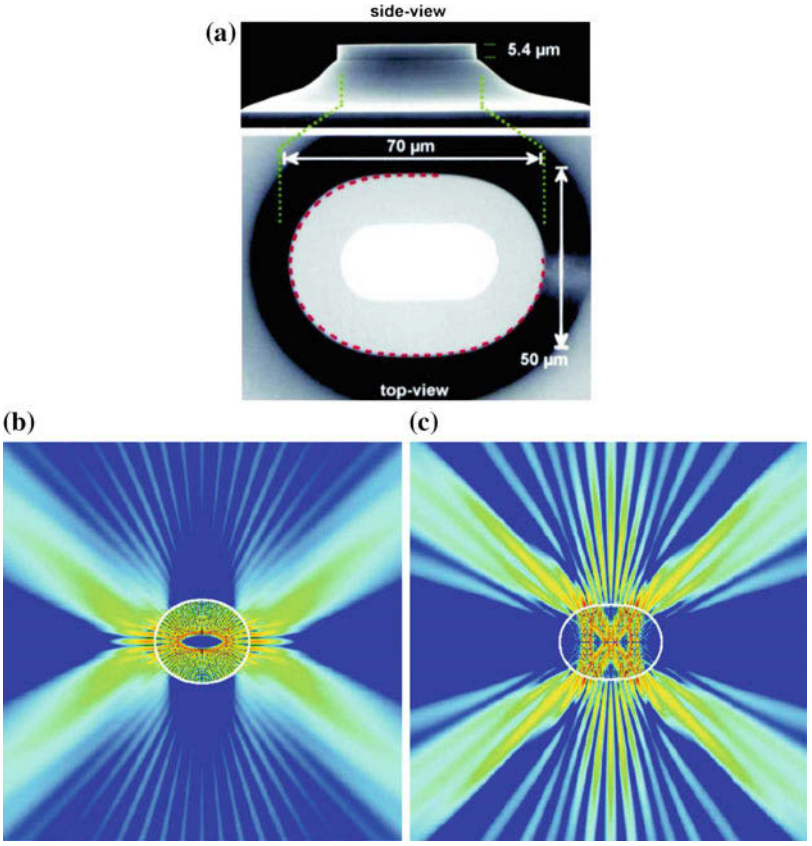


Fig. 19.25 (**a**) SEM image of a quadrupolar cylinder laser with deformation parameter $\epsilon \approx 0.16$ on a sloped InP pedestal. The *light grey area* in the top view is the electrical contact. (**b**) Simulated near-field intensity pattern of a chaotic whispering gallery mode for $\epsilon = 0.06$ and $n = 3.3$. (**c**) Simulated near-field intensity pattern of a bow-tie mode for $\epsilon = 0.15$. The length of the minor axis for (**b**) and (**c**) is 50 µm. Reprinted with permission from [1378], ©1998 AAAS

$$r(\phi) = R\left[1 + \frac{\epsilon}{2\pi}\,\phi\right]. \tag{19.51}$$

The experimental emission pattern is displayed in Fig. 19.26b. It exhibits a maximum along the direction of the tangent at the radius step. The simulated near-field intensity of such an emission mode is shown in Fig. 19.26c. In a spiral laser, ray dynamics is also chaotic [1380].

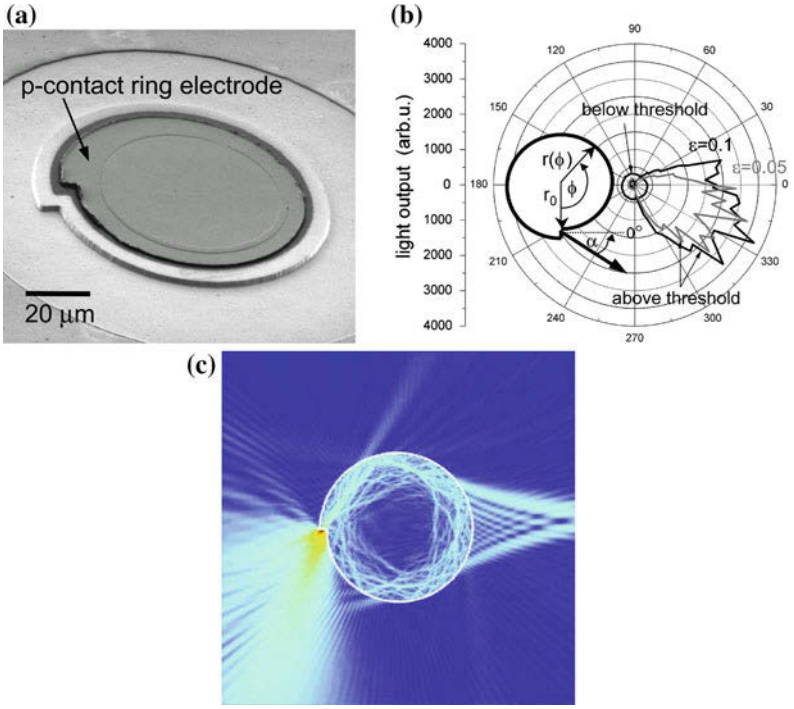


Fig. 19.26 (**a**) SEM image of a microcavity disc laser diode with a disc radius of 50 µm. The p-contact ring electrode defines the areas through which carriers are injected into the microdisc and where stimulated emission can take place. (**b**) Radial distribution of the light output from the spiral-shaped microdisc laser diode measured below and above threshold. The radius of the spiral microdisc was $r_0 = 250\,\mu\text{m}$ and the deformation parameters were $\epsilon = 0.05$ (*grey*) and $\epsilon = 0.10$ (*black*). An emission beam at an angle of $\alpha = 0°$ corresponds to a direction normal to the notch surface as shown in the *inset*. Below the laser threshold, the emission pattern is essentially isotropic and independent of the deformation parameter. Above the threshold, directional emission is clearly observed with the emission direction at a tilt angle $\alpha \approx 25°$. The measured divergence angle of the far-field pattern is $\sim 75°$ for $\epsilon = 0.10$ and $\sim 60°$ for $\epsilon = 0.05$. Reprinted with permission from [1381], ©2004 AIP. (**c**) Simulated near-field intensity pattern of an emission mode with $nkR \approx 200$ for deformation $\epsilon = 0.10$. Reprinted with permission from [1379], ©2003 AIP

19.2.4 Hexagonal Cavities

Hexagonal cavities develop, e.g., in microcrystals of wurtzite semiconductors (with the *c*-axis along the longitudinal axis of the pillar). In Fig. 19.27a, a ZnO tapered hexagonal resonator (needle) is shown. Whispering gallery modes modulate the

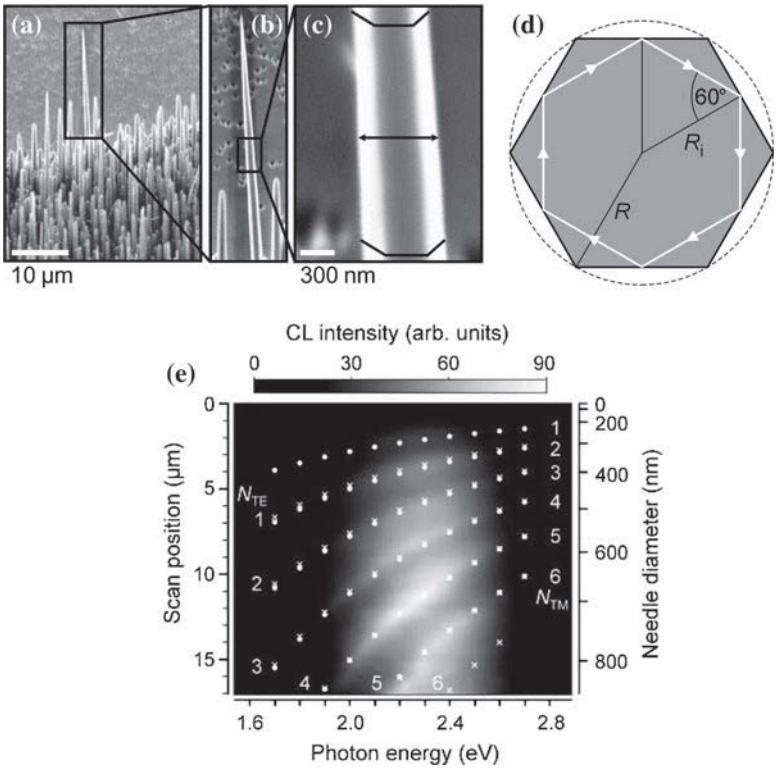


Fig. 19.27 (**a**)–(**c**) SEM images of ZnO nanoneedle fabricated by pulsed laser deposition. (**d**) Schematic geometry of cross-sectional plane. R_i (R) is the radius of the incircle (circumscribing circle). The circumference of the inscribed *white hexagon*, representing the path of a whispering gallery mode, has a length of $6R_i$. (**e**) Two-dimensional plot of spectra recorded along a linescan along the needle's longitudinal axis. The left vertical axis shows the linescan position x, the right one refers to the respective needle diameter D. The spectral maxima, i.e. the measured WGM energies, appear as bright belts going from the bottom left corner to the right upper one. With decreasing diameter, all resonances shift systematically to higher energies. The *white dots* give theoretical TM-resonance energy positions obtained from (19.52), *white crosses* give the same for TE-polarization. Reprinted with permission from [1382], ©2004 APS

intensity of the green ZnO luminescence [1382].[4] In a simple plane-wave model, the resonance condition is given by

$$6R_i = \frac{h\,c}{n\,E}\left[N + \frac{6}{\pi}\arctan\left(\beta\sqrt{3n^2-4}\right)\right], \tag{19.52}$$

[4]We note that besides the green luminescence as in Fig. 10.20, an unstructured green band also occurs that is observed here. Its origin may be linked to the oxygen vacancy [1383].

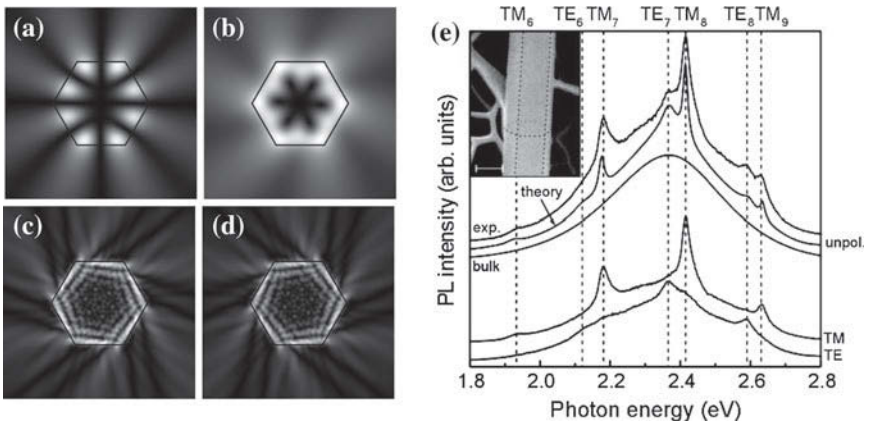


Fig. 19.28 Simulated near-field intensity pattern of modes in a cavity with hexagonal cross section (absolute value of electric field in linear grey scale): Modes ($N = 4$) with (**a**) symmetry $-a$ and (**b**) mode 4+ (nomenclature from [1384]) for $n = 2.1$ and $kR = 3.1553 - \mathrm{i}0.0748$. Modes (**c**) 26− and (**d**) 26+ for $n = 1.466$ and $kR = 22.8725 - \mathrm{i}0.1064$. The displayed modes have a chiral pattern. Emission originates mostly from the corners. (**e**) Micro-photoluminescence spectra of a single ZnO nanopillar. The three topmost curves are unpolarized. The curve labeled 'bulk' shows the unmodulated luminescence of the green luminescence in bulk. The line labeled 'exp.' shows the experimental μ-PL spectrum of the investigated nanopillar. The experimental spectra recorded for TM- and TE-polarization, respectively, are shown in the lowest two curves. The curve labeled 'theory' displays the theoretical luminescence spectra. *Dashed vertical lines* are guides to the eye referring to the spectral position of the dominating WGMs. The *inset* shows a SEM image of the investigated pillar, the scale bar has a length of 500 nm. The *dotted lines* show the position of the edges of the hexagonal resonator obtained from topography contrast

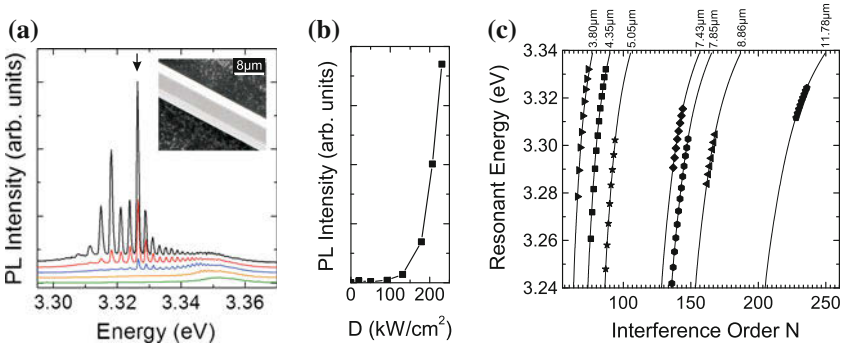


Fig. 19.29 (**a**) Photoluminescence spectra of a ZnO microwire with hexagonal cross section for various pump power densities (*lowest curve*: $D = 60\,\mathrm{kW/cm^2}$, *top curve*: $D = 250\,\mathrm{kW/cm^2}$) at $T = 10\,\mathrm{K}$. The inset shows the scanning electron microscopy image of a typical microwire ($d = 6.40\,\mu\mathrm{m}$). (**b**) Dependence of the emitted PL intensity of a selected lasing peak (denoted by an *arrow* in the spectrum in part (**a**)) on the excitation density D. *Lines* are guide to the eyes. (**c**) Dependence of the resonant energies on the interference order N for wires with different diameters as labeled on top of the graph. *Lines* are the predicted theoretical values calculated from (19.52) using diameter values obtained from SEM measurements; the *symbols* represent the experimentally observed peaks. Adapted from [1385]

where R_i is the radius of the inner circle (Fig. 19.27d), n is the index of refraction, N is the mode number and β is given by $\beta_{TM} = 1/n$ ($\beta_{TE} = n$) for TM (TE) polarization, respectively. Due to birefringence, $n_{||}$ ($n_{\perp}$) has to be used as the index of refraction for TM (TE) polarization.

A $N = 26$ whispering gallery mode of a hexagonal resonator is shown in Fig. 19.28c, d. The 6-fold symmetric emission stems from the edges of the hexagon. While whispering gallery resonators have typically mode numbers $N \gg 1$, in such hexagonal resonators the whispering gallery modes could be followed down to $N = 1$ [1382] as shown in Fig. 19.27a, b, e.

Under high optical pumping laser action occurs on the whispering gallery modes. The peak positions, close to the band gap in the spectral region of the electron-hole plasma, follow (19.52) [1385], as shown for various diameters in Fig. 19.29. Pumping threshold even at room temperature is below 100 kW/cm^2 [1386].

Chapter 20
Transparent Conductive Oxide Semiconductors

Many of the most important semi-conductors are oxides.
A.H Wilson, 1939 [65]

Abstract Typical materials in the class of transparent conductive oxides are mentioned. Their physical properties and the limits of conductivity versus transparency are discussed.

20.1 Introduction

Transparent conductive oxides (TCO) are semiconductors that are simultaneous transparent and highly conductive. Therefore they can serve as transparent contacts, e.g. as a solar cell front contact or in display applications. The materials are typically fabricated in the form of thin films on glass, polymers or similar substrates and devices. The crystallographic structure is polycrystalline or amorphous. The first TCO investigated was CdO in pressed powder [1387] and thin film form [31]. The recent historic development of the resistivity of the most important TCO materials in the last 30 years is shown in Fig. 20.1. Further information on TCO films can be found in [1388–1390].

20.2 Materials

Any wide-gap ($E_g > 3\,eV$) semiconductor that is conductive, e.g. due to intrinsic defects or by chemical impurities (doping), can be considered to be a TCO. Practically only a few, non-toxic materials that can be easily deposited are of importance. The first TCO application was heating of air plane windows. As always in semiconductor technology, price drives the suitability for applications. This in particular true for TCO applications since they include large area devices such as solar

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_20

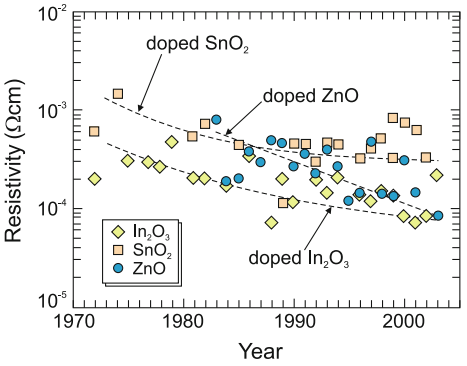


Fig. 20.1 Historical development of the resistivity of doped ZnO (*circles*), In_2O_3 (*squares*) and SnO_2 (*diamonds*) TCO films. Adapted from [1390]

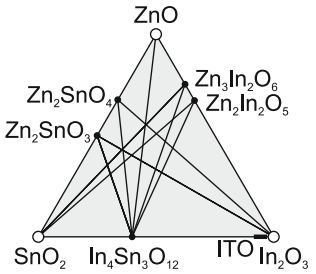


Fig. 20.2 Practical TCO materials

cells, displays and also large glass panes for electromagnetic shielding and architectural heat and IR transparency management. Therefore large quantities of TCO are needed. The popular ITO (indium-tin-oxide) suffers from large indium price and potential indium scarcity, opening the field for aluminum-doped ZnO (ZAO) which is abundant. Possible other compounds including Cd are of no practical interest due to toxicity. Conductive GaN has not been considered due to its large processing temperatures. All practical TCO materials contain either Zn, Sn or In (Fig. 20.2). A number of TCO materials is reviewed in [1391]. TCO are mostly n-conducting. Also p-conducting TCOs have been reported, e.g. $CuAlO_2$ delafossite [1392] with room temperature conductivity of $\sigma = 1$ S/cm, $ZnIr_2O_4$ [1393] with $\sigma = 2$ S/cm and $ZnCo_2O_4$ spinel [1394] with $\sigma = 20$ S/cm. However, there has been so far no report on a practical transparent highly conductive p-type electrode.

The term ITO stands for a variety of Sn-doped indium oxide (In_2O_3) materials, the Sn content being typically in the 5–10 % range but not strictly defined. The crystal structure of In_2O_3 [1395] is shown in Fig. 20.3. The effect of the replacement of indium by tin atoms in In_2O_3 on mechanical, electrical and optical properties of ITO has been calculated in [1396] using DFT. In Fig. 20.4 the band structures of pure In_2O_3 and $(Sn_{0.065}In_{0.935})_2O_3$ (one out of 16 indium atoms was replaced in the calculation) are compared [1397]. The fundamental band gap is slightly lowered, the high doping introduces a gigantic Burstein-Moss shift. Additionally another band gap opens that splits the lowest conduction band. Also the conduction mechanism in amorphous oxides has been discussed [1398].

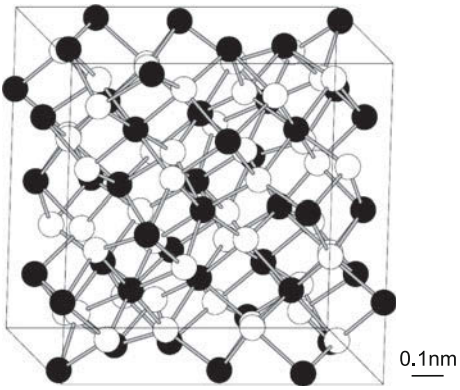


Fig. 20.3 The bixbyite crystal structure of indium oxide ($In_{32}O_{48}$) showing one unit cell where the indium and oxygen atoms are represented by the *full* and *empty circles*, respectively. Adapted from [1396]

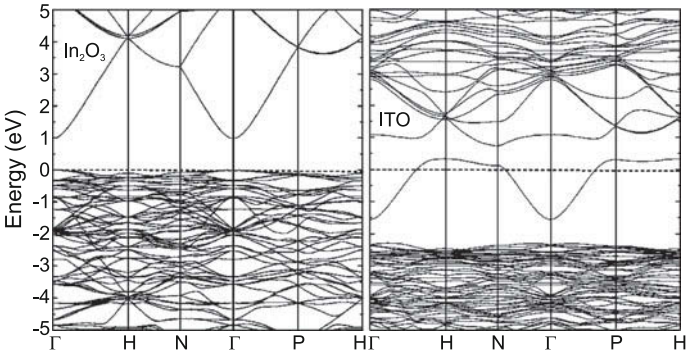


Fig. 20.4 Comparison of the band structure of In_2O_3 (*left*) and $(Sn_{0.065}In_{0.935})_2O_3$ (*right*). The position of the Fermi level is for both cases at $E = 0$. Adapted from [1397]

20.3 Properties

The best conductivity of TCOs is in the range of $10^{-4}\,\Omega$cm for the specific resistivity. Such value is about three orders of magnitude smaller than that of metals. However, the TCO meanwhile is highly transparent in the visible region, while metals become transparent only in the UV region because of their high plasma frequency (cf. Sect. 9.8.1).

The conduction mechanism in ZnO:Al is band transport in a highly doped semiconductor. The carrier concentration is typically around $10^{21}\,\text{cm}^{-3}$. One of the best results is a (Hall) mobility of $47.6\,\text{cm}^2/\text{Vs}$, leading to a specific resistivity of $8.5 \times 10^{-5}\,\Omega$cm [1399]. The mobility is limited by (doping level dependent) ionized impurity scattering (Sect. 8.3.3) as shown in Fig. 20.5 for various films. A detailed discussion of ionized impurity scattering in doped ZnO films can be found in [1400].

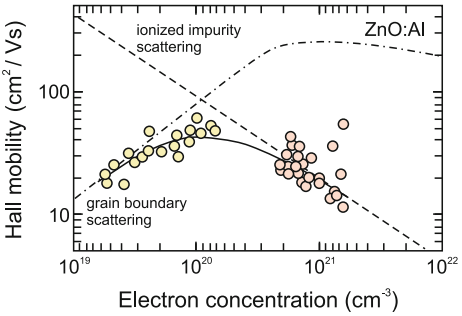


Fig. 20.5 Hall mobility for various ZnO:Al TCO thin films as a function of carrier (electron) concentration. The *dashed line* is the Brooks–Hering theory of ionized impurity scattering (taking into account non-parabolicity of the conduction band). The *dash-dotted line* is mobility in the presence of grain boundaries, the *solid line* is combined theory. *Symbols* are experimental data from two sets of samples. Adapted from [1390]

Also the mobility is correlated (Fig. 20.6) with scattering at structural defects such as grain boundaries [1401] (cf. Fig. 8.5). We note that the carrier mobility in a (polycrystalline) TCO is not very different from that in highly doped (crystalline) silicon (Fig. 8.7).

Conductivity and transparency are, however, linked. The high doping of the TCO leads to shifts in the band gap (renormalization and Burstein-Moss shift), band tails and the like that can introduce absorption in the visible spectral region. Also the infrared transparency is related to the conductivity by free carrier absorption and the plasma edge (Sect. 9.8.1). With increasing carrier density, the plasma edge shifts into the visible spectral range (Fig. 20.7a), limiting the possible maximum carrier density to several $10^{21}\,\mathrm{cm}^{-3}$, the exact value depending on the carrier mass. In Fig. 20.7b the transparency spectra of two SnO_2 films with different conductivity are compared. The higher conductivity due to larger carrier concentration leads to reduced IR

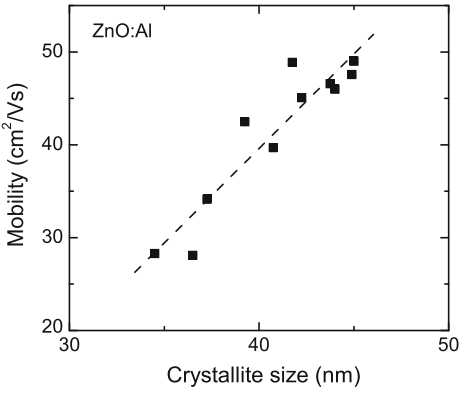


Fig. 20.6 Hall mobility of ZnO:Al films (of varying thickness) versus the crystallite size. Experimental data (from [1390]) are shown as *symbols*. The *dashed line* is a guide to the eyes

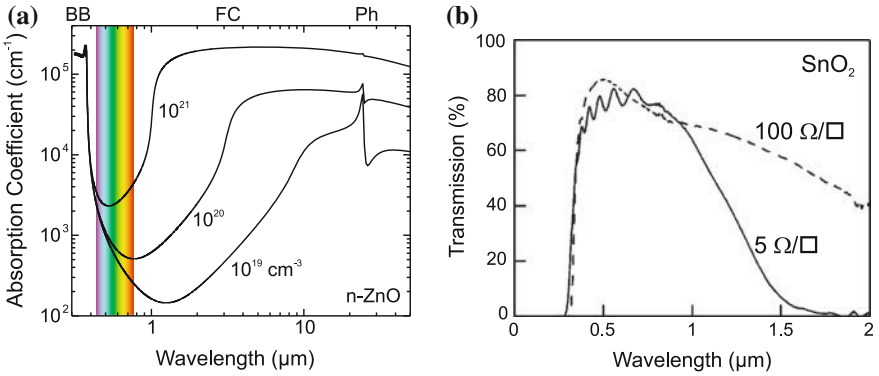


Fig. 20.7 (**a**) Absorption coefficient of n-type ZnO for various electron concentrations as labeled. Spectra synthesized from ellipsometric data of thin films. The visible spectral range is indicated. The contributing absorption processes are band-band transitions (BB), free carrier absorption (FC) and phonon-related absorption (Ph) as labeled. (**b**) Transparency versus wavelength spectra of two SnO_2 films with different conductivity as labeled. Adapted from [1402]

transparency. Generally, the transmission of a TCO is limited by the band edge on the high energy side of the spectrum and the plasma edge on the low energy side of the spectrum. The limited carrier concentration due to the onset of free carrier absorption in the visible spectral range ($\sim 3 \times 10^{21}\,\mathrm{cm}^{-3}$) and the limited mobility in the presence of such high impurity concentration (max. $50\,\mathrm{cm}^2/\mathrm{Vs}$) restrict the minimum resistivity of a TCO (or any transparent conductor) to about $4 \times 10^{-5}\,\Omega\,\mathrm{cm}$ [1403].

Part III
Applications

Chapter 21
Diodes

Abstract A thorough treatment of Schottky (metal–semiconductor) diodes, MIS (metal–insulator–semiconductor) diodes and (bipolar) pn-diodes is given, focussing on suitable materials, the formation of space charge layers and the forward and reverse current–voltage characteristics. Applications of such devices based on their rectifying properties are discussed.

21.1 Introduction

One of the simplest[1] semiconductor devices is the diode. It is a so-called two-terminal device, i.e. a device with two leads. The most prominent property of a diode is the rectifying current–voltage (I–V) characteristic. This function was initially realized with vacuum tubes (Fig. 21.1); a heated filament emits electrons that are transferred through vacuum to the anode if it is on a positive potential. The semiconductor diode technology led to a tremendous miniaturization, integration with other devices (in planar technology) and cost reduction.

We distinguish[2] *unipolar* diodes, for which the *majority* carriers cause the effects (e.g. metal–semiconductor diodes), and *bipolar* diodes in which *minority* carriers play the decisive role, e.g. in the pn-junction diode.

21.2 Metal–Semiconductor Contacts

The metal–semiconductor contact was investigated in 1874 by F. Braun (see Sect. 1.1). For metal sulfides, e.g. $CuFeS_2$, he found nonohmic behavior. We remark here that we treat first metal–semiconductor contacts with rectifying properties. Later it becomes understandable that metal–semiconductor contacts can also be used as

[1] The simplest device is a resistor made from a homogeneous piece of semiconductor, used, e.g., as a part of an integrated circuit or as a photoresistor as discussed in Sect. 22.2.

[2] This distinction is not only made for diodes but also many other semiconductor devices such as transistors, see Chap. 24.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_21

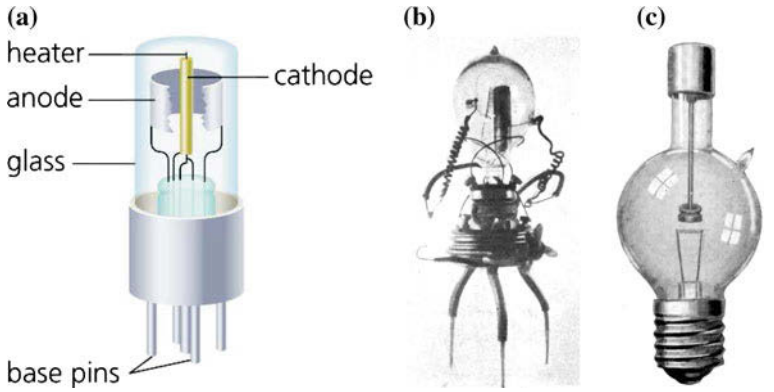


Fig. 21.1 (**a**) Schematic image of a vacuum diode. The electron current flows from the heated cathode to the anode when the latter is at a positive potential. (**b**) One of John A. Fleming's first diode 'valves', 1904. (**c**) Commercial 'tungar' rectifier, around 1916 [1405]. (**b**, **c**) adapted from [1404]

ohmic contacts, i.e. contacts with a very small contact resistance. Rectifying metal–semiconductor contacts are also called Schottky diodes. A very important variation are metal–insulator–semiconductor diodes for which an insulator, mostly an oxide, is sandwiched between the metal and the semiconductor. Such diodes are treated in Sect. 21.3. Reviews on Schottky diodes can be found in [1406–1412].

21.2.1 Band Diagram in Equilibrium

The metal and the semiconductor have generally different positions of the Fermi levels relative to the vacuum level. When the metal is in contact with the semiconductor, charges will flow in such a way that in thermodynamic equilibrium the Fermi level is constant throughout the structure.[3] In the following we treat two limiting cases: The contact of a metal with a semiconductor without any surface states (Schottky–Mott model) and a contact where the semiconductor has a very high density of surface states (Bardeen model).

The position of the Fermi level in the metal is given by the work function W_m that is shown in Fig. 21.2 for various metals (see also Table 21.1). A recent review of published values of metal work functions is given in [1413]. The work function reflects the atomic shell structure; minima of the work function exist for group-I elements. The work function is the energy difference between the vacuum level (an electron is at rest in an infinite distance from the metal surface) and the metal Fermi level ($W_\mathrm{m} > 0$). Since the electron density in the metal conduction band is very high,

[3] This situation is similar to the heterostructure interface (Sect. 12.3.4), with the metal, however, having a very short screening length.

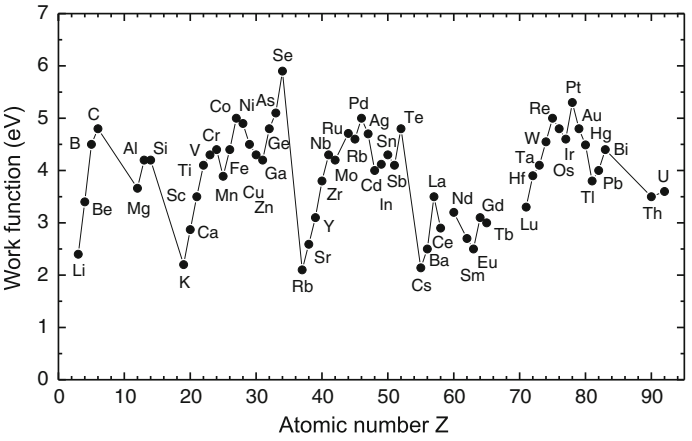


Fig. 21.2 Work function W_m of various metals

Table 21.1 Values of the work function W_m of various metals

Z	Element	W_m (eV)	Z	Element	W_m (eV)	Z	Element	W_m (eV)
3	Li	2.4	37	Rb	2.1	64	Gd	3.1
4	Be	3.4	38	Sr	2.59	65	Tb	3.0
5	B	4.5	39	Y	3.1	66	Dy	–
6	C	4.8	40	Zr	3.8	67	Ho	–
12	Mg	3.66	41	Nb	4.3	68	Er	–
13	Al	4.2	42	Mo	4.2	69	Tm	–
14	Si	4.2	44	Ru	4.71	70	Yb	–
19	K	2.2	45	Rh	4.6	71	Lu	3.3
20	Ca	2.87	46	Pd	5.0	72	Hf	3.9
21	Sc	3.5	47	Ag	4.7	73	Ta	4.1
22	Ti	4.1	48	Cd	4.0	74	W	4.55
23	V	4.3	49	In	4.12	75	Re	5.0
24	Cr	4.4	50	Sn	4.3	76	Os	4.8
25	Mn	3.89	51	Sb	4.1	77	Ir	4.6
26	Fe	4.4	52	Te	4.8	78	Pt	5.3
27	Co	5.0	55	Cs	2.14	79	Au	4.8
28	Ni	4.9	56	Ba	2.5	80	Hg	4.49
29	Cu	4.5	57	La	3.5	81	Tl	3.8
30	Zn	4.3	58	Ce	2.9	82	Pb	4.0
31	Ga	4.2	59	Pr	–	83	Bi	4.4
32	Ge	4.8	60	Nd	3.2	90	Th	3.5
33	As	5.1	62	Sm	2.7	92	U	3.6
34	Se	5.9	63	Eu	2.5			

the position of the metal Fermi level does not change considerably when charge is exchanged between the metal and the semiconductor.

Since in a semiconductor the Fermi level depends strongly on the doping and temperature it is not useful to characterize the material itself. For semiconductors the electron affinity $\chi_{sc} = E_{vac} - E_C > 0$ is defined as the energy difference between the vacuum level and the conduction-band edge (see Fig. 12.2).

Ideal Band Structure

When the metal and the semiconductor are not in contact (Fig. 21.3a), the metal is characterized by its work function $W_m = E_{vac} - E_F$ and the semiconductor by its electron affinity χ_{sc}. First, we assume that $W_m > \chi_{sc}$. For an n-type semiconductor, the energy difference between the Fermi level and the conduction band is denoted as

$$-e\,V_n = E_C - E_F \tag{21.1}$$

being negative, $V_n < 0$, for nondegenerate semiconductors. Thus, the position of the semiconductor Fermi level is given as

$$E_F = E_{vac} - \chi_{sc} + e\,V_n. \tag{21.2}$$

If the metal and semiconductor are brought into contact the Fermi levels will equilibrate. For the case of Fig. 21.3 ($E_{F,sc} > E_{F,m}$) electrons will flow from the semiconductor to the metal. The negative surface charge of the metal is compensated by a positive charge (due to D^+) in the semiconductor in the vicinity of the surface. Eventually a (Schottky) barrier of height[4] F_{Bn}

$$F_{Bn} = W_m - \chi_{sc} \tag{21.3}$$

forms at the interface. The subscript 'n' stands for the contact on an n-type semiconductor. Surface/interface effects such as non-matching bonds, surface states, etc. are neglected at this point. In the semiconductor there exists a positively charged region that is called the *depletion layer* or space-charge region [1414]. Its extension (w in Fig. 21.3b) and properties will be discussed in Sect. 21.2.2. The space charge region in the metal is very thin due to the small screening length.

For a contact on a p-type semiconductor the barrier F_{Bp} (to the valence band) is (see Fig. 21.4d)

$$F_{Bp} = E_g - (W_m - \chi_{sc}). \tag{21.4}$$

Between the surface of the metal and the bulk part of the semiconductor there is a potential drop

$$V_{bi} = \frac{F_{Bn}}{e} + V_n = \frac{W_m - \chi_{sc}}{e} + V_n, \tag{21.5}$$

[4] We denote the energy barrier height with $F_B = -e\phi_B$.

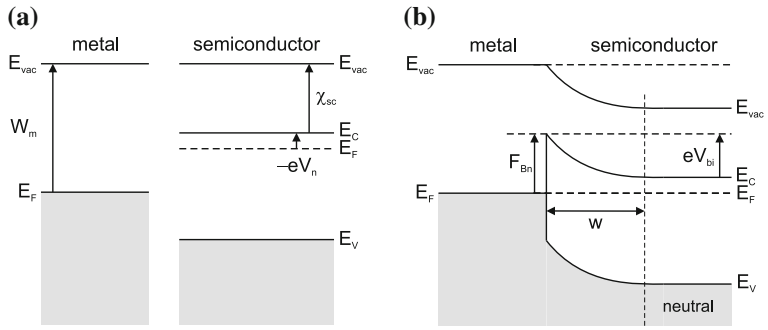


Fig. 21.3 Schematic band structure of a metal–semiconductor junction that is dominated by bulk properties of the semiconductor. (**a**) no contact, (**b**) metal and semiconductor in contact. w denotes the width of the depletion layer. Outside the depletion layer the semiconductor is neutral. $F_{\mathrm{B,n}}$ denotes the Schottky barrier height, V_{bi} denotes the built-in voltage (here $V_{\mathrm{bi}} > 0$)

which is termed the built-in potential (or diffusion voltage). The exact form of the voltage drop, the so-called band bending, will be discussed in Sect. 21.2.2.

The surface index is defined as

$$S = \frac{\partial F_{\mathrm{Bn}}}{\partial W_{\mathrm{m}}}. \tag{21.6}$$

From the present consideration (21.3), the same semiconductor with metals of varying work function should result in $S = 1$.

Band Structure in the Presence of Surface States

Experimental data shown in Fig. 21.5a, however, show a different behavior with smaller slope. For GaAs, e.g., the barrier height is almost independent of the metal work function. Thus, a different model is needed for realistic Schottky diodes. A rule of thumb for the dominantly covalent semiconductors is that for n-type material the barrier height is 2/3 of the band gap and for p-type material 1/3 of the band gap, such that $E_{\mathrm{C}} - E_{\mathrm{F}} \approx 2E_{\mathrm{g}}/3$ (Fig. 21.5b). Only for ionic semiconductors $S \approx 1$ holds (Fig. 21.5c) [1419].

If the semiconductor has a large density of states at its surface ($\gtrsim 10^{12}\,\mathrm{cm}^{-2}$), there is a space-charge region already without the metal [1420]. Surface traps are filled up to the Fermi level (Fig. 21.6a). The size of the band bending in the semiconductor will be denoted as F_{Bn} since it will turn out below as the Schottky barrier height. If the density of surface states is very high, the charge carriers moving from the semiconductor into the metal upon contact formation are accommodated in the surface states and the position of the Fermi level at the semiconductor surface changes only very little. Thus, the space-charge region is not modified and it is identical to the surface depletion region. The Schottky barrier height is then given by the band bending at the (bare) semiconductor surface F_{Bn} (Fig. 21.6d) and does not depend on the metal work

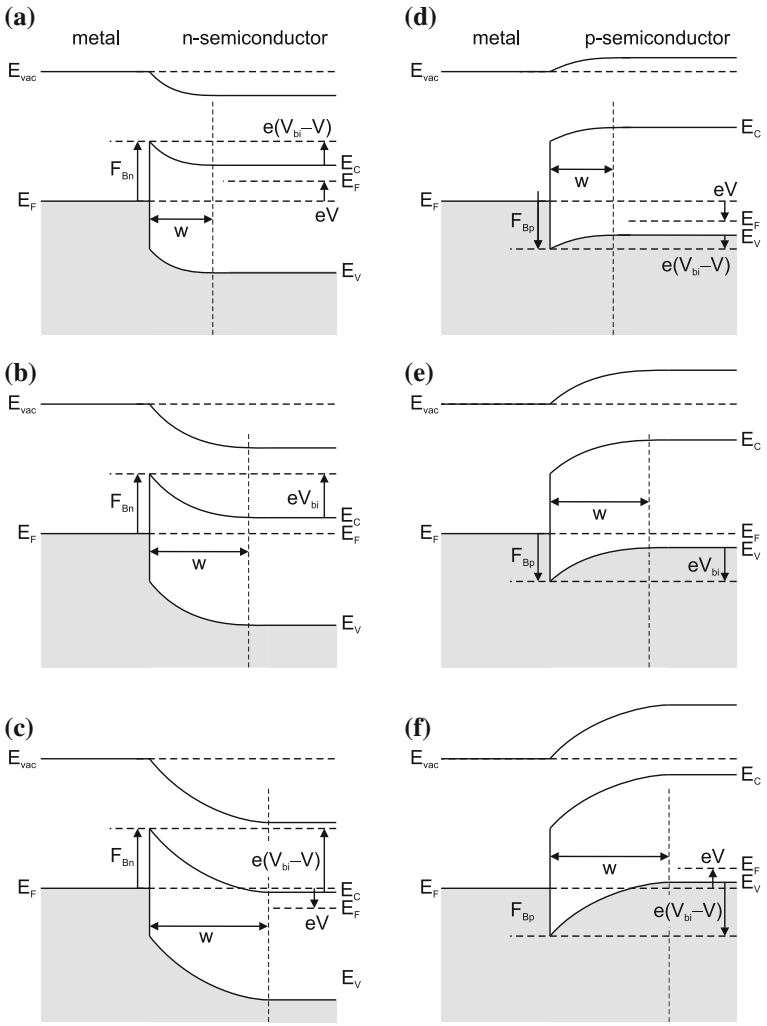


Fig. 21.4 Band structures of metal–semiconductor junctions for (**a**–**c**) an n-type semiconductor and (**d**–**f**) a p-type semiconductor (here $F_{\mathrm{Bp}} > 0$). (**b**, **e**) in thermodynamic equilibrium, (**a**, **d**) with forward bias ($V > 0$), (**c**, **f**) with reverse bias ($V < 0$)

function at all (Bardeen model [1420]). For this limiting case we find for the surface index $S = 0$.

For actual metal–semiconductor contacts the surface index S takes values between 0 and 1. A theory involving the semiconductor band structure and midgap (surface) states (MIGS) is needed [617, 1421]. For Si, the experimental result is $S = 0.27$; the corresponding density of surface states is $D_{\mathrm{s}} = 4 \times 10^{13}\,\mathrm{cm}^{-2}\,\mathrm{eV}^{-1}$.

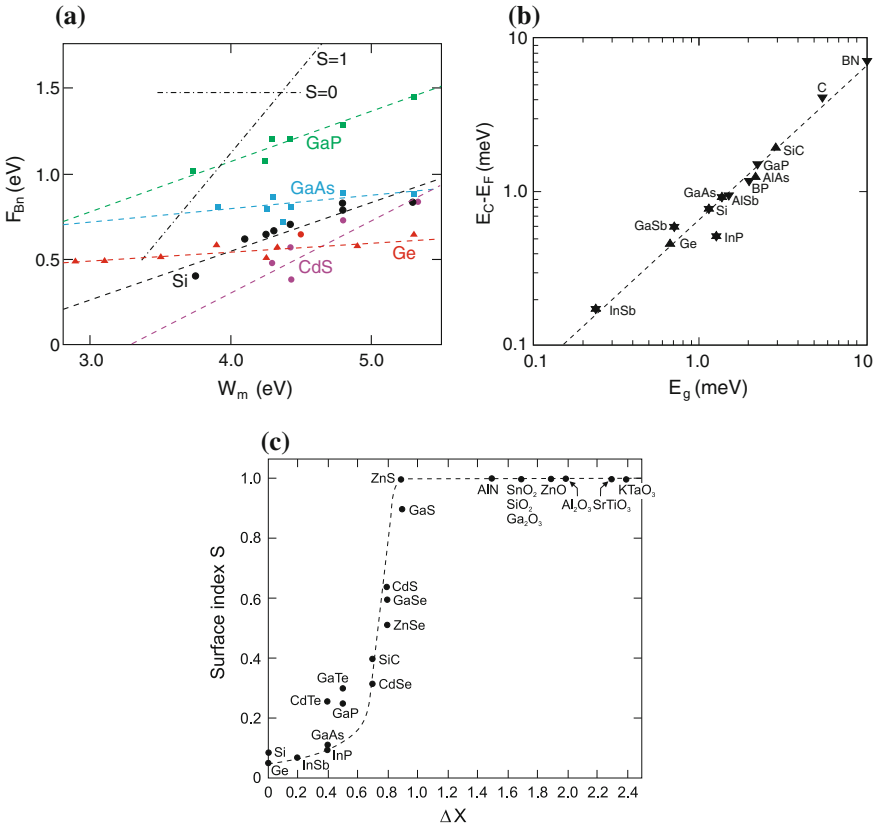


Fig. 21.5 (**a**) Experimental Schottky barrier heights F_{Bn} versus metal work function W_m for various metal–semiconductor junctions as labeled. *Dashed lines* are guides to the eye, *dash-dotted lines* indicate dependencies for $S = 1$ and $S = 0$. Data from [1415, 1416]. (**b**) $E_C - E_F$ at the metal–semiconductor interface versus the band gap E_g for Au Schottky contacts on various semiconductors. The *dashed line* represents $E_C - E_F = 2E_g/3$. Data from [1417]. (**c**) Surface index S versus the electronegativity difference ΔX between the species of compound semiconductors. *Dashed line* is a guide to the eye. Data from [1416, 1418]

21.2.2 Space-Charge Region

The width w of the space-charge region is calculated next. First, we make the so-called abrupt approximation. In this approximation (Schottky–Mott model), the charge density ρ in the space-charge region ($0 \leq x \leq w$) is given by the doping, i.e. $\rho = +eN_D$. Outside the space-charge region the semiconductor is neutral, i.e. $\rho = 0$ and the electric field is zero, i.e. $d\varphi/dx = 0$. As further boundary conditions the potential at the interface is $\varphi(0) = -V_{bi} < 0$. The potential drop in the space-charge region is determined by the one-dimensional Poisson equation

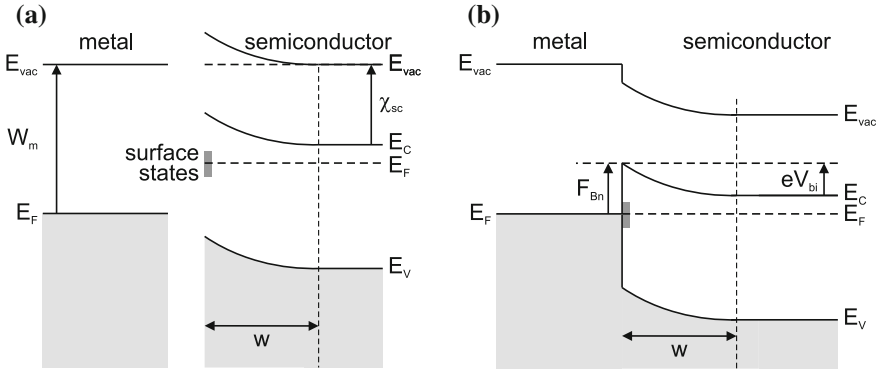


Fig. 21.6 Schematic band structure of a metal–semiconductor junction that is dominated by surface properties of the semiconductor. (**a**) No contact; due to pinning of the Fermi level at surface states of the semiconductor, a depletion-layer of width w is already present. (**b**) Metal and semiconductor in contact

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}x^2} = -\frac{\rho}{\epsilon_\mathrm{s}}, \tag{21.7}$$

where $\epsilon_\mathrm{s} = \epsilon_\mathrm{r}\,\epsilon_0$ is the static dielectric constant of the semiconductor. Using the ansatz $\varphi(x) = \varphi_0 + \varphi_1 x + \varphi_2 x^2$ we find

$$\varphi(x) = -V_\mathrm{bi} + \frac{e\,N_\mathrm{D}}{\epsilon_\mathrm{s}}\left(w_0\,x - \frac{1}{2}\,x^2\right), \tag{21.8}$$

w_0 being the depletion layer width at zero bias.

The spatial dependence $E(x)$ of the electric field strength is

$$E(x) = -\frac{e\,N_\mathrm{D}}{\epsilon_\mathrm{s}}(w_0 - x) = E_\mathrm{m} + \frac{e\,N_\mathrm{D}}{\epsilon_\mathrm{s}}\,x, \tag{21.9}$$

with the maximum field strength $E_\mathrm{m} = -eN_\mathrm{D}\,w_0/\epsilon_\mathrm{s}$ at $x = 0$. From the condition $\varphi(w_0) = 0$ we obtain w_0 as

$$w_0 = \sqrt{\frac{2\,\epsilon_\mathrm{s}}{eN_\mathrm{D}}\,V_\mathrm{bi}}. \tag{21.10}$$

The charge density and the potential in the abrupt approximation are shown in Fig. 21.7 for GaAs material parameters.

Beyond the abrupt approximation, the thermal distribution of the majority carriers must be treated with more care. The dependence of the charge density $\rho = e(N_\mathrm{D}^+ - n)$ on the potential φ (within the Boltzmann approximation) is ($\beta = e/kT$)

$$\rho = e\,N_\mathrm{D}\left[1 - \exp\left(\beta\,\varphi\right)\right]. \tag{21.11}$$

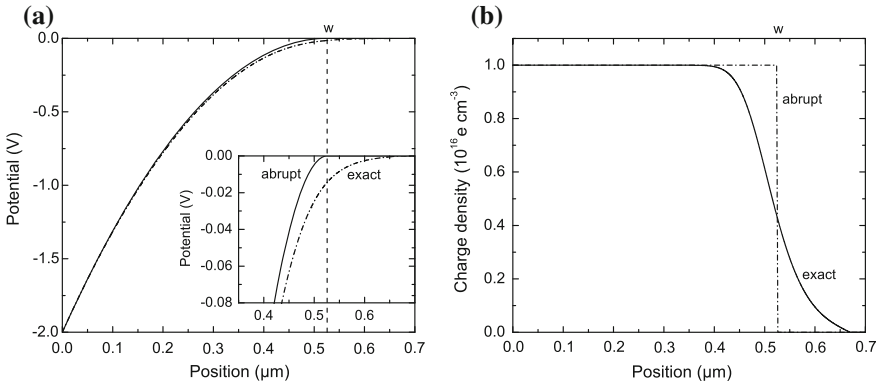


Fig. 21.7 (**a**) Potential φ and (**b**) charge density ρ across the depletion layer of a Schottky n-GaAs diode. Calculation parameters are $\epsilon_s = 12.5\,\epsilon_0$, $V_{bi} - V_{ext} = 2\,V$ (small reverse bias), $N_D = 1 \times 10^{16}\,cm^{-3}$, $T = 300\,K$. Abrupt approximation is shown as *solid lines*, exact (numerical) calculation as *dash-dotted line*. *Dashed line* in (**a**) indicates depletion layer width w_0 in the abrupt approximation

The actual charge density and the potential, obtained from a numerical solution of (21.7), are shown in Fig. 21.7 in comparison with the abrupt approximation. Clearly, at the Schottky depletion layer width w_0, the charge varies continuously and the potential does not drop to zero.

We note that for the depletion layer $\varphi \leq 0$ and $n \leq N_D$. The charge difference $\Delta\rho$ (due to the tail of the thermal distribution of the majority charge carriers in the depletion layer) between the real distribution (21.11) to the abrupt approximation model with constant charge density ($\rho_0 = eN_D$) in the depletion layer is

$$\Delta\rho(x) = \rho(x) - \rho_0 = -e\,N_D\,\exp\left(\beta\,\varphi(x)\right). \tag{21.12}$$

The integration of $\Delta\rho$ over the depletion layer yields that the voltage drop V_{bi} across the depletion layer needs to be corrected by ΔV

$$\Delta\varphi = \int_0^{w_0}\left[\int_0^x \frac{-\Delta\rho(x')}{\epsilon_s}\,dx'\right]dx = \frac{1}{\beta}\left[1 - \exp(-\beta\,V_{bi})\right] \approx \beta^{-1}. \tag{21.13}$$

The approximation is valid for $\beta V_{bi} \gg 1$. Therefore, (21.10) is corrected to

$$w_0 = \sqrt{\frac{2\,\epsilon_s}{e\,N_D}\left(V_{bi} - \beta^{-1}\right)}. \tag{21.14}$$

When a potential difference V_{ext} is applied externally to the diode, (21.14) is modified by the change in the interface boundary condition, $\varphi(0) = -V_{bi} + V_{ext}$. The band structure is shown schematically in Fig. 21.4a for a forward bias and in Fig. 21.4c for a reverse bias. Therefore, we obtain for the depletion-layer width

(within the abrupt approximation)

$$w(V_{\text{ext}}) = \sqrt{\frac{2\,\epsilon_{\text{s}}}{e\,N_{\text{D}}}\left(V_{\text{bi}} - V_{\text{ext}} - \beta^{-1}\right)}. \tag{21.15}$$

Now we can also give explicitly the value of the maximum electric field E_{m} (at $x = 0$)

$$\begin{aligned} E_{\text{m}} &= -\sqrt{\frac{2\,e\,N_{\text{D}}}{\epsilon_{\text{s}}}\left(V_{\text{bi}} - V_{\text{ext}} - \beta^{-1}\right)} \\ &= -\frac{2}{w}\left(V_{\text{bi}} - V_{\text{ext}} - \beta^{-1}\right). \end{aligned} \tag{21.16}$$

We note that so far the barrier height is independent of the applied bias voltage. In the next section, it is shown that this is actually not the case.

An alternative approach to solve the Poisson equation is to consider the field as a function of the potential, $E(\phi)$. Then,

$$\frac{\mathrm{d}E^2}{\mathrm{d}\phi} = 2\,E\,\frac{\mathrm{d}E}{\mathrm{d}\phi}, \tag{21.17}$$

and

$$\frac{\mathrm{d}E}{\mathrm{d}\phi} = \frac{\mathrm{d}E}{\mathrm{d}x}\,\frac{\mathrm{d}x}{\mathrm{d}\phi} = -\frac{\mathrm{d}E}{\mathrm{d}x}\,\frac{1}{E} = -\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2}\,\frac{1}{E}. \tag{21.18}$$

Combining these two equations and using (21.7) and (21.11) we find

$$\frac{\mathrm{d}E^2}{\mathrm{d}\phi} = -\frac{2\,e\,N_{\text{D}}}{\epsilon_{\text{s}}}\left(1 - \exp(\beta\,\phi)\right). \tag{21.19}$$

The integration of this equation yields with the boundary condition $E(\phi = 0) = 0$,

$$E^2(\phi) = -\frac{2\,e\,N_{\text{D}}}{\epsilon_{\text{s}}}\left(\phi - \frac{\exp(\beta\,\phi)}{\beta} + \frac{1}{\beta}\right). \tag{21.20}$$

At the interface $\phi(x = 0) = -(V_{\text{bi}} - V)$, and for sufficient voltage drop in the semiconductor $\exp\beta\,\phi \ll 1$, and therefore (21.16) for E_{m} is recovered without using the abrupt approximation or its correction (21.13).[5] Without approximation,

[5]The functional integration method is limited to bijective potentials $\phi(x)$, i.e. strictly monotonously falling or rising potentials [1422] and thus covers monotonously varying doping density within the depletion layer.

$$E_\mathrm{m} = -\sqrt{\frac{2\,e\,N_\mathrm{D}}{\epsilon_\mathrm{s}}\left(V_\mathrm{bi} - V_\mathrm{ext} - \frac{1}{\beta}\left[1 - \exp(-\beta(V_\mathrm{bi} - V_\mathrm{ext}))\right]\right)}. \tag{21.21}$$

21.2.3 Schottky Effect

The barrier height is reduced by the image-charge effect that has been neglected so far. This effect has been worked out for metals by Schottky [1423] and has been adapted to semiconductors [1424, 1425]. An electron (charge $q = -e$) at position x in the semiconductor is facing a metal surface. The metal surface is at zero position (Fig. 21.8a). The potential distribution of the free charge is modified since the metal surface is an equipotential surface. The potential distribution outside the metal is identical to that if an *image* charge $-q$ were located at $-x$. This image charge exerts a force (image force F_if) on the electron

$$F_\mathrm{if} = -\frac{q^2}{16\,\pi\,\epsilon_\mathrm{s}\,x^2}, \tag{21.22}$$

where ϵ_s is the relative dielectric constant of the semiconductor. In order to bring an electron to x from infinity the work E_if

$$E_\mathrm{if} = \int_\infty^x F_\mathrm{if}\,\mathrm{d}x = -\frac{q^2}{16\,\pi\,\epsilon_\mathrm{s}\,x} \tag{21.23}$$

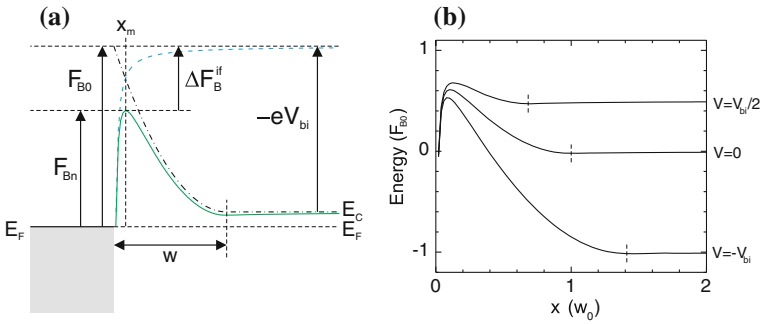


Fig. 21.8 (**a**) Energy of a particle with respect to the metal surface (*dashed line*), conduction band in semiconductor depletion layer (*dash-dotted line*) and combined effect (*solid line*). The image charge energy lowers the potential barrier F_{B_0} by the amount $\Delta F_\mathrm{B}^\mathrm{if}$ to F_Bn. (**b**) Conduction band on the semiconductor side of a metal–semiconductor junction at various bias voltages ($V = 0, +V_\mathrm{bi}/2$, and $-V_\mathrm{bi}$ as labeled) taking into account the Schottky effect. The width of the depletion layer is indicated with a *short vertical dashed line*. The barrier height without Schottky effect is F_B0. The *dashed line* is the situation without Schottky effect for zero bias

is needed. This image potential energy is shown in Fig. 21.8a. The total energy E_{tot} (solid line in Fig. 21.8a) of the electron in the presence of an electric field E is given by

$$E_{\text{tot}} = q\,E\,x - \frac{q^2}{16\,\pi\,\epsilon_{\text{s}}\,x}. \tag{21.24}$$

We note that the divergence at $x = 0$ stems from the continuum idealization of the problem. It can be removed by shifting the position of the image charge slightly against the interface or by using an extended charge [1050].

The maximum of this function ($\mathrm{d}E_{\text{tot}}/\mathrm{d}x = 0$) is at x_{m}

$$x_{\text{m}} = \sqrt{\frac{e}{16\,\pi\,\epsilon_{\text{s}}\,E}}. \tag{21.25}$$

The energy barrier (21.3) (without image charge effect now labeled as F_{B_0}) is reduced by ΔF_{B} given by

$$\Delta F_{\text{B}}^{\text{if}} = e\sqrt{\frac{e\,E}{4\pi\,\epsilon_{\text{s}}}} = 2\,e\,E\,x_{\text{m}}. \tag{21.26}$$

With the field in the vicinity of the interface given by E_{m} from (21.16), the barrier reduction is[6]

$$\Delta F_{\text{B}}^{\text{if}} = e\left[\frac{e^3\,N_{\text{D}}}{8\,\pi^2\,\epsilon_{\text{s}}^3}\left(V_{\text{bi}} - V_{\text{ext}} - \beta^{-1}\right)\right]^{1/4}. \tag{21.27}$$

For $\epsilon_{\text{s}} = \epsilon_0$ (vacuum) and a field strength of 10^5 V/cm the maximum position is at $x_{\text{m}} = 6$ nm and the barrier reduction is $\Delta F_{\text{B}}^{\text{if}} = 0.12$ eV. For 10^7 V/cm, $x_{\text{m}} = 1$ nm and $\Delta F_{\text{B}}^{\text{if}} = 1.2$ eV. For semiconductors with $\epsilon_{\text{r}} \sim 10$ the effect is smaller (Fig. 21.9). The Schottky effect depends on the bias voltage as visualized in Fig. 21.8b and therefore the barrier height depends on the applied bias voltage.

21.2.4 Capacitance

The total space charge Q (per unit area) in the semiconductor is ($V = V_{\text{ext}}$)

$$Q(V) = e\,N_{\text{D}}\,w = \sqrt{2\,e\,N_{\text{D}}\,\epsilon_{\text{s}}\left(V_{\text{bi}} - V - \beta^{-1}\right)}. \tag{21.28}$$

and depends on the external voltage.

[6]The term ϵ_{s}^3 is technically $\epsilon_{\text{s}}\epsilon_{\text{d}}^2$ where ϵ_{d} is the image-force dielectric constant. ϵ_{d} is equal to ϵ_{s} if the transit time of an electron from the metal to the maximum of the potential energy is sufficiently long to build up the dielectric polarization of the semiconductor [1406].

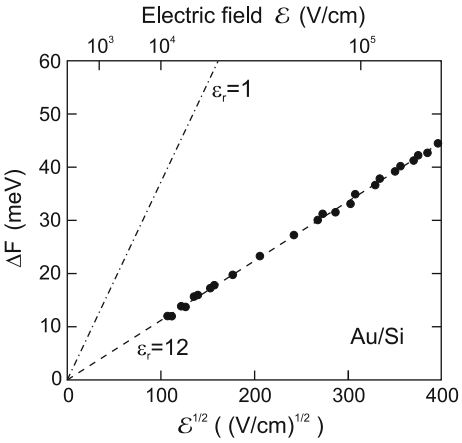


Fig. 21.9 Electric-field dependence of the image charge lowering of the Schottky barrier. *Dash-dotted* line is for vacuum dielectric constant, *dashed line* is (21.26) for $\epsilon_r = 12$. Adapted from [1426]

For measurement of the depletion layer capacitance first the external dc bias voltage V is set which defines the extension of the depletion layer. The (differential) capacitance is probed by an ac voltage with small amplitude $\delta V \ll V$. First we assume that the ac frequency ω is small compared to characteristic time constants of the electrically active impurities (quasi-static capacitance) and discuss the bias dependence $C(V)$. Following we discuss the frequency and temperature dependence of the capacitance[7] $C(\omega, T)$, in particular when the measurement frequency is in the range of the (temperature dependent) electron capture or emission rate (10.42, 10.43).

Bias Dependence

From (21.28) the capacitance $C = |\mathrm{d}Q/\mathrm{d}V|$ (per unit area) of the space charge region is given by

$$C = \sqrt{\frac{e\,N_\mathrm{D}\,\epsilon_\mathrm{s}}{2\left(V_\mathrm{bi} - V - \beta^{-1}\right)}} = \frac{\epsilon_\mathrm{s}}{w}. \tag{21.29}$$

Equation (21.29) can also be written as

$$\frac{1}{C^2} = \frac{2\left(V_\mathrm{bi} - V - \beta^{-1}\right)}{e\,N_\mathrm{D}\,\epsilon_\mathrm{s}}. \tag{21.30}$$

If $1/C^2$ is measured as a function of the bias voltage (C–V spectroscopy), it should be linearly dependent on the bias voltage if the doping concentration is homogeneous (Fig. 21.10a). The doping concentration can be determined from the slope via

$$N_\mathrm{D} = -\frac{2}{e\,\epsilon_\mathrm{s}}\left[\frac{\mathrm{d}}{\mathrm{d}V}\left(\frac{1}{C^2}\right)\right]^{-1}, \tag{21.31}$$

[7] Probing the capacitance as a function of the ac frequency is called admittance spectroscopy.

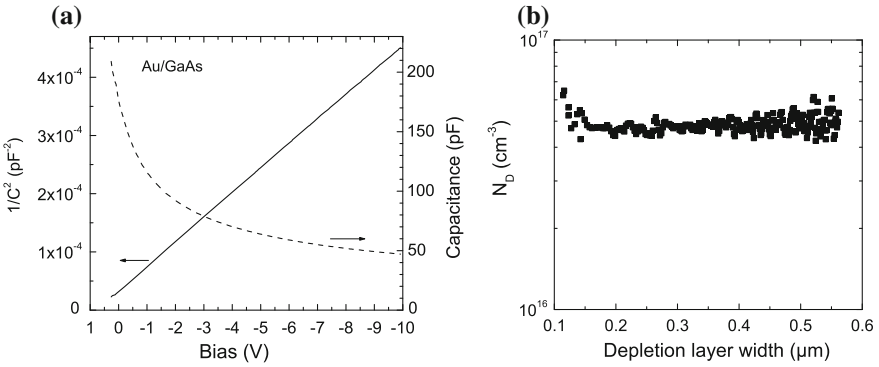


Fig. 21.10 (**a**) Capacitance C (*dashed line*) and $1/C^2$ (*solid line*) versus bias voltage dependence for an Au/GaAs Schottky diode (2-μm MOVPE-grown GaAs:Si on a n-GaAs substrate) at room temperature. From the extrapolation to $1/C^2 = 0$ and (21.30) we obtain $V_{\mathrm{bi}} = 804 \pm 3\,\mathrm{mV}$. (**b**) Donor concentration ($N_{\mathrm{D}} = 4.8 \times 10^{16}\,\mathrm{cm}^{-3}$) determined via (21.31) from the $1/C^2$ plot versus the depletion layer width (calculated using (21.29))

(see Fig. 21.10b) and the built-in voltage V_{bi} from the extrapolation to $V = V'$ such that $1/C^2 = 0$, $V_{\mathrm{bi}} = V' + kT/e$. Using (21.5) the Schottky barrier height can be determined from this [1427] via

$$F_{\mathrm{B_n}} = e\,V' - e\,V_{\mathrm{n}} + kT - \Delta F_{\mathrm{B}}^{\mathrm{if}}, \tag{21.32}$$

where $\Delta F_{\mathrm{B}}^{\mathrm{if}}$ is the barrier lowering (21.27) due to the image force effect between the flat-band and the zero-bias cases.

We note that for inhomogeneous doping the depth profile of the doping can be determined by C–V spectroscopy. The $1/C^2$ versus bias curve is then no longer a straight line and exhibits a varying slope. $N_{\mathrm{D}}(w)$ is evaluated according to (21.31) using $w = \epsilon_{\mathrm{s}}/C$ from (21.29) [1428],

$$N_{\mathrm{D}}\left(w = \frac{\epsilon_{\mathrm{s}}}{C}\right) = -\frac{2}{e\,\epsilon_{\mathrm{s}}}\left[\frac{\mathrm{d}}{\mathrm{d}V}\left(\frac{1}{C^2}\right)\right]^{-1}, \tag{21.33}$$

Using functional integration, the capacitance of a depletion layer can be expressed in terms of $\varphi(0) = V_{\mathrm{bi}} - V$ without explicit knowledge of the potential $\varphi(x)$ [1422].[8] In the approximation $e\varphi(0) \gg kT$ we obtain for homogeneous doping as refinement of (21.29)

$$C = \sqrt{\frac{e\,N_{\mathrm{D}}\,\epsilon_{\mathrm{s}}}{2\left(V_{\mathrm{bi}} - V - \beta^{-1}\left(\frac{n_0}{N_{\mathrm{D}}} - \ln\frac{n_0}{N_{\mathrm{D}}}\right)\right)}}, \tag{21.34}$$

[8] This is valid as long as $\varphi(x)$ is strictly monotonous.

where n_0 is the electron concentration in the neutral region. Within this general treatment, the validity of (21.31) has been confirmed. Also, C does not diverge in the flatband case, for $V \to V_{bi}$, as for the abrupt approximation but exhibits a maximum [1422].

At a given bias voltage, the charge (ionized donors or acceptors) at the boundary of the space-charge region is tested by the capacitance measurement. However, this principle works only if the depth of the space-charge region actually changes with the bias voltage. The method can therefore not be directly applied to such systems like δ-doped layers or quantum wells.

Frequency and Temperature Dependence

The release of carriers from (and capture on) a donor occurs with a characteristic emission rate g_c (10.43) (capture rate r_c). This is true similarly for acceptors. Therefore, the capacitance depends on the sampling frequency (Fig. 21.11a). If the capacitance is probed with a frequency much smaller than the release rate, the system appears to be in equilibrium and has the (quasi-)static capacitance C_0. If the probe frequency is much higher, the system cannot follow and the donor does not contribute to the capacitance. The characteristic frequency $\hat{f}$ at the turning point of $C(f)$ is [895]

$$2\pi\hat{f} = 2\,g_c, \tag{21.35}$$

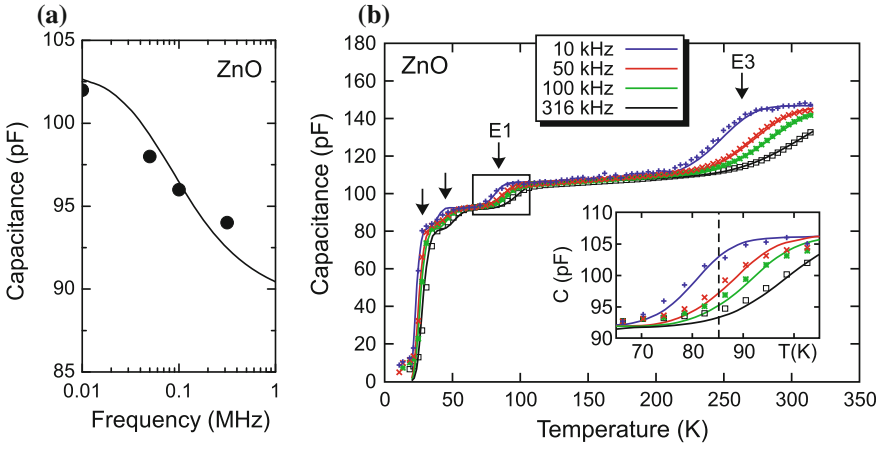


Fig. 21.11 (**a**) Capacitance versus probing frequency f for a Pd/ZnO Schottky diode (zero bias) at $T = 85$ K. Theoretical dependence (*solid line*) and experimental data (*circles*). (**b**) Capacitance of the same diode as a function of temperature (thermal admittance spectroscopy, TAS) for four different probing frequencies f =10, 50, 100 and 316 kHz (ac amplitude 50 mV). *Arrows* denote the release of carriers from four different defect levels, two shallow ones and the well-known defects E1 and E3 [655, 1432]. The *inset* shows the contribution of the E1 defect, indicated by a *rectangle* in the main graph, in more detail. *Symbols* are experimental data, *lines* are fit with four-level model (E1: $E_D = 116$ meV, E3: $E_D = 330$ meV) [1433]

with corrections of the simple factor of 2 discussed in [1429]. Since the emission rate depends exponentially on temperature, for a given frequency the capacitance depends on temperature [1430]. This is shown in Fig. 21.11b for ZnO which exhibits several donor levels. At low temperature shallow levels release their carriers, at higher temperature the deeper levels start to contribute. The DX center in AlGaAs (see Sect. 7.7.6) has been investigated in [1431] with this technique.

21.2.5 Current–Voltage Characteristic

The current transport through a metal–semiconductor junction is dominated by the majority charge carriers, i.e. electrons (holes) in the case of an n-type (p-type) semiconductor, respectively.

In Fig. 21.12, the possible transport mechanisms are visualized for an n-type semiconductor. Thermionic emission 'above' the barrier involves the hot electrons from the thermal distribution and will be important at least at high temperatures. Tunneling 'through' the barrier will be important for thin barriers, i.e. at high doping ($w \propto N_D^{-1/2}$, cf. (21.15)). 'Pure' tunneling for electrons close to the (quasi-) Fermi level, also called field emission, and thermionic field emission, i.e. tunneling of electrons with higher energies, are distinguished. Also, recombination in the depletion layer and hole injection from the metal are possible.

The transport of electrons above the barrier can be described with diffusion theory [1434, 1435] or thermionic-emission theory [1436]. A detailed treatment can be found in [1437, 1438]. In both cases the barrier height is large compared to kT. For thermionic emission (typically relevant for semiconductors with high mobility) the current is limited by the emission process and an equilibrium (constant electron quasi-Fermi level) is established throughout the depletion layer and ballistic transport is considered. In diffusion theory (for low mobility) a thermal equilibrium between metal and semiconductor electrons is established in the interface plane and the current is limited by diffusion and drift in the depletion region.

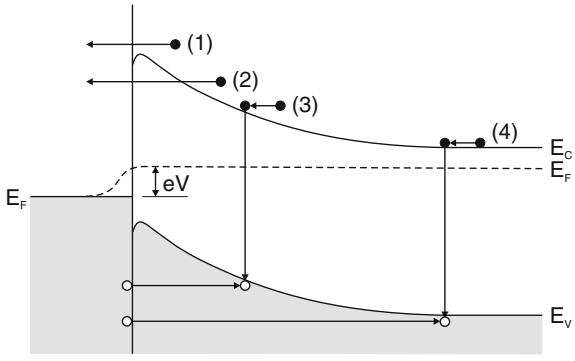


Fig. 21.12 Transport mechanisms at metal–semiconductor junctions. (1) Thermionic emission ('above' the barrier) (2) tunneling ('through' the barrier), (3) recombination in the depletion layer, (4) hole injection from metal

Thermionic Emission

The current density per unit area $j_{s\to m}$ of electrons that flow from the semiconductor into the metal is due to the hot electrons from the thermal distribution function

$$j_{s\to m} = \int_{E_F+F_{Bn}}^{\infty} (-e)\, v_x \, dn. \tag{21.36}$$

The integral starts at the lowest possible energy, the top of the Schottky barrier (no tunneling allowed in this model!). The electron density dn in a small energy interval dE is

$$dn = D(E) f(E)\, dE. \tag{21.37}$$

For a bulk semiconductor density of states (6.67) and the Boltzmann distribution ($F_B \gg kT$)

$$dn = \frac{1}{2\pi^2} \left(\frac{2m^*}{\hbar^2}\right)^{3/2} \sqrt{E - E_C}\, \exp\left(-\frac{E - E_F}{kT}\right) dE. \tag{21.38}$$

For a given energy E, the carrier velocity v is determined by

$$E = E_C + \frac{1}{2} m^* v^2. \tag{21.39}$$

Thus we obtain

$$\sqrt{E - E_C} = v \sqrt{m^*/2} \tag{21.40}$$

and

$$dE = m^* v\, dv. \tag{21.41}$$

Also, with (21.39) and (21.1),

$$E - E_F = (E - E_C) + (E_C - E_F) = \frac{1}{2} m^* v^2 - e V_n. \tag{21.42}$$

Therefore we write (21.38) as

$$dn = 2 \left(\frac{m^*}{h}\right)^3 \exp(\beta V_n)\, \exp\left(-\frac{m^* v^2}{2kT}\right) 4\pi v^2\, dv. \tag{21.43}$$

Then the one-dimensional integral over $4\pi v^2\, dv$ is converted into a three-fold integral over $dv_x\, dv_y\, dv_z$. Integration over all velocities in y and z directions yields a factor $2\pi kT/m^*$. The integration over v_x runs from the minimum velocity $v_{min,x}$ necessary to pass the barrier,

$$\int_{v_{\mathrm{min,x}}}^{\infty} \exp\left(-\frac{m^* v_x^2}{2\,kT}\right) v_\mathrm{x}\,\mathrm{d}v_\mathrm{x} = \frac{kT}{m^*} \exp\left(-\frac{m^* v_{\mathrm{min,x}}^2}{2\,kT}\right), \tag{21.44}$$

with the minimum velocity determined by

$$\frac{1}{2} m^* v_{\mathrm{min,x}}^2 = e\,(V_\mathrm{bi} - V). \tag{21.45}$$

Thus the current density is given as, using (21.5),

$$\begin{aligned} j_{\mathrm{s}\to\mathrm{m}} &= \frac{4\pi e\, m^* k^2}{h^3} T^2 \exp\left(-\beta(V_\mathrm{bi} - V_\mathrm{n})\right) \exp\left(\beta V\right) \\ &= A^* T^2 \exp\left(-\frac{F_\mathrm{Bn}}{kT}\right) \exp\left(\beta V\right), \end{aligned} \tag{21.46}$$

with A^* being the Richardson constant given by

$$A^* = \frac{4\pi e\, m^* k_\mathrm{B}^2}{h^3} = \frac{e\, N_\mathrm{C}\, \bar{v}}{4T^2}, \tag{21.47}$$

where $\bar{v}$ is the average thermal velocity in the semiconductor. A^* for electrons in vacuum is $120\,\mathrm{A\,cm^{-2}\,K^{-2}}$. A similar result is obtained for the thermionic emission of electrons from a metal (overcoming the work function) into vacuum.

If the bias is changed, the current from the semiconductor to the metal increases in the forward direction because the energy difference between the quasi-Fermi level and the top of the barrier is reduced. The current is reduced for reverse bias. The barrier from the metal into the semiconductor remains constant (except for the Schottky effect whose impact on the current–voltage characteristic is discussed next). Therefore the current $j_{\mathrm{m}\to\mathrm{s}}$ from the metal into the semiconductor is constant and can be obtained from the condition $j = j_{\mathrm{s}\to\mathrm{m}} + j_{\mathrm{m}\to\mathrm{s}} = 0$ for zero bias. Therefore the current–voltage characteristic in the thermionic-emission model is

$$\begin{aligned} j &= A^* T^2 \exp\left(-\frac{F_\mathrm{Bn}}{kT}\right) \left[\exp\left(\beta V\right) - 1\right] \\ &= j_\mathrm{s} \left[\exp\left(\beta V\right) - 1\right]. \end{aligned} \tag{21.48}$$

The pre-factor

$$j_\mathrm{s} = A^* T^2 \exp\left(-\frac{F_\mathrm{Bn}}{kT}\right) \tag{21.49}$$

is called the saturation current density. The saturation current increases with increasing temperature.

The dependence (21.48) represents the ideal diode characteristic and is plotted in Fig. 21.13.

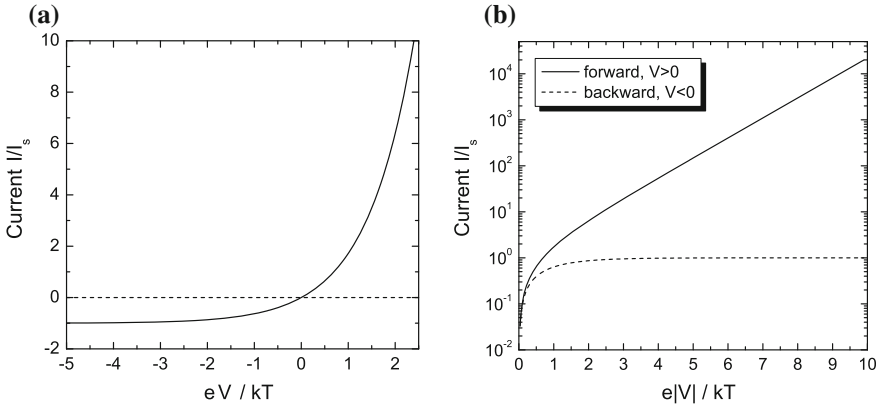


Fig. 21.13 Ideal diode I–V characteristics $I = I_s\,(\exp(eV/kT) - 1)$ (**a**) in linear plot and (**b**) semilogarithmic plot

The temperature dependence of the saturation current j_s can be written as

$$\ln\left(\frac{j_s}{T^2}\right) = \ln A^* - \frac{F_{Bn}}{kT} \tag{21.50}$$

by transforming (21.49). The plot of $\ln\left(j_s/T^2\right)$ to versus $1/T$ is called a Richardson plot and allows the barrier height and the Richardson constant to be determined from a linear fit (Fig. 21.14b).

Ideality Factor

If the Schottky (image force) effect, i.e. the change of barrier height with bias voltage, is considered, the semilogarithmic slope of the forward I–V characteristic is no longer $V_0^{-1} = e/kT$ but can be expressed as $V_0^{-1} = e/nkT$, n being a dimensionless parameter termed the ideality factor,[9]

$$j = j_s\left[\exp\left(\frac{e\,V}{n\,kT}\right) - 1\right]. \tag{21.51}$$

n is given by

$$n = \left(1 - \frac{1}{e}\frac{\partial F_B}{\partial V}\right)^{-1} \approx 1 + \frac{1}{e}\frac{\partial F_B}{\partial V}. \tag{21.52}$$

The ideality factor due to the image force effect n_{if} (using (21.27) and the image force barrier lowering $\Delta F_B^{if,0}$ at zero bias) [1439] is

[9] Obviously $n = 1$ for the ideal characteristic (21.48). Otherwise $n \geq 1$.

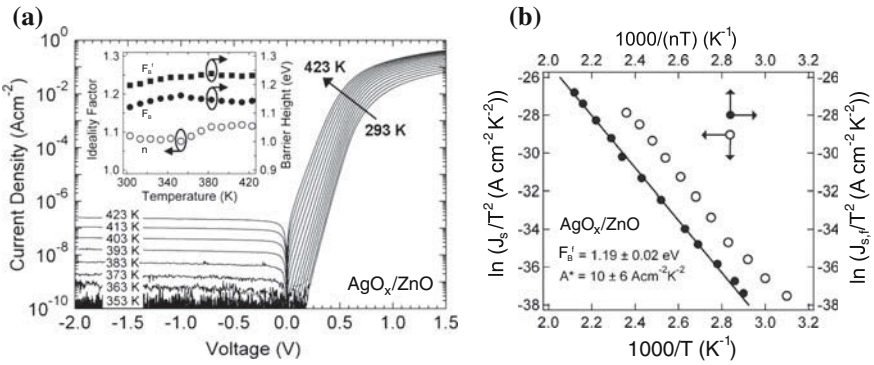


Fig. 21.14 Temperature dependent behavior of an AgO_x/ZnO Schottky diode. (**a**) I–V-characteristics. *Inset* depicts extracted parameters, ideality factor n, barrier height F_B and flatband barrier height F_B^f (21.54). (**b**) Richardson plots for the extracted barrier height and the calculated flatband barrier height with linear fit. Adapted from [1440]

$$n_{if} = 1 + \frac{\Delta F_B^{if,0}}{4\,e\,V_{bi}} = 1 + \frac{x_m}{w_0}. \tag{21.53}$$

Typical values are smaller than 1.03. For GaAs and $N_D = 10^{17}\,cm^{-3}$, $n = 1.02$. With regard to V_0 and its temperature dependence, we refer also to Fig. 21.17 and the related discussion.

In Fig. 21.14a the temperature dependence of the I–V-characteristic of an almost ideal ZnO Schottky diode [1440] is depicted. The extracted barrier height is almost independent of temperature as it should be. Also the ideality factor is rather small (about 1.1) and also independent of temperature. The Richardson plot Fig. 21.14b according to (21.50) for this diode is a straight line. The extracted constant $A^* = (10 \pm 6)\,A\,cm^{-2}\,K^{-2}$ is reasonably close to the theoretical value of $32\,A\,cm^{-2}\,K^{-2}$ (using $m^* = 0.32$).

In order to gain a barrier height independent from non-idealities, in [1441] the flatband barrier height F_B^f is discussed which is calculated from the diode parameters via

$$F_B^f = n\,F_B - (n-1)\,\frac{kT}{e}\,\ln\left(\frac{N_C}{N_D}\right), \tag{21.54}$$

assuming all donors are ionized (otherwise N_D shall be replaced by the electron density n, not to be confused here with the ideality factor). Accordingly, the saturation current density (21.49) is rewritten as

$$j_{s,f} = A^*\,T^2 \exp\left(-\frac{F_B^f}{kT}\right) \tag{21.55}$$

Laterally Inhomogeneous Barrier

The increase of the ideality factor due to a spatially inhomogeneous barrier height has been proposed early on [1439, 1442–1446]. The barrier height $F_{\mathrm{Bn}}(y, z)$ across the contact area is typically assumed to have a Gaussian probability distribution [1447] $p(F_{\mathrm{Bn}})$ with a mean value $\bar{F}_{\mathrm{Bn}}$ and a standard deviation σ_F, It turns out that the barrier height $F_{\mathrm{Bn}}^{\mathrm{C}}$ responsible for the capacitance, and thus the diffusion voltage determined by C–V spectroscopy, is given by the spatial average, i.e. $F_{\mathrm{Bn}}^{\mathrm{C}} = \bar{F}_{\mathrm{Bn}}$. The barrier height $F_{\mathrm{Bn}}^{\mathrm{j}}$ determining the current–voltage characteristics (cf. (21.48)) via

$$\begin{aligned} j &= A^* T^2 \left[\exp(\beta V) - 1\right] \int \exp\left(-\frac{F_{\mathrm{Bn}}}{kT}\right) p(F_{\mathrm{Bn}})\, \mathrm{d}F_{\mathrm{Bn}} \\ &= A^* T^2 \exp\left(-\frac{F_{\mathrm{Bn}}^{\mathrm{j}}}{kT}\right) \left[\exp(\beta V) - 1\right] \end{aligned} \tag{21.56}$$

is given by

$$F_{\mathrm{Bn}}^{\mathrm{j}} = \bar{F}_{\mathrm{Bn}} - \frac{\sigma_{\mathrm{F}}^2}{2\,kT}. \tag{21.57}$$

Thus, the barrier height determined from the current–voltage characteristic underestimates the spatial average of the barrier height.[10] The Richardson plot (21.50) is now modified (and is nonlinear in $1/T$) to

$$\ln\left(\frac{j_{\mathrm{s}}}{T^2}\right) = \ln A^* - \frac{F_{\mathrm{Bn}}}{kT} + \frac{\sigma_{\mathrm{F}}^2}{2\,k^2\,T^2}. \tag{21.58}$$

In [1448] the barrier nonuniformity due to the random distribution and discreteness of impurity charges in the depletion region is evaluated. This mechanism, yielding increasing barrier inhomogeneity for larger doping, represent a fundamental limit to the ideality of a Schottky diode.

If the distribution of barrier heights is discrete, at low voltages the current will first flow at certain spots with low barrier ('hot spots'). This leads to kinks in the I–V characteristics as found for SiC [1449] or ZnO [1450] (Fig. 21.15).

Temperature Dependence

Figure 21.16a shows the temperature-dependent I–V characteristics of a Pd/ZnO Schottky diode. A straightforward evaluation according to (21.48) results in a barrier height of about 700 meV and a Richardson constant that is orders of magnitude smaller than the theoretical value of $32\,\mathrm{A\,K^{-2}\,cm^{-2}}$ (for $m_{\mathrm{e}}^* = 0.27$). A fit of the temperature-dependent data with (21.57), as shown in Fig. 21.16b, results in $\bar{F}_{\mathrm{Bn}} = 1.1\,\mathrm{eV}$, in agreement with the (temperature-independent) value obtained from CV spectroscopy, and $\sigma_F = 0.13\,\mathrm{eV}$ [1451].

[10] This phenomenon is similar to the red-shift of luminescence lines (Stokes shift) due to thermalization in the presence of disorder, see Sect. 12.4.

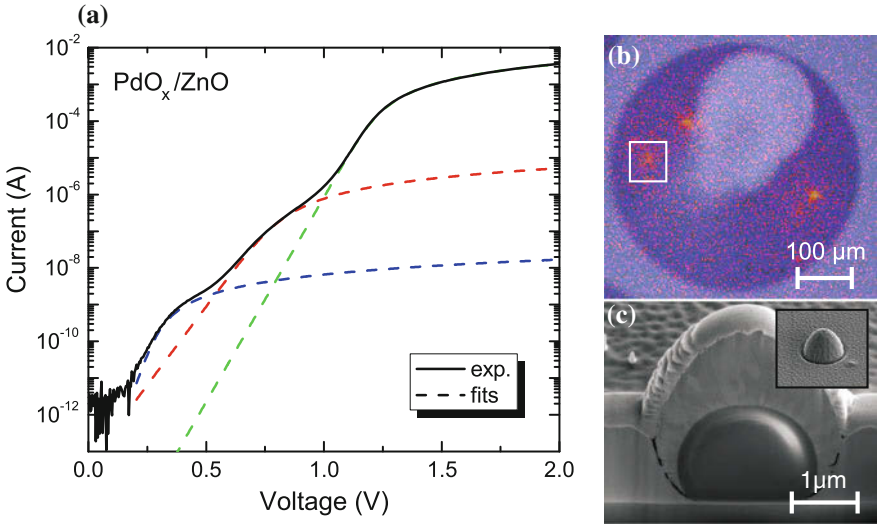


Fig. 21.15 (**a**) I–V characteristic of ZnO Schottky diode (*solid line*) with fits due to regions with three different barrier heights (*dashed lines*). (**b**) Thermography image at bias of 1 V. The *white square* marks the area where one hot spot was investigated in detail. (**c**) SEM image of FIB prepared cross section of a defect (Al_2O_3 particle) causing the lowest barrier hot spots, *inset* plane view of such defect before preparation. Adapted from [1450]

The temperature dependence of the ideality factor n is given by [1452]

$$n = \frac{1}{1 - \rho_2 + \rho_3/(2kT)}, \tag{21.59}$$

where ρ_2 (ρ_3) is the (temperature-independent) proportionality coefficient of the bias dependence of the mean barrier height (standard deviation), i.e.

$$\rho_2 = \frac{1}{e}\frac{\partial \bar{F}_{\mathrm{Bn}}}{\partial V} \tag{21.60a}$$

$$\rho_3 = \frac{1}{e}\frac{\partial\, \sigma_F^2}{\partial V}. \tag{21.60b}$$

The fit of $1/n - 1$ versus $1/T$ in Fig. 21.16c yields $\rho_2 = -0.025$ and $\rho_3 = -0.028$ eV for the ZnO diodes under investigation.

The forward I–V characteristic of an Au/GaAs Schottky diode reported in [1453, 1454] is shown in Fig. 21.17a at various temperatures. The current amplitude decreases with decreasing temperature due to the temperature dependence of the saturation current (21.48). Also, the slope V_0^{-1} of the characteristic $j = j_s \exp(V/V_0)$ varies with temperature. Looking at the temperature dependence of V_0, it is described as $V_0 = k(T + T_0)/e$ rather than with an ideality factor n in the form of $V_0 = nkT/e$.

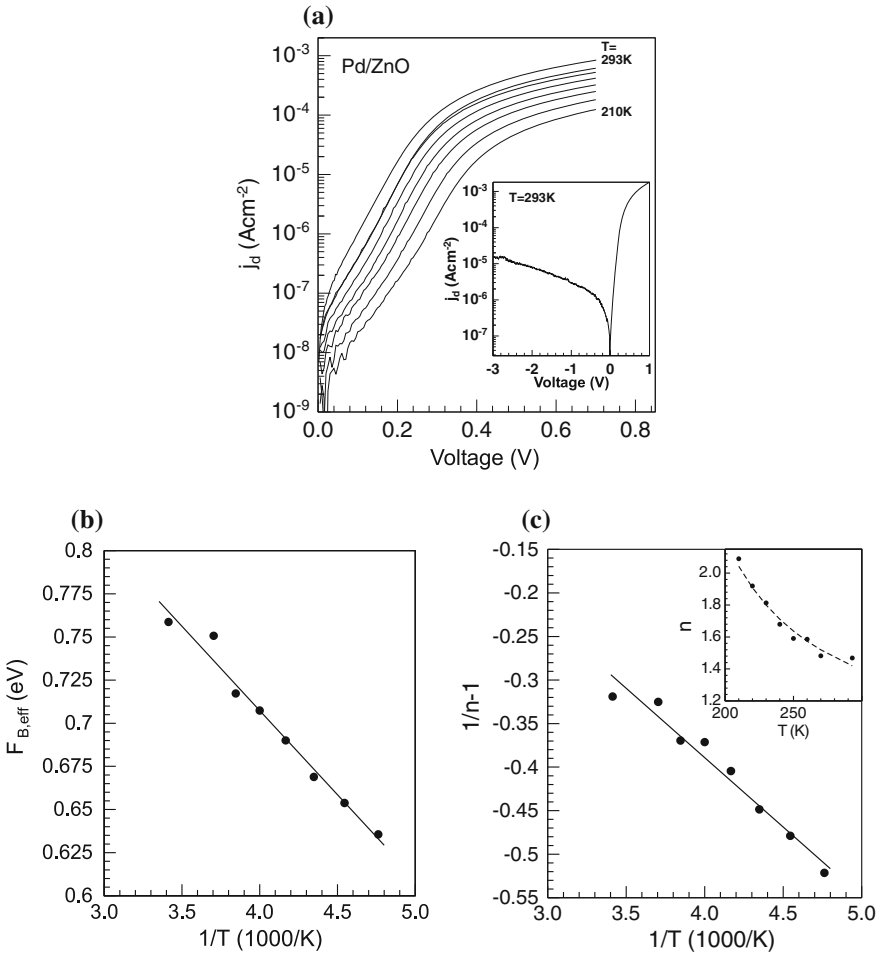


Fig. 21.16 (**a**) Forward I–V characteristic of Pd/ZnO Schottky diodes for various temperatures. Diode temperatures are 210, 220, 230, 240, 250, 260, 270, and 293 K. The *inset* shows the current density versus voltage for 293 K on a semilogarithmic scale. (**b**) Effective barrier height $F_{\mathrm{Bn}}^{\mathrm{j}}$ versus the inverse temperature. The *solid line* is a linear fit according to (21.57) yielding the standard deviation $\sigma_F = 0.13\,\mathrm{eV}$ and the mean barrier height $\bar{F}_{\mathrm{Bn}} = 1.1\,\mathrm{eV}$. (**c**) Plot of $1/n - 1$ versus the inverse temperature. The *solid line* is a linear fit of the data yielding the voltage deformation coefficients $\rho_2 = -0.025$ and $\rho_3 = -0.028\,\mathrm{eV}$. The *inset* shows the experimentally determined n factors and the n factors calculated from (21.59) using the voltage-deformation coefficients obtained from the linear fit (*dashed line*). Adapted from [1451]

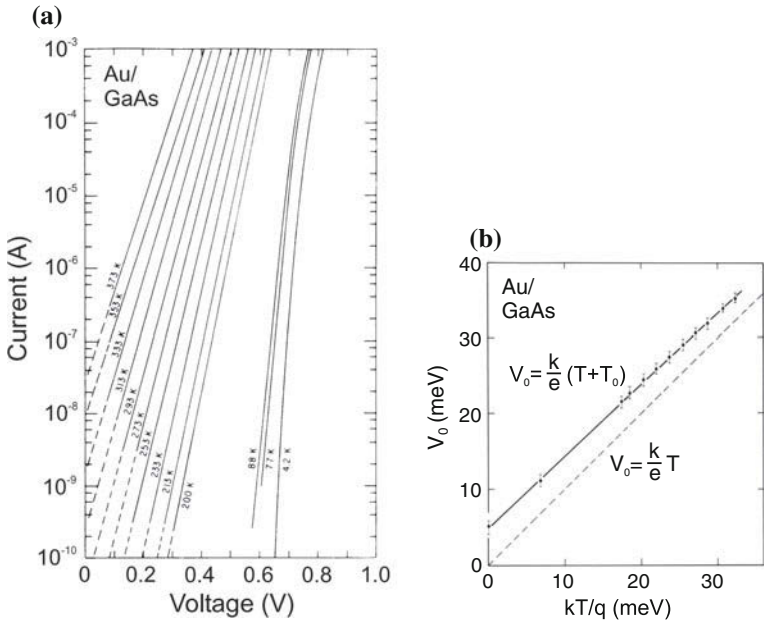


Fig. 21.17 (**a**) Forward I–V characteristic of Au/GaAs diode for various temperatures. (**b**) Temperature dependence of the voltage V_0. The experimental data are fitted with $T_0 = 45 \pm 8\,\mathrm{K}$. Adapted from [1453]

In other words, the ideality factor follows a temperature dependence $n = 1 + T_0/T$. In view of (21.59), such behavior means for small T_0 that $n \approx 1/(1 - T_0/T)$ and thus $\rho_2 = 0$ and $\rho_3 = 2kT_0$. For $T_0 = 45\,\mathrm{K}$, ρ_3 is 0.008 eV, which is a fairly small value. Thus, the temperature behavior of the diode is due to the narrowing of the Gaussian distribution of barrier height with bias voltage [1447].

Correlation of Barrier Height and Ideality Factor

In a set of similar diodes with varying magnitude of barrier inhomogeneities, it is found that the effective barrier height and the ideality factor correlate [1454]. The extrapolation to $n = n_{\mathrm{if}}$ yields the limit of the barrier height for a homogeneous barrier (Fig. 21.18). For silicon, it is found that the surface orientation has a minor influence on the Schottky barrier height (Fig. 21.18b) for an unreconstructed surface. The presence of a reconstruction lowers the barrier height[11] (Fig. 21.18a).

[11] Reconstructions are accompanied by redistributions of the valence charge with respect to the undisturbed bulk (Sect. 11.4). The subsequent extra interface dipoles alter the barrier heights of reconstructed interfaces [1455].

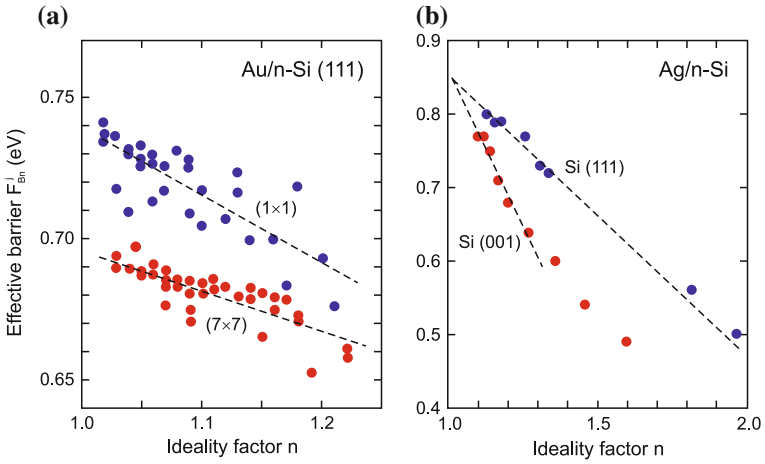


Fig. 21.18 (**a**) Effective barrier height versus ideality factor for Ag/n-Si Schottky diodes prepared on Si(111) surface with (7 × 7) reconstruction or unreconstructed (1 × 1) as labeled. The more ideal (1 × 1) surface exhibits the higher barrier. The *dashed lines* are linear fits. (**b**) Effective barrier height versus ideality factor for Au/n-Si Schottky diodes prepared on HF-dipped (1 × 1) unreconstructed (001) and (111) surfaces. The *dashed lines* are linear fits and extrapolation to $n = n_{if}$. Both surface orientations exhibit the same extrapolated homogeneous barrier height. Based on data compiled in [1455]

Diffusion Theory

In diffusion theory the current density is considered in the presence of a carrier-density and electric-field gradient. In the Boltzmann approximation the electron current is given by (8.59a). In stationary equilibrium the current density is constant, i.e. independent of x. Assuming that the carrier density has its equilibrium values at $x = 0$ and $x = w$, we find after integration and using (21.8)

$$\begin{aligned} j &= -e\,\mu_n\,N_C\,E_m\,\exp\left(-\frac{F_{Bn}}{kT}\right)\left[\exp\left(\beta V\right) - 1\right] \\ &= j_s\,\left[\exp\left(\beta V\right) - 1\right]. \end{aligned} \tag{21.61}$$

Therefore, also in this case the ideal diode characteristic is obtained, however, with a different saturation current. The ideality factor in diffusion theory is $n = 1.06$ (for $F_{Bn} \gtrsim 15\,kT$) [1409].

Combined Theory

A combination of both theories [1456] considers both mechanisms to be in series. The current can then be expressed as

$$j = \frac{eN_C v_r}{1 + v_r/v_D}\exp\left(-\frac{F_{Bn}}{kT}\right)\left[\exp\left(\beta V\right) - 1\right] \tag{21.62a}$$

$$= A^{**}\, T^2 \left[\exp(\beta V) - 1\right] \tag{21.62b}$$
$$= j_s \left[\exp(\beta V) - 1\right].$$

Here $v_r = \bar{v}/4$ is a 'recombination velocity' [1457] at the top of the barrier according to $j = v_r(n - n_0)$, n_0 being the equilibrium electron density at the top of the barrier and $\bar{v}$ is the average thermal velocity in the semiconductor. v_D is an effective diffusion velocity describing the transport of electrons from the edge of the depletion layer ($x = w$) to the top of the barrier ($x = x_m$). It is defined as

$$v_D^{-1} = \int_{x_m}^{w} \frac{-e}{\mu_n kT} \exp\left(-\frac{F_{Bn} - E_C(x)}{kT}\right) dx. \tag{21.63}$$

In [1456] μ_n has been assumed to be independent of the electric field. This assumption is potentially not realistic. If $v_D \gg v_r$, thermionic theory applies and we obtain (21.48). The case $v_r \gg v_D \sim \mu_n E_m$ relates to diffusion theory and we recover (21.61).

The constant A^{**} in (21.62b) is called the effective Richardson constant. Its calculated dependence on the electric field is shown in Fig. 21.19 for Si. At room temperature for most Ge, Si and GaAs Schottky diodes the thermionic emission of majority carriers is the dominating process. The effect of lateral barrier height inhomogeneity in a diffusion model has been discussed in [1458].

Tunneling Current

At high doping the width of the depletion layer becomes small and tunneling processes become more probable. Also at low temperatures, when thermionic emission is very small, tunneling processes can dominate the transport between metal and semiconductor. One process is tunneling of electrons close to the Fermi level of

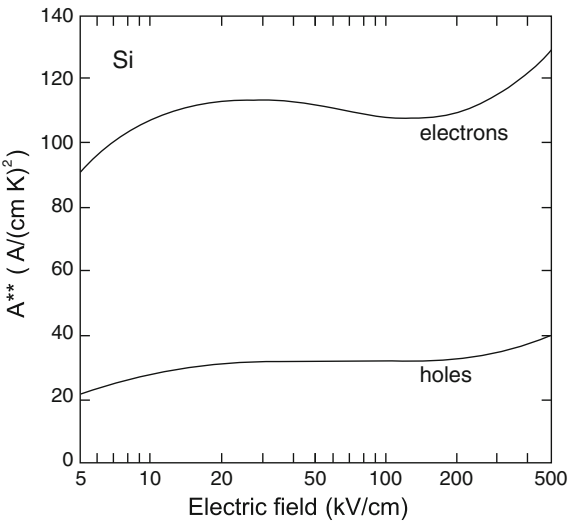


Fig. 21.19 Calculated effective Richardson constant A^{**} as a function of the electric field for a metal–Si diode at $T = 300\,\mathrm{K}$ for a (n-type or p-type) doping of $10^{16}\,\mathrm{cm}^{-3}$. Adapted from [1459]

the semiconductor. This process is called field emission (F) and is at least important for degenerate semiconductors at very low temperatures. If the temperature is raised, electrons are excited to higher energies where they encounter a thinner barrier. The tradeoff between thermal energy and barrier width selects an electron energy E_m above the conduction-band edge for which the current is largest. This process is known as thermionic field emission (TF). For very high temperatures enough carriers can overcome the barrier completely and we are back in the thermionic emission regime. The validity of the two regimes is shown in Fig. 21.20 for Au/GaAs Schottky diodes as a function of doping concentration (n-type) and temperature.

In the field-emission regime the forward current is given by [1409]

$$j = j_s \exp\left(\frac{eV}{E_{00}}\right), \tag{21.64}$$

with the characteristic energy parameter E_{00} given by

$$E_{00} = \frac{e\hbar}{2}\sqrt{\frac{N_D}{m^*\epsilon_s}}. \tag{21.65}$$

The saturation current is

$$j_s \propto \exp\left(-\frac{F_{Bn}}{E_{00}}\right). \tag{21.66}$$

In Fig. 21.21, the forward characteristic of a highly doped Au/Si is shown. The experimental value of $E_{00} = 29\,\text{meV}$ agrees well with the theoretical expectation of $E_{00} = 29.5\,\text{meV}$.

In the reverse direction the I–V characteristic under field emission is given by [1409]

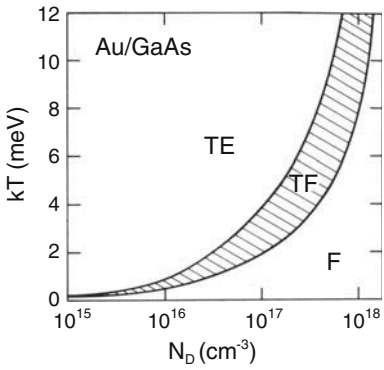


Fig. 21.20 Calculated conditions for thermionic field ('TF'), field ('F') and thermionic ('TE') emission in a Au/GaAs Schottky diode as a function of temperature and doping concentration. Adapted from [1409]

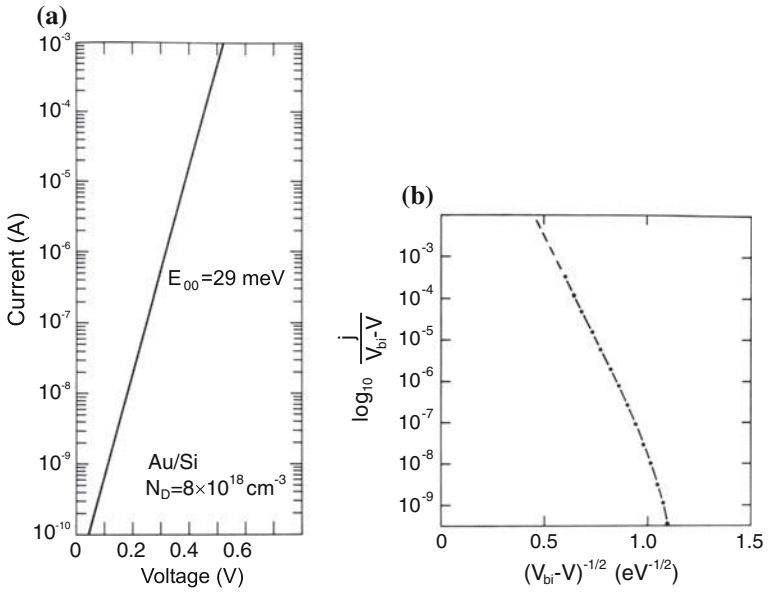


Fig. 21.21 (**a**) Forward and (**b**) reverse I–V characteristic of a Au/Si Schottky diode at 77 K. The doping concentration of the Si was $N_D = 8 \times 10^{18}\,\mathrm{cm}^{-3}$. Adapted from [1409]

$$j = \frac{4\pi\, e\, m^*}{h^3}\, E_{00}^2\, \frac{e\,(V_{\mathrm{bi}} - V)}{F_{\mathrm{Bn}}}\, \exp\left(-\frac{2\,F_{\mathrm{Bn}}^{3/2}}{3E_{00}\,\sqrt{e(V_{\mathrm{bi}} - V)}}\right). \tag{21.67}$$

From Fig. 21.21b, a barrier height of 0.79 eV is deduced.

In the TF-emission regime the current–voltage characteristic is given by

$$j = j_{\mathrm{s}}\, \exp\left(\frac{eV}{E_0}\right), \tag{21.68}$$

with

$$E_0 = E_{00}\, \coth\left(\frac{E_{00}}{kT}\right), \tag{21.69}$$

where E_{00} is given by (21.65). The energy for maximum TF emission E_{m} is given by $E_{\mathrm{m}} = e(V_{\mathrm{bi}} - V)/\cosh^2(E_{00}/kT)$. The coth-dependence of E_0 is shown in Fig. 21.22 for an Au/GaAs diode.

A Schottky diode can suffer from nonideality such as series and parallel ohmic resistance [1452]. These effects are discussed in some detail below for pn-diodes in Sect. 21.4.4 and apply similarly to Schottky diodes.

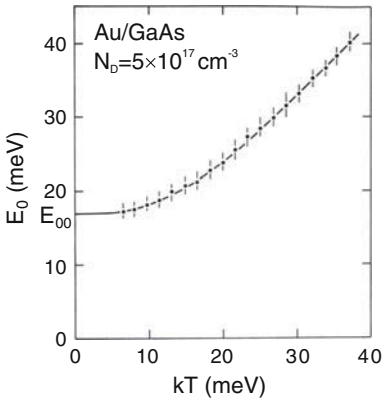


Fig. 21.22 Temperature dependence of E_0 of an Au/GaAs diode with $N_D = 5 \times 10^{17}\,cm^{-3}$. The *solid line* is the theoretical dependence for thermionic emission according to (21.69) with $N_D = 6.5 \times 10^{17}\,cm^{-3}$ and $m^* = 0.07$. Adapted from [1409]

21.2.6 Ohmic Contacts

Although an ohmic contact does not have a diode characteristic, it can be understood from the previous remarks. An ohmic contact will have a small contact resistance for both current directions. The voltage drop across the contact will be small compared to the voltage drop in the active layer (somewhere else). Details on electric contacts can be found in [1460].

The contact resistance R_c is defined as the differential resistance at $V = 0$

$$R_c = \left(\frac{\partial I}{\partial V}\right)^{-1}_{V=0}. \tag{21.70}$$

At low doping, the transport is dominated by thermionic emission (21.48). In this case R_c is given by

$$R_c = \frac{k}{e A^* T} \exp\left(\frac{F_{Bn}}{kT}\right). \tag{21.71}$$

A small barrier height (Fig. 21.23a) will lead to small contact resistance. A negative Schottky barrier height, i.e. $W_m < E_{vac} - E_F$ for a n-type semiconductor, leads to an accumulation layer without a barrier for carrier transport (Fig. 21.23b).

For high doping R_c is determined by the tunneling current (Fig. 21.23c) and is proportional to

$$R_c \propto \exp\left(\frac{F_{Bn}}{E_{00}}\right). \tag{21.72}$$

The contact resistance decreases exponentially with the doping. A theoretical calculation and experimental data are compared in Fig. 21.24 for contacts on Si.

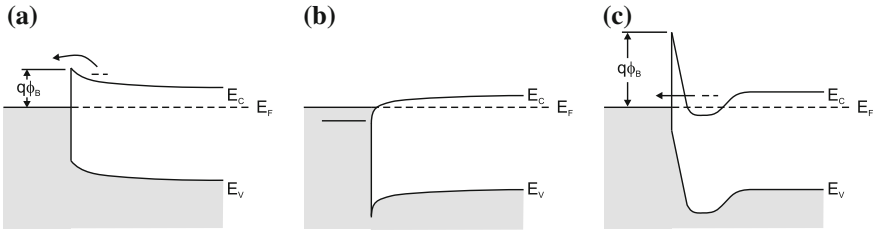


Fig. 21.23 Schematic conditions for the formation of an Ohmic contact, (**a**) low barrier height, (**b**) accumulation layer (**c**) high doping (thin depletion layer)

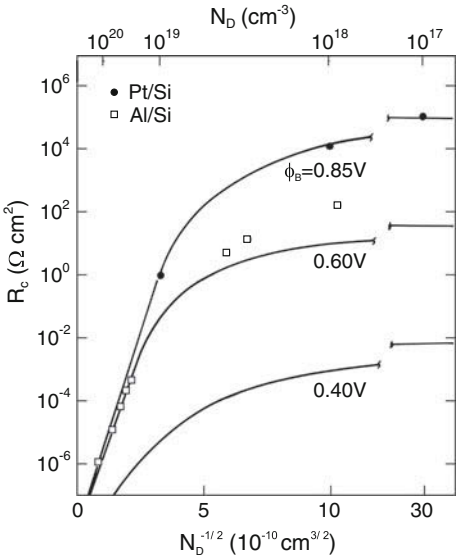


Fig. 21.24 Theoretical and experimental values of specific contact resistances at $T = 300\,\mathrm{K}$ for Al/n-Si [1461] and PtSi/n-Si [1462] contacts as a function of donor concentration. *Solid lines* are theoretical dependencies for different values of the barrier height as labeled. Adapted from [500]

The three mechanisms, low barrier height, accumulation layer and high doping, for the formation of Ohmic contacts are summarized schematically in Fig. 21.23. Ohmic contacts on wide band gap semiconductors are difficult, since metals with sufficiently small (large) work function for contacting n-type (p-type) material are mostly not available.

Although Schottky contact devices have their place in semiconductor technology, Ohmic contacts are indispensable for almost all devices.[12] Ohmic contacts are typically prepared by evaporating a contact metal containing the doping material for the semiconductor, e.g. Au/Zn for a contact on p-type GaAs [1463] and Au/Ge for a contact [1464] on n-type GaAs. The contact is alloyed around 400–500 °C (see Fig. 21.25) above the eutectic temperature of $T_{\mathrm{eu}} = 360\,^{\circ}\mathrm{C}$ (for Au/Ge) to form a

[12] Also a Schottky diode has an Ohmic back contact.

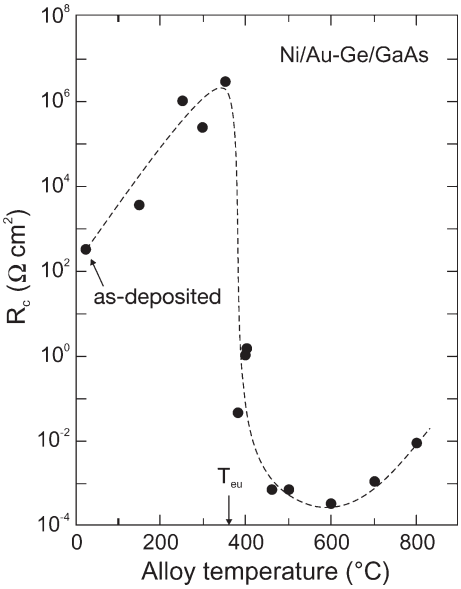


Fig. 21.25 Specific contact resistance for Ni/Au-Ge on n-type epitaxial GaAs for varying alloying temperatures (2 min). *Arrow* at T_{eu} denotes the eutectic temperature of Au-Ge. Adapted from [1468]

eutectic liquid in which the dopant can quickly diffuse. When the eutectic liquid cools it forms a solid, a highly doped semiconductor layer underneath the metal. The liquid-phase reactions can lead to inhomogeneous contacts. On n-type GaAs Pd/Ge/Au contacts have been reported to have superior structural quality [1465]. Ohmic contacts for a number of different semiconductors are reviewed in [1466, 1467].

21.2.7 Metal Contacts to Organic Semiconductors

Also, for organic semiconductors the metal contact plays a vital role, either for carrier injection or for manipulation of the space-charge region. The position of the Fermi level has been determined for various organic semiconductors as shown in Fig. 21.26. These data have been obtained from measurements on metal–semiconductor–metal structures (MSM, see also Sect. 22.3.5) as shown in Fig. 21.27a. The thin (50 nm) organic layer is fully depleted, thus the built-in field inside the semiconductor is constant. The built-in field is measured by applying an external dc bias and finding the external potential at which the electroabsorption signal vanishes. Figure 21.27b shows the measured electroabsorption signal $\Delta T/T$ (relative change of transmission T) and the optical density of an MEH-PPV film as a function of photon energy for an Al/MEH-PPV/Al structure. The exciton absorption peak is found at 2.25 eV. The bias

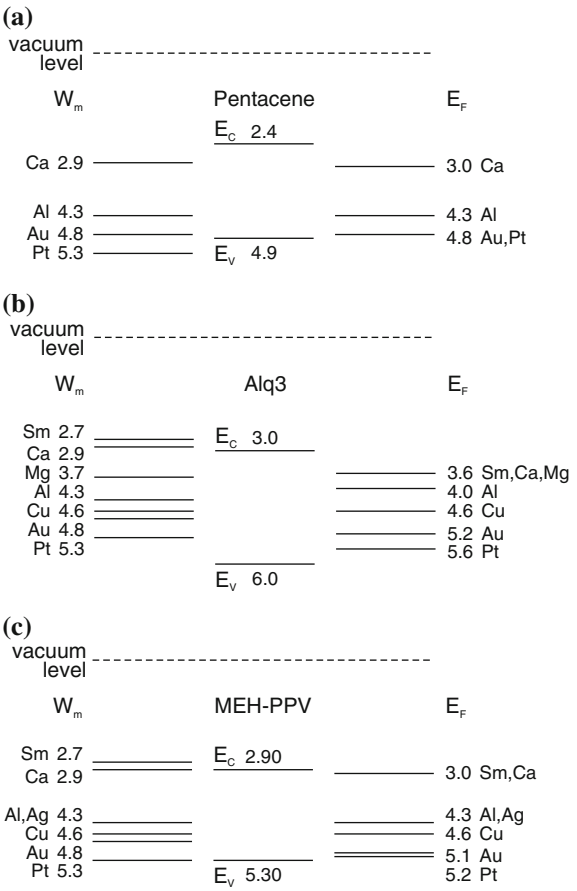


Fig. 21.26 Measured Fermi energies E_F (*labeled data* in eV) and the work functions W_m of various metals contacting (**a**) pentacene, (**b**) Alq_3 and (**c**) MEH-PPV. E_C (E_V) denotes the energy position of the electron (hole) polaron. Measured data for E_F for MEH-PPV from [1470], other from [1471]. Data for W_m from Table 21.1

at which the built-in field vanishes can then be determined for various other metals in metal/MEH-PPV/Al structures. Figure 21.26 summarizes such results for various metals and three organic semiconductors. The plot of the Fermi level position versus the metal work function (Fig. 21.27d, e) shows that the metals investigated do not introduce interface states in the single-particle gap that pin the Schottky barrier (see Fig. 21.5 for inorganic semiconductors). An electron trap, such as C_{60} in MEH-PPV, can pin the Fermi level of the n-contact metal and leads to a change of the built-in potential [1469].

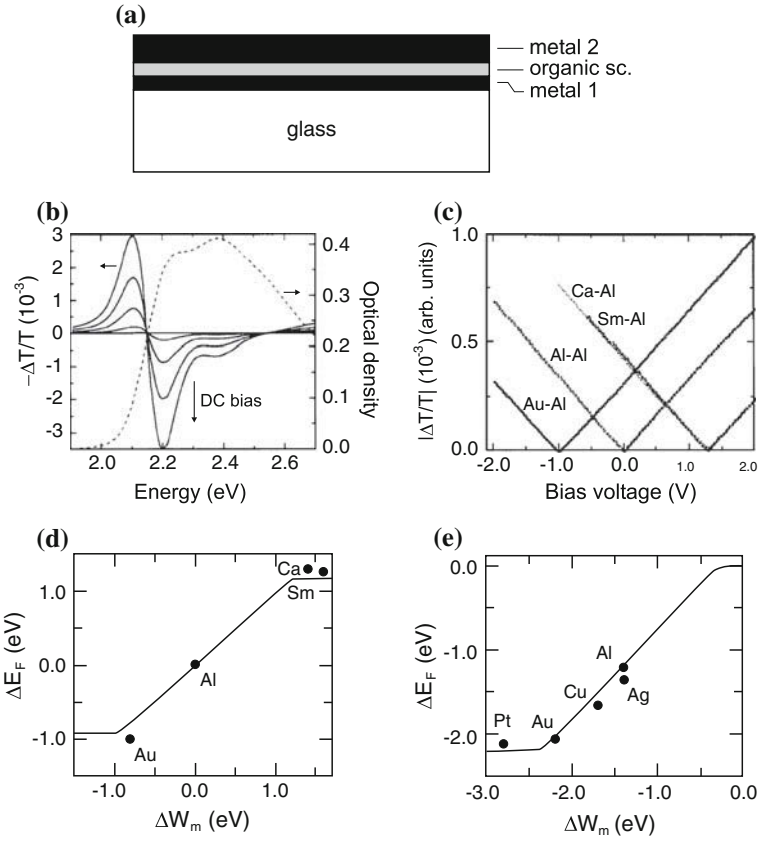


Fig. 21.27 (**a**) Schematic MSM structure with organic semiconductor (sc.) on transparent glass substrate. Metal 1 is thin and semitransparent. Thickness of organic semiconductor (polymer or small molecules) is about 50 nm. (**b**) Electroabsorption spectra of Al/MEH-PPV/Al structure at four dc bias voltages (*solid lines*) and optical density spectrum (*dashed line*). (**c**) Magnitude of the electroabsorption response at 2.1 eV as a function of bias for metal/MEH-PPV/Al structures. (**d**, **e**) Calculated (*solid lines*) and experimental (*points*) potential difference across (**d**) metal/MEH-PPV/Al structures and (**e**) metal/MEH-PPV/Ca structures as a function of the work-function difference of the contacts. Parts (**b**)–(**e**) adapted from [1470]

21.3 Metal–Insulator–Semiconductor Diodes

In a metal–insulator–semiconductor (MIS) diode an insulator is sandwiched between the metal and the semiconductor. Subsequently, a MIS contact has zero dc conductance. The semiconductor typically has an ohmic back contact. As insulator, often the oxide of the respective semiconductor, is used. In particular SiO_2 on Si has been technologically advanced (Fig. 21.28). In the latter case, the diode is called a MOS (metal–oxide–semiconductor) diode. This structure has great importance for

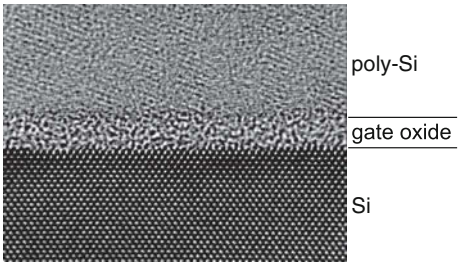


Fig. 21.28 High-resolution transmission electron microscopy image of a 1.6-nm thick gate oxide between poly-Si (cf. Sect. 24.5.4) and crystalline Si. From [1472]

the investigation of semiconductor surfaces and overwhelming importance for semiconductor technology (planar integration of electronic circuits, CMOS technology). Also, CCDs (Sect. 22.3.8) are based on MIS diodes.

21.3.1 Band Diagram for Ideal MIS Diode

An ideal MIS diode has to fulfill the following three conditions:

(i) (as shown in Fig. 21.29) without external bias the energy difference ϕ_{ms} between the work function of the metal and the semiconductor

$$\phi_{\mathrm{ms}} = W_{\mathrm{m}} - \left(\chi_{\mathrm{sc}} + \frac{E_{\mathrm{g}}}{2} \pm e\,\Psi_{\mathrm{B}}\right) \tag{21.73}$$

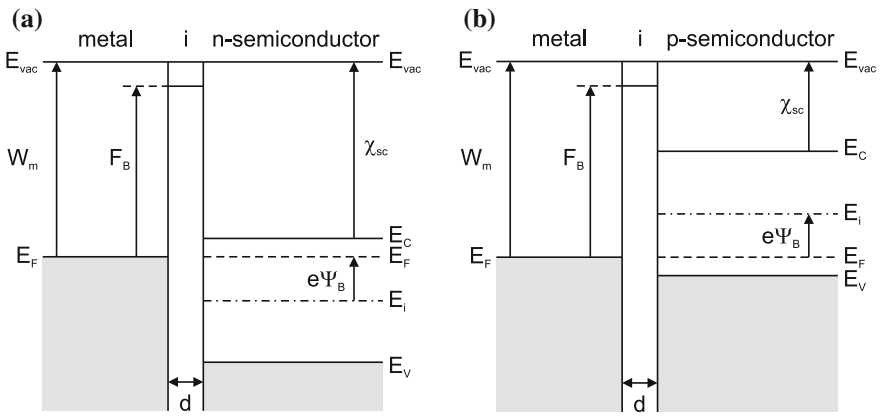


Fig. 21.29 Band diagram of an ideal MIS diode with (**a**) n- and (**b**) p-type semiconductor at external bias $V = 0$. The insulator ('i') thickness is d as labeled. The *dash-dotted line* represents the intrinsic Fermi level E_{i}

is zero ($\phi_{\rm ms} = 0$). The '+' ('−') sign in (21.73) applies to a p-type, Fig. 21.29b (n-type, Fig. 21.29a) semiconductor. $\Psi_{\rm B}$ is the potential difference between the intrinsic and actual Fermi level, $\Psi_{\rm B} = |E_{\rm i} - E_{\rm F}|/e > 0$.

(ii) The only charges present are those in the semiconductor and the opposite charge is on the metal surface close to the insulator.

(iii) There is no dc current between the metal and the semiconductor, i.e. the conductivity of the insulator is zero.

The quantity $\Psi_{\rm B}$ (Fig. 21.31) is given as (for a p-type material, using (7.18), the high temperature approximation (7.30) and (7.15))

$$e\,\Psi_{\rm B} = E_{\rm i} - E_{\rm F} = \left[\frac{E_{\rm C} + E_{\rm V}}{2} + \frac{kT}{2}\,\ln\left(\frac{N_{\rm V}}{N_{\rm C}}\right)\right] - \left[E_{\rm V} - kT\,\ln\left(\frac{N_{\rm A}}{N_{\rm V}}\right)\right]$$
$$= \frac{E_{\rm g}}{2} + kT\,\ln\left(\frac{N_{\rm A}}{\sqrt{N_{\rm C}\,N_{\rm V}}}\right) = kT\,\ln\left(\frac{N_{\rm A}}{n_{\rm i}}\right). \qquad (21.74)$$

When an ideal MIS diode is biased, three general cases—accumulation, depletion and inversion—can occur (Fig. 21.30). We discuss these first for the p-type semiconductor.

Figure 21.30d shows the accumulation case for a negative voltage at the metal.[13] Part of the voltage drops across the insulator, the rest across the semiconductor. The valence band is bent upwards towards the Fermi level. The quasi-Fermi level in the semiconductor, however, is constant since no dc current flows.[14]

Since the charge-carrier (hole) density depends exponentially on the energy separation $E_{\rm F} - E_{\rm V}$, a charge accumulation (of holes) occurs in the (p-type) semiconductor in the vicinity of the interface to the insulator.

In Fig. 21.30e the depletion case is shown. Now a moderate reverse voltage, i.e. a positive bias to the metal, is applied. A depletion of majority charge carriers occurs in the semiconductor close to the insulator. The quasi-Fermi level in the semiconductor remains beneath the intrinsic level ($E_{\rm i} \approx E_{\rm C} + E_{\rm g}/2$), i.e. the semiconductor remains p-type everywhere. If the voltage is increased further to large values, the quasi-Fermi level intersects the intrinsic level and lies *above* $E_{\rm i}$ close to the insulator (Fig. 21.30f). In this region, the electron concentration becomes larger than the hole concentration and we have the inversion case. The inversion is called 'weak' if the Fermi level is still close to $E_{\rm i}$. The inversion is called 'strong' when the Fermi level lies close to the conduction-band edge.

[13] This poling is a forward bias of the respective Schottky diode since the positive pole is at the p-type semiconductor.

[14] We note that in order to reach the situations shown in Fig. 21.30 from the zero bias case of Fig. 21.29, a current must have flowed since charge carriers are redistributed. Figure 21.30 depicts the stationary equilibrium after transient voltage switch-on effects have subsided. The time, however, that is needed in order to reach such stationary equilibrium from zero bias (thermal equilibrium) may be very long (e.g. days, cf. Sect. 22.3.8).

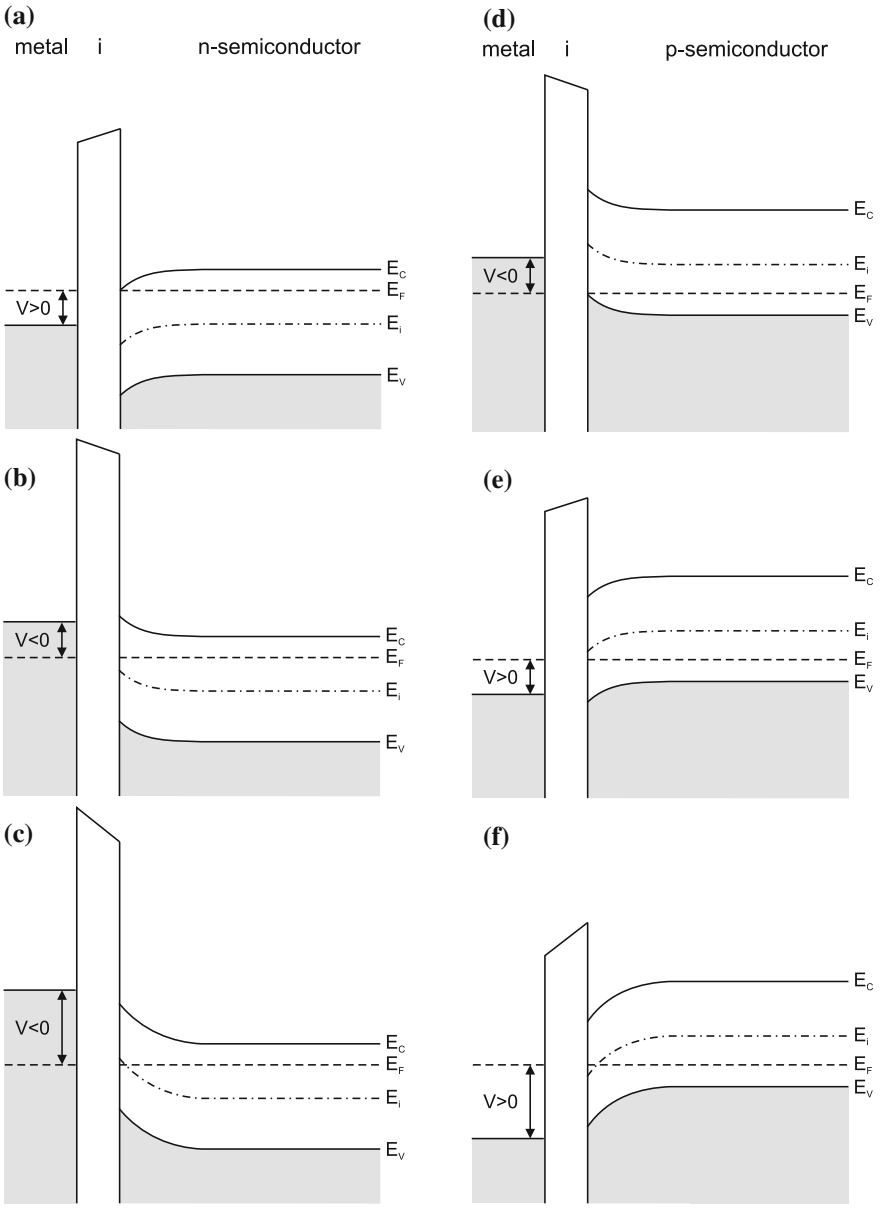


Fig. 21.30 Band diagram of ideal MIS diodes with (**a**–**c**) n-type and (**d**–**f**) p-type semiconductors for $V \neq 0$ in stationary equilibrium for the cases (**a**, **d**) accumulation, (**b**, **e**) depletion and (**c**, **f**) inversion

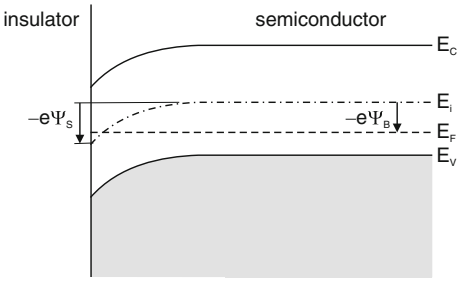


Fig. 21.31 Band diagram at the surface of a p-type semiconductor of a MIS diode. Accumulation occurs for $\Psi_s < 0$, depletion for $\Psi_s > 0$ and inversion (as shown here) for $\Psi_s > \Psi_B > 0$

The corresponding phenomena occur for n-type semiconductors for the opposite signs of the voltage with electron accumulation and depletion. In the inversion case, $p > n$ close to the insulator (Fig. 21.30a–c).

21.3.2 Space-Charge Region

Now we calculate the charge and electric field distribution in an ideal MIS diode, following the treatment in [1473]. We introduce the potential Ψ that measures the separation of the intrinsic bulk Fermi level and the actual intrinsic level E_i, i.e. $-e\Psi(x) = E_i(x) - E_i(x \to \infty)$ (see Fig. 21.31). Its value at the surface is termed Ψ_s, the surface potential. The value is positive, i.e. $\Psi_s > 0$, if the intrinsic Fermi level at the surface is below the bulk Fermi level.

The electron and hole concentrations are given (for a p-type semiconductor) as

$$n_p = n_{p_0} \exp(\beta\Psi) \tag{21.75a}$$

$$p_p = p_{p_0} \exp(-\beta\Psi), \tag{21.75b}$$

where n_{p_0} (p_{p_0}) are the bulk electron (hole) concentrations, respectively, and $\beta = e/kT > 0$.

Therefore, the net free charge is given by

$$n_p - p_p = n_{p_0} \exp(\beta\Psi) - p_{p_0} \exp(-\beta\Psi). \tag{21.76}$$

The electron and hole concentrations at the surface are denoted with an index 's' and are given by[15]

$$n_s = n_{p_0} \exp(\beta\Psi_s) \tag{21.77a}$$

$$p_s = p_{p_0} \exp(-\beta\Psi_s). \tag{21.77b}$$

[15] Ψ_s represents the voltage drop across the semiconductor that will be discussed in more detail in Sect. 21.3.3. In this sense, Ψ_s for the MIS diode is related to $V_{bi} - V$ for the Schottky contact.

We use the Poisson equation $\frac{d^2\Psi}{dx^2} = -\frac{\rho}{\epsilon_s}$ with the charge given by

$$\rho(x) = e\left[p(x) - n(x) + N_D^+(x) - N_A^-(x)\right]. \tag{21.78}$$

As boundary condition we employ that far away from the surface (for $x \to \infty$) there is charge neutrality (cf. (7.41)), i.e.

$$n_{p_0} - p_{p_0} = N_D^+ - N_A^-, \tag{21.79}$$

and that $\Psi = 0$. We note that $N_D^+ - N_A^-$ must be constant throughout the completely ionized, homogeneous semiconductor. Therefore (21.79) (but not charge neutrality) holds everywhere in the semiconductor. Using (21.76) the Poisson equation reads

$$\frac{\partial^2\Psi}{\partial x^2} = -\frac{e}{\epsilon_s}\left\{p_{p_0}\left[\exp(-\beta\Psi) - 1\right] - n_{p_0}\left[\exp(\beta\Psi) - 1\right]\right\}. \tag{21.80}$$

The Poisson equation is integrated and with the notations

$$\mathcal{F}(\Psi) = \sqrt{\left[\exp(-\beta\Psi) + \beta\Psi - 1\right] + \frac{n_{p_0}}{p_{p_0}}\left[\exp(\beta\Psi) - \beta\Psi - 1\right]} \tag{21.81a}$$

$$L_D = \sqrt{\frac{\epsilon_s\, kT}{e^2\, p_{p_0}}} = \sqrt{\frac{\epsilon_s}{e\,\beta\, p_{p_0}}}, \tag{21.81b}$$

with L_D being the Debye length for holes, the electric field can be written as

$$E = -\frac{\partial\Psi}{\partial x} = \pm\frac{\sqrt{2}\,kT}{e\,L_D}\mathcal{F}(\Psi). \tag{21.82}$$

The positive (negative) sign is for $\Psi > 0$ ($\Psi < 0$), respectively. At the surface, Ψ_s will be taken as the value for Ψ. The total charge Q_s per unit area creating the surface field

$$E_s = -\left.\frac{\partial\Psi}{\partial x}\right|_{x=0} = \pm\frac{\sqrt{2}\,kT}{e\,L_D}\mathcal{F}(\Psi_s) \tag{21.83}$$

is given by Gauss's law as $Q_s = -\epsilon_s E_s$.

The dependence of the space-charge density from the surface potential[16] is depicted in Fig. 21.32. When Ψ_s is negative, $\mathcal{F}$ is dominated by the first term in (21.81a) and the space charge is positive (accumulation) and proportional to $Q_s \propto \exp(\beta|\Psi|/2)$. For $\Psi_s = 0$ the (ideal) MIS diode is under flat-band condition and the space charge is zero. For $\Psi_s > 0$ the space charge is negative. For $0 < \Psi_s \leq \Psi_B$

[16] We note that we discuss the space-charge region now only with regard to Ψ_s, the voltage drop across the semiconductor, and the dependence of Ψ_s on the bias of the diode will be discussed in the next section.

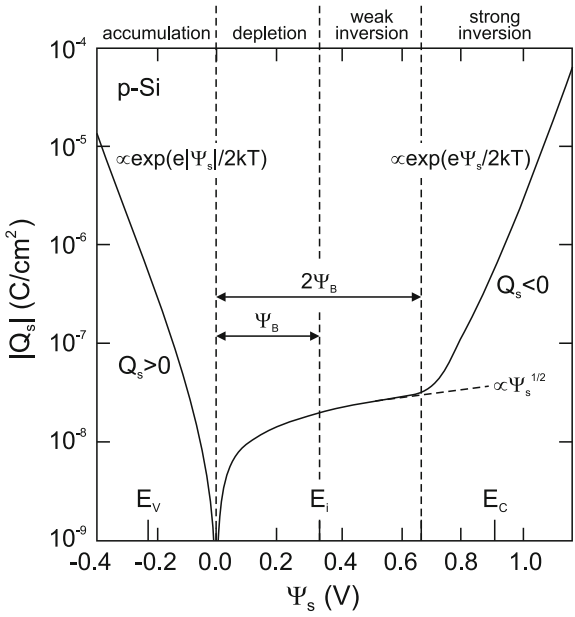


Fig. 21.32 Dependence of the space charge on the surface potential Ψ_s for p-type silicon with $N_A = 4 \times 10^{15}\,\mathrm{cm}^{-3}$ at $T = 300\,\mathrm{K}$. The flat-band condition is present for $\Psi_s = 0$, strong inversion for $\Psi_s > 2\,\Psi_B$

the space charge is due to ionized impurities (depletion) and $\mathcal{F}$ is dominated by the second term in (21.81a), i.e. $\Psi_s \propto \sqrt{\Psi_s}$. For $\Psi_B \leq \Psi_s \leq 2\,\Psi_B$ the diode is in the weak inversion regime and still $\Psi_s \propto \sqrt{\Psi_s}$.

Eventually the dominating term of the second bracket in (21.81a) $f_1 = n_{p_0}/p_{p_0}$ $\exp(\beta\Psi) = (n_i/N_A)^2 \exp(\beta\Psi)$ becomes comparable and exponentially larger than the dominating term of the first bracket $f_2 = \beta\Psi$. Solving $f_1 = f_2$ for $\Psi = \gamma\,\Psi_B$ yields the equation $(N_A/n_i)^{\gamma-2} = \gamma \ln(N_A/n_i)$ and a value γ larger than and close[17] to 2. Thus for $\Psi_s > \Psi_s^{inv} \approx 2\,\Psi_B$ (using (21.74)),

$$\Psi_s^{inv} \approx \frac{2\,kT}{e} \ln\left(\frac{N_A}{n_i}\right), \tag{21.84}$$

strong inversion starts and the space charge is given as $Q_s \propto -\exp(\beta\Psi/2)$.

For the case of strong inversion the band diagram is shown in Fig. 21.33 together with the charge, field and potential. The total voltage drop V across the MIS diode is

$$V = V_i + \Psi_s, \tag{21.85}$$

with V_i being the voltage drop across the insulator. In the case of inversion, the charge (per unit area) in the space-charge region

$$Q_s = Q_d + Q_n \tag{21.86}$$

[17] For $N_A/n_i = 10^4$, 10^6 and 10^8, we find $\gamma = 2.33$, 2.25 and 2.20, respectively.

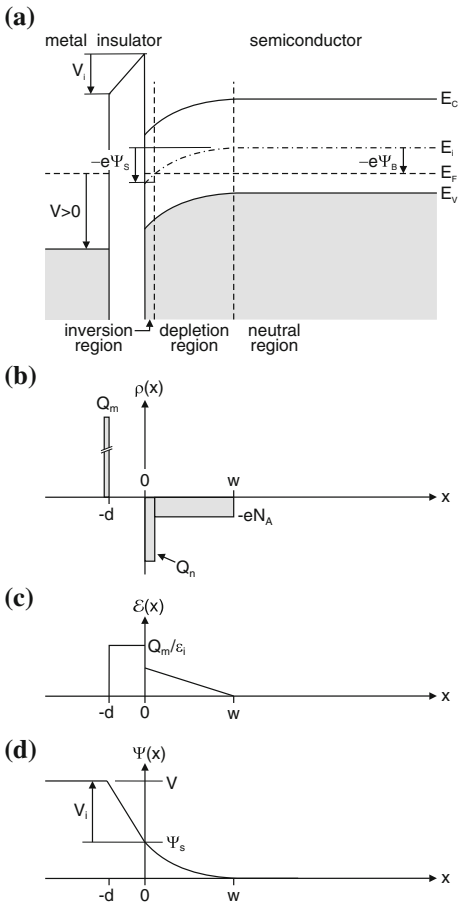


Fig. 21.33 Ideal MIS diode at inversion: (**a**) band diagram, (**b**) charge distribution, (**c**) electric field and (**d**) potential

is composed of the depletion charge (ionized acceptors)

$$Q_\mathrm{d} = -e\, w\, N_\mathrm{A}, \tag{21.87}$$

with w being the width of the depletion region, and the inversion charge Q_n, which is present only close to the interface.

The metal surface carries the opposite charge

$$Q_\mathrm{m} = -Q_\mathrm{s} \tag{21.88}$$

due to global charge neutrality. The insulator itself does not contribute charges in the case of an ideal MIS diode.

21.3.3 Capacitance

The insulator represents a capacitor with the dielectric constant ϵ_i and a thickness d. Therefore, the capacitance is

$$C_i = \frac{\epsilon_i}{d}. \tag{21.89}$$

Between the charges $-Q_s$ and Q_s the field strength E_i in the insulator is

$$E_i = \frac{|Q_s|}{\epsilon_i}. \tag{21.90}$$

The voltage drop V_i across the insulator is given by

$$V_i = E_i\, d = \frac{|Q_s|}{C_i}. \tag{21.91}$$

The total capacitance C of the MIS diode is given by the insulator capacitance in series with the capacitance C_d of the depletion layer

$$C = \frac{C_i\, C_d}{C_i + C_d}. \tag{21.92}$$

The capacitance of the space-charge region varies with the applied bias (Fig. 21.34). For forward bias (accumulation), the capacitance of the space-charge region is high. Therefore, the total capacitance of the MIS diode is given by the insulator capacitance $C \approx C_i$. When the voltage is reduced, the capacitance of the space-charge region drops to $C_d = \epsilon_s/L_D$ for the flat-band case ($\Psi_s = 0$). For a high reverse voltage, the semiconductor is inverted at the surface and the space-charge region capacitance is high again. In this case, the total capacitance is given by $C \approx C_i$ again.

The previous consideration assumes that the charge density in the semiconductor can follow changes of the bias sufficiently fast.[18] The inversion charge must disappear via recombination that is limited by the recombination time constant τ. For frequencies around τ^{-1} or faster, the charge in the inversion layer cannot follow and the capacitance of the semiconductor is given by the value $C_d \cong \epsilon_s/w_m$. w_m (Fig. 21.35) is the maximum depletion-layer width present at the beginning of inversion (cf. (21.10) and (21.84))

$$w_m \cong \sqrt{\frac{2\,\epsilon_s}{e\,N_A}\,\Psi_s^{inv}} = \sqrt{\frac{4\,\epsilon_s\,kT}{e^2\,N_A}\,\ln\left(\frac{N_A}{n_i}\right)}. \tag{21.93}$$

[18] Typically, a dc bias voltage V is set and the capacitance is sampled with a small ac voltage of amplitude ΔV, with $\delta V \ll V$.

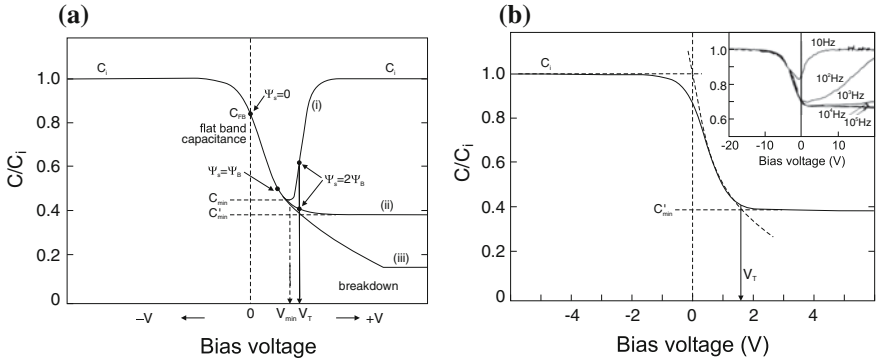


Fig. 21.34 (**a**) Schematic dependence of the capacitance of a MIS diode on the bias for (i) low frequencies, (ii) high frequencies and (iii) deep depletion. (**b**) High-frequency capacitance of a Si/SiO_2 diode. The *inset* shows the frequency dependence. Adapted from [1474]

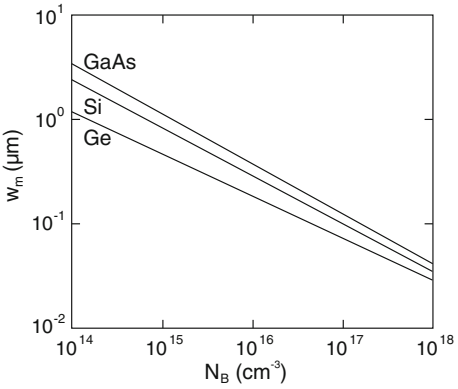


Fig. 21.35 Maximum width of the depletion layer w_m (21.93) at room temperature for deep depletion for GaAs, Si and Ge diodes as a function of bulk doping level

For further increased voltage (into the inversion regime), the electric field is screened by the inversion charge and the width of the depletion layer remains constant. Therefore, the total capacitance in the inversion regime is given by

$$C \cong \frac{\epsilon_i}{d + w_m\, \epsilon_i/\epsilon_s}. \tag{21.94}$$

21.3.4 Nonideal MIS Diode

In a real, i.e. nonideal, MIS diode, the difference ϕ_{ms} in the work functions of the metal and semiconductor (cf. (21.73)) is no longer zero. Therefore, the capacitance

versus voltage relation is shifted with respect to the ideal MIS diode characteristic by the *flat-band voltage shift* V_{FB}

$$V_{FB} = \phi_{ms} - \frac{Q_{ox}}{C_i}. \tag{21.95}$$

Additionally, the flat-band voltage can be shifted by charges Q_{ox} in the oxide that have been neglected so far. Such charges can be trapped, i.e. fixed with regard to their spatial position, or mobile, e.g. ionic charges such as sodium.

For Al as metal ($\phi_m = 4.1\,eV$) and n-type Si ($\phi_s = 4.35\,eV$), the flat-band voltage shift is $\phi_{ms} = -0.25\,V$, as shown schematically in Fig. 21.36a for zero bias. V_{FB} is split into 0.2 eV and 0.05 eV for the oxide and the silicon, respectively. In Fig. 21.36b, the dependence of ϕ_{ms} on the doping, conductivity type and metal is shown for various SiO_2–Si MIS diodes. An Au-SiO_2–Si diode with p-type Si and $N_A \approx 10^{15}\,cm^{-3}$ fulfills the condition of an ideal MIS diode with regard to $\phi_{ms} = 0$.

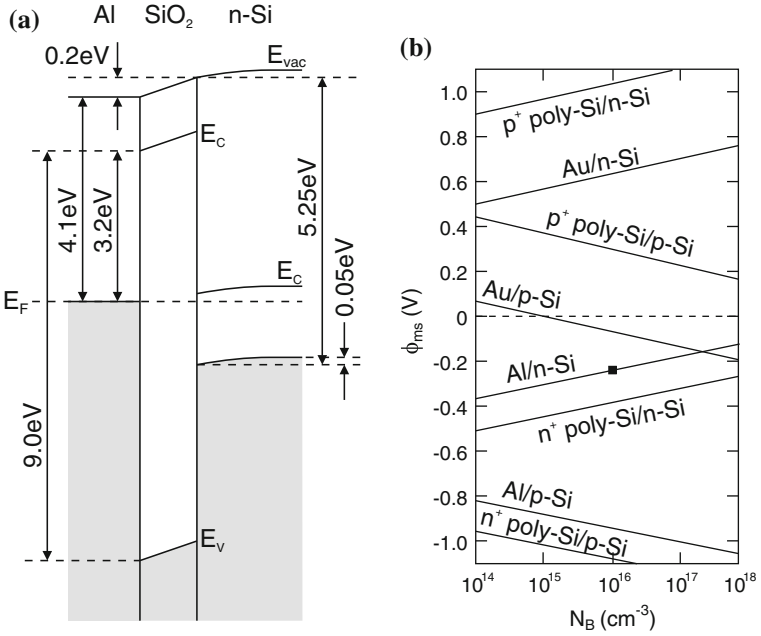


Fig. 21.36 (**a**) Schematic band diagram of an Al–SiO_2–Si (n-type) diode with 50 nm oxide thickness and $N_D = 10^{16}\,cm^{-3}$ for zero bias. Based on data from [1475]. (**b**) Difference of work functions ϕ_{ms} for SiO_2–Si MIS diodes and various doping levels and electrode materials (Al, Au and polycrystalline Si). The *square* represents the situation depicted in panel (**a**). Based on data from [1476]

21.4 Bipolar Diodes

A large class of diodes is based on pn-junctions. In a homo pn-junction an n-doped region is next to a p-doped region of the same semiconductor. Such a device is called *bipolar*. At the junction a depletion region forms. The transport properties are determined by the minority carriers. An important variation is the pin-diode in which an intrinsic (or lowly doped) region is between the doped region (Sect. 21.5.8). If the differently doped regions belong to different semiconductor materials, the diode is a heterostructure pn-diode (Sect. 21.4.6). Various schemes have been used to fabricate pn diodes (Fig. 21.37).

21.4.1 Band Diagram

If the doping profile is arbitrarily sharp, the junction is called *abrupt*. This geometry is the case for epitaxial pn-junctions where the differently doped layers are grown on top of each other.[19] For junctions that are fabricated by diffusion, the abrupt approximation is suitable for alloyed, ion-implanted and shallow-diffused junctions. For deep-diffused junctions a linearly graded approximation is better (Fig. 21.38), which is treated in more detail in [500]. If one doping level is much higher that the other, the junction is termed a *one-sided* (abrupt) junction. If $n \gg p$ ($p \gg n$), the junction is denoted as an n^+p-diode (p^+n-diode).

The thermodynamical equilibrium of a pn-diode is considered here only for the electronic system. The thermodynamic stability of the atomic doping distribution is discussed in [1477]. Typically, thermodynamics works to randomize the chemical concentration gradient; the existence of a pn-junction is due to the extremely low diffusion coefficient of dopants in the semiconductor lattice. Elevated temperatures can cause the destruction of the pn-diode via enhanced dopant diffusion (Sect. 4.2.3). However, a thermodynamically stable concentration gradient and thus a built-in field can exist in a multi-component system [1477].

21.4.2 Space-Charge Region

In thermodynamical equilibrium the Fermi level is constant ($\nabla E_\mathrm{F} = 0$). The built-in voltage V_bi is given by (see Fig. 21.39c)

$$e\,V_\mathrm{bi} = E_\mathrm{g} + e\,V_\mathrm{n} + e\,V_\mathrm{p}, \tag{21.96}$$

[19] The choices of dopant and the growth conditions, in particular the temperature, need to be made such that no interdiffusion of the dopants takes place.

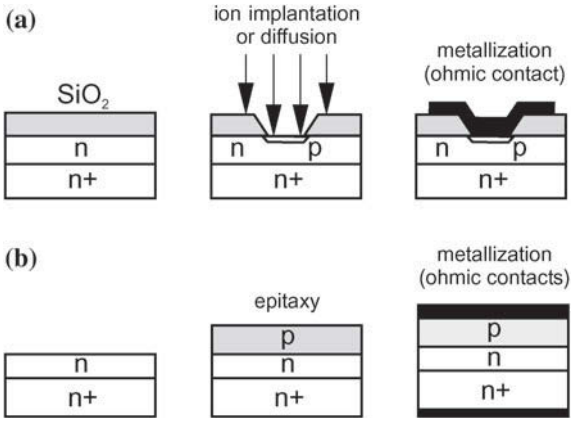


Fig. 21.37 Schematic fabrication technologies for bipolar diodes: (**a**) Planar junction with local impurity incorporation (diffusion from gas phase or ion implantation) through mask and contact metallization, (**b**) epitaxial junction

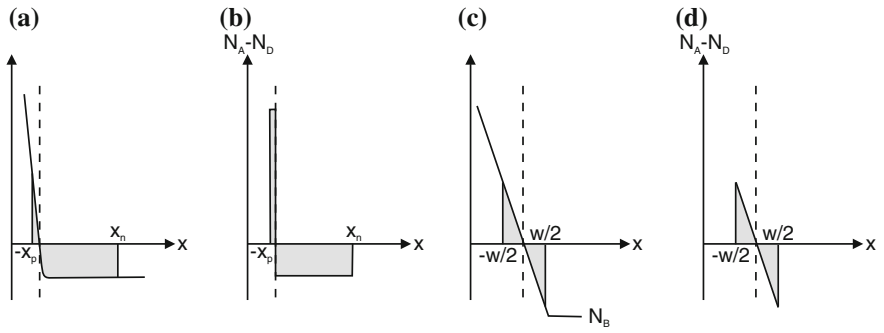


Fig. 21.38 Description of doping distribution with (**a**, **b**) abrupt approximation and with (**c**, **d**) linearly graded junction. (**a**, **c**) show real impurity concentration, (**b**, **d**) idealized doping profile

where V_n is the difference between conduction band and Fermi level on the n-side, $-eV_n = E_C - E_F$. V_p is the difference between valence band and Fermi level on the p-side, $-eV_p = E_F - E_V$. For the nondegenerate semiconductor V_n, $V_p < 0$ and (using (7.12), (7.10) and (7.11))

$$
\begin{aligned}
e\,V_{bi} &= kT\,\ln\left(\frac{N_C N_V}{n_i^2}\right) - \left[kT\,\ln\left(\frac{N_C}{n_{n_0}}\right) + kT\,\ln\left(\frac{N_V}{p_{p_0}}\right)\right] \\
&= kT\,\ln\left(\frac{p_{p_0}\,n_{n_0}}{n_i^2}\right) = kT\,\ln\left(\frac{p_{p_0}}{p_{n_0}}\right) = kT\,\ln\left(\frac{n_{n_0}}{n_{p_0}}\right) \qquad (21.97a) \\
&\cong kT\,\ln\left(\frac{N_A\,N_D}{n_i^2}\right). \qquad (21.97b)
\end{aligned}
$$

The electron and hole densities on either side of the junction (n_{p_0} and p_{p_0} at $x = -x_p$ and n_{n_0} and p_{n_0} at $x = x_n$) are related to each other by (from rewriting (21.97a))

$$n_{p_0} = n_{n_0} \exp(-\beta V_{bi}) \tag{21.98a}$$

$$p_{n_0} = p_{p_0} \exp(-\beta V_{bi}) . \tag{21.98b}$$

Microscopically, the equilibration of the Fermi levels on the n- and p-side occurs via the diffusion of electrons and holes to the p- and n-side, respectively. The electrons and holes recombine in the depletion layer. Therefore, on the n-side the ionized donors and on the p-side the ionized acceptors remain (Fig. 21.39a). These charges build up an electric field (Fig. 21.39d) that works against the diffusion current. At thermal equilibrium the diffusion and drift currents cancel and the Fermi level is constant.

Values for the built-in potential are depicted in Fig. 21.40 for Si and GaAs diodes. The spatial dependence of the potential in the depletion layer is determined by the Poisson equation.

We assume here the complete ionization of the donors and acceptors. Also, we neglect at first majority carriers in the depletion layers on the n- and p-sides.[20] With these approximations, the Poisson equation in the depletion layers on the n- and p-side reads

$$\frac{\partial^2 V}{\partial x^2} = -\frac{e N_D}{\epsilon_s}, \quad 0 \le x \le x_n \tag{21.99a}$$

$$\frac{\partial^2 V}{\partial x^2} = \frac{e N_A}{\epsilon_s}, \quad -x_p \le x \le 0. \tag{21.99b}$$

One integration yields (together with the boundary conditions that the field is zero at the boundaries of the depletion layer) the electric field in the two regions

$$E(x) = \frac{e}{\epsilon_s} N_D (x - x_n), \quad 0 \le x \le x_n \tag{21.100a}$$

$$E(x) = -\frac{e}{\epsilon_s} N_A (x + x_p), \quad -x_p \le x \le 0. \tag{21.100b}$$

The maximum field strength E_m is present at $x = 0$ and is given by

$$E_m = -\frac{e N_D x_n}{\epsilon_s} = -\frac{e N_A x_p}{\epsilon_s}. \tag{21.101}$$

The continuity of the field at $x = 0$ is equivalent to the overall charge neutrality

$$N_D x_n = N_A x_p. \tag{21.102}$$

[20] An abrupt decrease of the majority carrier density at the border of the space-charge region corresponds to zero temperature.

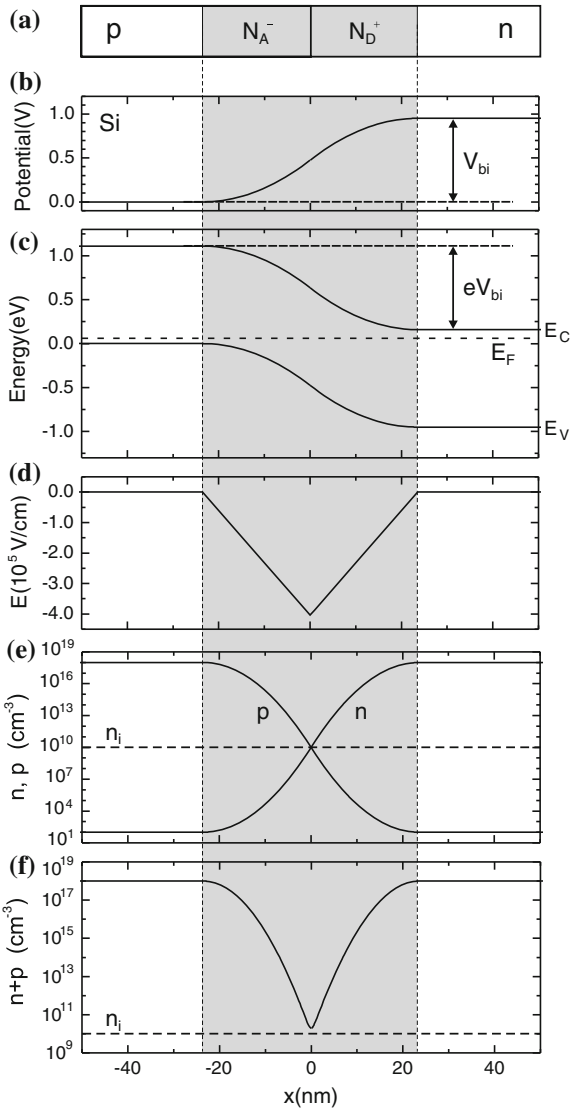


Fig. 21.39 pn-junction (abrupt approximation) in thermal equilibrium (zero bias) for Si at room temperature and $N_A = N_D = 10^{18}\,\mathrm{cm}^{-3}$. (**a**) Schematic representation of p-doped and n-doped region with depletion layer (*grey area*) and fixed space charges, (**b**) diffusion potential, (**c**) band diagram with Fermi level (*dashed line*), (**d**) electric field, (**e**) free-carrier concentrations n and p and (**f**) total free-carrier density $n + p$

Another integration yields the potential (setting $V(x = 0) = 0$)

$$V(x) = -E_\mathrm{m}\left(x - \frac{x^2}{2x_\mathrm{n}}\right),\ 0 \le x \le x_\mathrm{n} \tag{21.103a}$$

$$V(x) = -E_\mathrm{m}\left(x + \frac{x^2}{2x_\mathrm{p}}\right),\ -x_\mathrm{p} \le x \le 0. \tag{21.103b}$$

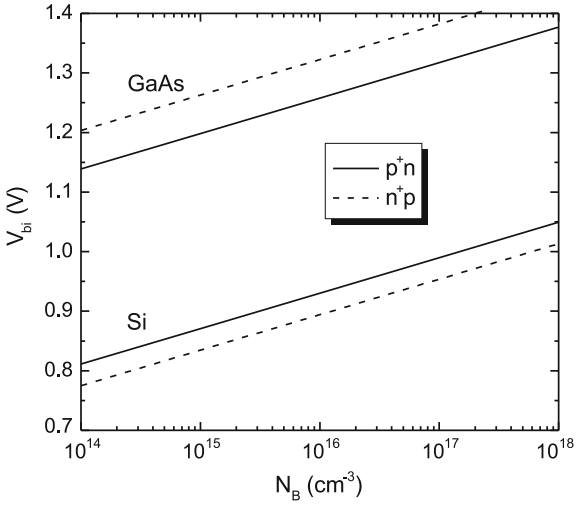


Fig. 21.40 Built-in voltage as a function of doping for one-sided Si and GaAs pn-diodes

The built-in potential $V_{\rm bi} = V(x_{\rm n}) - V(-x_{\rm p}) > 0$ is related to the maximum field via

$$V_{\rm bi} = -\frac{1}{2} E_{\rm m}\, w, \tag{21.104}$$

where $w = x_{\rm n} + x_{\rm p}$ is the total width of the depletion layer. The elimination of $E_{\rm m}$ from (21.101) and (21.104) yields

$$w = \sqrt{\frac{2\,\epsilon_{\rm s}}{e}\left(\frac{N_{\rm A}+N_{\rm D}}{N_{\rm A}\,N_{\rm D}}\right) V_{\rm bi}} = x_{\rm n} + x_{\rm p}; \tag{21.105}$$

the two parts of the depletion layer are given by

$$x_{\rm n} = \frac{N_{\rm A}}{N_{\rm A}+N_{\rm D}} \sqrt{\frac{2\,\epsilon_{\rm s}}{e}\,\frac{N_{\rm A}+N_{\rm D}}{N_{\rm A}\,N_{\rm D}}\,V_{\rm bi}} \tag{21.106a}$$

$$x_{\rm p} = \frac{N_{\rm D}}{N_{\rm A}+N_{\rm D}} \sqrt{\frac{2\,\epsilon_{\rm s}}{e}\,\frac{N_{\rm A}+N_{\rm D}}{N_{\rm A}\,N_{\rm D}}\,V_{\rm bi}}. \tag{21.106b}$$

For p^+n and n^+p junctions, the width of the depletion layer is determined by the lowly doped side of the junction

$$w = \sqrt{\frac{2\,\epsilon_{\rm s}}{e\,N_{\rm B}}\,V_{\rm bi}}, \tag{21.107}$$

where $N_{\rm B}$ denotes the doping of the lowly doped side, i.e. $N_{\rm A}$ for a n^+p diode and $N_{\rm D}$ for a p^+n diode.

If the spatial variation of the majority carrier density is considered in more detail (and for finite temperature, cf. (21.13)), an additional term $-2kT/e = -2/\beta$ is added [1473] to V_{bi}

$$w = \sqrt{\frac{2\,\epsilon_{\mathrm{s}}}{e}\left(\frac{N_{\mathrm{A}}+N_{\mathrm{D}}}{N_{\mathrm{A}}\,N_{\mathrm{D}}}\right)\left(V_{\mathrm{bi}}-V-2\beta^{-1}\right)}. \tag{21.108}$$

Also, the external bias V has been included in the formula. If w_0 denotes the depletion layer width at zero bias, the depletion layer width for a given voltage V can be written as

$$w(V) = w_0\sqrt{1-\frac{V}{V_{\mathrm{bi}}-2/\beta}} \approx w_0\sqrt{1-\frac{V}{V_{\mathrm{bi}}}}. \tag{21.109}$$

Using the Debye length (cf. (21.81b))

$$L_{\mathrm{D}} = \sqrt{\frac{\epsilon_{\mathrm{s}}\,kT}{e^2\,N_{\mathrm{B}}}}, \tag{21.110}$$

the depletion layer width for a one-sided diode can be written as (with $\beta = e/kT$)

$$w = L_{\mathrm{D}}\sqrt{2\,(\beta V_{\mathrm{bi}} - \beta V - 2)}. \tag{21.111}$$

The Debye length is a function of the doping level and is shown for Si in Fig. 21.41. For a doping level of $10^{16}\,\mathrm{cm}^{-3}$ the Debye length in Si is 40 nm at room temperature.

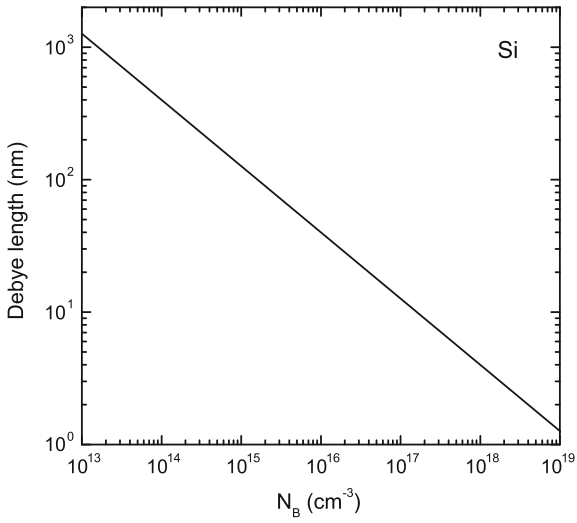


Fig. 21.41 Debye length in Si at room temperature as a function of the doping level N_{B} according to (21.110)

For one-sided junctions the depletion layer width is about $6\,L_D$ for Ge, $8\,L_D$ for Si and $10\,L_D$ for GaAs.

The external bias is counted positive if the '+' ('−') pole is at the p-side (n-side). The reverse voltage has opposite polarity. If a reverse bias is applied, the depletion layer width is increased (Fig. 21.42).

21.4.3 Capacitance

The capacitance of the depletion layer is the charge change upon a change of the external bias. It is given as

$$C = \left|\frac{\mathrm{d}Q}{\mathrm{d}V}\right| = \frac{\mathrm{d}(eN_B w)}{\mathrm{d}(w^2 eN_B/2\epsilon_s)} = \frac{\epsilon_s}{w} = \frac{\epsilon_s}{\sqrt{2}\,L_D}\sqrt{\beta V_{bi} - \beta V - 2}. \tag{21.112}$$

Therefore, the capacitance of the depletion layer is inversely proportional to the depletion-layer width (see the two scales in Fig. 21.42). A detailed treatment has been given in [1478]. $1/C^2$ is proportional to the external bias

$$\frac{1}{C^2} = \frac{2\,L_D^2}{\epsilon_s^2}\,(\beta V_{bi} - \beta V - 2)\,. \tag{21.113}$$

From C–V spectroscopy the doping level can be obtained from the slope

$$\frac{\mathrm{d}(1/C^2)}{\mathrm{d}V} = \frac{2\,\beta\,L_D^2}{\epsilon_s^2} = \frac{2}{e\,\epsilon_s\,N_B}. \tag{21.114}$$

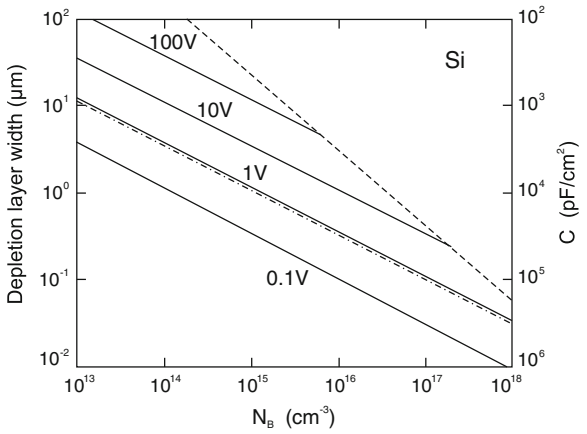


Fig. 21.42 Width of the depletion layer and capacitance per area for one-sided, abrupt Si junctions for various values of $V_{bi} - V - 2kT/e$ as labeled. The *dash-dotted line* is for zero bias, the *dashed line* is the limit due to avalanche breakdown. Adapted from [500]

From the extrapolation to the voltage for which $1/C^2 = 0$ the built-in voltage can be obtained.

21.4.4 Current–Voltage Characteristics

Ideal Current–Voltage Characteristics

Now, the currents in thermodynamical equilibrium ($V = 0$) and under bias are discussed. A diode characteristic will be obtained. We work at first with the following assumptions: abrupt junction, Boltzmann approximation, low injection, i.e. the injected minority-carrier density is small compared to the majority-carrier density, and zero generation current in the depletion layer, i.e. the electron and hole currents are constant throughout the depletion layer. In the presence of a bias, electrons and holes have quasi-Fermi levels and the carrier densities are given by (cf. (7.56a, b))

$$n = N_{\mathrm{C}} \exp\left(\frac{F_{\mathrm{n}} - E_{\mathrm{C}}}{kT}\right) \tag{21.115a}$$

$$p = N_{\mathrm{V}} \exp\left(-\frac{F_{\mathrm{p}} - E_{\mathrm{V}}}{kT}\right). \tag{21.115b}$$

Using the intrinsic carrier concentration, we can write

$$n = n_{\mathrm{i}} \exp\left(\frac{F_{\mathrm{n}} - E_{\mathrm{i}}}{kT}\right) = n_{\mathrm{i}} \exp\left[\beta(\psi - \phi_{\mathrm{n}})\right] \tag{21.116a}$$

$$p = n_{\mathrm{i}} \exp\left(-\frac{F_{\mathrm{p}} - E_{\mathrm{i}}}{kT}\right) = n_{\mathrm{i}} \exp\left[\beta(\phi_{\mathrm{p}} - \psi)\right], \tag{21.116b}$$

where ϕ and ψ are the potentials related to the (quasi-) Fermi level and the intrinsic Fermi levels, $-e\phi_{\mathrm{n,p}} = F_{\mathrm{n,p}}$ and $-e\psi = E_{\mathrm{i}}$. The potentials ϕ_{n} and ϕ_{p} can also be written as

$$\phi_{\mathrm{n}} = \psi - \beta^{-1} \ln\left(\frac{n}{n_{\mathrm{i}}}\right) \tag{21.117a}$$

$$\phi_{\mathrm{p}} = \psi + \beta^{-1} \ln\left(\frac{p}{p_{\mathrm{i}}}\right). \tag{21.117b}$$

The product np is given by

$$np = n_{\mathrm{i}}^2 \exp\left[\beta(\phi_{\mathrm{p}} - \phi_{\mathrm{n}})\right]. \tag{21.118}$$

Of course, at thermodynamical equilibrium (zero bias) $\phi_{\mathrm{p}} = \phi_{\mathrm{n}}$ and $np = n_{\mathrm{i}}^2$. For forward bias $\phi_{\mathrm{p}} - \phi_{\mathrm{n}} > 0$ (Fig. 21.43a) and $np > n_{\mathrm{i}}^2$. For reverse bias $\phi_{\mathrm{p}} - \phi_{\mathrm{n}} < 0$ (Fig. 21.43b) and $np < n_{\mathrm{i}}^2$.

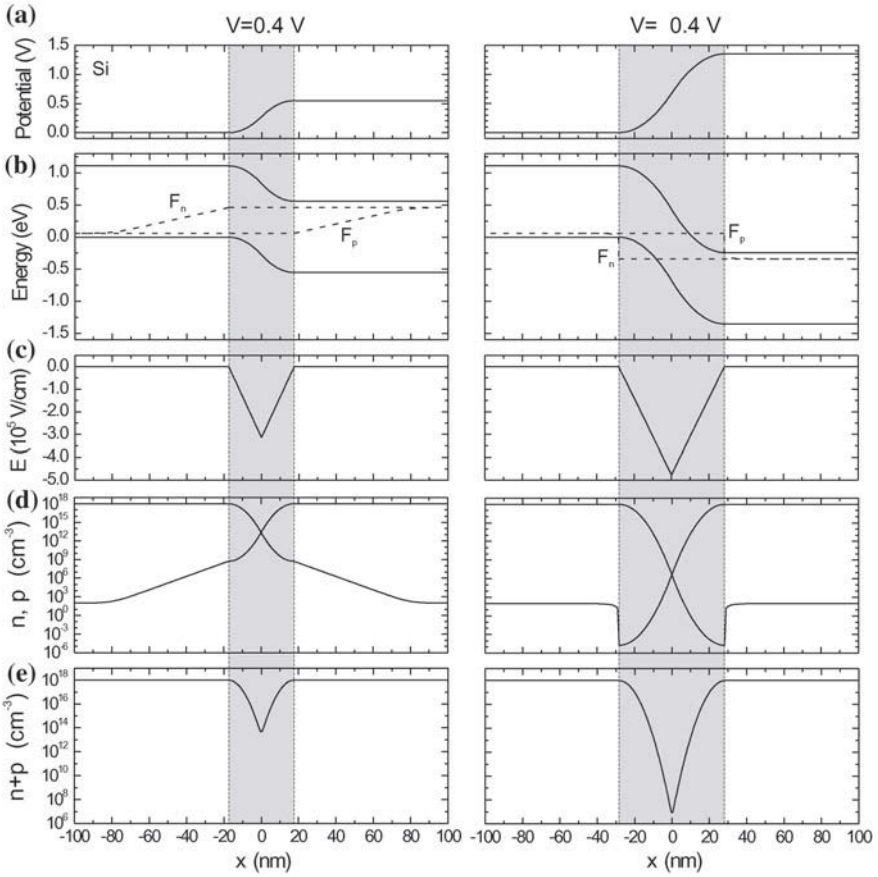


Fig. 21.43 (**a**) Diffusion potential, (**b**) band diagram, (**c**) electric field, (**d**) electron and hole concentrations and (**e**) $n + p$ under forward bias $+0.4\,\mathrm{V}$ (*left panel*) and reverse bias (*right panel*) $-0.4\,\mathrm{V}$ for a silicon pn-diode at room temperature with $N_\mathrm{A} = N_\mathrm{D} = 10^{18}\,\mathrm{cm}^{-3}$ (same as in Fig. 21.39). The *dashed lines* in (b) are the electron and hole quasi-Fermi levels F_n and F_p. The depletion layer is shown as the *grey area*. The diffusion length in the n- and p-type material is taken as 4 nm. This value is much smaller than the actual diffusion length (μm-range) and is chosen here only to show the carrier concentration in the depletion layer and the neutral region in a single graph

The electron current density (per unit area) is given by (8.59a) that reads here with $\mathbf{E} = -\nabla\psi$ and n given by (21.116a) as[21]

$$\mathbf{j}_\mathrm{n} = -e\,\mu_\mathrm{n}\,\left(n\,\mathbf{E} + \beta^{-1}\,\nabla n\right) = e\,n\,\mu_\mathrm{n}\,\nabla\phi_\mathrm{n}. \tag{21.119}$$

[21] We remind the reader that μ_n was defined as a negative number.

Similarly, we obtain for the hole current density (using (8.59b) and (21.116b))

$$\mathbf{j}_\mathrm{p} = e\,\mu_\mathrm{p}\,\left(p\,\mathbf{E} - \beta^{-1}\,\nabla p\right) = -e\,p\,\mu_\mathrm{p}\,\nabla\phi_\mathrm{p}. \tag{21.120}$$

The current through the depletion layer is constant since no recombination/generation is assumed for now. The gradient of the quasi-Fermi levels in the depletion layer is very small and the quasi-Fermi levels $\phi_{\mathrm{n,p}}$ are practically constant. The electron and hole currents are shown in Fig. 21.44 together with the carrier densities. The change of carrier density is mostly due to the variation of ψ (or E_i).

Therefore, the voltage drop across the depletion layer is $V = \phi_\mathrm{p} - \phi_\mathrm{n}$ and (21.118) reads

$$n\,p = n_\mathrm{i}^2\,\exp\left(\beta V\right). \tag{21.121}$$

The electron density at the boundary of the depletion layer on the p-side (at $x = -x_\mathrm{p}$) is (using (21.121))

$$n_\mathrm{p} = \frac{n_\mathrm{i}^2}{p_\mathrm{p}}\,\exp\left(\beta V\right) = n_{\mathrm{p}_0}\,\exp\left(\beta V\right). \tag{21.122}$$

Similarly, the hole density on the n-side at $x = x_\mathrm{n}$ is given by

$$p_\mathrm{n} = p_{\mathrm{n}_0}\,\exp\left(\beta V\right). \tag{21.123}$$

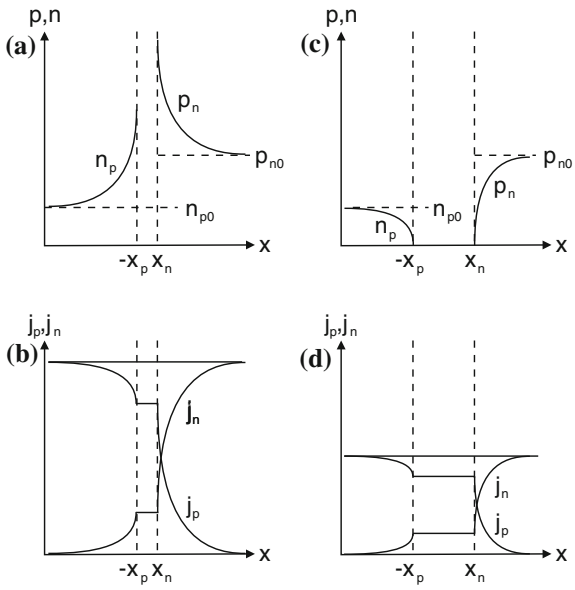


Fig. 21.44 Carrier densities (**a**, **c**) and current densities (**b**, **d**) (linear scales) in a pn-diode under (**a**, **b**) forward bias and (**c**, **d**) reverse bias

From the continuity equation and the boundary condition that far away from the depletion layer the hole density is p_{n_0}, the hole density on the n-side is given by

$$p_n(x) - p_{n_0} = p_{n_0} \left[\exp(\beta V) - 1\right] \exp\left(-\frac{x - x_n}{L_p}\right), \tag{21.124}$$

where $L_p = \sqrt{D_p \tau_p}$ is the hole (minority-carrier) diffusion length.

The hole current density at the boundary of the depletion layer on the n-side is

$$j_p(x_n) = -e D_p \frac{\partial p_n}{\partial x}|_{x_n} = \frac{e D_p p_{n_0}}{L_p} \left[\exp(\beta V) - 1\right]. \tag{21.125}$$

Similarly, the electron current in the depletion layer is

$$j_n(-x_p) = \frac{e D_n n_{p_0}}{L_n} \left[\exp(\beta V) - 1\right]. \tag{21.126}$$

The total current due to diffusion is

$$j_d = j_p(x_n) + j_n(-x_p) = j_s \left[\exp(\beta V) - 1\right], \tag{21.127}$$

with the saturation current given by

$$j_s^d = \frac{e D_p p_{n_0}}{L_p} + \frac{e D_n n_{p_0}}{L_n}. \tag{21.128}$$

This dependence is an ideal diode characteristic and the famous result from Shockley. For a one-sided (p^+n-) diode, the saturation current is

$$j_s^d \cong \frac{e D_p p_{n_0}}{L_p} \cong e \left(\frac{D_p}{\tau_p}\right)^{1/2} \frac{n_i^2}{N_B}. \tag{21.129}$$

The saturation depends via D_p/τ_p weakly on the temperature. The term n_i^2 depends on T, proportional to $T^3 \exp(-E_g/kT)$, which is dominated by the exponential function.

If the minority carrier lifetime is given by the radiative recombination (10.19), the hole diffusion length is

$$L_p = \sqrt{\frac{D_p}{B n_{n_0}}}. \tag{21.130}$$

For GaAs (Tables 8.2 and 10.1) with $N_D = 10^{18}\,\mathrm{cm}^{-3}$, we find $\tau_p = 10\,\mathrm{ns}$ and $L_p \approx 3\,\mu\mathrm{m}$. For L_n we find $14\,\mu\mathrm{m}$, however, the lifetime at room temperature can be significantly shorter due to nonradiative recombination and subsequently also the

diffusion length will be shorter (by about a factor of 10). For $L \sim 1\,\mu\text{m}$, the diffusion saturation current is $j_s^d \sim 4 \times 10^{-20}\,\text{A/cm}^2$.

The radiative recombination rate (band–band recombination, b–b) in the neutral n-region (as relevant for LEDs, see Sect. 23.3) is $B(np - n_i^2) \approx Bn_{n_0}(p_n(x) - p_{n_0})$. Therefore, the recombination current $j_{d,n}^{b-b}$ in the neutral n-region is (using (21.124))

$$\begin{aligned} j_{d,n}^{b-b} &= e \int_{x_n}^{\infty} B\, n_i^2 \left[\exp(\beta V) - 1\right] \exp\left(-\frac{x - x_n}{L_p}\right) dx \\ &= e\, B\, L_p\, n_i^2 \left[\exp(\beta V) - 1\right]. \end{aligned} \tag{21.131}$$

For the neutral region on the p-side a similar expression results. The total radiative recombination current from the neutral regions is

$$j_d^{b-b} = e\, B\, (L_n + L_p)\, n_i^2 \left[\exp(\beta V) - 1\right]. \tag{21.132}$$

For GaAs, the saturation current for the radiative recombination in the neutral region

$$j_s^{r,b-b} = e\, B\, (L_n + L_p)\, n_i^2 \tag{21.133}$$

is (Tables 7.1 and 10.1) for a diffusion length of $1\,\mu\text{m}$ of $j_s^{r,b-b} \sim 4 \times 10^{-21}\,\text{A/cm}^2$.

Since the (radiative) minority-carrier lifetime is inversely proportional to the majority-carrier density, the relevant diffusion length is that of the side with the lower doping level and is given by

$$L = \frac{1}{n_i} \sqrt{\frac{D_B\, N_B}{B}}, \tag{21.134}$$

where D_B is the minority-carrier diffusion coefficient on the lowly doped side. The radiative recombination current from the neutral region can be written as

$$j_d^{b-b} = e \sqrt{B\, D_B\, N_B}\; n_i \left[\exp(\beta V) - 1\right]. \tag{21.135}$$

The I–V characteristic for pn-diodes from two semiconductors with different band gap are shown in Fig. 21.45a (for Ge and Si). The Si diode has the much smaller saturation current. The saturation current increases at higher temperature (Fig. 21.46). A wide-gap pn diode (GaN) exhibits small saturation currents (The $n = 1.5$ part of the GaN in Fig. 21.45b belongs to a saturation current density of $7 \times 10^{-27}\,\text{A/cm}^2$, the $n = 1.5$ part to $\approx 10^{-34}\,\text{A/cm}^2$) and displays sizable current (density) only for larger voltages.

Real I–V Characteristics

Besides the ideal I–V characteristics due to diffusion several effects contribute to the characteristics of real bipolar diodes:

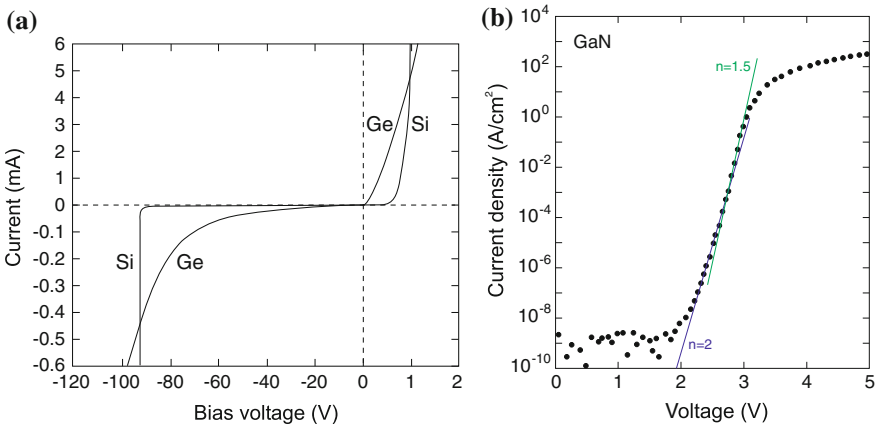


Fig. 21.45 Comparison of the characteristics of (**a**) Ge and Si and (**b**) GaN pn diodes at room temperature. Note the different scales in the forward and reverse regime in (**a**). The *lines* in (**b**) are related to ideality factors 2 and 1.5 as labeled. Adapted from [1479]

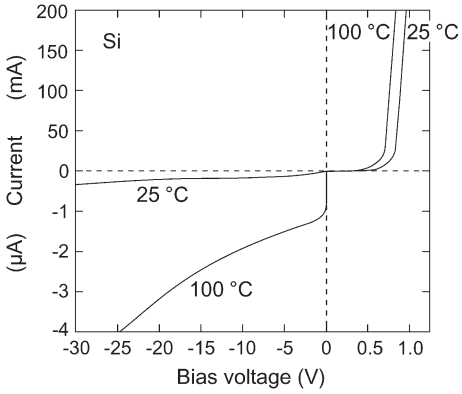


Fig. 21.46 Characteristics of a Si power diode at two temperatures, 25 and 100 °C

- a generation–recombination (G–R) current is present due to traps in the depletion layer
- already for fairly small forward voltages, high injection conditions are present, i.e. $p_n \ll n_n$ is no longer valid
- the series resistance R_s of the diode is finite (ideally $R_s = 0$)
- the diode has a finite parallel (shunt) resistance R_p (ideally $R_p = \infty$)
- at high reverse voltage the junction breaks down; this phenomenon is treated in Sect. 21.4.5

First, we consider the generation–recombination current due to band–impurity (b–i) processes (see Sect. 10.9). Such recombination is nonradiative or at least does not produce photons with an energy close to the band gap. The impact on the I–V diode characteristic has first been considered in [1480].

The net b–i recombination rate is given by (10.52). For reverse voltage, the generation dominates the G–R current. For $n < n_\mathrm{i}$ and $p < n_\mathrm{i}$, the net recombination rate r is

$$r \cong \frac{\sigma_\mathrm{n}\,\sigma_\mathrm{p}\,v_\mathrm{th}\,N_\mathrm{t}}{\sigma_\mathrm{n}\,\exp\left(\frac{E_\mathrm{t}-E_\mathrm{i}}{kT}\right)+\sigma_\mathrm{p}\,\exp\left(\frac{E_\mathrm{i}-E_\mathrm{t}}{kT}\right)}\,n_\mathrm{i} \equiv \frac{n_\mathrm{i}}{\tau_\mathrm{e}}, \tag{21.136}$$

where τ_e is the effective electron lifetime. The generation current density is given by

$$j_\mathrm{g} = \frac{e\,n_\mathrm{i}\,w}{\tau_\mathrm{e}}. \tag{21.137}$$

Since the width of the depletion layer varies with the applied reverse bias V, we expect a dependence

$$j_\mathrm{g} \propto \sqrt{V_\mathrm{bi}+|V|}. \tag{21.138}$$

The saturation current is given by the sum of the diffusion and generation parts

$$j_\mathrm{s} = e\sqrt{\frac{D_\mathrm{p}}{\tau_\mathrm{p}}}\,\frac{n_\mathrm{i}^2}{N_\mathrm{D}} + \frac{e\,n_\mathrm{i}\,w}{\tau_\mathrm{e}}. \tag{21.139}$$

In semiconductors with large n_i (narrow band gap, e.g. Ge) the diffusion current will dominate; in Si (larger band gap) the generation current can dominate.

The maximum of the recombination rate is present for $E_\mathrm{t} \approx E_\mathrm{i}$ (10.57). Then $n_\mathrm{t} = p_\mathrm{t} = n_\mathrm{i}$ in (10.52). Assuming $\sigma = \sigma_\mathrm{n} = \sigma_\mathrm{p}$, the recombination rate is

$$r_\mathrm{b-i} = \sigma\,v_\mathrm{th}\,N_\mathrm{t}\,\frac{n\,p-n_\mathrm{i}^2}{n+p+2\,n_\mathrm{i}}. \tag{21.140}$$

Using (21.118) we can write

$$r_\mathrm{b-i} = \sigma\,v_\mathrm{th}\,N_\mathrm{t}\,n_\mathrm{i}\,\frac{n_\mathrm{i}}{n+p+2\,n_\mathrm{i}}\left[\exp\left(\beta V\right)-1\right]. \tag{21.141}$$

The term $\zeta = \frac{n_\mathrm{i}}{n+p+2n_\mathrm{i}}$ is maximal for $n = p$, which is given (from (21.121)) by

$$n_\mathrm{mr} = p_\mathrm{mr} = n_\mathrm{i}\,\exp\left(\beta V/2\right). \tag{21.142}$$

The function $\zeta(x)$ cannot be integrated analytically. In order to evaluate the integral of ζ over the depletion layer

$$\chi = \int_{-x_\mathrm{p}}^{x_\mathrm{n}} \zeta\,\mathrm{d}x, \tag{21.143}$$

the maximum rate

$$\zeta_{\mathrm{mr}} = \frac{n_{\mathrm{i}}}{n_{\mathrm{mr}} + p_{\mathrm{mr}} + 2\,n_{\mathrm{i}}} = \frac{1}{2}\,\frac{1}{1 + \exp\left(\beta V/2\right)} \tag{21.144}$$

can be integrated over the depletion layer as an approximation [500], $\chi \approx \zeta_{\mathrm{mr}}\, w$. This approach yields a recombination current

$$j_{\mathrm{mr}} = \frac{e\,\sigma\,v_{\mathrm{th}}\,N_{\mathrm{t}}\,w\,n_{\mathrm{i}}}{2}\,\frac{\exp\left(\beta V\right) - 1}{\exp\left(\beta V/2\right) + 1} \cong j_{\mathrm{s}}^{\mathrm{mr}}\,\exp\left(\frac{\beta V}{2}\right), \tag{21.145}$$

with $j_{\mathrm{s}}^{\mathrm{mr}} = e\sigma v_{\mathrm{th}} N_{\mathrm{t}} w n_{\mathrm{i}}/2$ and the approximation being valid for $eV/kT \gg 1$. Thus the nonradiative band–impurity recombination is often said to cause an ideality factor of $n = 2$.

For a better approximation of χ, the dependence of the potential $\varphi(x)$ can be approximated as linear (constant-field approximation), i.e. using the local field E_{mr} at the position where $n = p$ [1481]. For a symmetric diode with $n_{\mathrm{n}_0} = p_{\mathrm{p}_0}$, this position is at $x = 0$; for a one-sided junction on the lower-doped side. E_{mr} is given for $p_{\mathrm{p}_0} \leq n_{\mathrm{n}_0}$ by

$$E_{\mathrm{mr}} = -\frac{\sqrt{2}}{w}\,\left(V_{\mathrm{bi}} - V\right)\,\sqrt{1 + \frac{1}{\beta\left(V_{\mathrm{bi}} - V\right)}\,\ln\frac{p_{\mathrm{p}_0}}{n_{\mathrm{n}_0}}}\,\sqrt{1 + \frac{p_{\mathrm{p}_0}}{n_{\mathrm{n}_0}}}. \tag{21.146}$$

For a symmetric diode (21.147a) holds, for a one-sided diode the approximation in (21.147b) holds

$$E_{\mathrm{mr}} = -\frac{2}{w}\,\left(V_{\mathrm{bi}} - V\right) \propto \sqrt{V_{\mathrm{bi}} - V} \tag{21.147a}$$

$$E_{\mathrm{mr}} \cong -\frac{\sqrt{2}}{w}\,\left(V_{\mathrm{bi}} - V\right) \propto \sqrt{V_{\mathrm{bi}} - V}. \tag{21.147b}$$

We note that for zero bias ($V = 0$), (21.104) is recovered from (21.147a). Using the above approximation ζ is given by

$$\zeta(x) = \frac{1}{2}\,\frac{1}{1 + \exp\left(\beta V/2\right)\,\cosh(\beta E_{\mathrm{mr}} x)}. \tag{21.148}$$

Since ζ decreases sufficiently fast within the depletion layer, the integration over the depletion layer can be extended to $\pm\infty$ and we obtain

$$\chi = \frac{2}{\beta E_{\mathrm{mr}}}\,\frac{1}{\sqrt{\exp(\beta V) - 1}}\,\arctan\left[\sqrt{\frac{\exp(\beta V/2) - 1}{\exp(\beta V/2) + 1}}\right]. \tag{21.149}$$

We note that for $V = 0$, the integral takes the value $\chi = (\beta E_{\rm mr})^{-1}$. The recombination current is now given by [1481]

$$j_{\rm r,b-i} = \frac{2\,\sigma\,v_{\rm th}\,N_{\rm t}\,n_{\rm i}\,kT}{E_{\rm mr}}\ \arctan\left[\sqrt{\frac{\exp(\beta V/2)-1}{\exp(\beta V/2)+1}}\right]\sqrt{\exp(\beta V)-1}. \quad (21.150)$$

For large voltage the arctan term becomes $\pi/4$. For $eV/kT \gg 1$ the nonradiative recombination current can be written as

$$j_{\rm r,b-i} = j_{\rm s}^{\rm r,b-i}\,\exp\left(\frac{\beta V}{n}\right), \quad (21.151)$$

with $j_{\rm s}^{\rm r,b-i} = e\sigma v_{\rm th} N_{\rm t} n_{\rm i} kT\pi/(2E_{\rm mr})$. The voltage-dependent ideality factor n (semi-logarithmic slope $n = \beta j_{\rm r}(V)/j_{\rm r}'(V)$) is close but not identical to 2 and is shown in Fig. 21.47 for various values of $V_{\rm bi}$. The built-in voltage influences the logarithmic slope via the factor $1/E_{\rm mr}$ in (21.150).

In the case of radiative band–band (b–b) recombination, the recombination rate is given by (10.14). Together with (21.121) and integrated over the depletion layer, the recombination current in the depletion layer is given by (cf. (21.132))

$$j_{\rm r,b-b} = e\,B\,w\,n_{\rm i}^2\,\left[\exp(\beta V)-1\right], \quad (21.152)$$

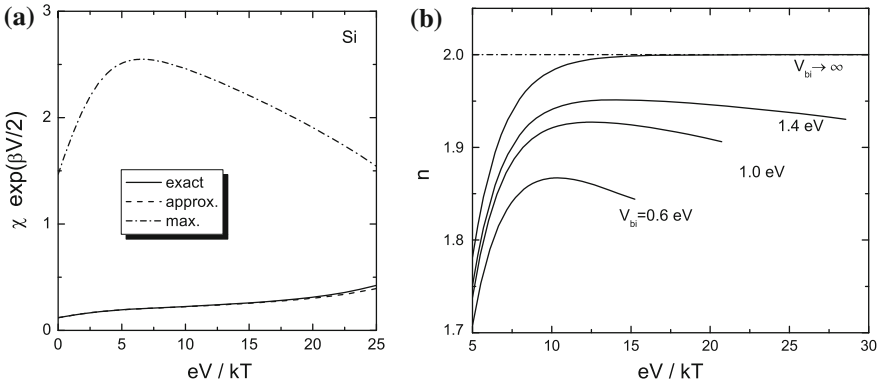


Fig. 21.47 (**a**) Integral χ (21.143) multiplied by $\exp(\beta V/2)$ in order to extract the differences on a linear scale. *Solid line* Exact numerical calculation, *dash-dotted line* standard approximation with constant maximum rate, *dashed line* this work (approximation with constant field). As material parameters we have used room temperature and $n_{\rm i} = 10^{10}\,{\rm cm}^{-3}$ (Si), $n_{\rm n_0} = 10^{18}\,{\rm cm}^{-3}$ and $p_{\rm p_0} = 10^{17}\,{\rm cm}^{-3}$. (**b**) Logarithmic slope of band–impurity recombination current in the forward bias regime for various values of the built-in voltage $V_{\rm bi} = 0.6, 1.0$, and 1.4 eV and in the limit $V_{\rm bi} \to \infty$. Adapted from [1481]

and exhibits an ideality factor of $n = 1$. Comparing (21.132) and (21.152), the dominating radiative-recombination current is determined by the ratio of w and $L_n + L_p$. Since in the forward direction, the depletion-layer width tends towards zero (for flat-band conditions), the radiative-recombination current is dominated by the recombination in the neutral region(s).

For high injection current (under forward bias), the injected minority-carrier density can become comparable with the majority-carrier density. In this case, diffusion *and* drift need to be considered. At large current density, the voltage drop across the junction is small compared to the ohmic voltage drop across the current path. In the simulation (Fig. 21.48), the high-injection effects start on the n-doped side because it has been modeled with the lower doping ($N_D < N_A$).

The series resistance R_s (typically a few Ohms) also effects the characteristic at low injection. The voltage drop across the junction is reduced by $R_s I$. Thus, the I–V characteristic taking into account the effect of the series resistance is

$$I = I_s \left(\exp \left[\frac{e\,(V - R_s I)}{n\,kT} \right] - 1 \right). \tag{21.153}$$

This equation is implicit with regard to I and can only be solved numerically. At high current, the resistance of the junction becomes very small (Fig. 21.49a); then the I–V characteristic deviates from the exponential behavior (Fig. 21.50a); finally it is dominated by the series resistance and becomes linear (Fig. 21.50b). Sometimes the

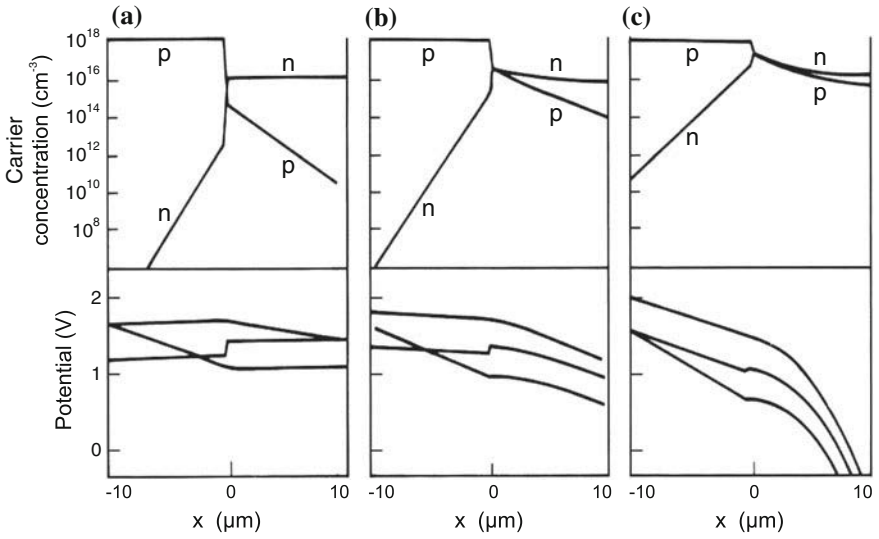


Fig. 21.48 Theoretical modeling of charge-carrier concentration, intrinsic Fermi level (potential) ψ and quasi-Fermi levels (with arbitrary offset) for a Si p^+n diode for various current densities: (**a**) $10\,A\,cm^{-2}$, (**b**) $10^3\,A\,cm^{-2}$ and (**c**) $10^4\,A\,cm^{-2}$. $N_A = 10^{18}\,cm^{-3}$, $N_D = 10^{16}\,cm^{-3}$, $\tau_n = 0.3\,ns$, $\tau_p = 0.84\,ns$. Adapted from [1482]

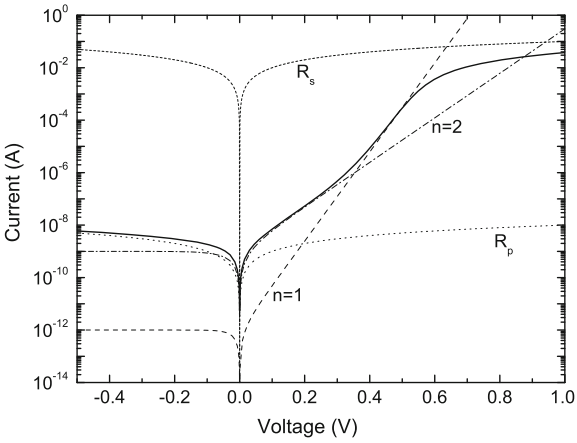


Fig. 21.49 Theoretical I–V characteristic of a diode at room temperature with saturation currents for the $n = 1$ and $n = 2$ processes of $I_s^{n=1} = 10^{-12}$ A and $I_s^{n=2} = 10^{-9}$ A and resistances $R_s = 10\,\Omega$, $R_p = 100\,\mathrm{M}\Omega$. *Dashed line* Ideal diode with $n = 1$ characteristic only, *dash-dotted line* only $n = 2$ process, *dotted line* only parallel ohmic resistance, *short dashed line* only series resistance, *solid line* all effects combined as in (21.155)

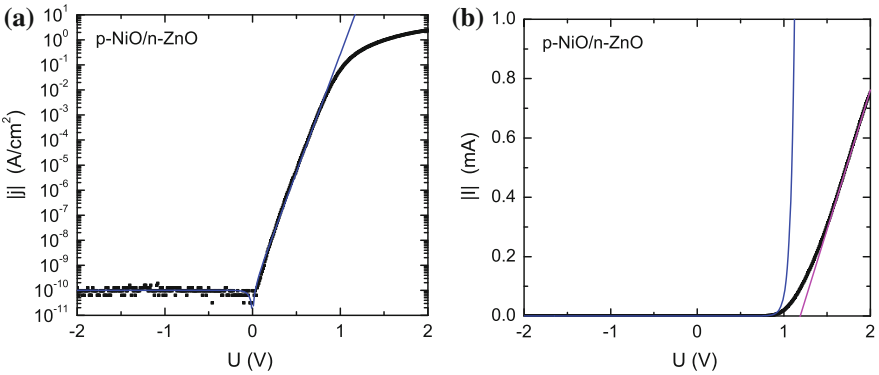


Fig. 21.50 (**a**) jV-characteristic of NiO/ZnO bipolar diode (cmp. Sect. 21.4.6) in half-logarithmic plot. The *blue line* is a fit with ideality factor $n = 1.8$. (**b**) IV-characteristic in linear plot. The *purple line* in (**b**) is a linear fit (with $R_s = 1.06\,\mathrm{k}\Omega$) for the range where the series resistance dominates

extrapolated voltage from the linear range (1.19 V in Fig. 21.50b) is called 'threshold' voltage, but this is a wrong term for this behavior.

The diode can also exhibit a parallel (shunt) resistance R_p, e.g. due to surface conduction between the contacts. Including the shunt resistance, the diode characteristic is

$$I = I_s \left(\exp\left[\frac{e\,(V - R_s\,I)}{n\,kT} \right] - 1 \right) + \frac{V - R_s\,I}{R_p}. \tag{21.154}$$

The shunt resistance can be evaluated best from the differential conductance in the reverse-voltage regime [1452]. Due to a high surface-state density, the passivation of GaAs diodes can be difficult. Si can be very well passivated with low leakage current and high reliability.

Often, a clear distinction between the $n = 1$ and $n = 2$ regimes cannot be made. In this case, an intermediate ideality factor $1 \leq n \leq 2$ is fitted to the I–V characteristic as in (21.154). If the current can be separated into a $n = 1$ (diffusion) and a $n = 2$ (recombination–generation) process, the characteristic is given by (see Fig. 21.49)

$$I = I_{\mathrm{s}}^{n=1}\left(\exp\left[\frac{e\,(V - R_{\mathrm{s}}\,I)}{kT}\right] - 1\right) + I_{\mathrm{s}}^{n=2}\left(\exp\left[\frac{e\,(V - R_{\mathrm{s}}\,I)}{2\,kT}\right] - 1\right) + \frac{V - R_{\mathrm{s}}\,I}{R_{\mathrm{p}}}. \tag{21.155}$$

In summary, the pn-diode has the equivalent circuit given in Fig. 21.51; the photocurrent source I_{ph} is discussed below in Sect. 22.3.

21.4.5 Breakdown

If a large voltage is applied in the reverse direction, the pn-junction breaks down. At breakdown, a small voltage increase leads to a dramatic increase of the current. There are three mechanisms that lead to breakdown: thermal instability, tunneling, and avalanche multiplication [1483–1485]. Defects cause localized pre-breakdown sites [1485, 1486].

Thermal Instability

The reverse current at large applied voltage leads to a power dissipation and heating of the junction. This temperature increase leads to a further increase of the saturation

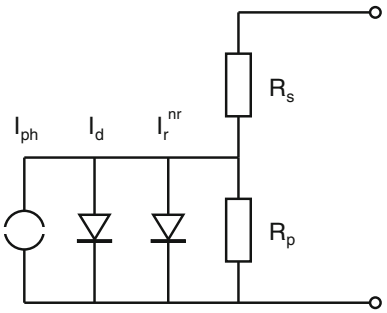


Fig. 21.51 Equivalent circuit of a pn-diode. Serial (R_{s}) and parallel (R_{p}) resistance and diode currents: I_{d} (due to diffusion, $n \approx 1$), $I_{\mathrm{r}}^{\mathrm{nr}}$ (nonradiative recombination due to band–impurity recombination, $n \approx 2$) and an ideal current source due to photogeneration (as discussed in Sect. 22.3)

current (21.129). If the heat sink, e.g. the mounting of the chip, is not able to transport the heat away from the device, the current increases indefinitely. If not limited by a resistor, such a current can destroy the device. The thermal instability is particularly important for diodes with high saturation current, e.g. Ge at room temperature.

Tunneling

At large reverse bias, charge carriers can tunnel between conduction and valence band through the junction. A more detailed discussion will be given below in Sect. 21.5.9 about the tunneling diode. Since for the tunneling effect a thin barrier is necessary, breakdown due to tunneling is important for diodes where both sides are highly doped. For Si and Ge diodes, tunneling dominates the breakdown if the breakdown voltage V_{br} is $V_{\mathrm{br}} < 4E_{\mathrm{g}}/e$. For $V_{\mathrm{br}} > 6E_{\mathrm{g}}/e$ avalanche multiplication dominates. The intermediate regime is a mixed case.

With increasing temperature, the tunneling current can be achieved already with a smaller field (since the band gap decreases with increasing temperature), thus the breakdown voltage *decreases* (negative temperature coefficient).

Avalanche Multiplication

Avalanche multiplication due to impact ionization is the most important mechanism for the breakdown of pn-diodes. It limits the maximum reverse voltage for most diodes and also the collector voltage in a bipolar transistor or the drain voltage in a field-effect transistor. Avalanche multiplication can be used for the generation of microwave radiation or for photon counting (cf. Sect. 22.3.6).

Impact ionization was discussed in Sect. 8.4.4. The most important parameters are the electron and hole ionization coefficients α_{n} and α_{p}. For discussion of the diode breakdown, we assume that at $x = 0$ a hole current I_{p_0} enters the depletion layer. This current is amplified by the field in the depletion layer and impact ionization. At the end of the depletion layer ($x = w$), it is $M_{\mathrm{p}} I_{\mathrm{p}_0}$, i.e. $M_{\mathrm{p}} = I_{\mathrm{p}}(w)/I_{\mathrm{p}}(0)$. Similarly, an electron current is increased on its way from w to $x = 0$. The incremental change of the hole current due to electron–hole pairs generated along a line element $\mathrm{d}x$ is

$$\mathrm{d}I_{\mathrm{p}} = (I_{\mathrm{p}}\,\alpha_{\mathrm{p}} + I_{\mathrm{n}}\,\alpha_{\mathrm{n}})\,\mathrm{d}x. \tag{21.156}$$

The total current in the depletion layer is $I = I_{\mathrm{p}} + I_{\mathrm{n}}$ and is constant in stationary equilibrium. Therefore,

$$\frac{\mathrm{d}I_{\mathrm{p}}}{\mathrm{d}x} - \left(\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}}\right) I_{\mathrm{p}} = \alpha_{\mathrm{n}}\, I. \tag{21.157}$$

The solution is

$$I_{\mathrm{p}}(x) = I\,\frac{\frac{1}{M_{\mathrm{p}}} + \int_0^x \alpha_{\mathrm{n}}\,\exp\left[-\int_0^x (\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}})\,\mathrm{d}x'\right]\mathrm{d}x}{\exp\left[-\int_0^x (\alpha_{\mathrm{p}} - \alpha_{\mathrm{n}})\,\mathrm{d}x'\right]}. \tag{21.158}$$

For $x = w$ we find for the multiplication factor

$$1 - \frac{1}{M_\mathrm{p}} = \int_0^w \alpha_\mathrm{n} \exp\left[-\int_0^x (\alpha_\mathrm{p} - \alpha_\mathrm{n})\,\mathrm{d}x'\right]\mathrm{d}x. \tag{21.159}$$

Avalanche breakdown is reached for $M_\mathrm{p} \to \infty$, i.e. when

$$\int_0^w \alpha_\mathrm{n} \exp\left[-\int_0^x (\alpha_\mathrm{p} - \alpha_\mathrm{n})\,\mathrm{d}x'\right]\mathrm{d}x = 1. \tag{21.160}$$

A corresponding and equivalent equation is obtained when the consideration is started with the electron current. If $\alpha_\mathrm{p} = \alpha_\mathrm{n} = \alpha$, (21.160) simplifies to

$$\int_0^w \alpha\,\mathrm{d}x = 1. \tag{21.161}$$

This means that per transit of one carrier through the depletion layer, on average another carrier is created such that the process just starts to diverge. The breakdown voltage for various semiconductor materials is shown in Fig. 21.52a as a function of the doping level. The depletion-layer width w at breakdown and the maximum electric field E_m are depicted in Fig. 21.52b.

The average impact ionization coefficient α has been given as [1488]

$$\alpha = AE^7. \tag{21.162}$$

for silicon with $A = 1.8 \times 10^{-35}\,(\mathrm{cm/V})^7\,\mathrm{cm}^{-1}$. For the breakdown condition (21.161), using (21.9) and (21.162) the depletion layer width at breakdown w_B (w_B in cm, N_D in cm^3) is found to be

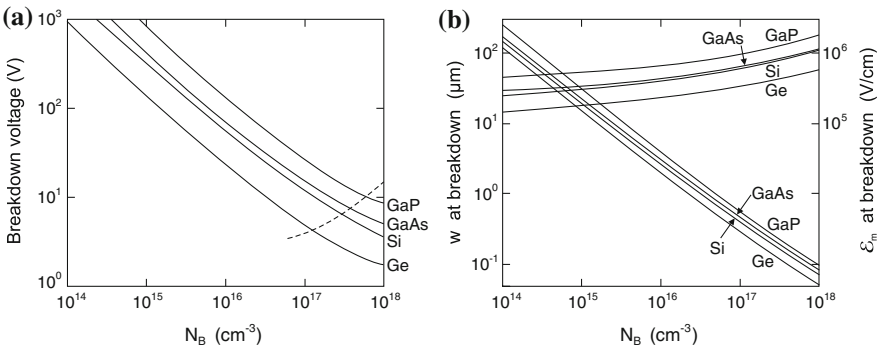


Fig. 21.52 (**a**) Avalanche breakdown voltage for one-sided abrupt junctions in Ge, Si, (100)-GaAs and GaP at $T = 300\,\mathrm{K}$. The *dashed line* indicates the limit of avalanche breakdown at high doping due to tunneling breakdown. (**b**) Depletion-layer width w at breakdown and maximum electric field E_m for the same junctions. Adapted from [1487]

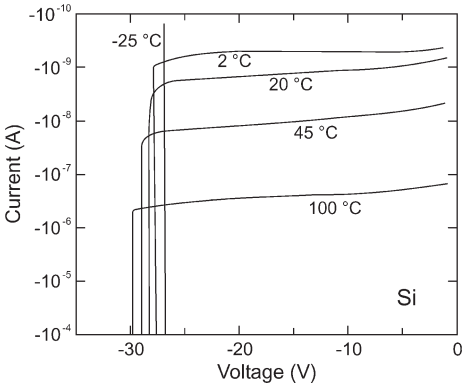


Fig. 21.53 Temperature dependence of a n^+p Si-diode with $N_B = 2.5 \times 10^{16}\,cm^{-3}$ and a guard-ring structure (see Fig. 21.54d). The temperature coefficient $\partial V_{br}/\partial T$ is 0.024 V/K. Adapted from [1491]

$$w_B = 2.67 \times 10^{10}\, N_D^{-7/8}. \tag{21.163}$$

From this the breakdown voltage can be calculated with (21.15) (V_B in V, N_D in cm^3) [1489]

$$V_B = 6.40 \times 10^{13}\, N_D^{-3/4}. \tag{21.164}$$

In GaAs, the impact-ionization coefficients and therefore the breakdown voltage are direction dependent. At a doping of $N_B = 10^{16}\,cm^{-3}$, the breakdown voltage is the same for (001) and (111) orientation; for smaller doping the breakdown voltage of (001)-oriented GaAs is smaller, for larger doping that of GaAs (111) [1490].

At higher temperatures, the charge carriers release their excess energy faster to the lattice.[22] Thus, less energy is available for impact ionization and the required electric field is higher. Therefore, the breakdown voltage *increases* with the temperature (Fig. 21.53). This behavior is opposite to tunneling diodes and the two processes can be distinguished in this way.

In planar structures (Fig. 21.54a), high electric fields as present in high power devices will occur at the parts with large curvature. At these sites breakdown will occur first and at much lower voltages than expected for a perfectly planar (infinite) structure [1492, 1493]. For devices that require high breakdown voltage, design changes have to be made. These include deep junctions (Fig. 21.54b) with a smaller curvature, a field-ring structure (Fig. 21.54c) in which an additional depletion layer is used to smooth the field lines and the often used guard ring (Fig. 21.54d) for which a circular region of low doping (and thus high breakdown voltage) is incorporated.

The decrease of breakdown voltage in a junction with cylindrical and spherical geometry has been numerically calculated as a function of curvature [1494]

[22] The scattering rate becomes higher with increasing temperature and, e.g., the mobility decreases, see Sect. 8.3.9, and the drift saturation velocity decreases, see Sect. 8.4.1.

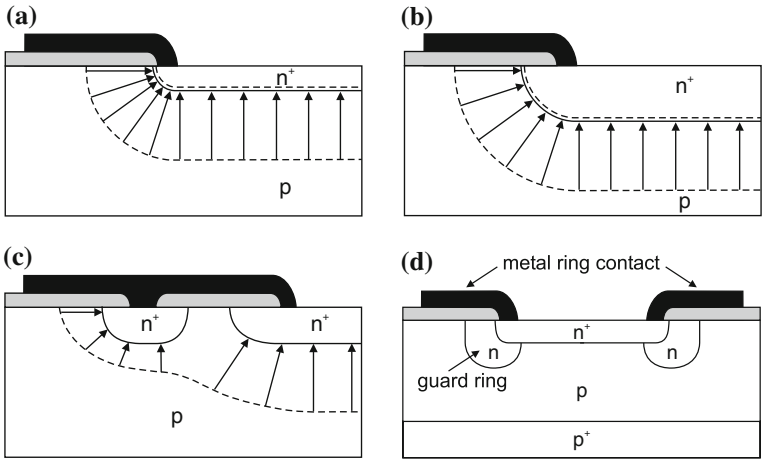


Fig. 21.54 (**a**) Large electric fields at large curvatures in a shallow junction. Avoidance of regions with large electric fields by (**b**) deep junction and (**c**) field-ring structure. (**d**) shows a guard-ring structure with circular, low-doped n region. *Grey area* denotes insulating material, arrows indicate field lines and the *dashed lines* indicate the extension of the depletion layers

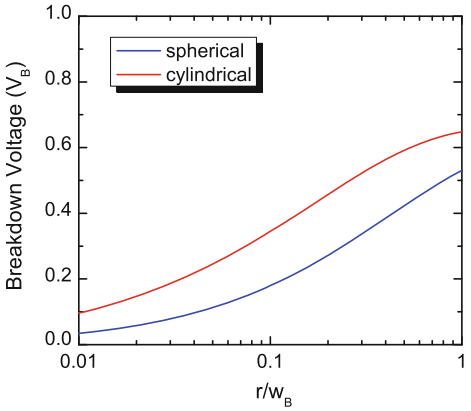


Fig. 21.55 Breakdown voltage (in units of the breakdown voltage V_B of a plane junction) for cylindrical and spherical junctions as a function of the curvature radius (in units of the depletion layer width at breakdown W_B for a plane junction). Data from [1489]

(Fig. 21.55). Analytical formulas in terms of the ratio of radius of curvature and the depletion layer width at breakdown for a plane junction r/w_B have been given later [1489].

Defects

In material with extended defects such as polycrystalline silicon breakdown can occur locally at lower voltage than in the corresponding defect-free bulk material ('pre-breakdown') [1486]. This effect typically occurs at certain grain boundaries and probably involves mid-gap states. It is also accompanied with electroluminescence by which it can be detected with high spatial resolution (Fig. 21.56).

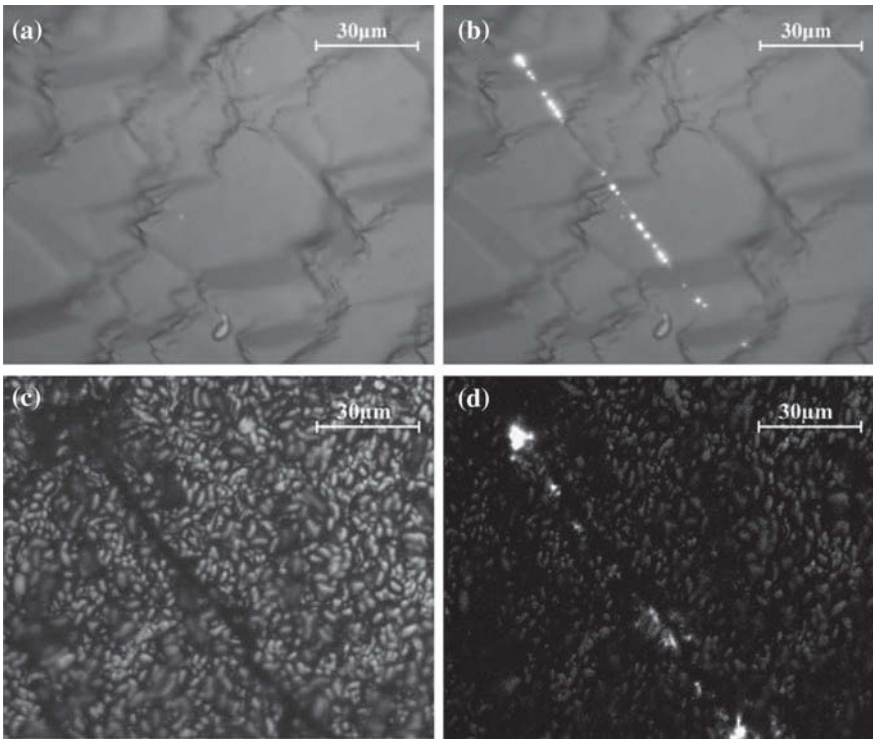


Fig. 21.56 Microscopic (**a**) forward bias electroluminescence and (**b**) microscopic reverse bias electroluminescence (µ-ReBEL) image of an alkaline texturized solar cell (cmp. Fig. 22.58) in a region without specific surface features at $U = -17\,\mathrm{V}$. The bulk defects at these positions are revealed topographically in the neighboring acidic texturized solar cell in the microscopic (**c**) image. Their ReBEL pattern appears similarly in the corresponding EL image (**d**). Adapted from [1486]

21.4.6 Heterostructure Diodes

In a heterostructure diode, the n- and p-regions are made of different semiconductors. Such structures are discussed in detail in [1495].

Type-I Heterostructure Diodes

In Fig. 21.57, the band diagram is shown for a type-I heterostructure with the n (p) region having the larger (smaller) band gap. Additionally to the built-in voltage, the barrier in the valence band is increased. Such a diode finds application in particular as an injection (emitter–base) diode in heterostructure bipolar transistors (Sect. 24.2.7). In such diode the (mostly undesired) hole current from the p- to the n-side is reduced. The peak in the conduction band poses potentially a greater barrier for electron transport across the interface than the diffusion potential itself. The spike can be reduced by grading the materials across the heterojunction and creating a smooth

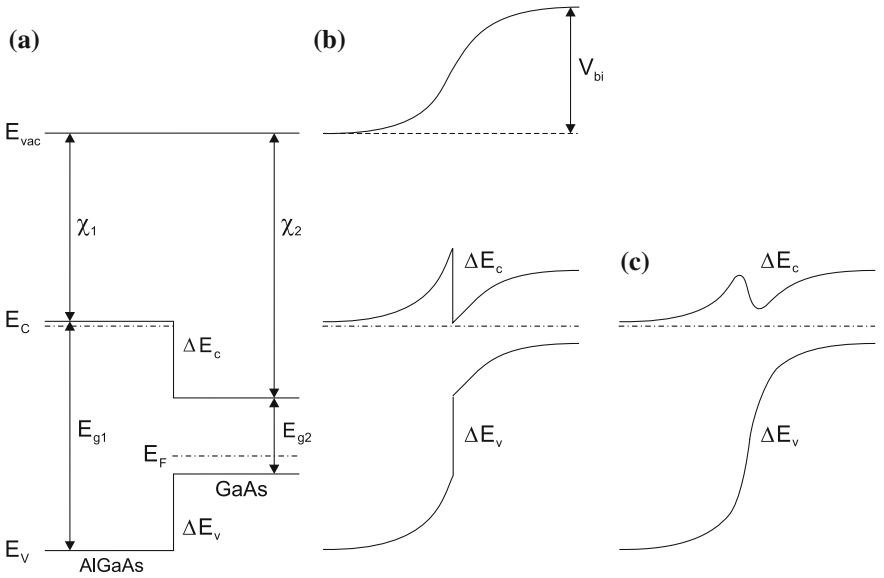


Fig. 21.57 Schematic band diagram of a n-AlGaAs/p-GaAs diode (**a**) without contact of the n- and p-materials, (**b**) in thermodynamic equilibrium, and (**c**) with graded Al composition at the heterointerface

transition of E_g between the materials. The effect of grading on the properties of the heterojunction is discussed in detail in [1496].

Type-II Heterostructure Diodes

Many semiconductors can only be doped n-type or p-type (cmp. Sect. 7.4.2). Bipolar heterostructures can be formed and, depending on the position of the bands relativ to the vacuum level, can form a type-II heterostructure. Examples are p-NiO/n-ZnO or p-CuI/n-ZnO; the band diagram of the latter is depicted in Fig. 21.58a for a typical situation (p^+-n diode) [494, 1497]. Rectification for such device can be very high ($>10^{10}$, Fig. 21.58b). Minority carrier injection is excluded due to the high barriers. The current across the interface is a recombination current, presumably via interface defects, which exhibits an ideality factor around 2 [1498].

21.4.7 Organic Semiconductor Diodes

A bipolar diode from organic semiconductors consists of the p-conductive hole transport layer (HTL) and the n-conductive electron transport layer (ETL). The low conductivity of organic semiconductors causes the applied voltage to drop over the entire structure [1499–1501] (Fig. 21.59c) while in a typical silicon diode for sufficient forward bias (and moderate injection, cmp. Fig. 21.48) flat-band conditions

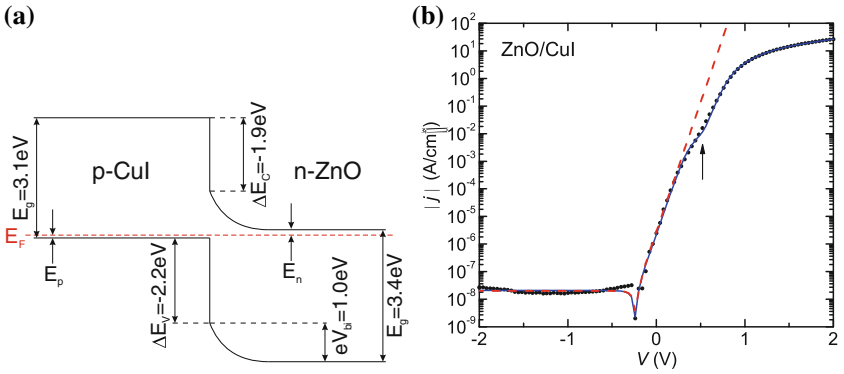


Fig. 21.58 (**a**) Schematic band diagram of a p-CuI/n-ZnO diode. Adapted from [1497]. (**b**) jV-characteristics (room temperature) of type-II epitaxial CuI/ZnO heterojunction bipolar diode (*symbols*) and fit (*red dashed line*) with $j_s = 2 \times 10^{-8}$ A/cm^2 and ideality factor $n = 1.8$. The other fit (*blue line*) takes into account additionally a barrier inhomogeneity and the series resistance. Courtesy of Ch. Yang

are present (Fig. 21.59b). The first organic homodiode was reported only in 2005 for zinc-phthalocyanine (ZnPc) doped with $[Ru(terpy)_2]^0$ (n-type) and F_4-TCNQ (p-type, cmp. Fig. 17.7b) [1502]. Deviation from ideal Shockley behavior is discussed in this report.

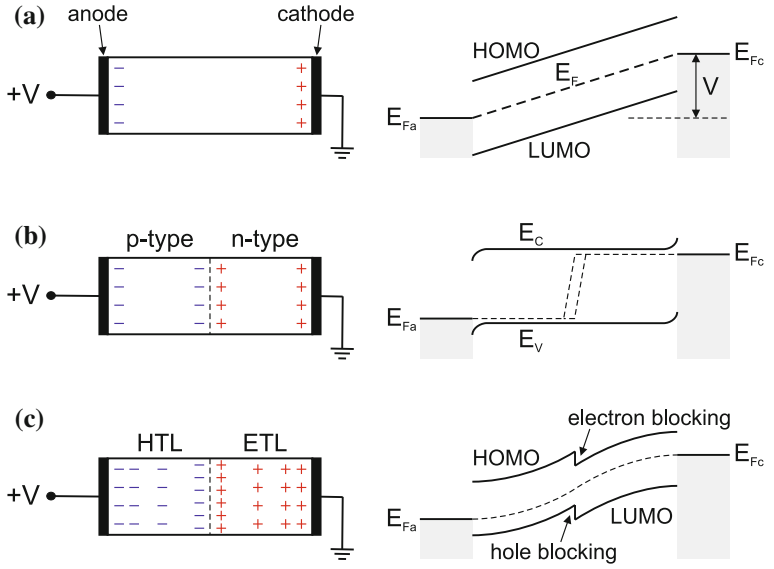


Fig. 21.59 Schematic sample geometry and charge distribution (*left*) and energy diagram (*right*) for (**a**) an ideal insulator, (**b**) a typical inorganic semiconductor pn-diode and (**c**) a double layer organic diode under forward bias V. Adapted from [1501]

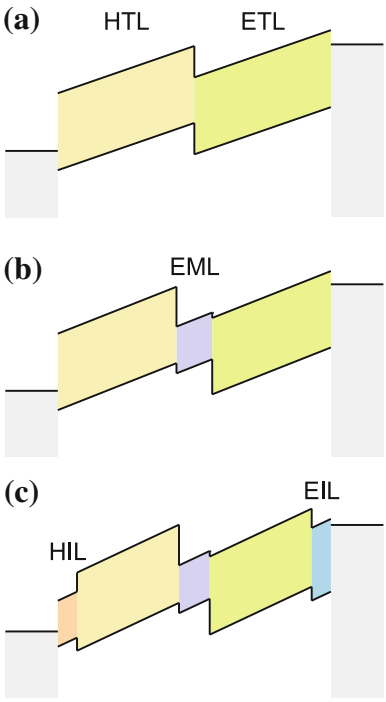


Fig. 21.60 Schematic energy diagram for (**a**) dual layer organic diode (HTL, ETL), (**b**) with additional emission layer (EML) and (**c**) with further hole and electron injection layers (HIL, EIL)

The hole injecting contact is often made from ITO, the electron injection contact from low work function metals such as Al, Mg or Ca. Specially designed layers for efficient charge injection (HIL, hole injection layer [1503] and EIL, electron injection layer [1504]) can be introduced between the contact metal and the transport layers (Fig. 21.60c). The particular level lineup of the HTL and HTL as shown in Fig. 21.59c (also Fig. 21.60a) leads to a barrier for electrons and holes. The hole and/or electron blocking is beneficial for recombination in the vicinity of the interface and prevents exciton diffusion to the contacts. In organic light emitting diodes (Sect. 23.3.7) an additional layer (EML, emission layer) designed for efficient radiative recombination (Sect. 17.6) is introduced between HTL and ETL (Fig. 21.60b).

21.5 Applications and Special Diode Devices

In the following, various electronic applications of diodes are discussed. The most important special diode types are introduced. Optoelectronic applications (involving absorption and emission of photons) are treated below (Chap. 22). The applications of Schottky diodes are discussed in [1505].

21.5.1 Rectification

In a rectifier, the diode has to supply a high resistivity for one polarity of the bias and a low one to the other polarity. In Fig. 21.61a, a single-path rectification method is shown. Only the positive half-wave can pass the load resistor R_L (Fig. 21.61b). In Fig. 21.61c, the characteristic of a Si diode is shown. Of course, the voltage drop across the diode can only range in the 1-V regime. In order to make the setup work, the load resistor has to be considered. The total current is given by $I = I_\mathrm{s}\left[\exp\left(eU_\mathrm{d}/nkT\right) - 1\right]$. The total voltage U is split between the voltage drop across the diode U_d and that over the load resistance $U_\mathrm{L} = R_\mathrm{L}I$. The current is therefore given by

$$I = \frac{U - U_\mathrm{d}}{R_\mathrm{L}}. \tag{21.165}$$

For sizeable currents the voltage drop across the diode U_d is between 0.7 and 1 V. The characteristic is linear between about 1 and 220 V (Fig. 21.61d). Typically, the voltage U_L is low-pass filtered with a capacitor parallel to the load resistor. The effective voltage is the peak voltage divided by 2.

The drawback of the single diode rectifier is that only the positive half-wave contributes to a dc signal. The setup in Fig. 21.61e (bridge rectifier) allows both half-waves to contribute to the dc signal. The effective voltage in this case is the peak voltage divided by $\sqrt{2}$.

The forward resistance in the static (R_f) and dynamic (r_f) case are (for $\beta V_\mathrm{f} > 3$)

$$R_\mathrm{f} = \frac{V_\mathrm{f}}{I_\mathrm{f}} \cong \frac{V_\mathrm{f}}{I_\mathrm{s}} \exp\left(-\frac{e\,V_\mathrm{f}}{n\,kT}\right) \tag{21.166a}$$

$$r_\mathrm{f} = \frac{\partial V_\mathrm{f}}{\partial I_\mathrm{f}} = \frac{n\,kT}{e\,I_\mathrm{s}} \exp\left(\frac{e\,V_\mathrm{f}}{n\,kT}\right) \cong \frac{n\,kT}{e\,I_\mathrm{f}}. \tag{21.166b}$$

For reverse bias we have ($\beta|V_\mathrm{r}| > 3$)

$$R_\mathrm{r} = \frac{V_\mathrm{r}}{I_\mathrm{r}} \cong \frac{V_\mathrm{r}}{I_\mathrm{s}} \tag{21.167a}$$

$$r_\mathrm{r} = \frac{\partial V_\mathrm{r}}{\partial I_\mathrm{r}} = \frac{nkT}{eI_\mathrm{s}} \exp\left(\frac{e\,|V_\mathrm{r}|}{n\,kT}\right). \tag{21.167b}$$

Thus, the dc and ac rectifications ratios are given by

$$\frac{R_\mathrm{r}}{R_\mathrm{f}} = \exp\left(\frac{e\,V_\mathrm{f}}{n\,kT}\right) \tag{21.168a}$$

$$\frac{r_\mathrm{r}}{r_\mathrm{f}} = \frac{I_\mathrm{f}}{I_\mathrm{s}\,\exp\left(\frac{e\,|V_\mathrm{r}|}{n\,kT}\right)}. \tag{21.168b}$$

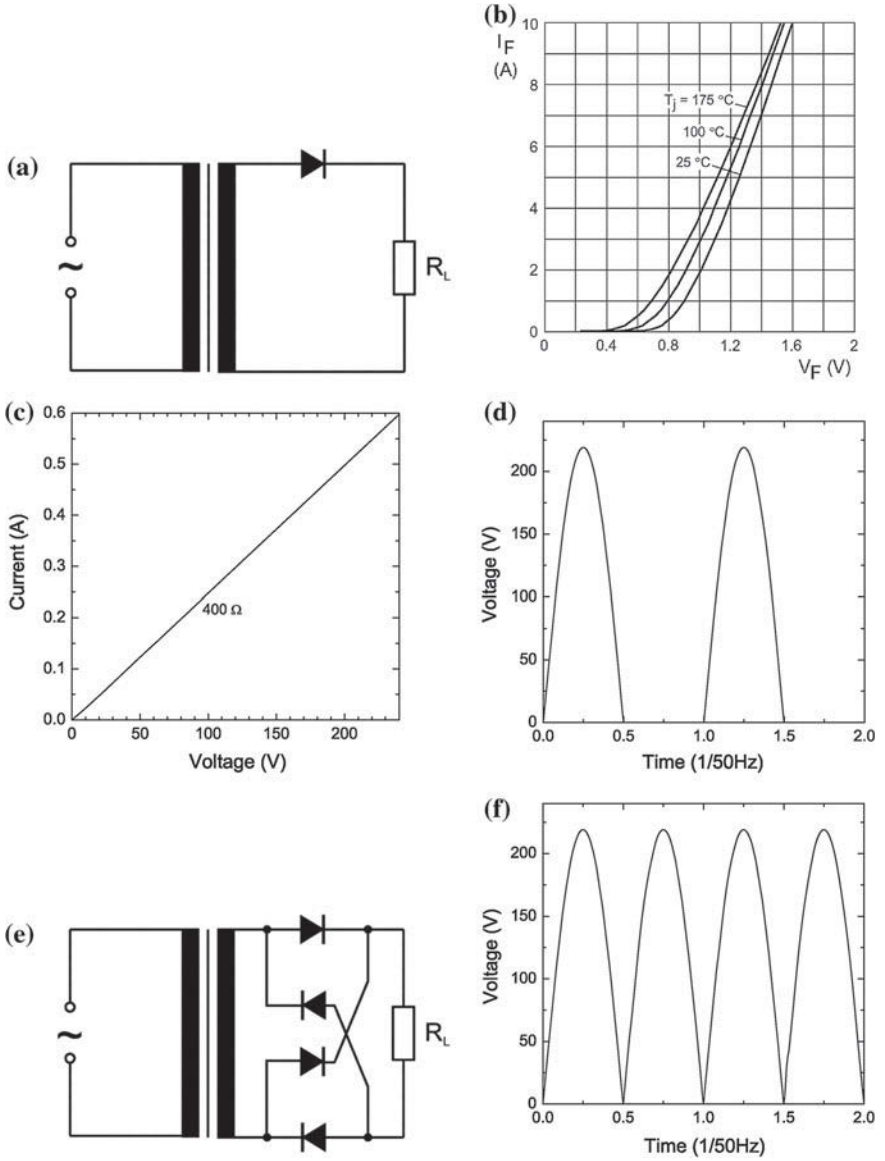


Fig. 21.61 (**a**) Single-path mains rectifier, (**b**) characteristics of Si diode (BYD127, Philips), (**c**) load characteristics of the mains rectifier ($R_L = 400\,\Omega$), (**d**) voltage output of single-path mains rectifier. (**e**) Depicts the schematic circuit of a bridge rectifier that works for both half-waves, (**f**) the resulting voltage output

Rectifiers generally have slow switching speeds. A significant time delay arises from the necessary charge-carrier recombination when the diode switches from low (forward) to high (reverse) impedance. This poses typically no problem for line-frequency (50–60 Hz) applications. For fast applications, however, the minority-carrier lifetime needs to be reduced, e.g. by doping Si with Au (see Sect. 10.9).

21.5.2 Frequency Mixing

The nonlinear characteristic of the diode allows the mixing of frequencies, e.g. for second- (or higher-) harmonic generation, upconversion or demodulating of radio-frequency (RF) signals. A single balanced mixer is shown in Fig. 21.62a, b. The RF signal consists of a RF carrier frequency f_0 modulated with an intermediate frequency (IF) signal $f_{\mathrm{IF}}(t)$. It is mixed with a local oscillator (LO) that has a constant frequency f_{LO} outside the RF modulation bandwidth $f_0 \pm f_{\mathrm{IF}}$. The IF signal can be detected from the setup in Fig. 21.62a if filtered through a low-pass filter to avoid loss of power to the IF amplifier. The temperature dependence of the diode characteristic (via j_{s} and β) on mixing efficiency is typically less than 0.5 dB for a 100-K change in temperature.

Problems of single-diode mixers are the radiation of local-oscillator power from the RF input port,[23] loss of sensitivity by absorption of input power in the local oscillator circuit, loss of input power in the intermediate frequency amplifier, and the generation of spurious output frequencies by harmonic mixing. Some of these problems can be solved by circuit techniques, but these circuits often introduce new problems. Most mixers therefore use multiple-diode techniques to provide a better solution of these problems. In Fig. 21.62c, the circuit diagram of a double balance mixer is shown. Even-order harmonics of both the LO and the signal frequency are rejected. This mixer does not require a low-pass filter to isolate the IF circuit. The three ports are isolated from each other by the symmetry of the circuit. These mixers usually cover a broader frequency band than the others. Ratios as high as 1000:1 are available. Microwave equivalents (working at $f \gg 1$ GHz) of such mixer circuits are available. Bandwidth ratios as high as 40:1 are available at microwave frequencies in MMICs (millimeter-wave integrated circuits).

The common drawback of MMIC diodes is that they are obtained from the Schottky barriers used in field-effect transistors, that have inferior performance compared to discrete diodes. The use of pHEMT technology[24] for millimeter-wave applications provides diodes that differently from regular Schottky diodes, since they consist of a Schottky barrier in series with a heterojunction. In Fig. 21.62d, a MMIC 45 GHz mixer is shown using fast GaAs-based pHEMTs.

[23] That in military applications could make the mixer detectable by the enemy.

[24] Pseudomorphic high electron mobility transistors, cf. Sect. 24.5.8.

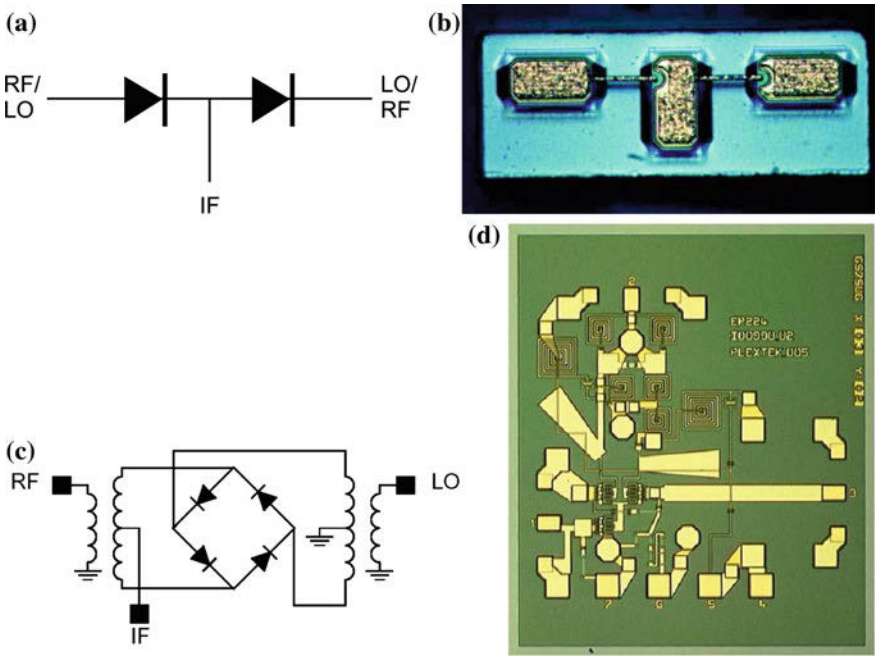


Fig. 21.62 (**a**) Schematic circuit of single balanced mixer with input (RF: radio frequency, LO: local oscillator) and output (IF: intermediate frequency). (**b**) Optical plan-view image ($300 \times 125\,\mu m^2$) of a high-speed single balanced mixer with two GaAs Schottky diodes with opposite poling. The device properties are $R_s = 5\,\Omega$, for $I = 1\,\mu A$ a forward and reverse voltage of 0.7 and 6 V, respectively; the capacitance of each diode is 8 fF. Reprinted with permission from [1506]. (**c**) Schematic circuit of a double balanced mixer. (**d**) Optical image ($1.65\,mm^2$) of 40–45 GHz MMIC (Gilbert cell) mixer on GaAs basis using pHEMTs. Reprinted with permission from [1507]

21.5.3 Voltage Regulator

In a voltage regulator, the large variation of resistance with applied bias is used. This effect occurs in the forward direction and close to the breakdown voltage.

In Fig. 21.63a, a simple circuit is shown. When the input voltage V_{in} is increased, the current increases. The preresistor $R_1 = 5\,k\Omega$ and the load resistor represent a voltage divider with $V_{in} = IR_1 + V_{out}$. The total current I is given by the two currents through the diode and the load resistor $I = I_s\left[\exp(\beta V_{out}/n) - 1\right] + V_{out}/R_L$. Therefore, the output voltage is implicitly given by

$$V_{out}\left(1 + \frac{R_1}{R_L}\right) = V_{in} - R_1 I_s\left[\exp\left(\frac{\beta V_{out}}{n}\right) - 1\right]. \tag{21.169}$$

A large current change is related to a fairly small change of the voltage across the diode, which at the same time is the output voltage. Therefore, a change in the input voltage causes only a small change in the output voltage.

We assume a diode with $n = 1$ and $I_S = 10^{-14}$ A with the characteristic shown in Fig. 21.63a. The numerical example in Fig. 21.63c is calculated for $R_L = 2\,\mathrm{k\Omega}$ and $4\,\mathrm{k\Omega}$, respectively. The output voltage varies by about 0.02 V if the input varies between 5 and 9 V. In Fig. 21.63d, the differential voltage change $\alpha = \frac{V_{in}}{V_{out}} \frac{\partial V_{out}}{\partial V_{in}}$ is shown.

In this way, voltage peaks can be filtered from the input voltage. If two antiparallel diodes are used, this principle works for both polarities. Instead of a diode in the forward direction, the very steep slope of the diode I–V characteristic at the breakdown can be used. Just before breakdown, the diode has a high resistance and the voltage drops at the load resistor. If the input voltage increases a little, the diode becomes conductive and shorts the additional voltage (the maximum allowed breakdown current needs to be observed!). Due to its small saturation current, typically Si

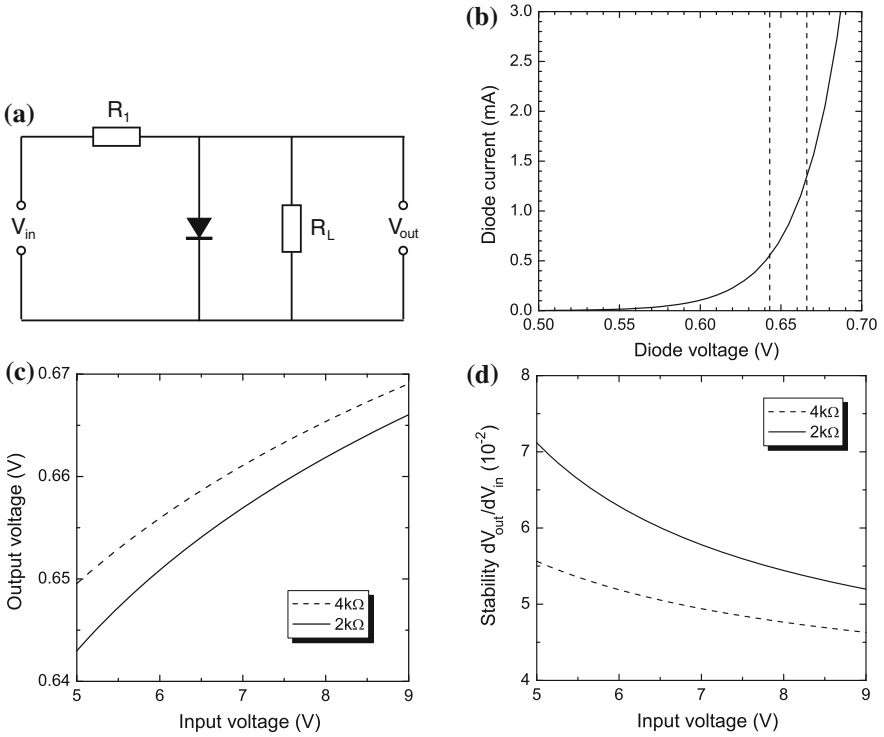


Fig. 21.63 (**a**) Circuit diagram of a voltage regulator, (**b**) diode characteristic ($n = 1$). The *vertical dashed lines* show the operation conditions for $R_L = 2\,\mathrm{k\Omega}$ and $U_E = 5$ and 9 V and thus the principle of voltage stabilization. (**c**) Output versus input voltage and (**d**) stability (differential voltage ratio α, see text) for input voltage between 5 and 9 V

diodes are used. The breakdown voltage can be designed via the diode parameters. Such diodes with defined breakdown voltage are called Z- or Zener diodes (see next section).

If the breakdown is due to tunneling (avalanche multiplication), the breakdown voltage decreases (increases) with temperature. If two diodes with positive and negative temperature coefficient are put in series, a very good temperature stability of the breakdown voltage of 0.002 %/K can be achieved. Such diodes can be used to realize a reference voltage.

21.5.4 Zener Diodes

A Zener diode is designed to have a defined breakdown voltage. Zener diodes are available with a number of different standard breakdown voltages. Their characteristic is shown for reverse bias with the current shown positive. The characteristics of various Zener diodes for different breakdown voltages are shown in Fig. 21.64.

21.5.5 Varactors

A diode exhibits a voltage-dependent capacitance. This effect can be used to tune an oscillator using the diode bias (voltage-controlled oscillator, VCO). The equivalent circuit is shown in Fig. 21.65. The capacitance consists of a parasitic capacitance C_p due to mounting and bonding. This effect also causes a parasitic inductance. The series resistance due to mounting can typically be neglected. The variable junction capacitance C_j and the ohmic resistance R_s are bias dependent.

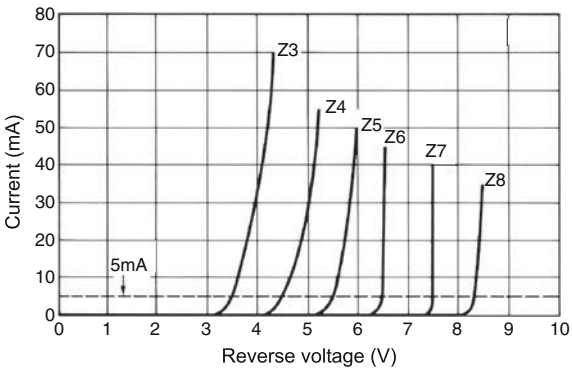


Fig. 21.64 Characteristics of a field of Zener diodes (at room temperature)

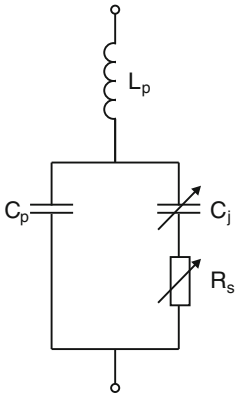


Fig. 21.65 Equivalent circuit of a varactor diode with parasitic capacitance C_p and inductance L_p and variable capacitance C_j and resistance R_s

The $C(V)$ dependence has generally a power law with an exponent γ (which itself may depend on the bias voltage)

$$C = \frac{C_0}{(1 + V/V_{bi})^{\gamma}}, \tag{21.170}$$

where C_0 is the-zero bias capacitance. Since the frequency f of an LC oscillator circuit depends on $C^{-1/2}$ the frequency, f depends on the voltage as $f \propto V^{\gamma/2}$. Therefore, a $\gamma = 2$ dependence is most desirable.

For uniformly doped profiles, the capacitance depends with an inverse square root law on the applied voltage (21.113), i.e. $\gamma = 0.5$. Hyperabrupt junctions are typically made by ion implantation or epitaxy with graded impurity incorporation to create a special nonuniform doping profile (Fig. 21.66a). For a doping profile

$$N_B(z) = \hat{N}_B \left(\frac{z}{z_0}\right)^m \tag{21.171}$$

the capacitance is given as

$$C = \left[\frac{e\,\hat{N}_B\,\epsilon_s^{m+1}}{(m+2)\,z_0^m\,(V_{bi} - V)}\right]^{\frac{1}{m+2}} = \frac{C_0}{(1 + V/V_{bi})^{1/(m+2)}}. \tag{21.172}$$

Ideally, $m = -3/2$ results is a linear frequency versus voltage relation. The C–V characteristic of an implanted, hyperabrupt diode exhibits a part that has an exponent $\gamma = 2$ (Fig. 21.66b). A $\gamma = 2$ $C(V)$ dependence and therefore a linear $f(V)$ curve can be achieved over more than one octave using computer-controlled variable epitaxial-layer doping (Fig. 21.66c).

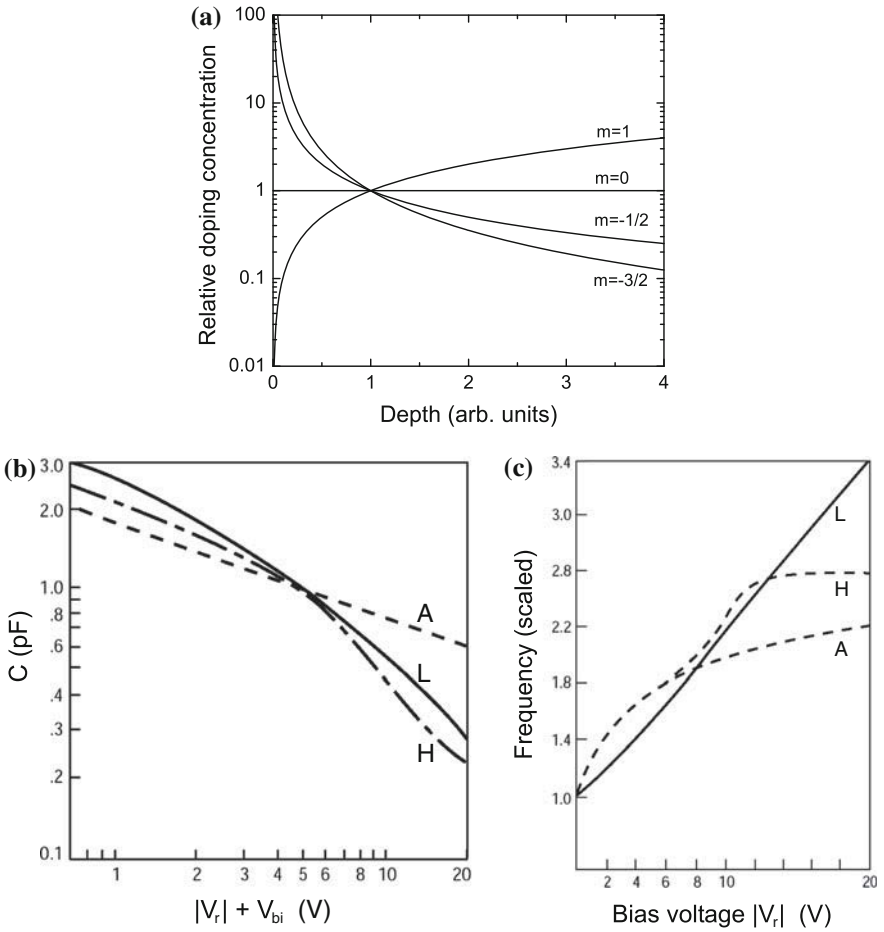


Fig. 21.66 (**a**) Donor-doping profile according to (21.171) in p^+n or Schottky diodes for $m = 0$ (abrupt junction), $m = 1$ (linearly graded junction) and two values with $m < 0$ (hyperabrupt junctions). (**b**) Bias dependence of capacitance for diodes with abrupt junction ('A', $\gamma = 0.5$), hyperabrupt junction ('H', $\gamma > 0.5$) and 'epilinear' junction ('L', $\gamma = 2$). (**c**) Frequency–voltage tuning characteristics (scaled to 1.0 for $V = 0$) for the three diode types. Parts (**b**) and (**c**) adapted from [1508], reprinted with permission

21.5.6 Fast-Recovery Diodes

Fast-recovery diodes are designed for high switching speeds. The switching speed from the forward to the reverse regime is given by the time $t_0 = t_1 + t_2$ with t_1 being the time to reduce the minority carrier density to the equilibrium value (e.g. $p_n \rightarrow p_{n_0}$) and t_2 being the time in which the current decreased exponentially (Fig. 21.67). The time t_1 can be drastically reduced by incorporation of deep levels that act as

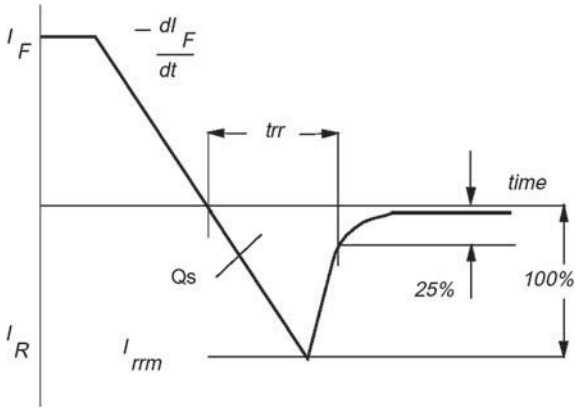


Fig. 21.67 Current versus time trace for a (soft) fast-recovery diode. Reprinted with permission from [1509]

recombination centers. A prominent example is Si:Au. However, this concept is limited since the reverse generation current, e.g. (21.150), depends on the trap density. For direct semiconductors, recombination times are short, e.g. 0.1 ns or less for GaAs. In silicon, they can be extremely long (up to ms) or at least 1–5 ns. Schottky diodes are suitable for high-speed applications since they are majority-carrier devices and minority-charge carrier storage can be neglected.

21.5.7 Step-Recovery Diodes

This type of diode is designed to store charge in the forward direction. If polarity is reversed, the charge will allow conductance for a short while, ideally until a current peak is reached (Fig. 21.68a), and then cutoff the current very rapidly during the so-called snapback time T_s (Fig. 21.68b). The cutoff can be quite rapid, in the ps regime. These properties are used for pulse (comb) generation or as a gate in fast sampling oscilloscopes. In Si, only 0.5–5 μs are reached (fast-recovery diode, see previous section) while GaAs diodes can be used in the several tens of GHz regime.

Using a heterostructure GaAs/AlGaAs diode (cf. Sect. 21.4.6), as shown schematically in Fig. 21.69a, a steepening of a 15-V, 70-ps (10 to 90 %) pulse to a fall time of 12 ps was observed (Fig. 21.69c). The forward current of the diode was 40 mA, supplied via a bias tee.

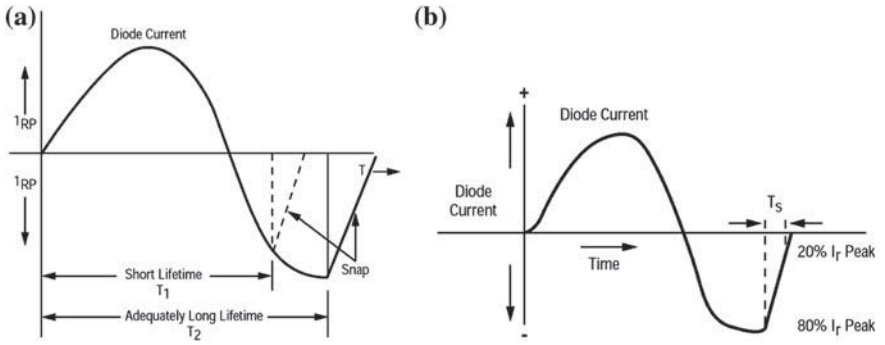


Fig. 21.68 (**a**) Current versus time trace for a step-recovery diode and sinusoidal voltage input. The lifetime must be sufficiently large such that a current peak is reached. (**b**) Definition of the snapback time T_S. Reprinted with permission from [1510]

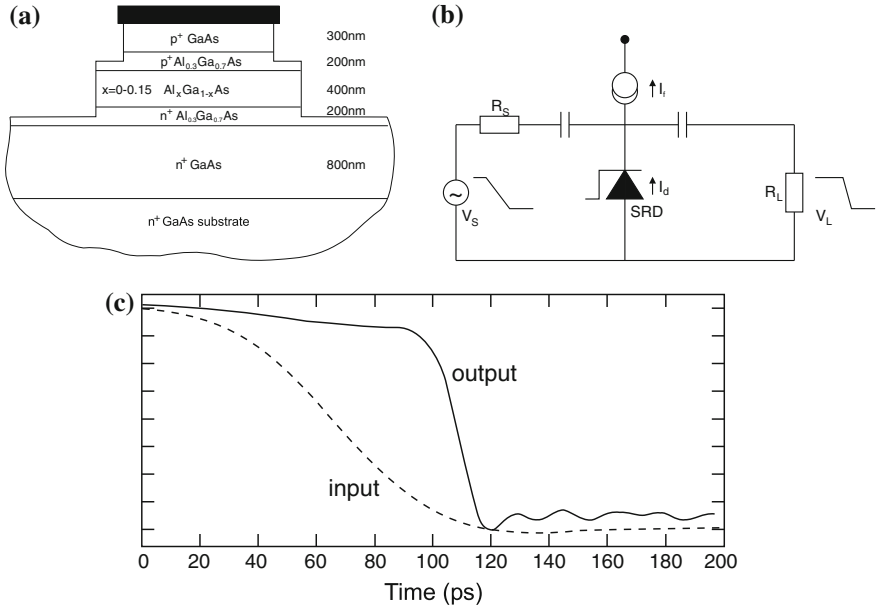


Fig. 21.69 (**a**) Schematic layer sequence for fast GaAs/AlGaAs step-recovery diode. (**b**) Circuit with input and output pulse. (**c**) Input (*dashed line*) and output (*solid line*) waveforms. Vertical division is 2 V. Adapted from [1511]

21.5.8 Pin-Diodes

In a pin-diode, an intrinsic (i), i.e. undoped region (with higher resistivity) is located between the n- and the p-regions (Fig. 21.70a). Often, also a region with low n- or p-doping is used. In this case, the center region is denoted as a ν- or π-region,

respectively. The fabrication of arbitrary doping profiles and an intrinsic region poses little problem for epitaxial diodes.

Via the Poisson equation, the charge in the intrinsic layer is related to the electric field. If no dopants are present, there is a constant (maximum) field in the i-region at zero bias (Fig. 21.70d). If there is low doping, a field gradient exists.

The capacitance for reverse bias is $\epsilon_s A/w$ and is constant starting at fairly small reverse bias (10 V). The series resistance is given by $R_s = R_i + R_c$. The contact resistance R_c dominates the series resistance for large forward bias.

21.5.9 Tunneling Diodes

For the invention of the tunneling diode and the explanation of its mechanism the 1973 Nobel Prize in Physics was awarded to L. Esaki. Eventually, the tunneling diode did not make the commercial breakthrough due to its high basis capacitance. It is used for special microwave applications with low power consumption, for frequency stabilization and possibly in tunneling field-effect transistors (Sect. 24.5.6).

First, the tunneling diode is a pn-diode. While the tunnel effect [1512] has already been discussed for Schottky diodes, it has not yet been considered by us for pn–diodes. We expect the tunnel effect to be important if the depletion-layer is thin, i.e. when the doping of both sides is high.

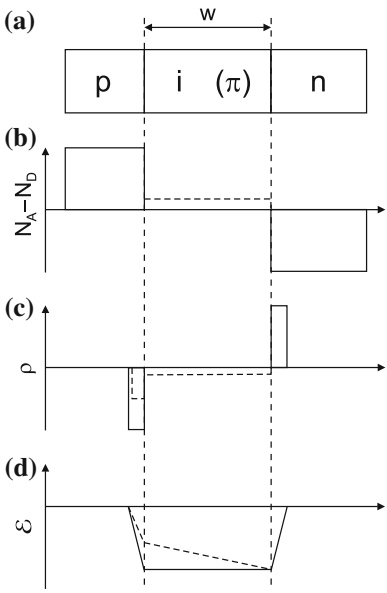


Fig. 21.70 (**a**) Schematic layer sequence of pin-diode ('i': intrinsic, 'π': lowly p-doped), (**b**) net impurity distribution $N_A - N_D$, (**c**) space charge and (**d**) electric field in a pin (*solid lines*) and a p–π–n (*dashed lines*) diode

The doping is so high that the quasi-Fermi levels lie within the respective bands (Fig. 21.71). The degeneracy is typically several kT and the depletion layer width is in the 10 nm range.

In the forward direction, the I–V characteristic of the tunneling diode exhibits a maximum followed by a minimum and subsequently an exponential increase (Fig. 21.72a). As shown in Fig. 21.72b, the total current consists of three currents, the band-to-band tunneling current, the excess current and the thermal (normal thermionic diode) current.

The $V = 0$ situation is again shown in Fig. 21.73b. Upon application of a small forward bias, electrons can tunnel from populated conduction-band states on the n-doped side into empty valance-band states (filled with holes) on the p-doped side (Fig. 21.73c). We note that this tunneling process is usually considered elastic. However, signatures at characteristic phonon and multi-phonon energies are found (at low

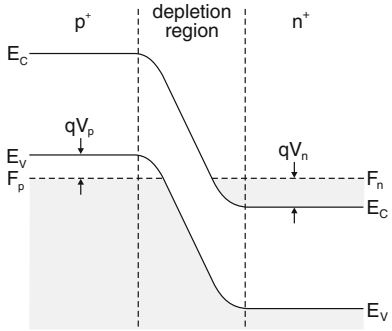


Fig. 21.71 Band diagram of a tunneling diode in thermodynamic equilibrium ($V = 0$). V_n and V_p characterize the degeneracies on the n- and p-side, respectively

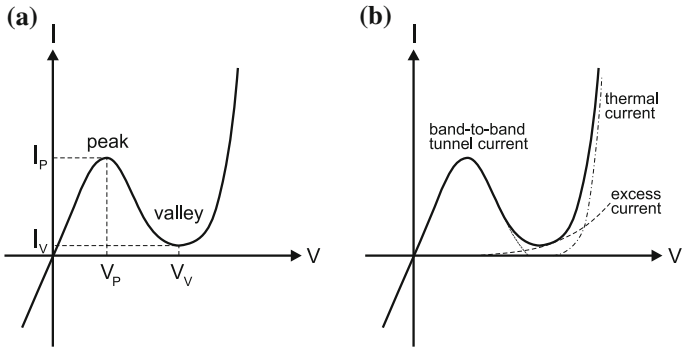


Fig. 21.72 (**a**) Static current–voltage characteristics of a typical tunneling diode. Peak and valley current and voltage are labeled. (**b**) The three components of the current (*short dashed* band-to-band tunnel current, *dashed* excess current, *dash-dotted* thermal current) are shown separately. Adapted from [500], ©1981 Wiley

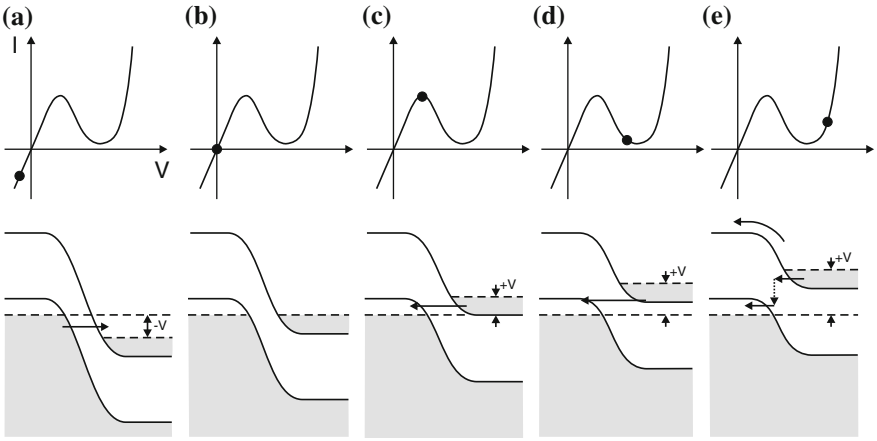


Fig. 21.73 I–V characteristics (*upper row*) and simplified band structure (*lower row*) of a tunneling diode at various bias voltages as indicated with a *dot* in the I–V plot. (**a**) Reverse bias, (**b**) in thermodynamic equilibrium ($V = 0$), (**c**) in the maximum of the tunneling current, (**d**) close to the valley and (**e**) forward bias with dominating thermal current. The tunneling current is indicated with *straight arrows*. In (**e**) the thermionic emission current (*curved arrow*) and the excess current with inelastic tunneling (*dotted arrow*) are shown

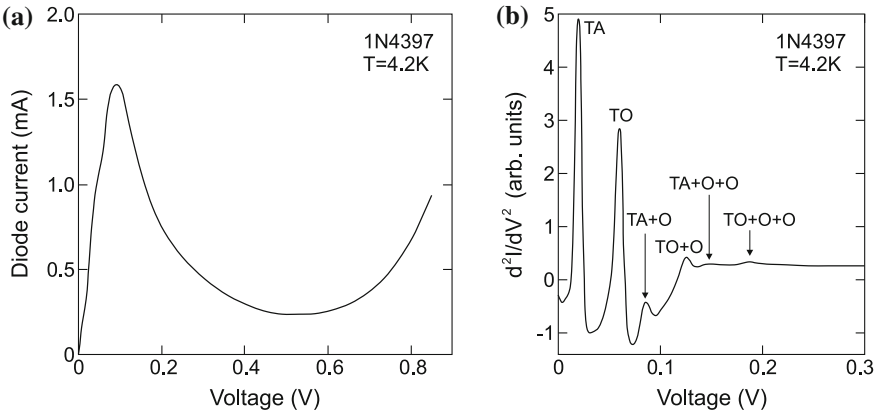


Fig. 21.74 (**a**) Current–voltage characteristic of silicon tunneling diode (model 1N4397 from American Microsemiconductor [1518]) at low temperature ($T = 4.2\,\mathrm{K}$). (**b**) $\mathrm{d}^2I/\mathrm{d}V^2$ for small forward voltages, exhibiting characteristic phonon energies. Adapted from [1516]

temperature) in the forward current [1513, 1514] and are best seen in a $\mathrm{d}^2I/\mathrm{d}V^2$-plot [1515, 1516] (Fig. 21.74). Peaks appear at typical phonon energies of silicon.[25]

[25] We note that for the 1N4396 silicon tunneling diode [1517] the phonon structure of germanium was found in [1516].

A similar situation, now with electrons tunneling for the valence band on the p-side into the conduction band on the n-side, is present for small reverse bias (Fig. 21.73a). Thus the rectifying behavior of the diode is lost. This property makes it suitable for the monolithic ohmic connection of two pn-diodes which is technologically used e.g. in multi-junction solar cells (cmp. Sect. 22.4.6).

For larger forward bias, the bands are separated so far that the electrons coming from the n-doped side do not find final states on the p-doped side. Thus the tunneling current ceases (Fig. 21.73d). The current minimum is at a voltage $V = V_n + V_p > 0$. The thermal current is the normal diode diffusion current (Fig. 21.73e). Therefore, a minimum is present in the I–V characteristic. The excess current is due to inelastic tunneling processes through states in the band gap and causes the minimum to not drop down to almost zero current.

The peak (V_p, I_p) and valley (V_v, I_v) structure of the characteristic leads to a region of negative differential resistance (NDR). I_p/I_v is termed the peak-to-valley ratio (Fig. 21.75). Peak-to-valley ratios of 8 (Ge), 12 (GaSb, GaAs), 4 (Si), 5 (InP) or 2 (InAs) have been reported (all at room temperature).

21.5.10 Backward Diodes

When the doping in a tunneling diode is nearly or not quite degenerate, the peak-to-valley ratio can be very small. Then the tunnel current flows mostly in the reverse direction (low resistance) and the forward direction has a higher resistance (with or without the NDR regime). Such diodes are called backward diodes. Since there is no minority-charge carrier storage, such diodes are useful for high-frequency applications.

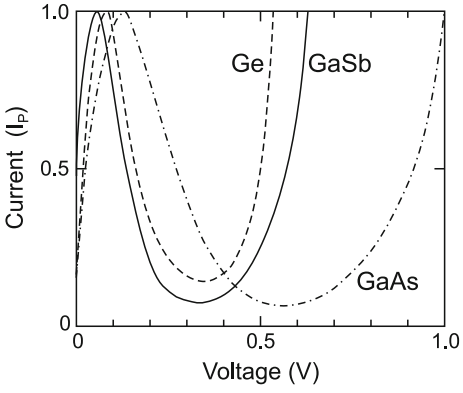


Fig. 21.75 Comparison of tunneling characteristic (room temperature) of diodes based on Ge, GaSb and GaAs. Peak-to-valley ratios are 8 (Ge) and 12 (GaSb, GaAs). Adapted from [500]

21.5.11 Gunn Diodes

The Gunn diode is not really a diode and thus more appropriately called *Gunn element*. It allows to generate microwave radiation in the frequency range of 1–100 GHz [1519] and beyond [1520] using GaAs and in the THz-regime using GaN [1521]. Extracting higher harmonics, frequencies of several 100 GHz can be realized [1522, 1523].

The Gunn element relies on the negative differential resistance (NDR) occurring in semiconductors with two valleys of different mobility such as GaAs or InP (cmp. Sect. 8.4.2). At high fields, the electrons are scattered from the Γ-valley into the upper valley (L-valley for most materials). Accordingly, the Gunn element is also called *transferred electron device* (TED). Details on Gunn elements can be found in [1524].

The Gunn effect was discovered by and named after J.B. Gunn [122]. The spontaneous oscillation of current in a n-type semiconductor occurs when a sufficiently large voltage (pulse) is applied that causes the electric field in the semiconductor to reach the NDR regime (Fig. 21.76). The applied average field of $E = 16\,\mathrm{V}/25\,\mu\mathrm{m} = 6.4\,\mathrm{kV/cm}$ [122] is larger than the threshold field of $E_\mathrm{T} = 3.2\,\mathrm{kV/cm}$ for NDR in GaAs (Table 8.4).

The self-started oscillations are due the inherent instability introduced by NDR. A homogeneous (uniform) electric field and electron distribution is unstable and a thin high-field domain with bipolar charge distribution (Gunn domain, predicted in [673]) can develop and drift through the device. After reaching the anode, another domain can form, causing a periodically fluctuating current. The highest frequencies can be achieved with the so-called limited-space-charge accumulation (LSA) mode which operates without domains. Optimal performance requires non-homogeneous doping profiles. A detailed discussion of the various oscillation mechanisms can be found in [674, 1519, 1525].

Discussion of further microwave diode devices such as the IMPATT diode can be found in [500, 1526].

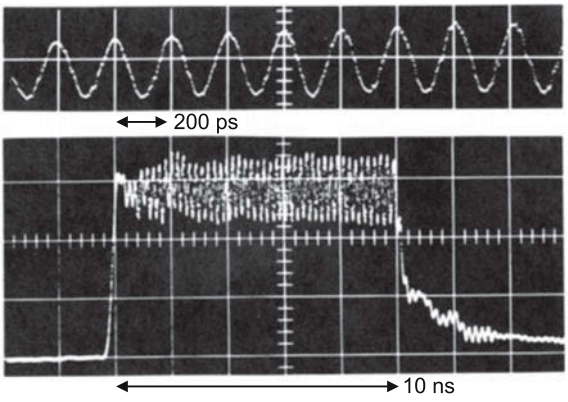


Fig. 21.76 Current traces upon excitation of a 25 μm thick piece of n-type GaAs with a voltage pulse (amplitude 16 V and duration 10 ns). The upper trace is an expanded view of the lower trace. The oscillation period is 4.5 GHz. Adapted from [122]

Chapter 22
Light-to-Electricity Conversion

Abstract The use of diodes for the conversion of electromagnetic radiation (light) into electrical signals is discussed with many examples for devices such as photoconductors, pn-, pin-, MSM- and avalanche photodiodes and charge-coupled devices. Energy conversion with solar cells is explained and standard and advanced solar cells concepts are introduced.

22.1 Photocatalysis

The absorption of light in a semiconductor across the band gap creates free electrons and holes. In particular, for small particle size in powders[1] these charge carriers can reach the surface of the semiconductor. At the surface they can react with chemicals. The hole can form •OH radicals from OH^- attached to the bead. The electron can form $O_2\bullet^-$. These radicals can subsequently attack and detoxify, e.g., noxious organic pollutants in the solution surrounding the semiconductor. Such photocatalytic activity has been found, e.g., for TiO_2 and ZnO powders. A review of photocatalysis, in particular with TiO_2 particles and their surface modifications with metals and other semiconductors, is found in [1527].

The efficiency of the photocatalytic activity depends on the efficiency of the charge separation (Fig. 22.1). Any electron–hole pair that recombines within the bulk or the surface of the particle is lost for the catalytic activity. Thus, surfaces must exhibit a small density of recombination centers. Surface traps, however, can be beneficial for charge-carrier separation when they 'store' the charge-carrier rather than letting it recombine. Small particles are expected to exhibit more efficient charge-carrier separation than larger ones. Electrons at the surface can be donated and reduce an electron acceptor, typically oxygen, $A \rightarrow A^-$. A hole at the surface can oxidize a donor species, $D \rightarrow D^+$.

An example of increased photocatalytic activity are TiO_2 powders with deposited metal particles (such as Pt) for H_2 evolution and metal-oxide particles (such as RuO_2) for O_2 evolution. Such a system behaves as a short-circuited microscopic

[1] 'Small' is here in relation to the diffusion length and does not need to be in the range where quantization effects (quantum dots) are present.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_22

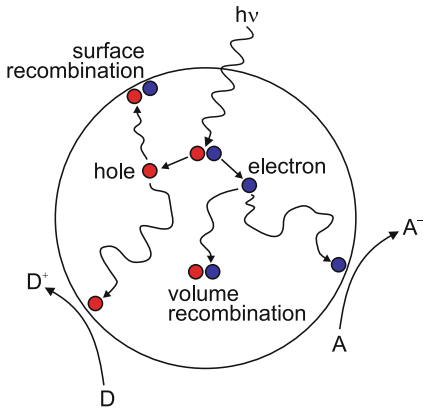


Fig. 22.1 Principle of photocatalytic activity. Light absorption creates an electron–hole pair. The electron and hole diffuse and can recombine in the bulk or at the surface. Free carriers can react at the surface with species from the surrounding solution, reducing an electron acceptor or oxidizing a donor species. Adapted from [1527]

photoelectrochemical cell in which Pt is the cathode and RuO_2 is the anode [1528]. Excitation with light energy above the band gap in the TiO_2 particle (3.2 eV) injects electrons into the Pt particles and holes into the RuO_2 particles. Trapped electrons in Pt reduce water to hydrogen and trapped holes in RuO_2 oxidize water to oxygen.

The photocatalytic activity is also tied to the geometrical shape of the semiconductor. Generally, powders with nanosized grains have much higher activity than those with microsized particles [1569]. In Fig. 22.2 it is shown that nanosized objects with high surface-to-volume ratio are more effective catalysts than rather compact surfaces.

In sun-protection cream only the UV absorption is wanted in UVA (330–420 nm) and UVB (260–330 nm) ranges. Subsequent photocatalysis on the skin and the presence of radicals are unwanted. Thus the semiconductor particles (∼10–200 nm diameter) are encapsulated in microbeads (∼1–10 μm diameter) of silica, PMMA or urethane, also improving ease of dispersion, aggregation, stability and skin feel.

22.2 Photoconductors

22.2.1 Introduction

Charge carriers can be generated in the semiconductor through the absorption of light with a photon energy above or below the band gap (Fig. 22.3). Absorption involving impurities occurs typically in the mid- and far-infrared spectral regimes (cf. Sect. 9.7). The additional charge carriers cause an increase in the conductivity (8.11).

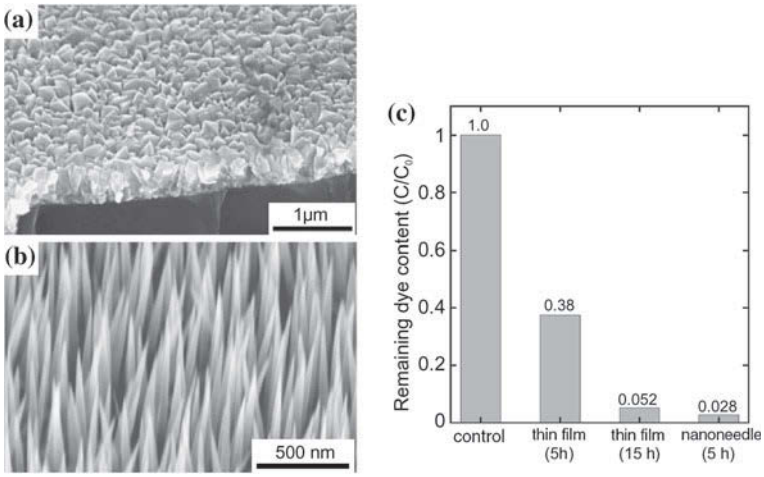


Fig. 22.2 SEM images of MOCVD-grown (**a**) ZnO thin film and (**b**) ZnO nanoneedle layer. (**c**) Comparison of the photocatalytic activity (decoloration of the dye Orange II in aqueous solution) of the ZnO thin film (irradiation with a Hg lamp for 5 h and 15 h) and the ZnO nanoneedles (irradiation 5 h). The sample labeled 'control' (scaled to 100 %) is the start situation (absorption of the dye Orange II) without photocatalytic process. Adapted from [1530], reprinted with permission

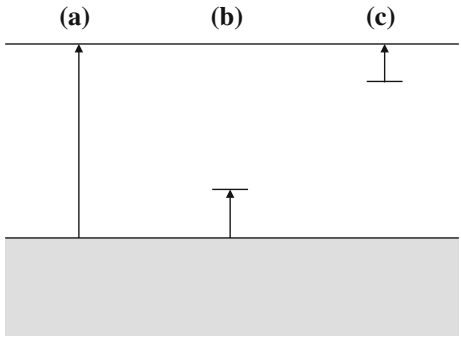


Fig. 22.3 Absorption and charge-carrier generation in a photoconductor: (**a**) band–band transition, (**b**) valence-band to acceptor and (**c**) donor to conduction-band transition

22.2.2 Photoconductivity Detectors

In stationary equilibrium for constant illumination of power P_{opt} and photon energy $E = h\nu$ the generation rate G is given by

$$G = \frac{n}{\tau} = \eta \frac{P_{\text{opt}}/h\nu}{V}, \tag{22.1}$$

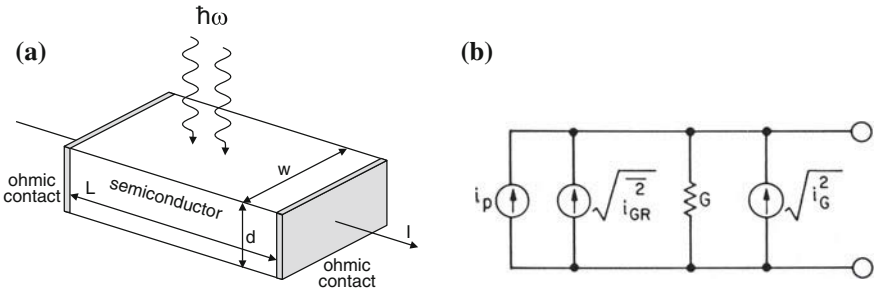


Fig. 22.4 (**a**) Scheme of photoconductor. (**b**) Equivalent circuit of photoconductor

where V is the volume ($V = wdL$, see Fig. 22.4) and τ denotes the charge-carrier lifetime. η is the quantum efficiency, i.e. the average number of electron–hole pairs generated per incoming photon. The photocurrent between the electrodes is

$$I_{\mathrm{ph}} = \sigma E w d \approx e \mu_{\mathrm{n}} n E w d, \tag{22.2}$$

assuming that $\mu_{\mathrm{n}} \gg \mu_{\mathrm{p}}$ and with $E = V/L$ denoting the electric field in the photoconductor, V being the voltage across the photoconductor. We can then also write

$$I_{\mathrm{ph}} = e \left(\eta \frac{P_{\mathrm{opt}}}{h\nu}\right) \left(\frac{\mu_{\mathrm{n}} \tau E}{L}\right) = g I_{\mathrm{p}}. \tag{22.3}$$

With the primary photocurrent $I_{\mathrm{p}} = e \left(\eta \frac{P_{\mathrm{opt}}}{h\nu}\right)$ we deduce a gain

$$g = \frac{I_{\mathrm{ph}}}{I_{\mathrm{p}}} = \frac{\mu_{\mathrm{n}} \tau E}{L} = \frac{\tau}{t_{\mathrm{r}}}, \tag{22.4}$$

where $t_{\mathrm{r}} = L/v_{\mathrm{d}}$ is the transit time through the photoconductor.

Now we consider a modulated light intensity

$$P(\omega) = P_{\mathrm{opt}} \left[1 + m \exp(\mathrm{i}\omega t)\right], , \tag{22.5}$$

where m is between 0 and 1. For $m = 0$ it is a constant light power, for $m = 1$ the intensity is sinusoidally modulated between 0 and $P_{\max} = 2P_{\mathrm{opt}}$. The rms optical power[2] is given by $\sqrt{2} m P_{\mathrm{opt}}$. In the case of $m = 1$ this is equal to $P_{\max}/\sqrt{2}$.

The rms photocurrent ($i^2 = \langle I^2 \rangle - \langle I \rangle^2$) is

$$i_{\mathrm{ph}} \approx \frac{e \eta m P_{\mathrm{opt}}}{\sqrt{2} h\nu} \frac{\tau}{t_{\mathrm{r}}} \frac{1}{\sqrt{1 + \omega^2 \tau^2}}. \tag{22.6}$$

[2]The rms value is the square root of the time average of the square of the power, $\sqrt{\langle P^2 \rangle}$.

Besides the photocurrent which is considered the signal, several sources of noise must be considered. Noise is in this case a fluctuation current i_n with $\langle i_n \rangle = 0$ (see Appendix J).

The thermal noise (Appendix J.3.1) at a conductivity $G = 1/R$ is[3]

$$i_G^2 = 4\,k\,T\,G\,B, \tag{22.7}$$

with B being the bandwidth over which the noise spectrum is integrated. The thermal noise at a resistor was experimentally found by Johnson [1531, 1532] and theoretically explained by Nyquist [1533].[4]

The statistical nature (Poisson statistics) of photon arrival (and absorption) and equally that of recombination of (photo-) excited electrons leads to fluctuations of the carrier density and consequently to fluctuating conductivity and gain. This so-called generation–recombination noise (Appendix J.3.4 is given by [1534]

$$i_{GR}^2 = 4\,e\,I_{ph}\,B\,g\,\frac{1}{1+\omega^2\,\tau^2} \tag{22.8}$$

for the modulation frequency ω, I_{ph} being the photocurrent in steady state (22.3). The equivalent circuit with the ideal photocurrent source and the noise currents is depicted in Fig. 22.4b. A detailed treatment can be found in [1535].

The signal-to-noise ratio of the power is then given by

$$S/N = \frac{i_{ph}^2}{i_G^2 + i_{GR}^2} = \frac{\eta\,m^2\,(P_{opt}/h\nu)}{8\,B}\left[1 + \beta^{-1}\frac{t_r}{\tau}(1+\omega^2\tau^2)\,\frac{G}{I_0}\right]^{-1}. \tag{22.9}$$

An important quantity is the noise equivalent power (NEP). This is the light power $(mP_{opt}/\sqrt{2})$ for which the S/N ratio is equal to 1 (for $B = 1$). The responsivity of a detector to light has been termed 'detectivity' and is the inverse of the noise equivalent power. It typically depends on the square root of the detector area A and the bandwidth B [1536]. Thus the detectivity D^* (D star) has been introduced defined by [1537]

$$D^* = \frac{\sqrt{A\,B}}{\mathrm{NEP}}, \tag{22.10}$$

in order to make various detectors comparable. The unit of D^* is $\mathrm{cm\,Hz^{1/2}\,W^{-1}}$, also known as Jones. The detectivity should be stated together with the modulation frequency. It can be given for monochromatic radiation of a particular wavelength λ or a blackbody spectrum of given temperature T. As refined measure for detectivity, D^{**} (D double star) has been defined to take into account the solid angle Ω from

[3] For frequencies $h\nu \ll kT$; at room temperature kT/h is in the THz regime.

[4] The formula for thermal noise (22.7) is the fluctuation-dissipation theorem in statistical physics, providing a general relation between the response of an equilibrium system to small external perturbations and its spontaneous fluctuations.

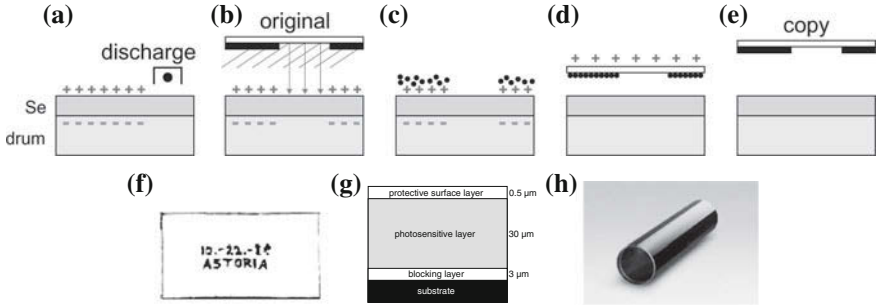


Fig. 22.5 Principle of xerography: (**a**) charging of the selenium-covered drum, (**b**) (reflection) exposure of the Se, exposed areas become uncharged, (**c**) toner addition, (**d**) toner transfer to paper for copy, and (**e**) fixation of the toner on the copy and preparation of drum for the next cycle. (**f**) First xerox copy (Oct. 22nd 1938). (**g**) Schematic cross section of coating of photosensitive drum. The indicated *thicknesses* are approximate. (**h**) Image of drum with photosensitive layer made from amorphous silicon. Part (**h**) from [1539]

which radiation can reach the detector [1538],

$$D^{**} = \sqrt{\Omega/\pi}\, D^{*}\ ; \tag{22.11}$$

for a Lambertian characteristic $D^{**} = D^{*}$.

22.2.3 Electrophotography

The principle of the Xerox copy machine is based on a photoconductive layer (Fig. 22.5). This layer is normally insulating such that both sides of the layer can be oppositely charged. If light hits the layer it becomes photoconductive and neutralizes locally. This requires a small lateral diffusion of charge carriers. Initially amorphous selenium ($E_g = 1.8\,\text{eV}$) was used. The conductivity in the dark of a-Se is $10^{16}\,\Omega$/cm. Se was subsequently replaced by organic material. The highest performance is currently achieved with amorphous silicon.

On the charged areas of the photosensitive layer toner can be attached. The toner pattern is subsequently transferred to the copy sheet and fixated. A copy takes typically more than one rotation of the drum. The principle was invented in 1938 by Chester F. Carlson (1906–1968) with sulfur as the photoconductor.[5]

[5] In 1947 the Haloid company bought the rights to this process, renamed itself XeroX and brought the first copy machine to the market in 1958 based on amorphous selenium. The word 'xerography' stems from the Greek word $\xi\acute{\epsilon}\rho o\varsigma$ (dry). The last 'X' in XeroX was added to mimic the name of the KodaK corporation.

22.2.4 QWIPs

Quantum-well intersubband photodetectors (QWIPs) are based on the absorption of photons between two quantum well subbands (Fig. 22.6). A review can be found in [1540]. Quantized electron or hole states can be used. Besides an oscillator strength for this transition, the lower level must be populated and the upper level must be empty in order to allow this process. The Fermi level is typically chosen by appropriate doping such that the lower subband is populated.

For infinite barrier height the energy separation between the first and second quantized levels (in the effective-mass theory) is (cf. (12.6))

$$E_2 - E_1 = 3\,\frac{\hbar^2}{2m^*}\,\frac{\pi^2}{L_z^2}. \tag{22.12}$$

For real materials the barrier height determines the maximum transition energy. Typical absorption and transmission spectra of a QWIP structure are shown in Fig. 22.7. The spectral response is in the mid- or far-infrared.

The dipole matrix element $\langle z\rangle = \langle \Psi_2|z|\Psi_1\rangle$ can be easily calculated to be

$$\langle z\rangle = \frac{16}{9\pi^2}\,L_z. \tag{22.13}$$

The oscillator strength is about 0.96. The polarization selection rule causes the absorption to vary $\propto \cos^2\phi$, where ϕ is the angle between the electric-field vector and the z direction (Fig. 22.8). This means that for vertical incidence ($\phi = 90°$) the absorption vanishes. Thus schemes have been developed to allow for skew entry of the radiation (Fig. 22.9a). The strict selection rule can be relaxed by using asymmetric potential wells (breaking of mirror symmetry/parity), strained materials (band mixing) or quantum dots (lateral confinement). Also, a grating can be used to create a finite angle of incidence (Fig. 22.9b).

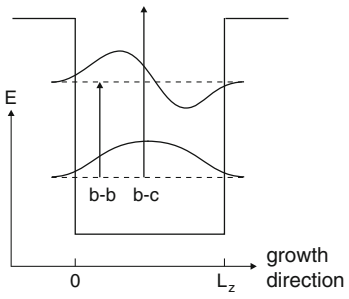


Fig. 22.6 Schematic level diagram of a quantum well. Optical intersubband transitions between the first and second quantized level (b–b) and the ground state and the continuum (b–c)

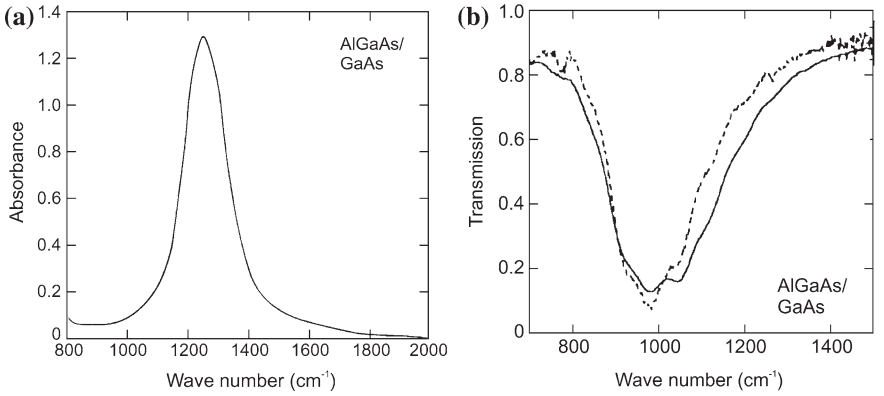


Fig. 22.7 (**a**) AlGaAs/GaAs QWIP absorption spectrum for multiple reflection geometry. Adapted from [1540]. (**b**) Transmission of AlGaAs/GaAs QWIP (100 QWs) in double reflection geometry (45° angle of incidence). The well doping is $1.0 \times 10^{12}\,\mathrm{cm}^{-2}$ (*dashed line*) and $1.5 \times 10^{12}\,\mathrm{cm}^{-2}$ (*solid line*). Adapted from [1541]

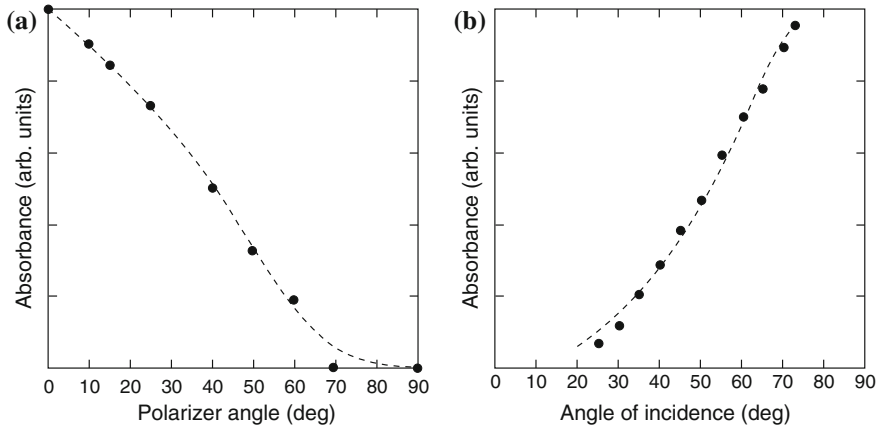


Fig. 22.8 Dependence of the QWIP response on (**a**) polarization and (**b**) angle of incidence. *Dashed lines* are guides to the eye. Adapted from [1540]

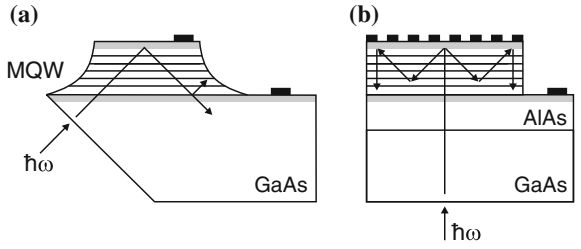


Fig. 22.9 QWIP geometries: (**a**) 45° edge coupled with multiple quantum-well (MQW) absorber and (**b**) grating coupled with GaAs substrate, AlAs reflector and metal grating on top. *Grey areas* are highly n-doped contact layers

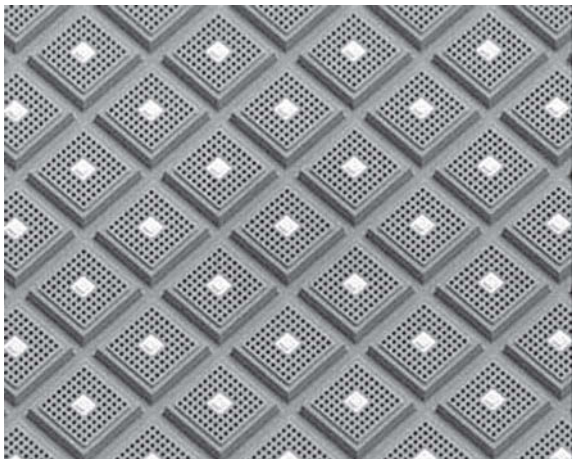


Fig. 22.10 Part of a 256×256 QWIP focal plane array (FPA) with grating coupler (area of one pixel: $37\,\mu m^2$). From [1543]

Besides a useful detectivity (2×10^{10} cm Hz$^{1/2}$/W at 77 K) QWIPs have the advantage, e.g. against HgCdTe interband absorbers, that the highly developed GaAs planar technology is available for the fabrication of focal plane arrays (FPA) as shown in Fig. 22.10. A FPA is an image sensor (in the focal plane of an imaging infrared optics) and is used, e.g., for the detection of heat leaks in buildings or night surveillance. In particular, night vision support in cars may become a major market. A competing technology are bolometric arrays with thermally insulated pixels based on MEMS technology. A review of FPA technologies can be found in [1542].

The carriers that have been optically excited into the upper state leave the QW by tunneling or thermionic emission. Also, a QWIP can be made based on the direct transfer from the (populated) subband into the continuum.

The incoming infrared radiation creates a photocurrent density of

$$i_{\mathrm{ph}} = e\,\eta_{\mathrm{w}}\,\Phi, \tag{22.14}$$

where η_{w} is the quantum efficiency of a single quantum well (including the escape rate) and Φ is the photon flux per time and unit area. During the transport of the charge carriers through the barrier they can be (re-)captured by the QW with the probability p_{c}. The capture probability decreases exponentially with the applied bias. The total photocurrent (including generation and recapture) is

$$I_{\mathrm{ph}} = (1 - p_{\mathrm{c}})\,I_{\mathrm{ph}} + i_{\mathrm{ph}} = \frac{i_{\mathrm{ph}}}{p_{\mathrm{c}}}. \tag{22.15}$$

If the quantum efficiency is small, the efficiency of N_{w} quantum wells $\eta \approx N_{\mathrm{w}} \times \eta_{\mathrm{w}}$. With this approximation the total photocurrent of N_{w} quantum wells is given by

$$I_{\mathrm{ph}} = e\,\eta\,\Phi\,g, \tag{22.16}$$

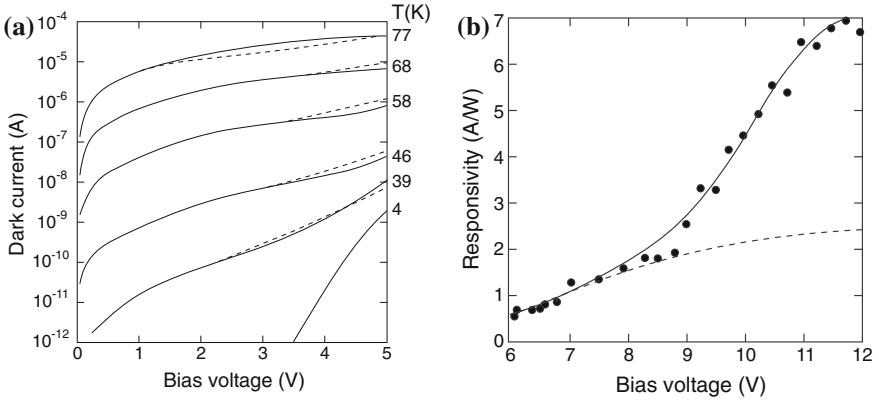


Fig. 22.11 (**a**) Dark current of a QWIP at 10.7 μm, experimental (*solid lines*) and theoretical (*dashed lines*) response. (**b**) QWIP responsivity as a function of the applied voltage. The *solid line* (*dashed line*) is the theoretical dependence with (without) the effect of avalanche multiplication. Adapted from [1540]

where g is termed the gain of the structure and is given by

$$g = \frac{1}{p_c} \frac{\eta_w}{\eta} \approx \frac{1}{N_w p_c}. \tag{22.17}$$

The dark current can be calculated from thermionic emission and agrees fairly well with experiment (Fig. 22.11a). When the voltage is increased further, avalanche multiplication can occur while the carriers are transported through the barrier(s). This mechanism provides further gain as shown in Fig. 22.11b.

22.2.5 Blocked Impurity-Band Detectors

Impurity absorption allows photoconductivity detectors in the mid- and far-infrared regions to be made. In particular, for THz spectroscopy in medicine and astronomy the extension to longer wavelengths is interesting. For conventional photoconductors the impurity concentration is well below the critical dopant concentration (cf. Sect. 7.5.7). Long-wavelength response can be achieved by going to impurity/host systems with smaller ionization energy, such as Si:B (45 meV) → Ge:As (12.7 meV) → GaAs:Te (5.7 meV). By applying stress to Ge the energy separation between impurity and conduction bands can be lowered and subsequently the detector response is shifted towards longer wavelengths.

For high doping the impurity level broadens to an impurity band and thus allows smaller ionization energy and thus stronger long-wavelength detector response. However, conduction in the impurity band leads to dark current and makes such detectors unfeasible. In a blocked impurity band (BIB) detector [1544–1546] an additional

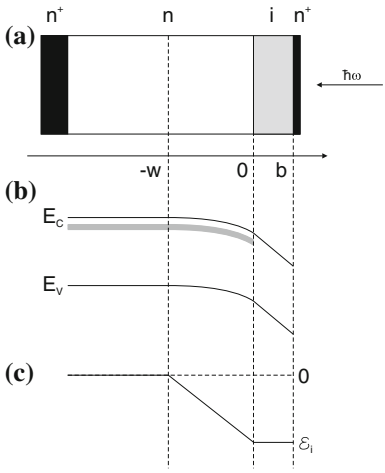


Fig. 22.12 (**a**) Structure of BIB photodetectors. Highly doped contact layers (*black*), doped semiconductor (*white*) and blocking (intrinsic) layer (*grey*). (**b**) Band diagram under small forward bias. *Shaded area* represents the donor impurity band. (**c**) Electric field in the structure

intrinsic blocking layer is sandwiched between the absorption layer and the contact (Fig. 22.12a). Such a structure is similar to a MIS diode, the insulator being the intrinsic semiconductor. We assume in the following an n-type semiconductor, such as Si:As or GaAs:Te, but also p-type BIBs can be made, e.g., from Ge:Ga.

The semiconductor is highly doped (N_D) and partly compensated (N_A). Typically, the acceptor concentration must be small, about $10^{12}\,\mathrm{cm}^{-3}$, and controls the formation of the electric field as shown below. The doping is so high that the impurities form an impurity band. Some of the electrons recombine with the acceptors $N_A^- = N_A$ and leave some donors charged $N_D^+ = N_A$. For GaAs, e.g., the donor concentration in the doped semiconductor is $>10^{16}\,\mathrm{cm}^{-3}$ and $\sim 10^{13}\,\mathrm{cm}^{-3}$ in the i-layer.

Under an external forward bias V, i.e. the positive pole is at the insulator, part of the applied voltage drops over the blocking layer of thickness b. If ideally no charges are present here, the electric field is constant. In the n-doped material electrons move in the impurity band towards the insulator, forming neutral donors in an electron accumulation layer of thickness w in the presence of the charged acceptors N_A^-. This layer is the absorption layer. The mechanism can also be considered as if positive charge (the charged donors, N_D^+) moves (via hopping conduction) towards the back contact. In the literature the layer close to the insulator is thus also termed a 'depletion layer'. The band diagram and the electric field are shown in Fig. 22.12b,c. Due to the blocking layer the carriers on the donors in the n-type material cannot spill via the impurity band into the contact but must be lifted (by photoabsorption) into the conduction band.

From the Poisson equation the electric field is given by

$$E(x) = -\frac{e}{\epsilon_s} N_A\,(w + x)\,, \quad -w \le x \le 0 \tag{22.18a}$$

$$E(x) = -\frac{e}{\epsilon_s} N_A\, w = E_i, \quad 0 \le x \le b. \tag{22.18b}$$

The voltage drops across the blocking layer V_b and the doped semiconductor V_s fulfill

$$V = V_b + V_s. \tag{22.19}$$

Integration of the fields yields

$$V_s = \frac{e}{\epsilon_s} N_A \frac{w^2}{2} \tag{22.20a}$$

$$V_b = \frac{e}{\epsilon_s} N_A \, w \, b. \tag{22.20b}$$

Substituting (22.20a, b) into (22.19) results in the width of the 'depletion layer'

$$w = \sqrt{\frac{2\,\epsilon_s\,V}{e\,N_A} + b^2} - b. \tag{22.21}$$

The high dopant concentration allows for much thinner absorption layers than in a conventional photoconductivity detector, making it less susceptible to background high-energy cosmic radiation. The recombination in the depletion layer is negligible. Detector performance is modeled in [1547].

22.3 Photodiodes

22.3.1 Introduction

The principle of the photodiode is the interband absorption of light in the depletion layer of a diode (or the i-zone of a pin-diode) and the subsequent separation of electrons and holes by the electric field. There are opposite requirements for fast detectors (thin depletion layer) and efficient detectors (complete light absorption, sufficiently thick depletion layer). For this reason generally semiconductors with high absorption coefficient are most suited (Fig. 22.13). In Fig. 22.14 the quantum efficiency and detectivity D^* of various semiconductor detectors are compared.

A diode can be operated without bias (photovoltaic mode) using the built-in field. An improvement in the speed of a pn-diode is achieved with a reverse bias since it increases the field strength in the depletion layer. However, the reverse bias is below the breakdown voltage. Operation near breakdown is exploited in the avalanche photodiode (APD). In the following we will discuss pn-, pin-, MS- (Schottky-), MSM- and heterostructure-diodes and APDs.

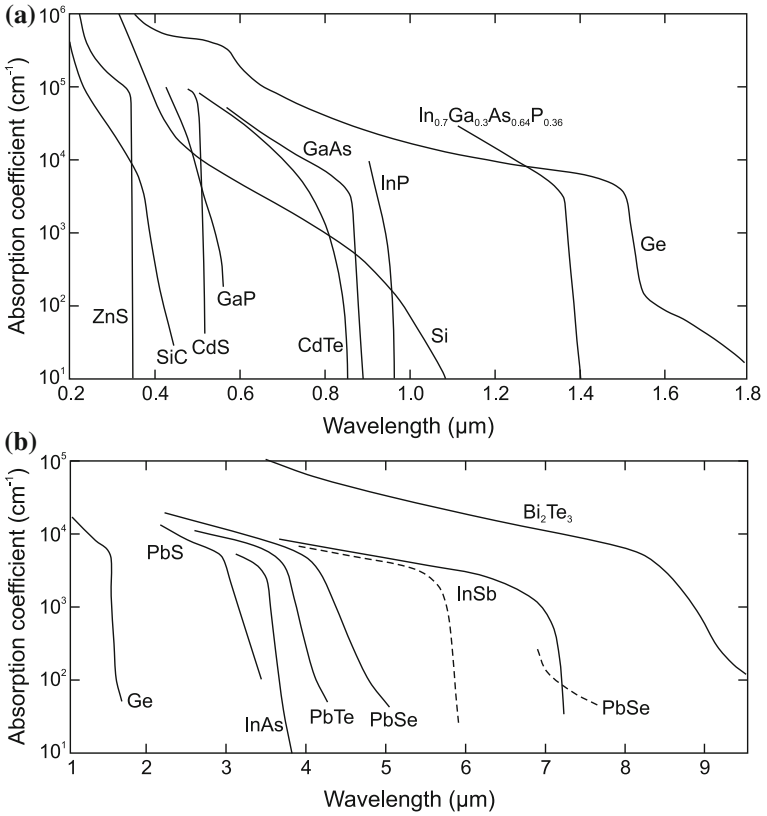


Fig. 22.13 Optical absorption coefficient of various semiconductor materials (as labeled) used for photodetectors (**a**) in the UV, visible and near-infrared range (at room temperature) and (**b**) in the mid-infrared spectral range at room temperature (*solid lines*) and at 77 K (*dashed lines*). Based on [1548]

22.3.2 pn Photodiodes

The most important figures of merit are the quantum efficiency, responsivity, noise equivalent power (NEP) and the response speed.

If the depletion layer is hit by a photon flux with a generation rate G_0 (i.e. electron–hole pairs per unit volume per unit time) the photogenerated current is added to the diffusion current. The photocurrent density j_p (per unit area) is

$$j_\mathrm{p} = -e\, G_0\, L_\mathrm{p} \tag{22.22}$$

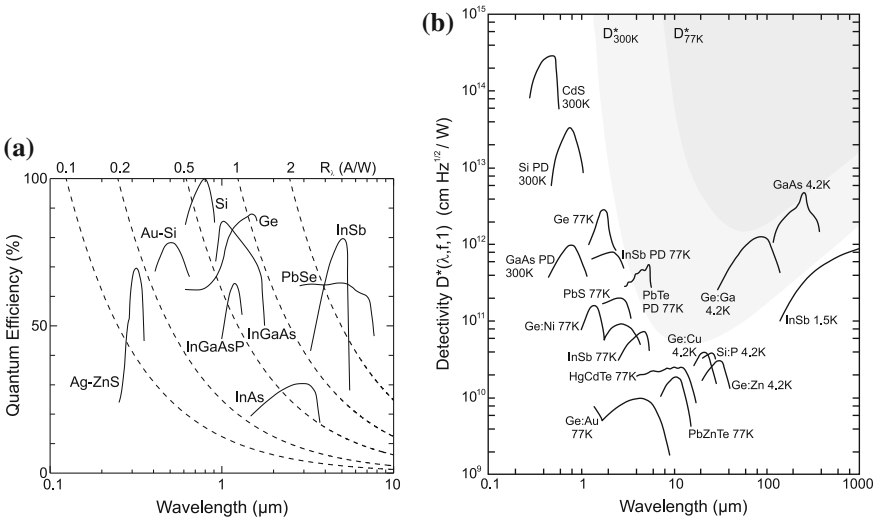


Fig. 22.14 (**a**) Quantum efficiency of various photodetectors. The *dashed lines* depict lines of equal responsivity (R_λ in A/W) as labeled on top of the panel. (**b**) Detectivity D^* of various photoconductors and photodiodes (PD). The *lighter* (*darker*) *shaded area* indicates the range unachievable at 300 K (77 K) due to background radiation. Adapted from [500]

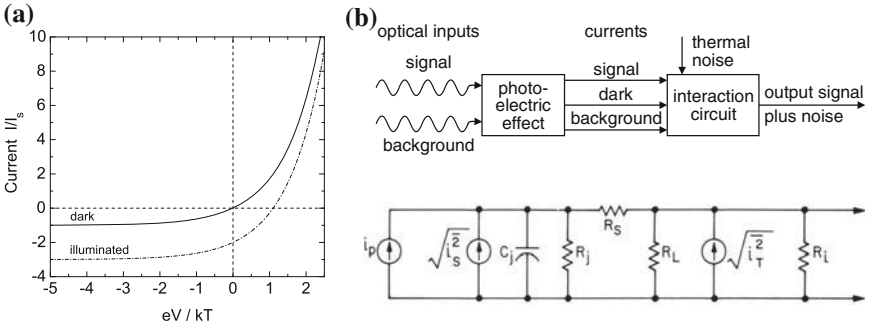


Fig. 22.15 (**a**) Schematic dark and illuminated I–V-characteristics of a photodiode (for the case $j_\mathrm{p} = -2j_\mathrm{s}$). (**b**) Schematic representation of currents in a photodiode and equivalent circuit. Part (**b**) adapted from [1549]

for a p^+n-diode. In order to obtain this result the diffusion and continuity equations have to be solved for the depletion region.[6] Equation (22.22) means that the dark I–V characteristic is shifted by j_p as shown in Fig. 22.15a. The number of electron–hole pairs that are generated per photon of energy $h\nu$ by the absorption of the (monochromatic) light power P_opt is

[6]This derivation is done in Sect. 22.3.3. Equation (22.22) is obtained from (22.33) for vanishing thickness w and $\alpha L_\mathrm{p} \ll 1$.

$$\eta = \frac{I_{\mathrm{ph}}/e}{P_{\mathrm{opt}}/(h\nu)}, \tag{22.23}$$

where $I_{\mathrm{ph}} = Aj_{\mathrm{ph}}$ is the photogenerated current over the surface A. The *responsivity* R_λ of the photodiode (for monochromatic radiation) is defined as

$$R_\lambda = \frac{I_{\mathrm{ph}}}{P_{\mathrm{opt}}} = \frac{e}{h\nu}\,\eta \approx \frac{\lambda}{1.24\,\mu\mathrm{m}}\,\eta. \tag{22.24}$$

For a modulated light intensity P_{opt} must be replaced by $mP_{\mathrm{opt}}/\sqrt{2}$. The equivalent circuit including noise sources for a photodiode is shown in Fig. 22.15b.

Random processes lead to shot noise $\langle i_{\mathrm{S}}^2 \rangle$ (Sect. J.3.3). Besides the photocurrent I_{ph} itself, the background radiation (I_{B}, in particular for infrared detectors) and the thermal generation (dark current, I_{D}) of carriers contribute:

$$\langle i_{\mathrm{S}}^2 \rangle = 2\,e\,\left(I_{\mathrm{ph}} + I_{\mathrm{B}} + I_{\mathrm{D}}\right)\,B, \tag{22.25}$$

with B being the bandwidth. Additionally, the parallel resistances cause thermal noise

$$\langle i_{\mathrm{T}}^2 \rangle = 4\,kT\,B/R_{\mathrm{eq}}. \tag{22.26}$$

The resistance R_{eq} is given by the resistance of the depletion layer (junction) R_{j}, the load R_{L} and the input of the amplifier R_{i} as $R_{\mathrm{eq}}^{-1} = R_{\mathrm{j}}^{-1} + R_{\mathrm{L}}^{-1} + R_{\mathrm{i}}^{-1}$. The series resistance R_{s} of the photodiode can be usually ignored in this context.

For a fully modulated signal the signal-to-noise ratio of the photodiode is given by

$$S/N = \frac{i_{\mathrm{ph}}^2}{\langle i_{\mathrm{S}}^2 \rangle + \langle i_{\mathrm{T}}^2 \rangle} = \frac{\left(e\,\eta\,P_{\mathrm{opt}}/h\nu\right)^2/2}{2e\,\left(I_{\mathrm{ph}} + I_{\mathrm{B}} + I_{\mathrm{D}}\right)\,B + 4\,kT\,B/R_{\mathrm{eq}}}. \tag{22.27}$$

Therefore the NEP is given by

$$\mathrm{NEP} = \frac{2\,h\nu\,B}{\eta}\left[1 + \sqrt{1 + \frac{I_{\mathrm{eq}}}{e\,B}}\right]. \tag{22.28}$$

The current I_{eq} is given by $I_{\mathrm{eq}} = I_{\mathrm{B}} + I_{\mathrm{D}} + 2kT/(eR_{\mathrm{eq}})$. If $I_{\mathrm{eq}}/eB \ll 1$, the NEP is determined by the shot noise of the signal itself. In the other limit $I_{\mathrm{eq}}/eB \gg 1$ the detection is limited by the background radiation or thermal noise. In this case, the NEP is (for $B = 1\,\mathrm{Hz}$, in $\mathrm{W\,cm^2\,Hz^{1/2}}$)

$$\mathrm{NEP} = \sqrt{2}\,\frac{h\nu}{\eta}\,\sqrt{\frac{I_{\mathrm{eq}}}{e}}. \tag{22.29}$$

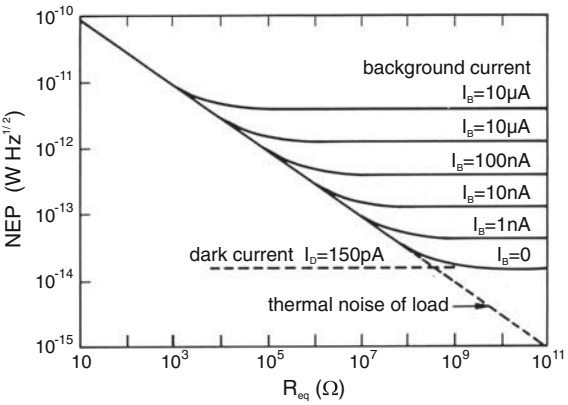


Fig. 22.16 NEP as a function of the resistance R_{eq} for a Si photodiode. From [1549]

In Fig. 22.16 the situation is shown for a silicon photodiode as a function of R_{eq}. The diode has a quantum efficiency of 75 % at $\lambda = 0.77\,\mu\text{m}$. A high value of $R_{\text{eq}} \sim 1\,\text{G}\Omega$ is necessary to ensure detection limited by dark current.

22.3.3 Pin Photodiodes

The depletion layer in pn-diodes is relatively thin such that the incident light is not completely absorbed. An almost complete absorption of light can be achieved by using a thick intrinsic absorption layer. The field in the intrinsic region is constant or slowly varying linearly (Fig. 21.70). The generation rate per unit area decreases exponentially following the Lambert–Beer law (9.16) as shown in Fig. 22.17c:

$$G(x) = G_0 \exp(-\alpha x). \tag{22.30}$$

The initial generation rate $G_0 = \Phi_0 \alpha$ is given by the incident photon flux per unit area Φ_0 and the reflectance of the surface R as $\Phi_0 = P_{\text{opt}}(1 - R)/(Ah\nu)$.

The drift current in the i-region collects all those carriers (if recombination in the depletion layer is neglected). The electron drift current is given by

$$j_{\text{dr}} = -e \int_0^w G(x)\,\text{d}x = e\,\Phi_0 \left[1 - \exp(-\alpha w)\right], \tag{22.31}$$

with w being the thickness of the depletion layer that is approximately the same as the thickness of the i-region. In the bulk (neutral) region ($x > w$) the minority-carrier density is determined by drift and diffusion[7] (10.76). The diffusion current density

[7] At the edge of the depletion layer, $x = w$, all photo-generated carriers are transported away instantly, thus the excess carrier density from photo-generated carriers is zero there and (10.76) applies.

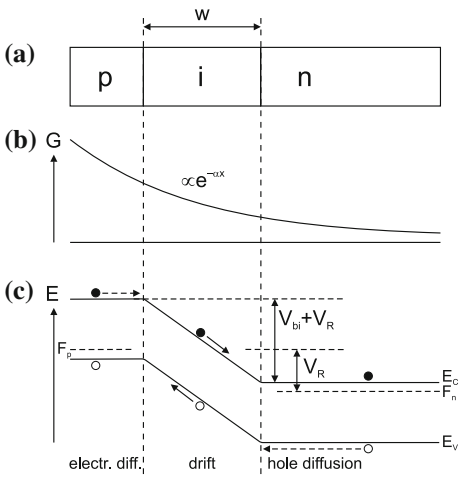


Fig. 22.17 (**a**) Schematic cross section of pin-diode, (**b**) profile of carrier generation due to light absorption and (**c**) schematic band structure under reverse bias. The generation of three electron-hole pairs is shown; subsequent drift (diffusive) transport is indicated with *solid* (*dashed*) *arrows*

at $x = w$ is thus given by

$$j_{\text{diff}} = e\,\Phi_0 \exp(-\alpha\, w)\;\frac{\alpha\, L_\text{p}}{1+\alpha\, L_\text{p}} + e\,p_{\text{n}_0}\,\frac{D_\text{p}}{L_\text{p}}. \tag{22.32}$$

The first term is due to the diffusion current of photo-generated carriers (10.76), the second term due to thermally generated carriers (21.129). The total current $j_{\text{tot}} = j_{\text{diff}} + j_{\text{dr}}$ is given by

$$j_{\text{tot}} = e\,\Phi_0 \left[1 - \frac{\exp(-\alpha\, w)}{1+\alpha\, L_\text{p}}\right] + e\,p_{\text{n}_0}\,\frac{D_\text{p}}{L_\text{p}}. \tag{22.33}$$

The first term is due to the photocurrent, the second term is due to the diffusion current known from the p^+n-diode. In normal operation, the second can be neglected compared to the first. The quantum efficiency is

$$\eta = \frac{j_{\text{tot}}/e}{P_{\text{opt}}/h\nu} = (1-R)\left[1 - \frac{\exp(-\alpha\, w)}{1+\alpha\, L_\text{p}}\right]. \tag{22.34}$$

For a high quantum efficiency, of course low reflectance and high absorption coefficient, i.e. $\alpha w \gg 1$, are necessary.

However, for $w \gg 1/\alpha$ the transit time through the depletion layer $t_\text{r} \approx w/v_\text{s}$ (at sufficiently high field, v_s being the drift-saturation velocity) increases too much. The 3 dB cutoff frequency $f_{3\,\text{dB}}$ (Fig. 22.18) is

$$f_{3\,\text{dB}} \cong \frac{2.4}{2\pi\, t_\text{r}} \cong \frac{0.4\, v_\text{s}}{w}. \tag{22.35}$$

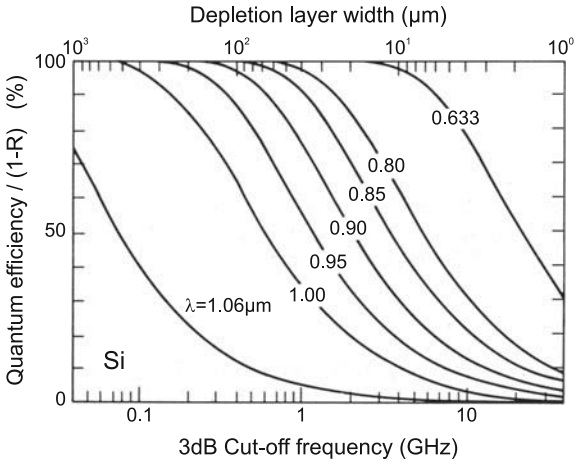


Fig. 22.18 Quantum efficiency and 3 dB cutoff frequency of a Si pin-diode at $T = 300\,\mathrm{K}$ for various wavelengths of input radiation. Adapted from [1549]

Therefore a tradeoff exists between the quantum efficiency and the response speed of the pin-photodiode (Fig. 22.18). Choosing $w \cong 1/\alpha$ is a good compromise.

22.3.4 Position-Sensing Detector

In a position-sensing detector (PSD) two electrodes are placed at opposite edges of a photodetector. The current output depends linearly on the beam position in between the electrodes, similar to a voltage divider. If two pairs of electrodes, one on the front and one on the back of the detector, are fabricated in orthogonal directions (Fig. 22.19a), the beam position can be measured in both x and y directions.

22.3.5 MSM Photodiodes

A MSM photodiode consists of a piece of semiconductor between two Schottky contacts (MS contacts). These are typically arranged laterally (as shown in Fig. 22.24b) but will first be considered at the front and back of the semiconductor [1551]. The band structure in thermodynamic equilibrium is shown in Fig. 22.20.

In the general case two different metals with two different barriers ϕ_{n1}, ϕ_{n2} and built-in voltage V_{D1}, V_{D2} are considered. If a voltage is applied across the MSM diode, one of the junctions is biased in the forward, the other in the reverse direction. We assume in Fig. 22.21 that the voltage biases the first contact in the reverse direction, i.e. the '+' pole is on the left contact. The applied voltage V is split between the two contacts, the larger voltage will drop at the reverse-biased contact (here: $V_1 > V_2$)

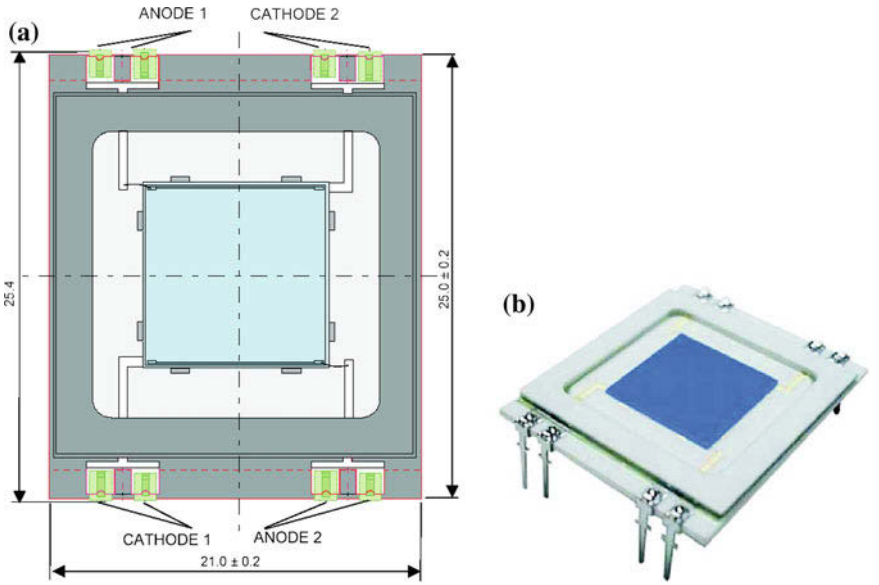


Fig. 22.19 (**a**) Scheme of two-dimensional position-sensing detector (PSD), (**b**) image of PSD. From [1550]

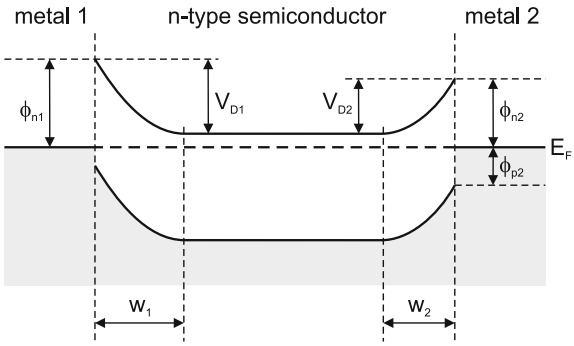


Fig. 22.20 Band diagram of a MSM structure with an n-type semiconductor in thermal equilibrium. In the general case, two different metals cause two different Schottky barrier heights and related depletion layer widths. Adapted from [1551]

$$V = V_1 + V_2. \tag{22.36}$$

The electron current arises from thermionic emission at contact 2. Due to current continuity (without recombination since we inject majority charge carriers) this is also the current through contact 1, i.e.

$$j_{n1} = j_{n2}. \tag{22.37}$$

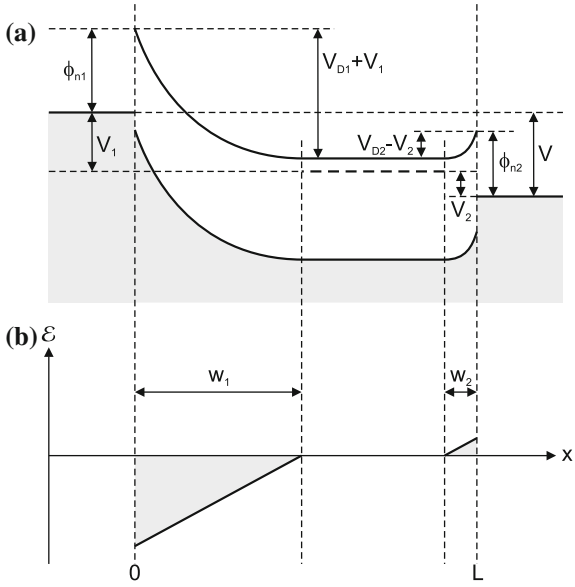


Fig. 22.21 (**a**) Band diagram for a MSM structure under bias ($V < V_{RT}$), (**b**) electric field distribution. Adapted from [1551]

The reverse current at contact 1 is

$$j_{n1} = A_n^* T^2 \exp(-\beta\phi_{n1}) \exp(\beta\Delta\phi_{n1}) \left[1 - \exp(-\beta V_1)\right], \tag{22.38}$$

where $\Delta\phi_{n1}$ is the barrier reduction due to the Schottky effect (Sects. 21.2.3 and (21.27)). The forward current at contact 2 is

$$j_{n2} = -A_n^* T^2 \exp(-\beta\phi_{n2}) \exp(\beta\Delta\phi_{n2}) \left[1 - \exp(\beta V_2)\right]. \tag{22.39}$$

For a symmetric structure, i.e. $\phi_{n1} = \phi_{n2}$ and $V_{D1} = V_{D2} = V_D$, (22.37)–(22.39) yield together with (21.27)

$$\left(\frac{e^3 N_D}{8\pi^2 \epsilon_s^3}\right)^{1/4} \left[(V_D + V_1)^{1/4} - (V_D - V_2)^{1/4}\right] = \frac{1}{\beta} \ln\left[\frac{\exp(\beta V_2) - 1}{1 - \exp(-\beta V_1)}\right]. \tag{22.40}$$

Together with (22.36) a numerical or graphical solution can be found. Initially (for small voltages) the injected hole current (from contact 2) is much smaller than the electron current and diffusion occurs in the neutral region.

The reach-through voltage V_{RT} is reached when the width of the neutral region is reduced to zero (Fig. 22.22a). At the juncture of the two depletion regions inside the semiconductor material the electric field is zero and changes sign. For a larger voltage V_{FB} flat-band conditions are present at contact 2, i.e. the electric field is zero at contact 2 (Fig. 22.22b). At even larger voltage V_B breakdown occurs.

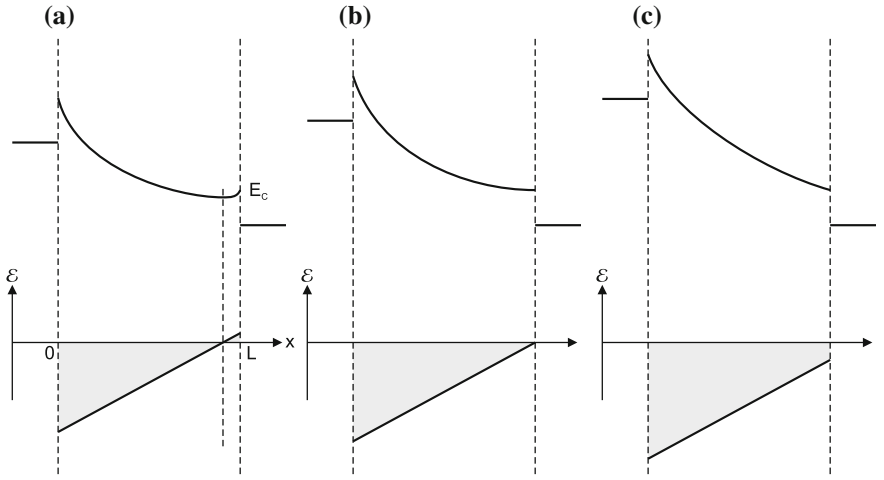


Fig. 22.22 Band diagram (*upper parts*) and electric field distribution (*lower parts*) in a MSM diode for various bias conditions: (**a**) at reach-through voltage V_{RT}, (**b**) at flat-band voltage V_{FB} and (**c**) for $V > V_{FB}$. Adapted from [1551]

At $V = V_{RT}$ we have

$$w_1 = \left[\frac{2\,\epsilon_s}{e\,N_D}\,(V_1 + V_{D1})\right]^{1/2} \tag{22.41a}$$

$$w_2 = \left[\frac{2\,\epsilon_s}{e\,N_D}\,(V_{D2} - V_2)\right]^{1/2} \tag{22.41b}$$

$$L = w_1 + w_2, \tag{22.41c}$$

and therefore (with (22.36))

$$V_{RT} = \frac{e\,N_D}{2\,\epsilon_s}\,L^2 - L\left[\frac{2\,e\,N_D}{\epsilon_s}\,(V_{D2} - V_2)\right] - \Delta V_D, \tag{22.42}$$

with $\Delta V_D = (V_{D1} - V_{D2})$, vanishing for a symmetric MSM structure. At and after reach-through the electric field varies linearly from 0 to L within the semiconductor. The point of zero electric field shifts towards contact 2. At the flat-band voltage this point has reached the contact 2 and the width of the depletion layer at contact 2 is zero. This condition leads (as long as no breakdown occurred) to

$$V_{FB} = \frac{e\,N_D}{2\,\epsilon_s}\,L^2 - \Delta V_D. \tag{22.43}$$

The maximum electric field is at contact 1 and is given (for $V > V_{FB}$) by

$$E_{m1} = \frac{V + V_{FB} + 2\,\Delta V_D}{L}. \tag{22.44}$$

If in a part of the structure the critical field E_B for impact ionization is reached (this will be at contact 1, since the field is highest there), the diode breaks down. Therefore the breakdown voltage is given by

$$V_B \approx E_B\,L - V_{FB} - 2\,\Delta V_D. \tag{22.45}$$

The current–voltage characteristic for a Si-MSM structure is shown in Fig. 22.23. At small voltages only small currents flow since one contact is in reverse bias. The hole current is much smaller than the electron current. Only those holes that diffuse through the neutral region contribute to the hole current. After reach-through the barrier $\phi_{p2} + V_{D2} - V_2$ for hole injection is strongly reduced that leads to strong hole injection. Beyond the flat-band voltage the hole current increases only weakly since a lowering of the barrier occurs only via the Schottky effect. For high fields ($V > V_{FB}$, before breakdown) the hole current is

$$j_{p1} = A_p^*\,T^2\,\exp\left(-\beta\phi_{p2}\right)\,\exp\left(\beta\Delta\phi_{p2}\right) = j_{p,s}\,\exp\left(\beta\Delta\phi_{p2}\right), \tag{22.46}$$

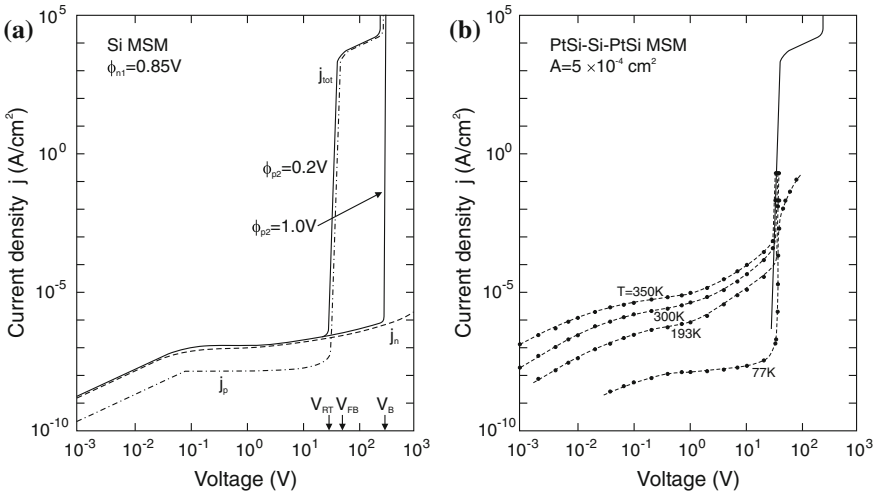


Fig. 22.23 Current–voltage characteristics of a Si MSM structure, $N_D = 4 \times 10^{14}\,cm^{-3}$, $L = 12\,\mu m$ (thin, polished, ⟨111⟩-oriented wafer), $T = 300\,K$. (**a**) Theory for two different values of ϕ_{p_2}, (**b**) experiment (for $\phi_{p_2} = 0.2\,V$). Adapted from [1551]

and the total current is

$$j = j_{\mathrm{n,s}} \exp\left(\beta \Delta\phi_{\mathrm{n1}}\right) + j_{\mathrm{p,s}} \exp\left(\beta \Delta\phi_{\mathrm{p2}}\right), \tag{22.47}$$

with $j_{\mathrm{n,s}} = A_{\mathrm{n}}^{*} T^{2} \exp\left(-\beta\phi_{\mathrm{n1}}\right)$ and $j_{\mathrm{p,s}} = A_{\mathrm{p}}^{*} T^{2} \exp\left(-\beta\phi_{\mathrm{p2}}\right)$.

In a MSM photodetector the metal contacts are typically formed in an interdigitated structure on the semiconductor surface (Fig. 22.24). These contacts shield some of the active area from photons. An increase in quantum efficiency can be achieved with transparent contacts (e.g. ZnO or ITO) and an antireflection (AR) coating.

The dark current is given by (22.47) and is minimal when electron and hole saturation currents are identical. This conditions leads to the optimal barrier height

$$\phi_{\mathrm{n}} = E_{\mathrm{g}} - \phi_{\mathrm{ph}} = \frac{1}{2}\beta^{-1} \ln\left(\frac{m_{\mathrm{e}}}{m_{\mathrm{hh}}}\right) + \frac{1}{2} E_{\mathrm{g}} \tag{22.48}$$

close to middle of the band gap. For InP and optimal barrier $\phi_{\mathrm{n}} = 0.645\,\mathrm{eV}$ a dark current of $0.36\,\mathrm{pA/cm^2}$ is expected for a field of $10\,\mathrm{V/\mu m}$. For deviating barrier height the current increases exponentially. The current–voltage characteristic of an InGaAs:Fe MSM photodetector is shown in Fig. 22.25 for a dark environment and various illumination levels.

The time-dependent response of a MSM photodetector depends on the drift time of the carriers, i.e. the time that a created electron and hole need to arrive at their respective contacts. In Fig. 22.26 a simulation is shown for a MSM detector. The current has two components, a fast one due to the electrons and a slow one due

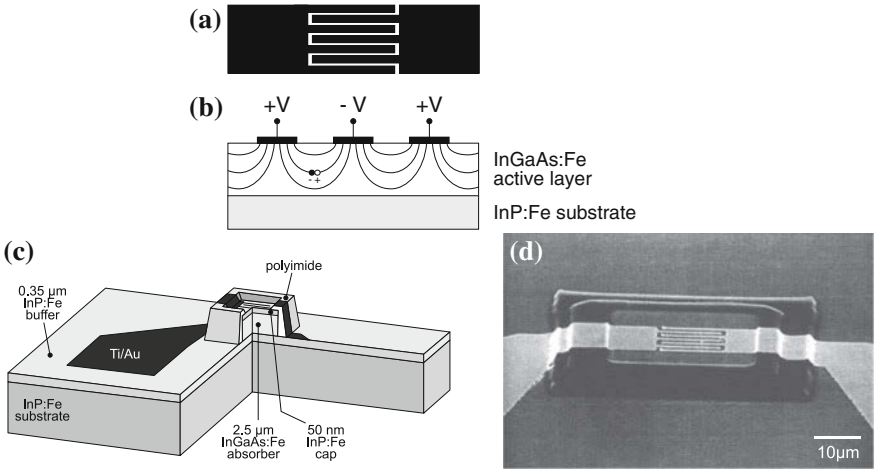


Fig. 22.24 Scheme of MSM photodetector with interdigital contacts in (**a**) plan view and (**b**) cross section. In part (**b**), the electric field lines are shown schematically together with an electron–hole pair ready to be separated. (**c**) Scheme of a MSM mesa structure, (**d**) SEM image of an InGaAs/InP MSM mesa photodetector. Parts (**c**) and (**d**) adapted from [1552]

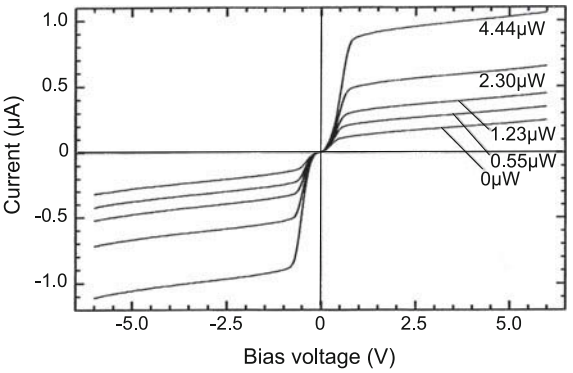


Fig. 22.25 dc I–V characteristic of an InGaAs/InP MSM photodetector (InP:Fe/InGaAs:Fe/InP:Fe, finger separation 1 μm, $\lambda = 1.3\,\mu$m) under illumination for dark environment (0 μW) and various illumination levels as labeled. Adapted from [1552]

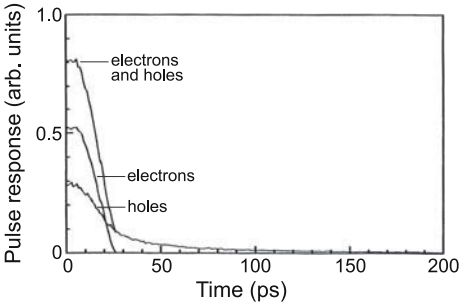


Fig. 22.26 Simulation of the time-dependent response of an InGaAs:Fe MSM photodetector to a short light pulse. Adapted from [1552]

to the holes that have the lower mobility and smaller drift saturation velocity. A similar dependence is found in experiment (Fig. 22.27a). For longer wavelengths the detector is slower since they penetrate deeper into the material and thus the charge carriers have a longer path to the contacts (cf. scheme in Fig. 22.24b). An important role is played by the finger separation; smaller finger separation ensures a more rapid carrier collection (Fig. 22.27b). In [1553] a bandwidth of 300 GHz was demonstrated for 100 nm/100 nm finger width and separation for LT-GaAs[8] and bulk GaAs, limited by the *RC* time constant. For 300 nm/300 nm fingers and a LT-GaAs a bandwidth of 510 GHz (pulsewidth of 0.87 ps) was reported, which is faster than the intrinsic transit time (1.1 ps) and not limited by the *RC* time constant (expected pulse width 0.52 ps), due to the recombination time (estimated to be 0.2 ps).

22.3.6 Avalanche Photodiodes

In an avalanche photodiode (APD) intrinsic amplification due to carrier multiplication (through impact ionization) in a region with high electric field is used to increase the photocurrent. The field is generated by a high reverse bias in the diode. In an ideal

[8] LT: grown at *low temperature*, i.e. containing many defects that reduce the carrier lifetime.

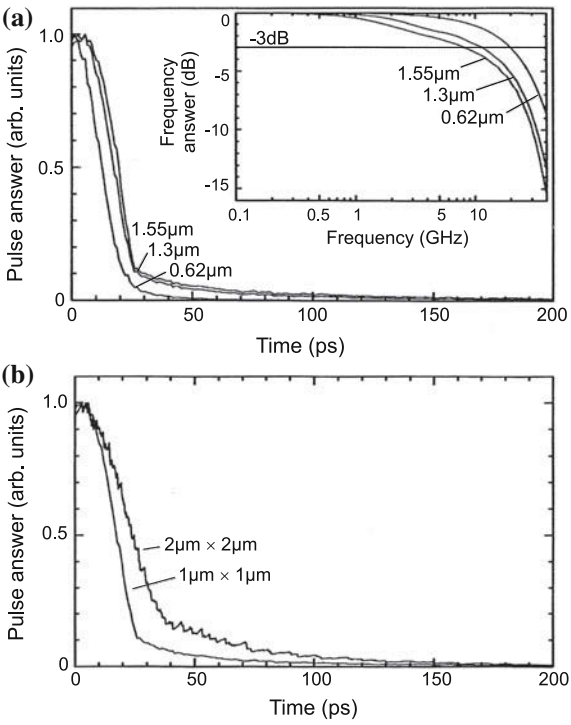


Fig. 22.27 (**a**) Experimental time-dependent response of an InGaAs:Fe MSM photodetector to a short light pulse for three different wavelengths, *inset* shows the frequency response from a Fourier transform. (**b**) Response of the MSM for two different finger widths and separations (both 1 or 2 μm, respectively), InGaAs layer thickness 2 μm, $\lambda = 1.3\,\mu$m and bias voltage 10 V. Adapted from [1552]

APD only one type of carrier is multiplied, resulting in the lowest noise. If electrons are injected into the field region at $x = 0$ (Fig. 22.28a), the multiplication factor for electrons is

$$M_\mathrm{n} = \exp(\alpha_\mathrm{n}\, w), \tag{22.49}$$

for $\alpha_\mathrm{p} = 0$. Typically, both carrier types suffer multiplication. If the electron and hole impact ionization coefficients are the same ($\alpha_\mathrm{n} = \alpha_\mathrm{p} = \alpha$), the multiplication factor for electrons and holes M is given by

$$M = \frac{1}{1 - \alpha\, w}. \tag{22.50}$$

The rms value of the current noise is the same as in the case of the pn-diode (22.25), only that now the gain M is added

$$\langle i_\mathrm{S}^2 \rangle = 2\, e\, \left(I_\mathrm{ph} + I_\mathrm{B} + I_\mathrm{D}\right)\, \langle M^2 \rangle\, B. \tag{22.51}$$

The term $\langle M^2 \rangle$ is written as $\langle M \rangle^2\, F(M)$ with $F(M) = \langle M^2 \rangle / \langle M \rangle^2$ being the excess noise factor that describes the additional noise introduced by the random nature of the impact ionization. For multiplication started with electron injection, it is given

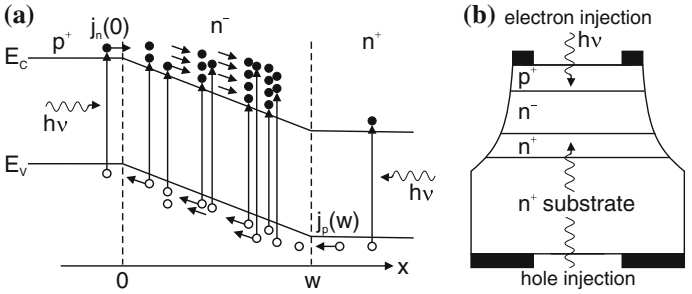


Fig. 22.28 Schematic band structure (**a**) and schematic device setup (**b**) of an avalanche photodiode (APD). From [1047]

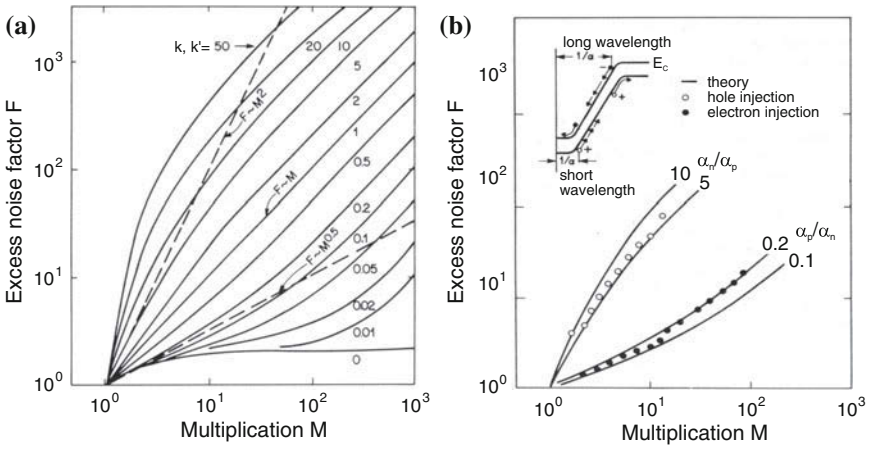


Fig. 22.29 (**a**) Excess noise factor for various values of the ratio of ionization coefficients k or k'. Adapted from [1554]. (**b**) Experimental results for F for a Si APD with 0.1 μA primary current. The *empty* (*full*) symbols are for short (long) wavelengths [primary hole (electron) current]. The *inset* shows the schematic band diagram of the np-diode under reverse bias. Adapted from [1555]

by [1554]

$$F(M) = k\,M + (1-k)\left(2 - \frac{1}{M}\right), \tag{22.52}$$

with $k = \alpha_\mathrm{p}/\alpha_\mathrm{n}$. For hole injection starting the multiplication (22.52) holds with k substituted by $k' = \alpha_\mathrm{n}/\alpha_\mathrm{p}$. In Fig. 22.29a the excess noise factor is shown versus the average multiplication for various values of k and k'.

Experimental data are shown in Fig. 22.29b for a Si APD. For short wavelengths absorption is preferential at the surface (n-region) and we have the case of hole injection. The data for the excess noise factor are fairly well fit with $k' \approx 5$. For longer wavelengths, the data for electron injection are fit by $k \approx 0.2 = 1/k'$.

For a fully modulated signal the signal-to-noise ratio is given by

$$S/N = \frac{(e\,\eta\,P_{\mathrm{opt}}/h\nu)^2/2}{2\,e\,\left(I_{\mathrm{ph}} + I_{\mathrm{B}} + I_{\mathrm{D}}\right)\,F(M)\,B + 4\,kT\,B/(R_{\mathrm{eq}}\,M^2)}. \tag{22.53}$$

If S/N is limited by thermal noise, the APD concept leads to a drastic improvement of noise.

The APD can used for single photon detection; then the variation in pulse height (with reasonable limits) plays no role for the count rate. The number of dark counts can be reduced by cooling the APD. Using constant fraction triggering, also the arrival time of the photon can be determined, allowing time resolution typically in the 100 ps regime.

A particular APD structure is known as a *solid-state multiplier*. It has separate absorption and amplification regions (SAM structure). In the low-field region the light is absorbed. One type of carrier is transported with the drift field E_{d} to the multiplication region in which a large field E_{m} is present and multiplication occurs. In Fig. 22.30a a homo-APD with SAM structure is shown. Regions with different electric field are created by a special doping profile.[9] A π-p-π structure leads to regions with homogeneous low and high field strengths.

The performance of a commercial silicon APD is depicted in Fig. 22.31. With increasing reverse bias, both the dark current and the multiplication factor increase. Before breakdown at about 77 V, the best ratio of M and I_{dark} is found. This particular diode has a typical excess noise factor of $F = 2$ for $M = 100$. Using (22.52), this shows that only one kind of carriers is multiplied.

In the case of a heterostructure-APD with SAM structure (Fig. 22.30b) absorption (of light with sufficiently long wavelength with an energy smaller the InP band gap) takes place only in the InGaAs layer. Since no light is absorbed in the multiplication region, the device functions similarly for front and back illumination. In [1557] a multi-stage InGaAs-based APD on InP is described that is optimized for electron multiplication in ten subsequent gain sections; a total gain of 10^3 is achieved with an excess noise factor F of about 40, belonging to an effective ionization ratio of $k = 0.036$.

22.3.7 Traveling-Wave Photodetectors

In a standard photodetector there was a tradeoff between the thickness of the absorption layer and the speed of the detector. In a traveling-wave photodetector the light absorption occurs in a waveguide such that for sufficient length L all incident light is absorbed. Complete absorption is achieved ('long' waveguide) if $L \gg (\Gamma\alpha)^{-1}$, α being the absorption coefficient and $\Gamma \leq 1$ being the optical confinement factor,

[9] Employing Poisson's equation $\partial(\epsilon_{\mathrm{s}}(x)E(x))/\partial x = \rho(x)$.

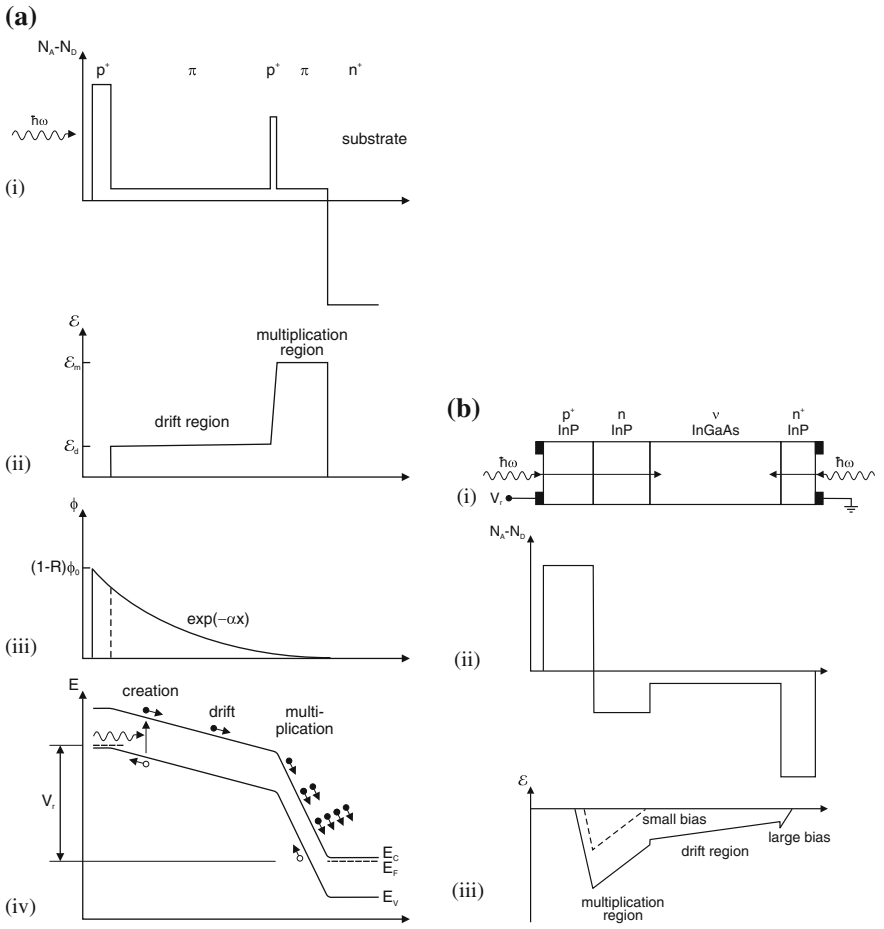


Fig. 22.30 (**a**) Homo-APD with SAM structure. (**i**) doping profile, (**ii**) electric field, (**iii**) photon flux or electron–hole pair generation rate and (**iv**) schematic band diagram under reverse voltage V_r with charge-carrier transport. The multiplication is for $\alpha_n \gg \alpha_p$. Adapted from [500]. (**b**) (**i**) Scheme of an InP/InGaAs hetero-APD with SAM structure, (**ii**) doping profile and (**iii**) electric field for small (*dashed line*) and large (*solid line*) reverse bias V_r. Adapted from [1047]

the geometrical overlap of the optical mode with the cross section of the absorbing medium (cf. also Sect. 23.4.4).

The electrical connections are designed along this waveguide on the sides (coplanar layout, Fig. 22.32). The bandwidth limitation due to a *RC* time constant is now replaced by the velocity match of the light wave $v_{opt} = c/n$ and the traveling electric wave in the contact lines $v_{el} \approx 1/\sqrt{LC}$. While the two waves travel along the waveguide, energy is transferred from the light wave to the electric wave. The 3 dB

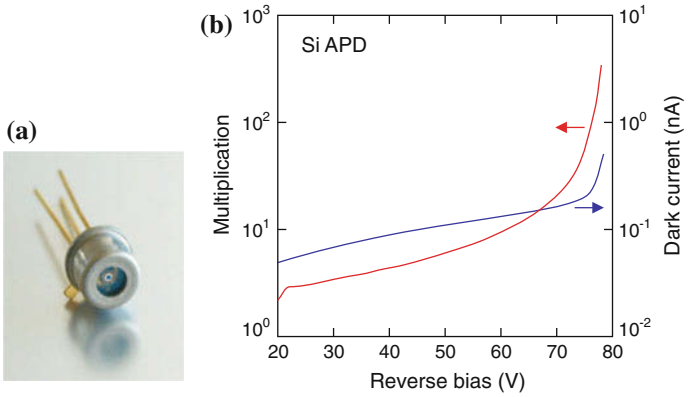


Fig. 22.31 (**a**) TO casing of Si APD. (**b**) Dark current (*blue*) (active area 0.2 mm^2) and multiplication factor (*red*) as a function of reverse bias (at 23 °C). Adapted from [1556]

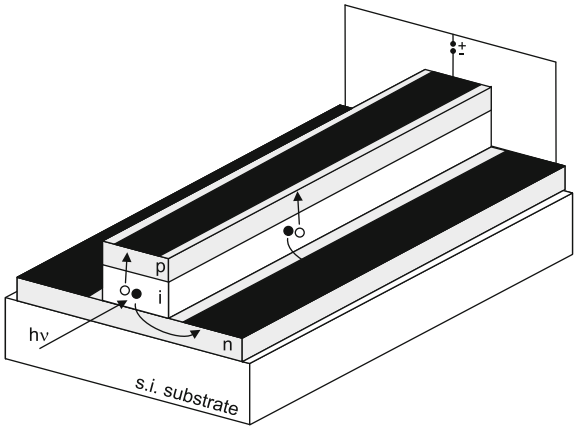


Fig. 22.32 Scheme of a traveling-wave photodetector with pin structure and coplanar contacts

bandwidth due to velocity mismatch B_{vm} (for impedance- matched, long waveguides) is given by [1558]

$$B_{\mathrm{vm}} = \frac{\Gamma\,\alpha}{2\pi}\,\frac{v_{\mathrm{opt}}\,v_{\mathrm{el}}}{v_{\mathrm{opt}} - v_{\mathrm{el}}}. \tag{22.54}$$

For a MSM structure, whose electrode separation has been designed with a self-aligned process (without extensive effort in lateral patterning) by an etch depth of a few 100 nm (Fig. 22.33), 3 dB cutoff frequencies beyond 500 GHz have been achieved (Fig. 22.34). The quantum efficiency of this detector was still 8.1 %.

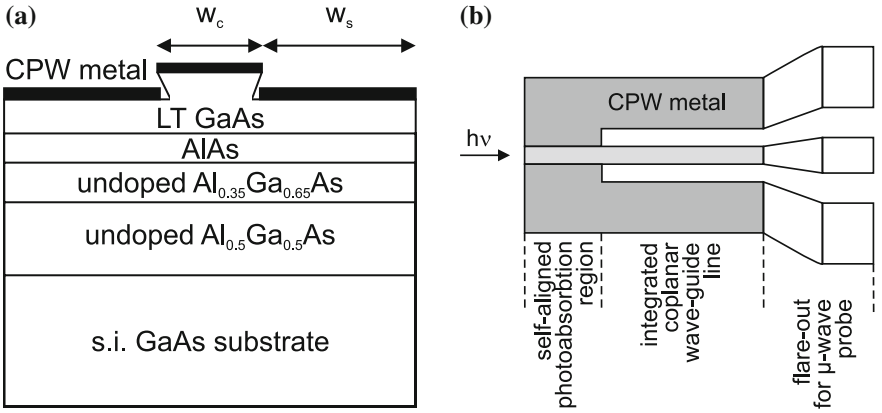


Fig. 22.33 Scheme of a MSM traveling-wave photodetector in (**a**) cross section and (**b**) plan view. Adapted from [1559]

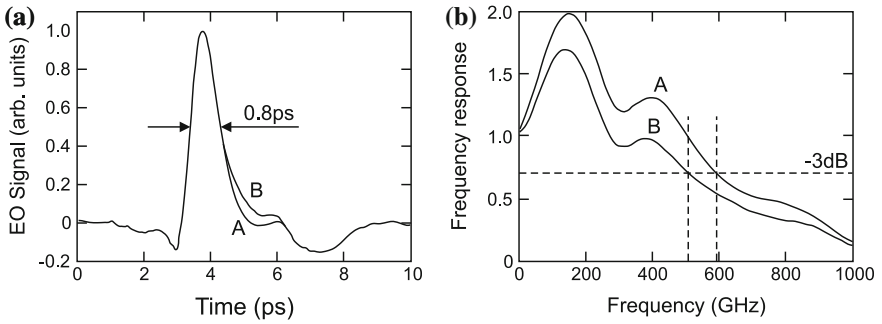


Fig. 22.34 (**a**) Pulse response (FWHM = 0.8 ps) and (**b**) frequency response (Fourier transform of time response) of a MSM traveling-wave photodetector (bias 5 V) for various illumination intensities, A: 1 mW, B: 2.2 mW. Adapted from [1559]

22.3.8 Charge Coupled Devices

The concept of the charge coupled device (CCD), an array of connected photodetectors serving as an image sensor, was devised by W.S. Boyle and G.E. Smith [1560] (Fig. 22.35) and realized [1561]. As textbook for further details [1562, 1563] may serve.

A MIS diode (mostly a silicon-based MOS diode) can be designed as a light detector. The diode is operated in deep depletion. When a large reverse voltage is applied, initially a depletion layer is formed and the bands are strongly bent as shown in Fig. 22.36b. We note that in this situation the semiconductor is not in thermodynamic equilibrium (as it is in Fig. 22.36d) when the quasi-Fermi level is constant throughout the semiconductor. The inversion charge has yet to build up.

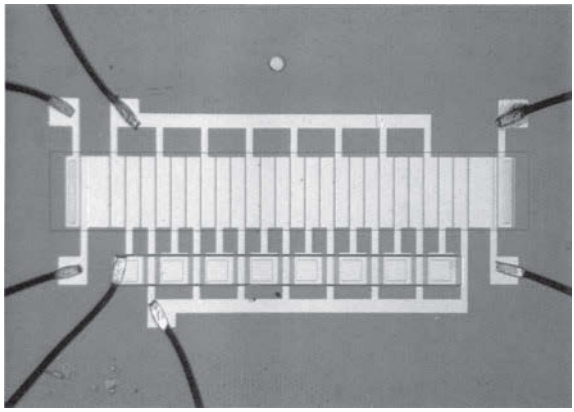


Fig. 22.35 First 8-bit charge coupled device (1970). The chip (size: $1.5 \times 2.5\,\mathrm{mm}^2$) consists of 24 closely packed MOS capacitors (*narrow rectangles* in the center grid). The *thick rectangles* at either end of the grid are input/output terminals

There are three mechanisms to generate the inversion charge. (a) generation–recombination, (b) diffusion from the depletion-layer boundary and (c) carrier generation by light absorption. Mechanisms (a) and (b) represent dark currents for the photodetector. The conductivity due to these two processes is shown in Fig. 22.37 and slowly builds up the inversion charge. Two temperature regimes are obvious; at low temperatures the generation dominates ($\propto n_\mathrm{i} \propto \exp\left(-E_\mathrm{g}/2kT\right)$), at high temperatures the diffusion ($\propto n_\mathrm{i}^2 \propto \exp\left(-E_\mathrm{g}/kT\right)$). The latter process can be strongly suppressed by cooling the device.

The gate voltage V_G and the surface potential Ψ_s are related to each other via

$$V_\mathrm{G} - V_\mathrm{FB} = V_\mathrm{i} + \Psi_\mathrm{s} = \frac{e\,N_\mathrm{A}\,w}{C_\mathrm{i}} + \frac{eN_\mathrm{A}w^2}{2\epsilon_\mathrm{s}}, \tag{22.55}$$

where w is the width of the depletion layer. w will be larger than w_m in thermodynamic equilibrium. The first term in the sum is $|Q_\mathrm{s}|/C_\mathrm{i}$ and the second is obtained by integrating the Poisson equation for the constant charge density $-eN_\mathrm{A}$ across the depletion layer. The elimination of w yields

$$V_\mathrm{G} - V_\mathrm{FB} = \Psi_\mathrm{s} + \frac{1}{C_\mathrm{i}}\sqrt{2e\epsilon_\mathrm{s}N_\mathrm{A}\Psi_\mathrm{s}}. \tag{22.56}$$

If light is absorbed in the depletion layer (process (c)), the hole (for p-Si) drifts towards the bulk material. The electron is stored as part of the signal charge Q_sig close to the oxide semiconductor interface (Fig. 22.36b).

$$V_\mathrm{G} - V_\mathrm{FB} = \frac{Q_\mathrm{sig}}{C_\mathrm{i}} + \frac{eN_\mathrm{A}w}{C_\mathrm{i}} + \Psi_\mathrm{s}. \tag{22.57}$$

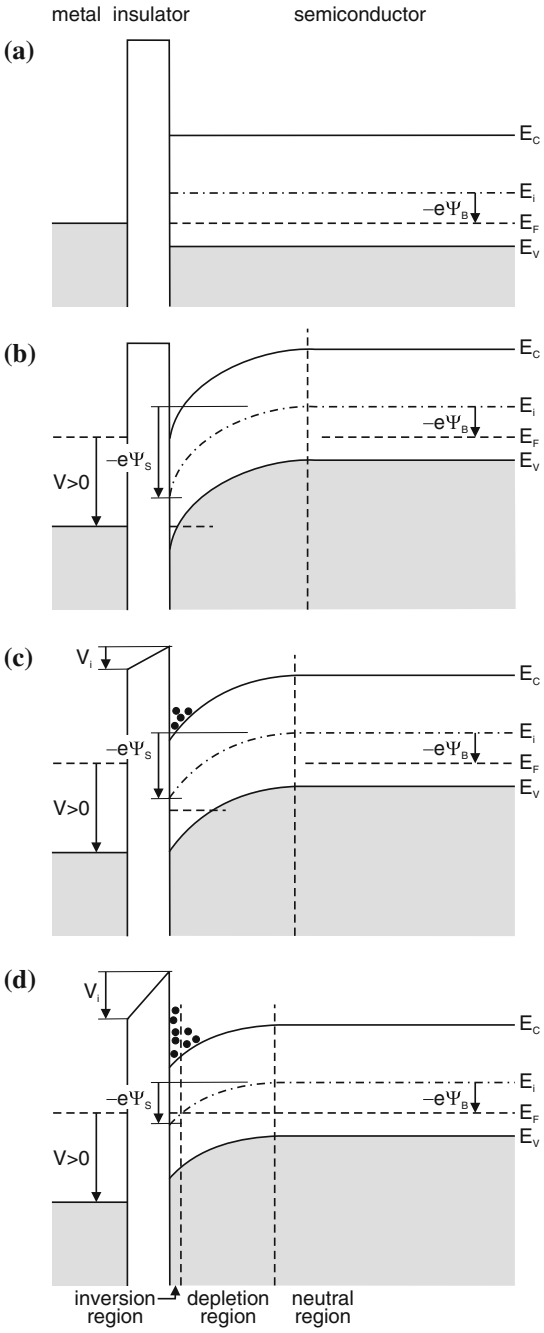


Fig. 22.36 Ideal MIS-diode (with p-type semiconductor) as photodetector (principle of a CCD pixel). (**a**) Without bias (see Fig. 21.29b). (**b**) Immediately after an external (reversely poled) voltage $V > 0$ has been applied, the surface potential is $\Psi_s = V$ and no charges have moved yet. (**c**) Strong depletion (still not in thermodynamic equilibrium) with signal charge and reduced surface potential $\Psi_s < V$. (**d**) The semiconductor in equilibrium (E_F is constant) with depletion and inversion layer (see Fig. 21.33). For all diagrams, $V = V - \mathrm{i} + \Psi_s$

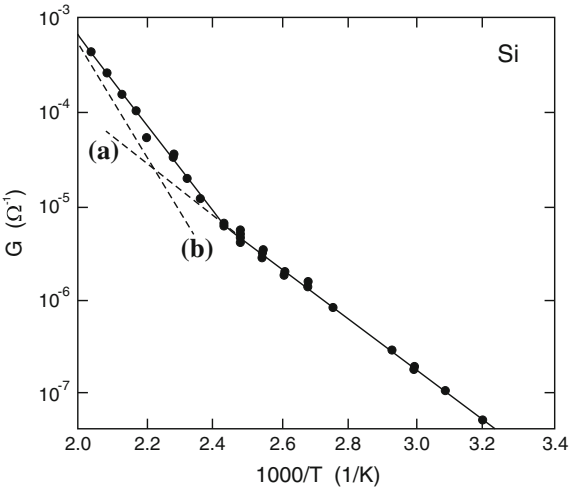


Fig. 22.37 Conductivity of a n-Si/SiO_2-diode as a function of temperature ($1/T$). The slope of the *dashed lines* is (a) ∼0.56 eV ($\approx E_g/2$) and (b) ∼1.17 eV ($\approx E_g$). Adapted from [1564]

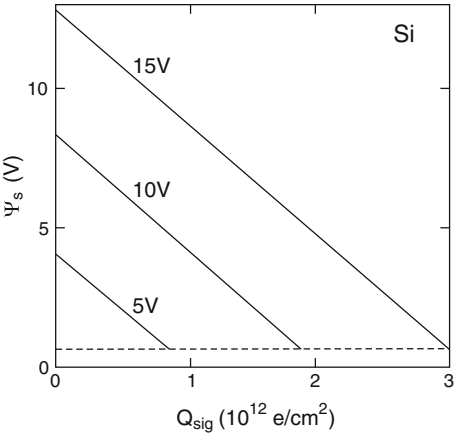


Fig. 22.38 Surface potential as a function of the signal charge Q_{sig} for various values of the bias $V_G - V_{FB}$ as labeled for a SiO_2/p-Si diode with $N_A = 10^{15}\,cm^{-3}$ and an oxide thickness of 100 nm. The *dashed line* represents the limit for inversion given by $\Psi_s = 2\Psi_B \approx 0.6\,V$. Adapted from [1565]

As a consequence of the increase in signal charge the potential well becomes shallower (22.57). For each gate voltage there is a maximum charge (well capacity). The maximum signal charge is reached for $\Psi_s \approx 2\Psi_B$ (Fig. 22.38).

In a charge coupled device (CCD) many light-sensitive MIS diodes, as described above, are fabricated in matrix form to create an image sensor. Upon application of a gate voltage they accumulate charge depending on the local exposure to light. The read out of this charge occurs by shifting the charge through the array to a read-out circuit. Therefore charge is transferred from one pixel to the next. Several schemes have been developed for this task. The three-phase clocking is shown schematically in Fig. 22.39. Other clocking schemes involve four, two or only one electrode per pixel [1566].

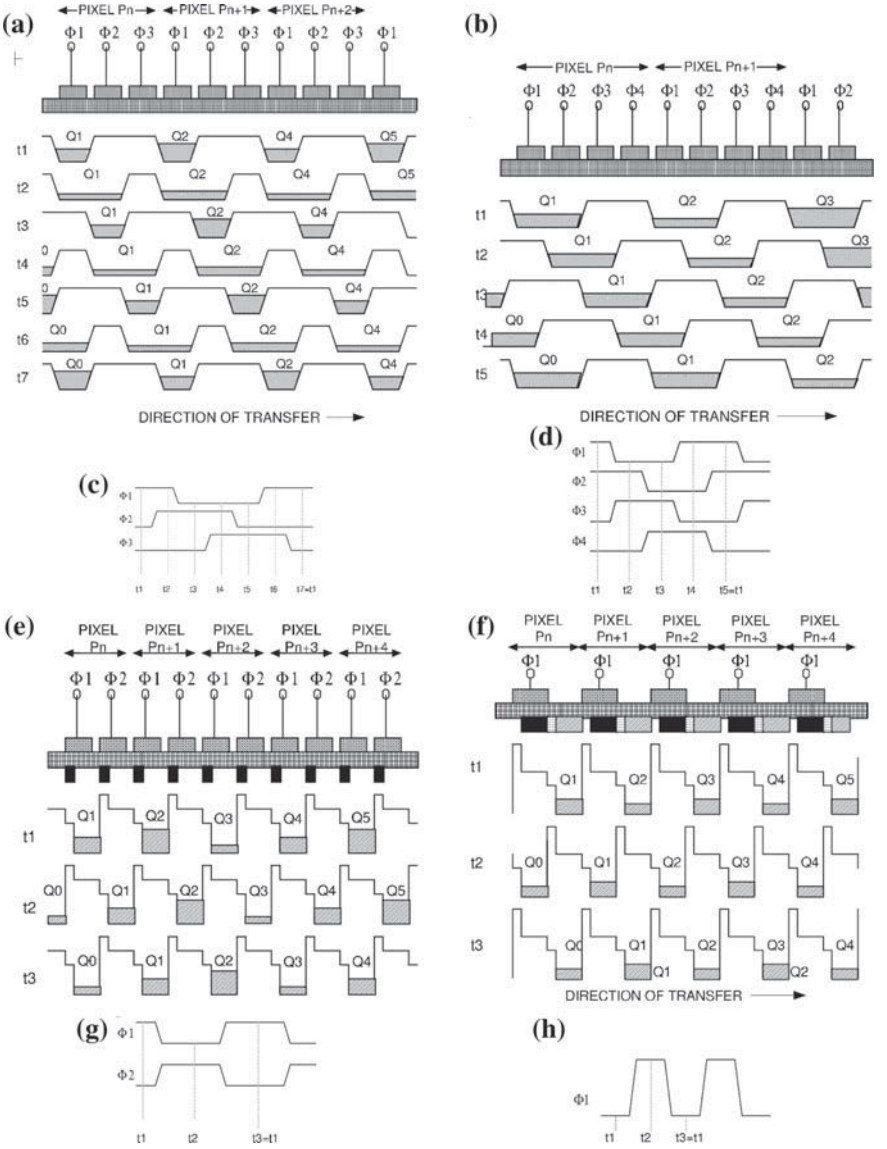


Fig. 22.39 (**a**) Three-phase CCD. Each pixel has three electrodes that can be switched independently (phases 1–3). (**b**, **e**, **f**) Schematic of CCDs with four, two or one phase, respectively. (**c**) (t_1) Charge accumulated after light exposure. A lateral potential well is formed along the row of pixels by the voltages at the three electrodes, e.g. $P_1 = P_3 = 5\,\mathrm{V}$, $P_2 = 10$,V. (t_2–t_7) transfer of charge, (t_7) has the same voltages as (t_1), the charge has been moved one pixel to the right. (**d**, **g**, **h**) Timing schemes for 4-, 2- and 1-phase CCDs, respectively. From [1566]

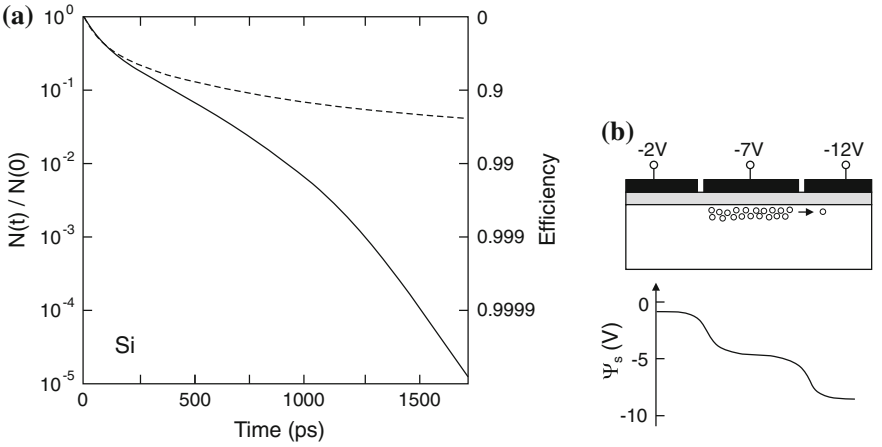


Fig. 22.40 (**a**) Efficiency of charge transfer with (*solid line*) and without (*dashed line*) the effect of the fringing field. (**b**) Schematic plot of the CCD electrodes and bias with the fringing field for a three-electrode CCD, oxide thickness 200 nm, doping $N_D = 10^{15}\,\mathrm{cm}^{-3}$. The electrode pitch is 4 μm and the gap between electrodes is 200 nm wide. Adapted from [1567]

Since the CCD sensor has many pixels (e.g. up to 4096) along a line, the charge transfer must be highly efficient. The transfer of charge carriers occurs via thermal (regular) diffusion, self-induced drift and the effect of the fringing field (inset of Fig. 22.40). The time constant with which the charge carriers move due to diffusion (in a p-type semiconductor) is

$$\tau_{\mathrm{th}} = \frac{4L^2}{\pi^2 D_{\mathrm{n}}}, \tag{22.58}$$

where L is the length of the electrode. For a sufficiently large charge packet the self-induced drift due to Coulomb repulsion is important. The decay of charge is then given by

$$\frac{Q(t)}{Q(0)} = \frac{t_0}{t + t_0}, \tag{22.59}$$

with $t_0 = \pi L^3 W_{\mathrm{e}} C_{\mathrm{i}}/(2\mu_{\mathrm{n}} Q(0))$. W_{e} is the width of the electrode. This dependence is shown as a dashed line in Fig. 22.40a. The last electrons are efficiently transferred by the drift induced by the fringing field of the electrodes (solid line in Fig. 22.40a). The origin of the fringing field is schematically shown in Fig. 22.40b; the minimum fringing field shown is 2×10^3 V/cm. In about 1–2 ns practically all $(1 - 10^{-5})$ charges are transferred. This enables clock rates of several 10 MHz.

For the clocking of the CCD the lateral variation of potential depth with the applied gate voltage is used. In Fig. 22.41 it is shown how a lateral variation of doping or oxide thickness creates a lateral potential well. Such structures are used to confine the row of pixels against the neighboring rows (channel stops, Fig. 22.42). In order

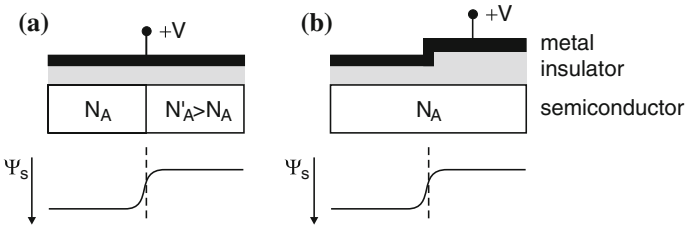


Fig. 22.41 Creation of a lateral potential well (barrier) in a MIS structure with (**a**) varying doping via diffusion or implantation and (**b**) varying (stepped) oxide thickness. *Upper row* shows schematic geometry, *lower row* depicts schematic lateral variation of the surface potential

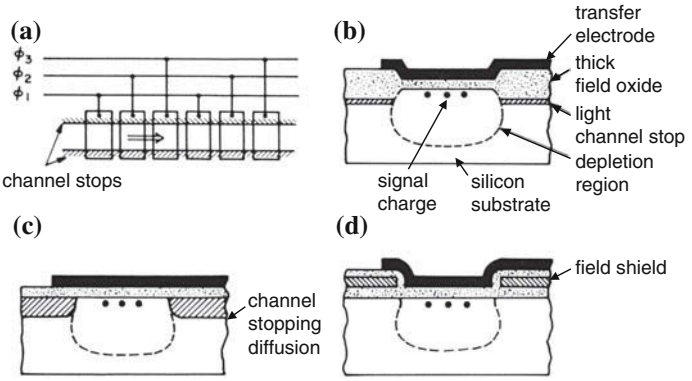


Fig. 22.42 (**a**) Schematic image of channel isolation. Cross section for channel isolation by (**b**) variation of oxide thickness, (**c**) highly doped region and (**d**) field effect. Adapted from [1568]

to avoid carrier loss at the interface between the oxide and the semiconductor a buried-channel structure is used (Fig. 22.43).

For front illumination parts of the contact electrodes shield the active area of the device. Higher sensitivity (in particular in the UV) is achieved for back illumination. For this purpose the chip is thinned (polished). This process is expensive and makes the chip mechanically less stable. For red/infrared wavelengths typically interference fringes occur for such thinned chips due to the small thickness. An increase in efficiency for front illumination can be achieved with an onchip microlens (Fig. 22.44).

For color imaging the CCD is covered with a three-color Bayer mask [1571] (Fig. 22.45a). On average there are one blue and one red pixel and two green pixels since green is the most prominent color in typical lighting situations. Thus each pixel delivers monochromatic information; RGB images are generated using suitable image software. Alternatives in high-end products are the use of a beam splitter, static color filters and three CCD chips, one for each color (Fig. 22.45b), or the time-sequential recording of three monochromatic images using one CCD chip and a rotating color-filter wheel (Fig. 22.45c). Another method is to shift

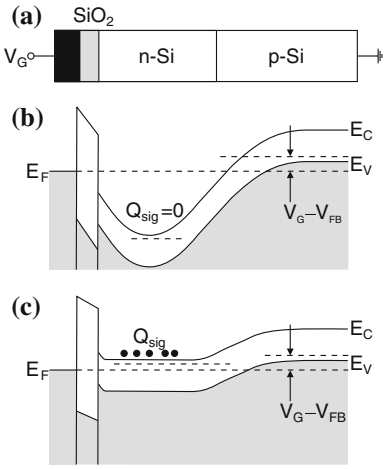


Fig. 22.43 (**a**) Schematic layers of MIS diode with buried-channel structure. Band diagram (**b**) after application of reverse voltage V_G and (**c**) with signal charge Q_{sig}. Adapted from [1569]

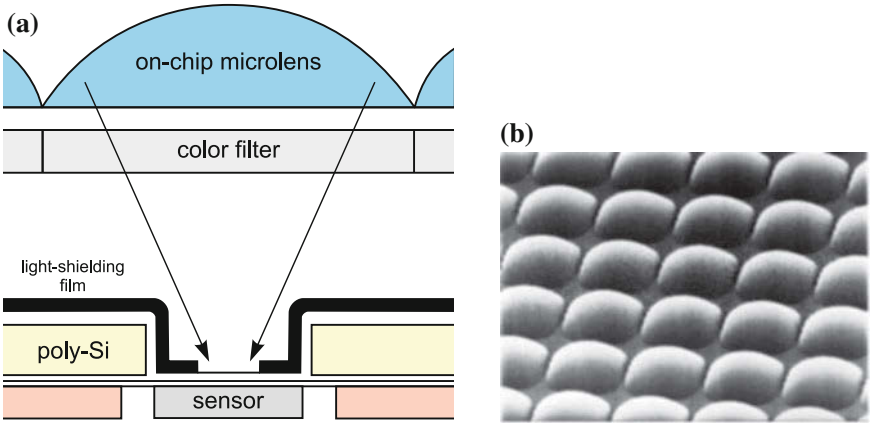


Fig. 22.44 (**a**) Scheme for enhancement of CCD efficiency for front illumination by application of an onchip microlens. (**b**) SEM image of an array of such microlenses. From [1570]

(by piezoelectric actors) the CCD by half or full pixels and take sequential exposures, increasing spatial and color resolution for still objects.

22.3.9 Photodiode Arrays

An array of photodiodes is also suitable to create an image sensor. During illumination each diode charges a capacitor that is read out with suitable electronics. Based on

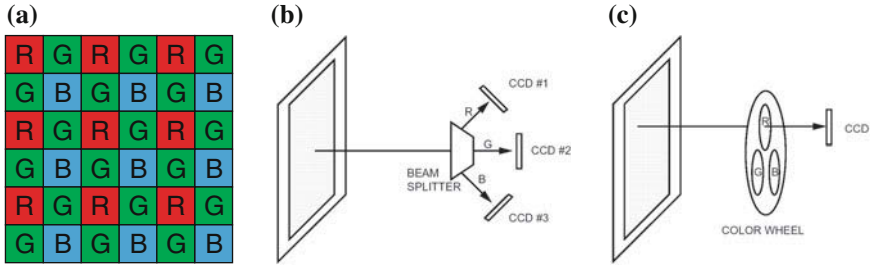


Fig. 22.45 (**a**) Arrangement of colors in a CCD Bayer [1571] color filter ('R': *red*, 'G': *green*, 'B': *blue*). Color splitting with (**b**) static color filters and (**c**) rotating color wheel. Parts (**b**) and (**c**) from [1566]

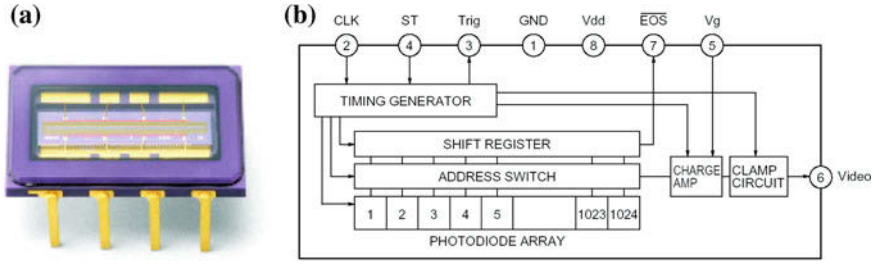


Fig. 22.46 (**a**) CMOS linear array sensor in a 8-pin package. (**b**) Block diagram, the built-in timing generator allows operation only with start and clock pulse inputs. Reprinted with permission from [1572]

CMOS technology (cf. Sect. 24.5.4) cheap image sensors can be made that show currently, however, inferior performance to CCDs. The built-in electronics allows simple outward connections (Fig. 22.46).

The three-color CCD image sensor did not offer RGB color information at each pixel. Therefore the spatial resolution of a color image is not directly given by the pixel distance. This is not a very dramatic drawback since human vision is more sensitive to intensity contrast than color contrast. However, RGB color information for each pixel would be desirable, giving higher resolution, in particular in professional photography. Such a sensor has been fabricated employing the wavelength dependence of the silicon absorption coefficient (Fig. 22.13). Blue light has the shortest and red light the largest penetration depth. By stacking three photodiodes on top of each other (Fig. 22.47) photocurrents at different penetration depth are recorded that can be used to generate a RGB value for each pixel.

In Fig. 22.48a a 16-channel array of silicon avalanche photodiodes is shown. It features a quantum efficiency of $>80\,\%$ between 760 and 910 nm. The pixel size is $648 \times 208\,\mu m^2$ on a $320\,\mu m$ pitch. The gain is 100 and the rise time 2 ns.

The InGaAs photodiode array in Fig. 22.48b is hybridized with CMOS read-out integrated circuits. It is useful for detection in the spectral range 0.8–1.7 µm. The

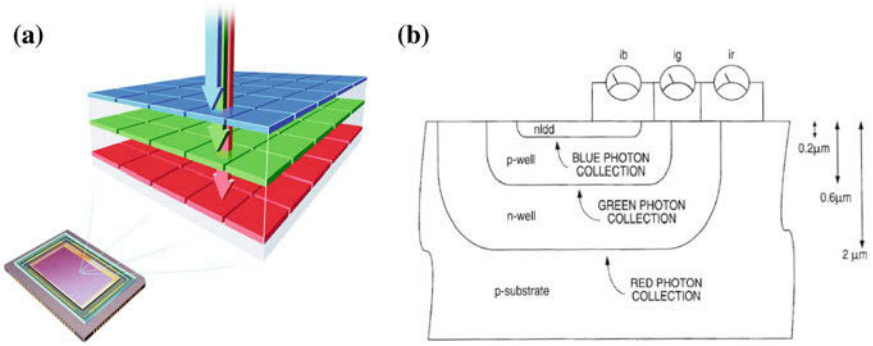


Fig. 22.47 (**a**) Scheme of image sensor with depth-dependent light detection. From [1573]. (**b**) Schematic layer sequence for three-color pixel. i_b, i_g and i_r denote the photocurrents for *blue*, *green* and *red* light, respectively. Adapted from [1574]

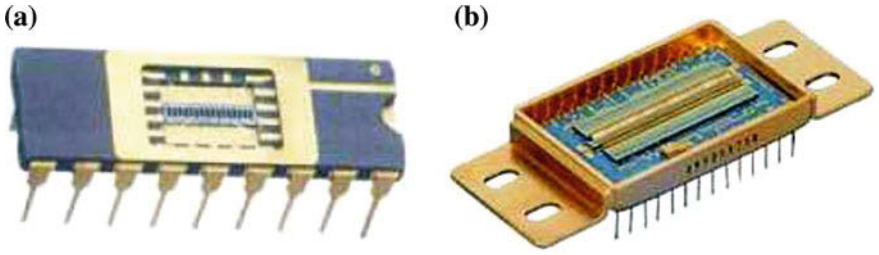


Fig. 22.48 (**a**) Array of 16 silicon APDs. From [1575]. (**b**) 1024-pixel InGaAs photodiode array. From [1576]

asymmetric diode size of $25 \times 500\,\mu\text{m}$ is designed for use in the focal plane of a monochromator.

Another special type of photodiode array is the four-quadrant detector. A light beam generates four photocurrents I_a, I_b, I_c, I_d of the respective parts (Fig. 22.49a). A beam deviation in the horizontal or vertical direction can be detected from the (signed) signals $(I_a + I_d) - (I_b + I_c)$ or $(I_a + I_b) - (I_c + I_d)$, respectively. We note that these signals can also be normalized to the total beam intensity $I_a + I_b + I_c + I_d$.

22.4 Solar Cells

Solar cells are light detectors, mostly photodiodes, that are optimized for the (large-area) conversion of solar radiation (light) into electrical energy. A 1993 review of the historic development of photovoltaics is given in [1578]. The latest data on solar cell efficiencies are compiled in the Solar Cell Efficiency Tables [1579].

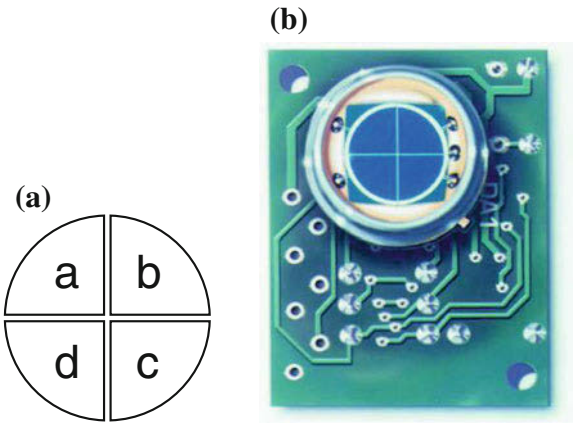


Fig. 22.49 (**a**) Scheme of four-quadrant photodetector with sections 'a', 'b', 'c' and 'd', (**b**) image of four-quadrant silicon photodetector with circuit board. Part (**b**) from [1577]

22.4.1 Solar Radiation

The sun has three major zones, the core with a temperature of 1.56×10^7 K and a density of 100 g/cm^3 in which 40 % of the mass is concentrated and 90 % of the energy is generated, the convective zone with a temperature of 1.3×10^5 K and a density of 0.07 g/cm^3, and the photosphere with a temperature of 5800 K and low density ($\sim 10^{-8}$ g/cm^3). The radius is 6.96×10^8 m and is about 100 times larger than that of the earth (6.38×10^6 m). The distance sun–earth is 1.496×10^{11} m. The angle under which the sun disk appears on earth is 0.54°. An energy density of 1367 ± 7 W/m^2 arrives at the earth in front of its atmosphere.

This value and the according spectrum of the sun's emission, which is similar to a blackbody with temperature 5800 K (Fig. 22.50), is termed air mass zero (AM0). The total energy that reaches the earth from the sun is 1.8×10^{17} W per year. This value is 10^4 times the world's primary energy need.

Air mass zero (AM0) is important for solar cells in satellites. When the solar spectrum passes the earth's atmosphere, it is changed with regard to its shape and the total energy density due to gas absorption (ozone, water, CO_2, ...). Depending on the meridian of the sun γ (Fig. 22.51), the spectrum on the surface of the earth is termed AMx with $x = 1/\sin\gamma$. In spring and fall (March 21st and September 21st), Leipzig (51°42′N latitude) has about AM1.61. At the summer (June 21st) and winter (December 21st) solstices the air mass in Leipzig is AM1.13 ($\gamma = 61.8°$) and AM3.91 ($\gamma = 14.8°$), respectively. Additionally, the duration of sunshine and thus the light power density is regionally different across the earth due to climate and weather (Fig. 22.52). For AM1.5, the incident power density is 844 W/m^2.

The global radiation reaching a photovoltaic cell has three components: (i) the direct radiation, (ii) diffuse radiation and (iii) reflected radiation. The relative amounts

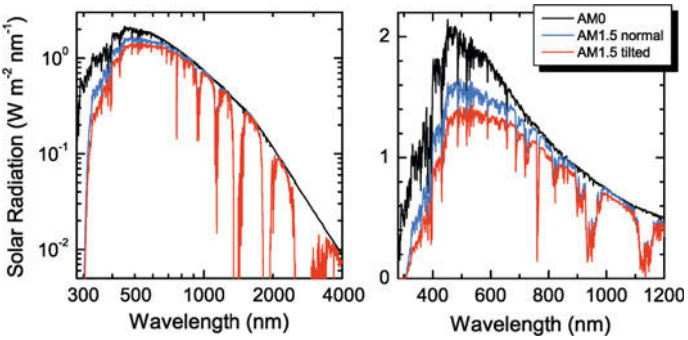


Fig. 22.50 Solar spectra (power per area and wavelength interval) for AM0 (*black line*, extraterrestrial irradiance) and AM1.5 (sun at 41.8° elevation above horizon) for direct normal irradiance (*blue line*) and global total irradiance (*red line*) on a sun facing surface (tilted 37° towards the equator). Left (right) graph in log–log (linear) scales

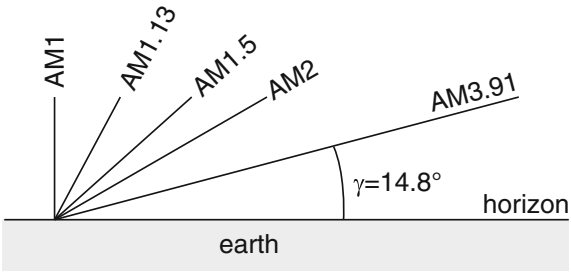


Fig. 22.51 Schematic path of sunlight through the atmosphere and definition of the air mass AM*x*

and their spectra depend on details such as the climate (e.g. humidity) or the environment (e.g. outdoors vs. urban).

22.4.2 Ideal Solar Cells

When a solar cell made from a semiconductor with a band gap E_g is irradiated by the sun, only photons with $h\nu > E_g$ contribute to the photocurrent and the output power. The I–V characteristic under illumination (Fig. 22.53) is given by

$$I = I_s \left[\exp(\beta V) - 1\right] - I_L, \tag{22.60}$$

with I_L being the current due to generation of excess carriers by the absorption of the sunlight. Assuming a simple n^+p-diode solar cell model, the current consists of two components: the depletion layer current j_{DL} from carriers absorbed in the depletion

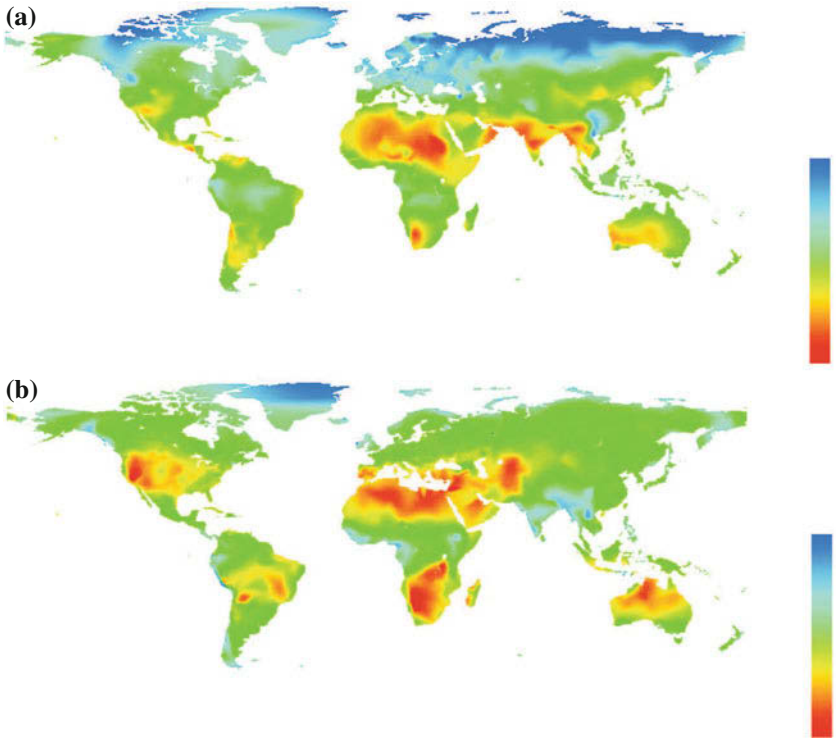


Fig. 22.52 Global sunshine distribution in (**a**) January and (**b**) July. The sunshine fraction is the actual number of bright sunshine hours over the potential number, and is thus expressed as a percentage figure. The color scale reaches from 0 to 100 %. The sunshine data are in a 0.5° grid based on data from [1580]

layer (field region) and the diffusion current j_D from absorption in the neutral region ($j = I/A$).

For the drift current out of the depletion layer of width w it can be assumed that it is collected fast and recombination plays no role. Thus (cmp. (22.31))

$$j_{DL}(\lambda) = e\, n_{ph}(\lambda)\,[1 - R(\lambda)]\,[1 - \exp(-\alpha(\lambda)\, w)], \tag{22.61}$$

where λ is the wavelength of the incident radiation, R the reflectance of the surface, α the absorption coefficient and $n_{ph}(\lambda)$ the photon flux (number of photons per area and time) at the given wavelength. For a solar spectrum, an integral needs to be performed over the spectral distribution:

$$j_{DL} = \int j_{DL}(\lambda)\, d\lambda, \tag{22.62}$$

the total photon flux being $n_{ph} = \int n_{ph}(\lambda)\, d\lambda$.

The diffusion current collected at the back contact is obtained from solving (10.71) (now for electrons in p-type material) with the appropriate boundary conditions (reversely bias depletion layer, $n_p(w) = 0$, $\Delta n_p(\infty) = n_p(\infty) - n_{p_0} = 0$ [1581]):

$$j_D(\lambda) = e\, n_{ph}(\lambda)\,[1 - R(\lambda)]\,\frac{\alpha\, L_n}{1 + \alpha\, L_n}\,\exp(-\alpha\, w) + e\, n_0\,\frac{D_n}{L_n}, \tag{22.63}$$

Dropping the wavelength dependence and neglecting the dark term, the usual formula is obtained,

$$j_L = e\, n_{ph}\,(1 - R)\left[1 - \frac{\exp(-\alpha\, w)}{1 + \alpha\, L_n}\right], \tag{22.64}$$

The last bracket represents the quantum efficiency [1582]. The model can be extended for taking into account a non-zero surface recombination velocity at the back contact at finite distance [1583].

Here a voltage *independent* photo-generated current I_L is assumed. If the diffusion length is small compared to the transport path, the carrier collection efficiency η_c becomes voltage dependent [1582]. The reduction of the diffusion potential for forward voltage decreases the carrier collection efficiency [1584], possibly to zero close to the built-in voltage.

The saturation current density is given by (21.128) and (21.129)

$$j_s = \frac{I_s}{A} = e\, N_C\, N_V \left(\frac{1}{N_A}\sqrt{\frac{D_n}{\tau_n}} + \frac{1}{N_D}\sqrt{\frac{D_p}{\tau_p}}\right)\exp\left(-\frac{E_g}{kT}\right), \tag{22.65}$$

with A being the cell area.

The voltage at $I = 0$ is termed the *open-circuit* voltage V_{oc}, the current at $V = 0$ is termed the *short-circuit* current $I_{sc} = I_L$ (Fig. 22.53). Only a part of the rectangle $I_{sc} \times V_{oc}$ can be used for power conversion. By choice of the load resistance R_L, the work point is set. At I_m and V_m, the generated power $P_m = I_m V_m$ is maximal. The filling factor F is defined as the ratio

$$F = \frac{I_m\, V_m}{I_{sc}\, V_{oc}}. \tag{22.66}$$

The open-circuit voltage is given by

$$V_{oc} = \frac{1}{\beta}\,\ln\left(\frac{I_L}{I_s} + 1\right) \cong \frac{1}{\beta}\,\ln\left(\frac{I_L}{I_s}\right) \tag{22.67}$$

and increases with increasing light power and decreasing dark current. The output power is

$$P = I\, V = I_s\, V\ (\exp(\beta V) - 1) - I_L\, V. \tag{22.68}$$

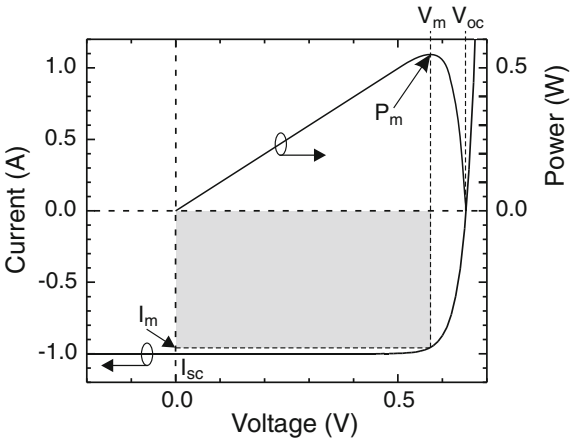


Fig. 22.53 Schematic I–V characteristics of a solar cell under illumination (*left scale*) and extracted power (*right scale*). The *grey area* is the maximum power rectangle with $P_\mathrm{m} = I_\mathrm{m} V_\mathrm{m}$

The condition $\mathrm{d}P/\mathrm{d}V = 0$ yields the optimal voltage at which the solar cell has to be operated and is given by the implicit equation

$$V_\mathrm{m} = \frac{1}{\beta} \ln\left(\frac{I_\mathrm{L}/I_\mathrm{s} + 1}{1 + \beta V_\mathrm{m}}\right) = V_\mathrm{oc} - \frac{1}{\beta} \ln\left(1 + \beta V_\mathrm{m}\right). \tag{22.69}$$

The current at maximum power is

$$I_\mathrm{m} = I_\mathrm{L}\left(1 - \frac{1 - \beta V_\mathrm{m}\, I_\mathrm{s}/I_\mathrm{L}}{1 + \beta V_\mathrm{m}}\right) \cong I_\mathrm{L}\left(1 - \frac{1}{\beta V_\mathrm{m}}\right). \tag{22.70}$$

E_m is the energy that is delivered per photon at the load resistor at the power maximum. The maximum power is $P_\mathrm{m} = I_\mathrm{L} E_\mathrm{m}/e$ and E_m is given by

$$E_\mathrm{m} \cong e\left[V_\mathrm{oc} - \frac{1}{\beta} \ln\left(1 + \beta V_\mathrm{m}\right) - \frac{1}{\beta}\right]. \tag{22.71}$$

The ideal solar cell has a (power) conversion efficiency $\eta = P_\mathrm{m}/P_\mathrm{in}$ that can be determined from Fig. 22.54a.

The right curve (1) in Fig. 22.54a shows the integral number n_ph of photons in the solar spectrum (per area and time) with an energy larger than a given one (E_g). For a given value of n_ph, the left curve (2) represents the value of E_m. The efficiency is the ratio of $E_\mathrm{m} n_\mathrm{ph}$ and the area under curve (1). The efficiency as a function of the band gap is shown in Fig. 22.55a. It has a fairly broad maximum such that many semiconductors can be used for solar cells, in principle. The maximum theoretical efficiency for a single junction is 31 % for nonconcentrated sunlight (AM1.5). This limit corresponds to the classic Shockley–Queisser limit [1586–1588], assuming radiative recombination as the only charge-carrier recombination mechanism. In [1589], the limit for a single material is found to be 43 % for an optimally tailored

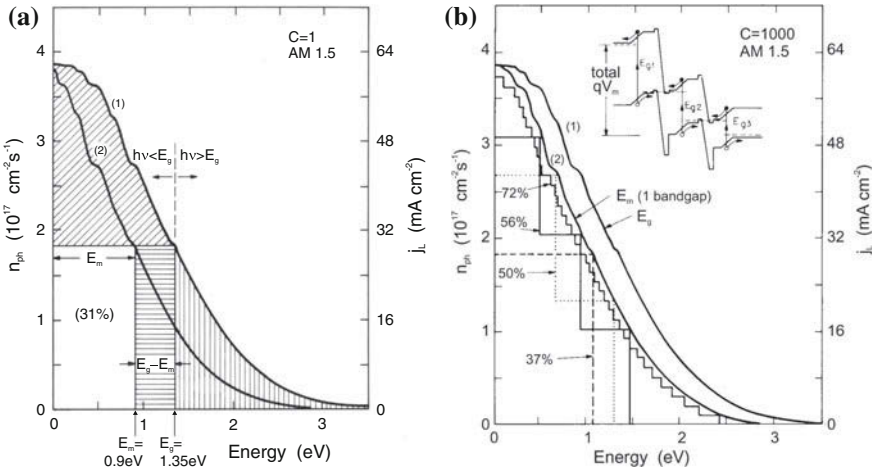


Fig. 22.54 (**a**) Number of photons n_{ph} per area and time in the sun spectrum (AM1.5, $C = 1$ sun) with an energy larger than a cutoff energy (*curve* 1) and graphical method to determine the quantum efficiency (from *curve* 2). Adapted from [1585]. (**b**) Number of photons in concentrated solar spectrum (AM1.5, $C = 1000$ sun) with an energy larger than a given energy and graphical method to determine the quantum efficiency of multi-junction solar cells. Adapted from [500] after [1585]

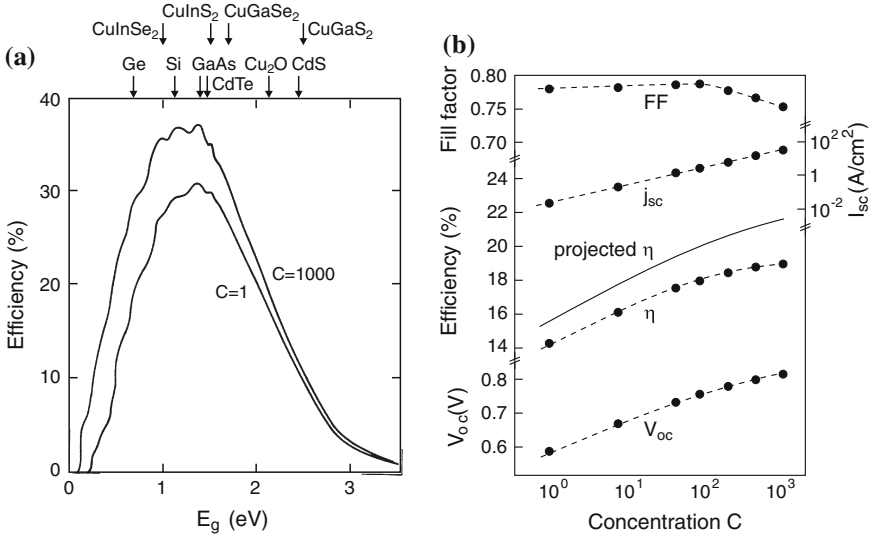


Fig. 22.55 (**a**) Ideal quantum efficiency of solar cells (single junction) as a function of band gap and light concentration C. The band gaps of some important semiconductors are denoted by *arrows*. Adapted from [1585]. (**b**) Properties of a silicon solar cell (water cooled) as a function of light concentration. *Solid line* is theoretically projected efficiency, *dashed lines* are guides to the eye. Adapted from [500]

band structure that allows carrier multiplication by optically excited hot carriers. The solar cell as a heat engine is dicussed in [1590].

When the sunlight is concentrated, e.g. by a lens, the efficiency increases (Fig. 22.55b). The short-circuit current increases linearly. The effect is mostly due to the increase of the open-circuit voltage. For $C = 1000$, the maximum theoretical efficiency for a single-junction solar cell is 38 %.

A further increase of efficiency can be achieved with multiple junctions using various materials for absorption. In a tandem cell (two junctions), the upper layer absorbs the higher-energy photons in a wide band gap material. The material with the lower band gap makes use of the low-energy photons. Thus, the cell works with two different values of E_m (Fig. 22.54b). With band gaps of 1.56 and 0.84 eV, an efficiency of 50 % can be reached theoretically. With three materials 56 %, and for a large number of materials 72 % is the limit. Between the junctions, tunneling diodes (Sect. 21.5.9) must be used to allow carrier transport through the entire structure. It is a nontrivial task to fabricate multiple heterojunctions due to incompatibilities of the lattice constants. Besides heteroepitaxy, wafer bonding can also be used for fabrication. A lattice-matched InGaP/GaAs/InGaAsN cell seems a viable solution for high-efficiency solar cells.

22.4.3 Real Solar Cells

For a real solar cell, the effect of parallel resistance R_{sh} (shunt resistance due to leakage current, e.g. by local shorts of the solar cell) and serial resistance R_s (due to ohmic loss) must be considered. The I–V characteristic is then (cf. (21.154))

$$\ln\left(\frac{I + I_L}{I_s} - \frac{V - IR_s}{I_s R_{sh}} + 1\right) = \beta\,(V - I\,R_s)\,. \tag{22.72}$$

The serial resistance affects the efficiency more strongly than the shunt resistance (Fig. 22.56). Therefore, it is frequently enough to consider R_s only and use (cf. Fig. 21.153)

$$I = I_s\,\exp\left(\beta(V - I\,R_s)\right) - I_L. \tag{22.73}$$

In the example of Fig. 22.56, a serial resistance of 5 Ω reduces the filling factor by a factor of about four.

At open circuit voltage the photo-generated carriers have nowhere to go; in an ideal solar cell, the only process is the radiative recombination and the photon escape. Certainly, the internal quantum efficiency should be high and the non-radiative recombination rate small compared to the radiative one (cf. Sect. 10.10). The open circuit voltage and thus the energy conversion efficiency depends also on the light extraction efficiency χ_{ex} which will be discussed in greater detail in the context of LEDs (Sect. 23.3.3). Based on [1592, 1593], the open circuit voltage V'_{oc} of a real solar cell is *reduced* ($\chi_{ex} \leq 1$) from the ideal value V_{oc} given in (22.67),

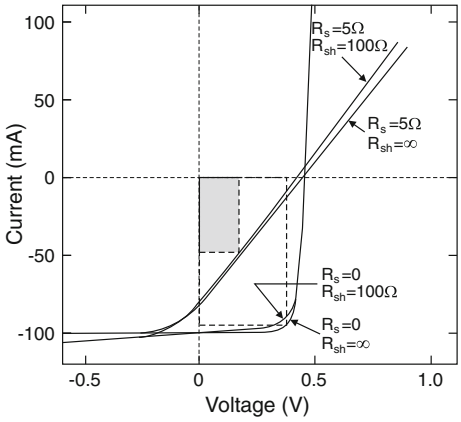


Fig. 22.56 *I–V* characteristics of a solar cell considering shunt and series resistances R_s and R_{sh}, respectively. The efficiency of the real cell (*shaded* power *rectangle*) is less than 30 % of that of the ideal cell. Adapted from [1591]

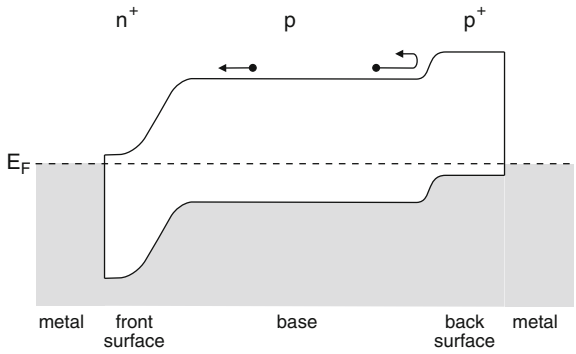


Fig. 22.57 Increase of the carrier collection efficiency by a back surface field. Adapted from [1594]

$$V'_{oc} = V_{oc} + \beta^{-1} \ln \chi_{ex}. \tag{22.74}$$

22.4.4 Design Refinements

In order to collect electrons most efficiently, a back surface field is used (Fig. 22.57). A higher-doped region at the back contact creates a potential barrier and reflects electrons back to the front contact.

An important point for optimization is the management of the reflection at the solar cell surface. First, a dielectric antireflection (AR) layer (or multi-layers) can be used. These layers should have a broad AR spectrum. Additionally, a textured surface reduces reflection (Fig. 22.58d), giving reflected photons a second chance for penetration (Fig. 22.58c). The reflectance of bare Si, 35 %, can be reduced to 2 %. An AM0 efficiency of over 15 % was reached using textured multi-crystalline cells. Alkaline KOH-based etches attack Si (001) anisotropically and yield pyramidal structures (Fig. 22.58b) with {111} facets. Recently an acidic HF/HNO_3-based

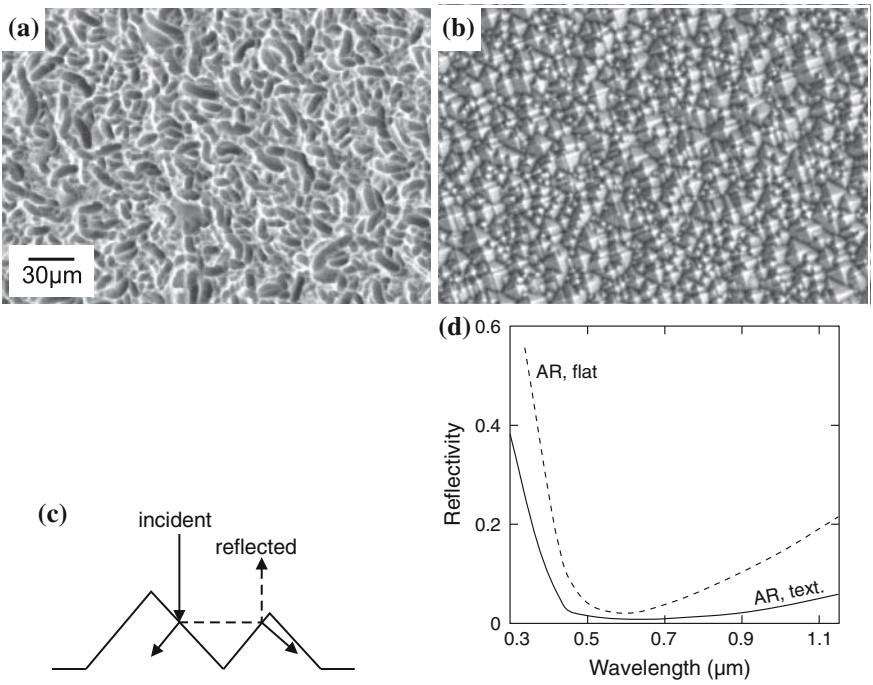


Fig. 22.58 (**a**) SEM image of topology of acidically etched multi-crystalline silicon wafer. (**b**) SEM image of alkaline etched mono-crystalline silicon wafer. (**c**) Exemplary light path. (**d**) Reflectance of antireflection-coated flat (*dashed line*) and textured (*solid line*) surface). Parts (**a**) and (**b**) adapted from [1596], part (**d**) adapted from [1597]

process has been established [1595], resulting in a worm-like surface pattern on multi-crystalline silicon wafers (Fig. 22.58a) with superior anti-reflection properties.

During its course over the sky during the day, the sun changes its angle towards a fixed solar cell.[10] A tracking mechanism can optimize the angle of incidence during the day and increase the overall efficiency of the solar cell (Fig. 22.59).

22.4.5 Modules

In order to cover a large area and supply certain values of output voltage and current, several solar cells are connected into modules. Arrays are built up from several modules (Fig. 22.60). If solar cells are connected in parallel, the total current increases; if they are connected in series, the output voltage increases. We note that in partially shadowed modules the reverse characteristics of solar cells are important [1600, 1601]; local breakdown can lead to hot spots and irreversible degradation.

[10] We are of course aware that the earth rather rotates around the sun.

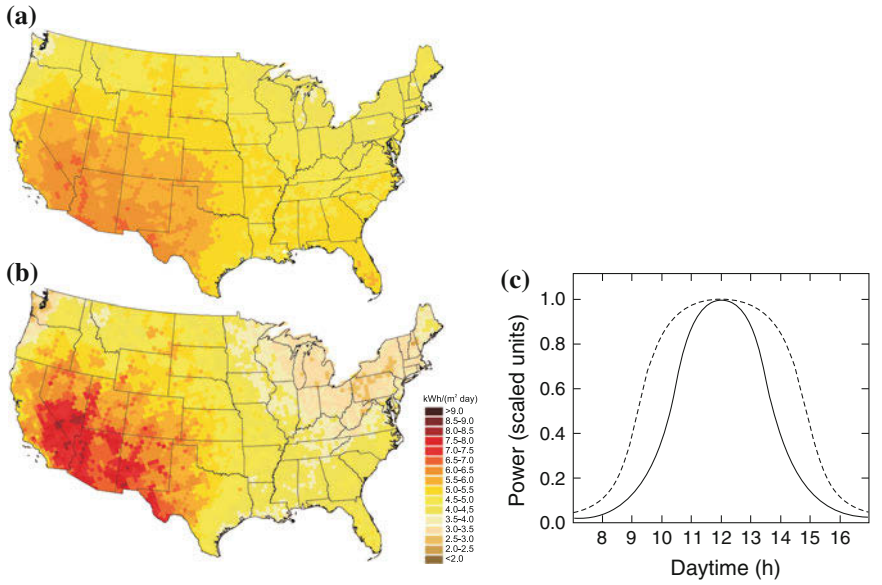


Fig. 22.59 Annual average solar energy (in kWh/(m^2 day)) for (**a**) an optimally tilted south-facing fixed panel and (**b**) an optimally two-axis tracked panel (in mainland US). Adapted from [1598]. (**c**) Power generation of a solar cell versus time (in daytime hours) for a stationary setup facing the sun at constant angle (*solid line*) and mounting with tracking (*dashed line*) to optimize the angle towards the sun. Adapted from [1599]

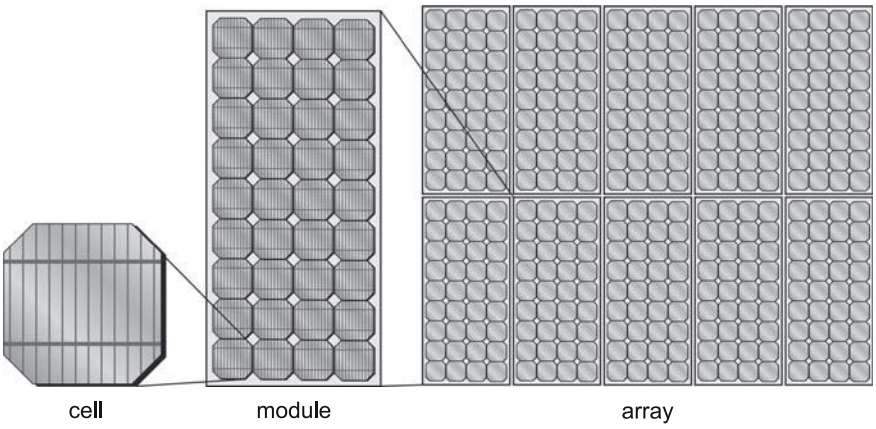


Fig. 22.60 Schematic drawing of a solar cell (with contact grid), a module (36 cells) and an array of ten modules

22.4.6 Solar-Cell Types

First Generation Photovoltaics

Silicon is the most frequently used material for solar cells. Cells based on single-crystalline silicon (wafers) have the highest efficiency but are the most expensive (Fig. 22.62a). The efficiencies of various solar cells are compiled in Table 22.1. Polycrystalline (for large grains also called multi-crystalline) silicon (Fig. 22.62b) is cheaper but offers less performance. Material design is oriented towards increasing the grain size and/or reducing their electrical activity. Grain boundaries act as recombination centers with a surface (i.e. interface) recombination velocity of 10^2 cm/s [1602] for particular, electrically fairly inactive grain boundaries, several 10^3 cm/s [1603], several 10^4 cm/s [1604] or even 10^5–10^7 cm/s [1605]. The grain boundaries reduce the effective diffusion length and thus carriers recombine before they can reach the contacts. A detailed theory of solar cell performance for polycrystalline material has been worked out in [1604] and explains the reduction of efficiency with decreasing grain size as shown in Fig. 22.61.

These solar cells are also called 'first-generation' photovoltaics. Thin sheets of crystalline silicon drawn from a melt between two seed crystals in a modified CZ growth (sheet silicon or ribbon silicon) allow cheaper production compared to cells based on 'traditional' polished wafers cut from a large silicon rod. Silicon made particularly for solar cell use is called 'solar-grade' silicon.

Second Generation Photovoltaics

Even cheaper are solar cells from amorphous silicon (Fig. 22.62c). Since silicon is an indirect semiconductor, a fairly thick layer is needed for light absorption. If direct band gap semiconductors are used, a thin layer ($d \approx 1\,\mu$m) is sufficient for complete light absorption. Such cells are called thin-film solar cells. A typical material class used in this type of cell are chalcopyrites, such as $CuInSe_2$ (CIS). The band gap is around 1 eV, which is not optimal. An improvement can be achieved by adding

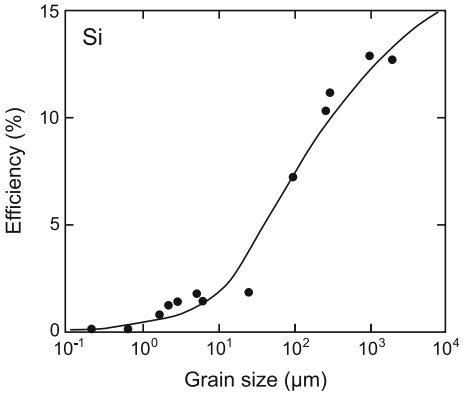


Fig. 22.61 Theoretical dependence of the effect of grain size on efficiency of polycrystalline solar cells (*solid line*) with experimental data points (*circles*). Adapted from [1604]

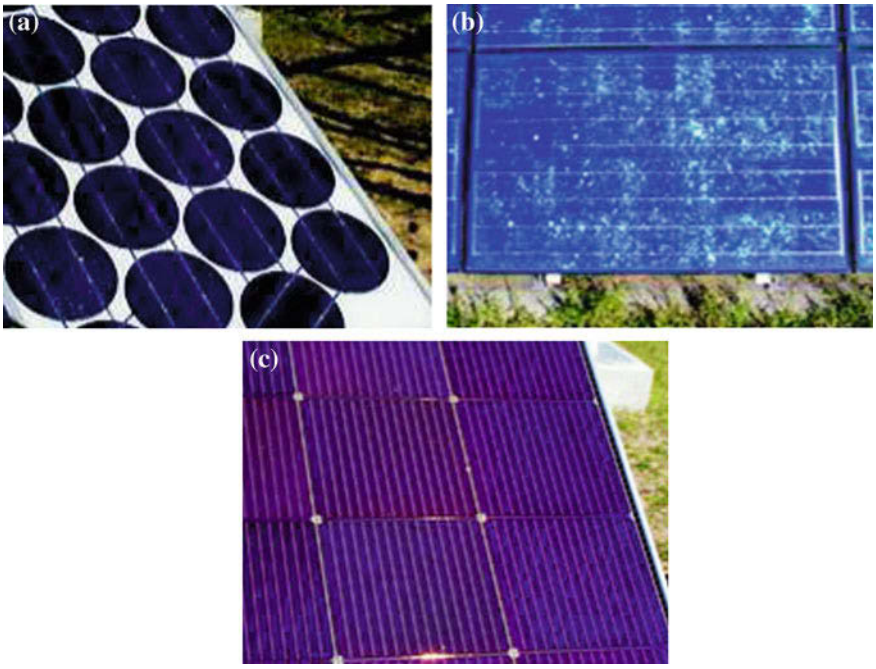


Fig. 22.62 Various types of solar cells: (**a**) monocrystalline silicon solar cell, (**b**) polycrystalline solar cell, (**c**) amorphous silicon solar cell. From [1700]

Table 22.1 Record efficiency of various solar cells (AM1.5, 1000 W/cm^2, 25° unless noted otherwise)

Cell material/type	Efficiency (%)	V_{oc} (V)	j_{sc} (mA/cm^2)	FF (%)	Date m/y
Si (crystalline)	25.6 ± 0.5	0.740	41.8	82.7	2/2014
Si (polycrystalline)	20.8 ± 0.6	0.6626	39.03	80.3	11/2014
Si (amorphous)	10.2 ± 0.3	0.896	16.36	69.8	7/2014
GaAs (crystalline)	25.9 ± 0.8	1.038	29.4	84.7	12/2007
GaAs (thin film)	28.8 ± 0.9	1.122	29.68	86.5	5/2012
GaAs (polycrystalline)	18.4 ± 0.5	0.994	23.2	79.7	11/1995
2J (GaInP/GaAs)	30.3	2.488	14.22	85.6	4/1996
3J (GaInP/GaAs/Ge)	32.0 ± 1.5	2.622	14.4	85.0	1/2003
3J (GaInP/GaAs/InGaAs)	37.9 ± 1.2	3.065	14.27	86.7	2/2013
3J (conc., 240 suns)	40.7 ± 2.4	2.911	3832	87.5	9/2006
4J (conc., 508 suns)	46.0 ± 2.2	4.227	6498	85.1	10/2014
CIGS	20.5 ± 0.6	0.752	35.3	77.2	3/2014
CdTe	21.0 ± 0.4	0.876	30.25	79.4	8/2014
Dye sensitized	11.9 ± 0.4	0.793	19.4	71.4	9/2014
Organic (thin film)	11.0 ± 0.3	0.793	19.4	71.4	9/2014

Most data from [1579], additional data for 3J (conc.) [1606]

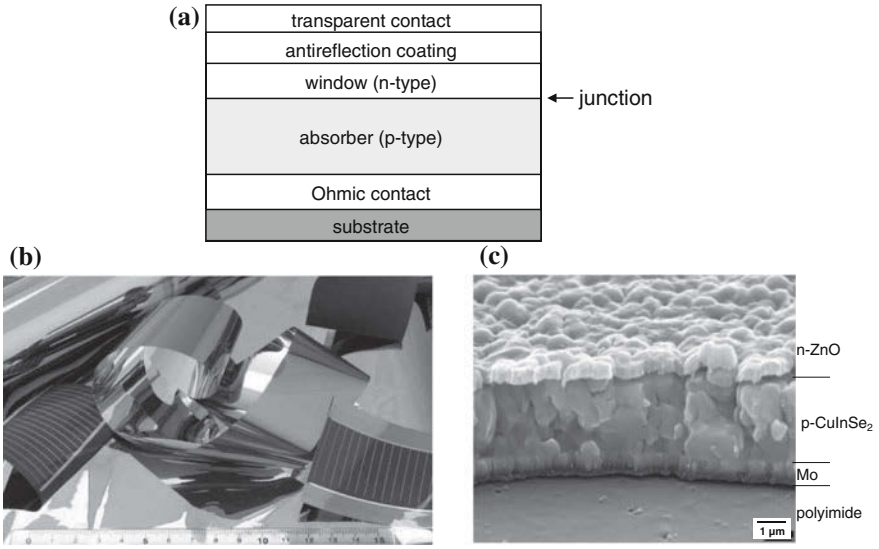


Fig. 22.63 (**a**) Schematic cross section of a polycrystalline thin-film solar cell. (**b**) Rolled sheets of CIS thin film solar cell on flexible Kapton foil. (**c**) SEM cross section of CIS thin-film solar cell. Parts (**b**) and (**c**) reprinted with permission from [1611]

Ga and/or S which increases the band gap, $Cu(In,Ga)(Se,S)_2$ (CIGS), to 1.2–1.6 eV. Using CIGS, an efficiency of over 19 % has been reported in laboratory samples; 12–13 % seeming realistic for production [1607]. Also CdTe is a viable absorber mostly sputtered on glass with over 16 % efficiency demonstrated and 9–10 % realistic in production. Thin-film solar cells can be fabricated on glass substrate or on flexible polymer substrate such as Kapton[11] (Fig. 22.63a, b). Also here, optimization of the grain size is important (Fig. 22.63b). As the front contact, a transparent conductive oxide (TCO), such as ITO ($InSnO_2$) or ZnO:Al, is used. If the front surface is given by the glass substrate, as can be the case for CdTe/glass solar cells, the glass is actually termed 'superstrate'. Thin-film and amorphous silicon solar cells are also termed 'second-generation' photovoltaics. Also organic materials can be used for solar cells [1608], promising low-cost production at acceptable performance, in 2004 2.5 % [1609] and in 2006 about 4 % maximum efficiency [1610].

Third Generation Photovoltaics

'Third-generation' photovoltaics attempt to go beyond the 30 % limit and comprise of multi-junction solar cells, concentration of sunlight, use of hot-carrier excess energy as discussed above and possibly other concepts including photon conversion [1612, 1613], intermediate band absorption [1614–1616], multi-exciton generation [1617] or the use of quantum dots [1618].

[11] Kapton® is a polyimide and a product and registered trademark of DuPont.

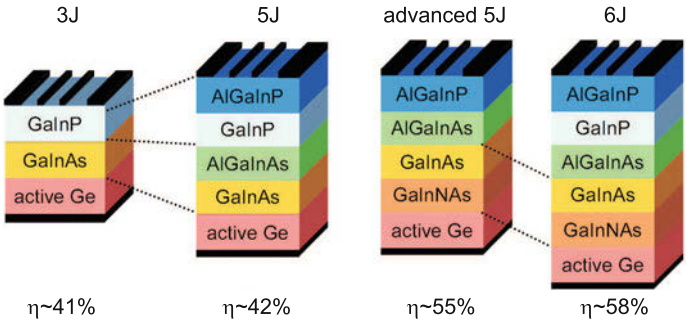


Fig. 22.64 Schematic layer stacking of multi-junction solar cells and expected efficiency. Adapted from [1619]

In multi-junction cells the different absorber layers are stacked on the substrate with increasing band gap and connected via (highly doped) tunneling junctions (Sect. 21.5.9). Under 500-fold AM1.5 illumination a three junction (3J) cell (GaInP/GaInAs/Ge) is expected to exhibit up to 41 % efficiency, for 5 J 42 % or up to 55 % using GaInNAs and with 6 J up to 59 % [1619, 1620] (Fig. 22.64). For a 3 J cell the record efficiency is 40.7 % (240 suns) using the layer structure as shown in Fig. 22.65 [1606]. Details on modeling of III–V multi-junction solar cells can be found in [1621]. In a monolithic cell the absorbers must be tuned such that the same current (Kirchhoff's law) can pass through all layers. Multi-junction solar cells are heteroepitaxial devices and thus expensive; the use of concentration is economically mandatory.

A recent novel route is a tandem cell comprising of a bottom silicon cell and a top perovskite cell. Organo-metal halide peroskites, e.g. of the ammonium trihalogen plumbates type, R-N_3PbI_3, have exhibited quite high conversion efficiencies [1624, 1625] and can be tuned with regard to their absorption range.

22.4.7 Commercial Issues

The cost[12] of producing photovoltaic (PV) modules, in constant dollars, has fallen from as much as $50 per peak watt in 1980 to as little as $3 per peak watt in 2004. A projected cost of 0.2 €/kWh in 2020, a third of the current cost, is realistic and competitive in many applications. In 2002, photovoltaic power of 560 MW was installed worldwide. By the end of 2003, a total photovoltaic power of about 350 MW was installed in Germany. Thus, the PV power is 0.33 % of the total installed electric power of 106 GW in Germany. The current market growth of 30 % (worldwide) is driven by crystalline silicon cells (95 % in 2002).

[12] The following information is taken from [1626, 1627].

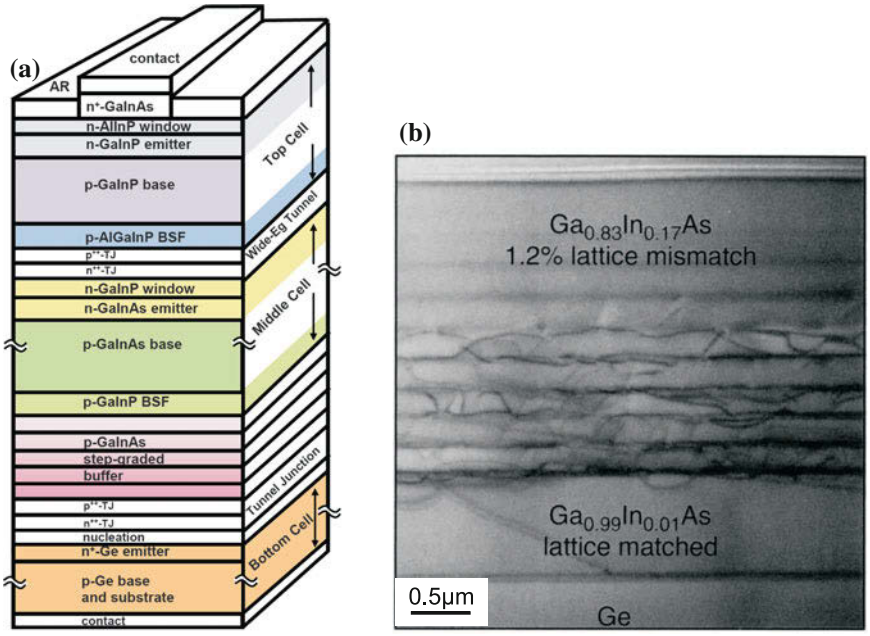


Fig. 22.65 (**a**) Schematic layer stacking of three-junction (3J) solar cell. The step-graded buffer (metamorphic buffer) changes the in-plane lattice constant for the following layers. Adapted from [1622]. (**b**) Cross-section TEM image of metamorphic InGaAs buffer on Ge. Adapted from [1623]

The energy payback period is also dropping rapidly. For example, it takes today's typical crystalline silicon module about 4 years to generate more energy than went into making the module in the first place. The next generation of silicon modules, which will employ a different grade of silicon and use thinner layers of semiconductor material, will have an energy payback of about 2 years. And thin-film modules will soon bring the payback down to one year or less. However, market growth of thin-film modules is currently slow. This means that these modules will produce 'free' and clean energy for the remaining 29 years of their expected life.

PV technology can meet electricity demand on any scale. The solar energy resource in a 100 mile-square area of Nevada could supply the United States with all its electricity (about 800 GW), using modestly efficient (10 %) commercial PV modules. A more realistic scenario involves distributing these same PV systems throughout the 50 states. Currently available sites, such as vacant land, parking lots, and rooftops, could be used. The land requirement to produce 800 GW would average out to be about 17 × 17 miles per state. Alternatively, PV systems built in the 'brownfields', the estimated 5 million acres of abandoned industrial sites in the US, could supply 90 % of America's current electricity. Solar power is expected to contribute 10 % of the US energy need in 2030. For Germany, more than 2 % in 2020 is probable.

In 2001, PV module shipments in the US approached the 400 MW mark, representing a $2.5 to $3 billion market. The US-based industry itself is now approaching $1 billion per year and provides 25,000 jobs. It is expected to grow to the $10–$15 billion level in the next 20 years, providing 300,000 jobs by 2025.

Chapter 23
Electricity-to-Light Conversion

Abstract Light emitting diodes (LEDs) and laser diodes are the focus of this chapter. For LEDs materials choices, the concepts of internal and external quantum efficiency as well as device design are treated. Special devices such as white LEDs, quantum dot and organic LEDs are introduced. For laser diodes the concepts of gain, loss and threshold, various heterostructures for modern device design and laser emission properties such as mode spectrum, far field, dynamics and tunability are discussed. Finally special devices such as the hot hole laser, the cascade laser and semiconductor optical amplifiers are mentioned.

23.1 Radiometric and Photometric Quantities

23.1.1 Radiometric Quantities

The radiometric quantities are derived from the radiant flux (power) Φ_e (or usually simply Φ) that is the total power (energy per time) emitted by a source, measured in Watts. The radiant intensity I_e is the radiant flux emitted by a point source into a solid angle,[1] measured in Watts per steradian (or W/sr). The irradiance E_e is the radiant flux per area incident on a given plane, measured in W/m^2. The radiance L_e is the radiant flux per area and solid angle as, e.g., emitted by an extended source, measured in W/(m^2 sr).

23.1.2 Photometric Quantities

The photometric quantities are related to the visual impression and are derived from the radiometric quantities by weighting them with the $V(\lambda)$ curve.

[1] A solid angle Ω is the ratio of the spherical surface area A and the square of the sphere's radius r, i.e. $\Omega = A/r^2$.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_23

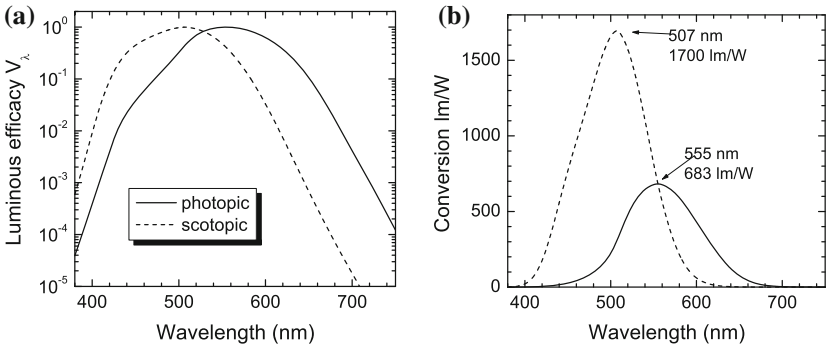


Fig. 23.1 (**a**) Relative eye sensitivity curves for photopic (light adapted, *solid line*) and (dark adapted, *dashed line*) vision. (**b**) Conversion of lumen to Watt for light- and dark-adapted vision

The luminous flux (luminosity or visible brightness) Φ_v of a source with the radiant flux (spectral power distribution) $\Phi(\lambda)$ is given by

$$\Phi_\mathrm{v} = K_\mathrm{m} \int_0^\infty \Phi(\lambda)\, V(\lambda)\, \mathrm{d}\lambda, \tag{23.1}$$

with $K_\mathrm{m} = 683\,\mathrm{lm/W}$. This formula is also the definition of the unit 'lumen'. In Fig. 23.1b, the conversion function[2] $V(\lambda)$ is shown for light and dark adapted vision.[3]

Further derived photometric quantities are luminous intensity (luminous flux per solid angle), measured in candela (cd), the illuminance (luminous flux per area), measured in lux (lx), and the luminance (luminous flux per area and solid angle). The latter is particularly important if the radiation enters an optical system, e.g. for refocusing. The radiometric and photometric quantities are summarized in Table 23.1.

[2]The $V(\lambda)$ curve has been experimentally determined by letting several observers adjust (decrease) the perceived brightness of a monochromatic light source at 555 nm to that of light sources of the same absolute radiation power at other wavelengths with so-called heterochromatic flicker photometry. The 'relative sensitivity curve for the CIE Standard Observer' was determined in 1924. The 'standard observer' is neither a real observer nor an average human observer. The curve has shortcomings, e.g., due to the used spectral band width (20–30 nm) of the light sources and the comparison of spectral power instead of the photon flux.

[3]While photopic vision is due to cones, the scotopic (dark-adapted) vision is due to rods. Rods are more than one thousand times as sensitive as the cones and can reportedly be triggered by individual photons under optimal conditions. Rods predominate in the peripheral vision and are not color sensitive.

Table 23.1 Radiometric and photometric quantities and units

Radiometric			Photometric		
Quantity	Symbol	Unit	Quantity	Symbol	Unit
Radiant flux	Φ_e	W	Luminous flux	Φ_v	lm
Radiant intensity	I_e	W/sr	Luminous intensity	I_v	cd
Irradiance	E_e	W/m^2	Illuminance	E_v	lx
Radiance	L_e	$W/m^2/sr$	Luminance	L_v	$lm/m^2/sr$

The photometric units are lumen (lm), lux (lx = lm/m^2) and candela (cd = lm/sr)

23.2 Scintillators

A scintillator (or phosphor) is a material that converts impacting high-energy radiation into photons [1628]. Besides a high conversion efficiency, the spectrum and decay time constant of the scintillator are important for display applications. For display purposes, the photons are directly used for forming the image for the observer. For radiation detection, the photons are fed to a photomultiplier tube and counted.

The most prominent applications, involving the conversion of electrons, are the screens of cathode ray tubes (CRT) (acceleration voltage $>10\,kV$) and of flat panel devices, such as field-effect displays (using a low voltage for excitation, typically $<1\,kV$) or plasma displays (using the UV light from the discharge of a plasma placed between two electrodes for excitation). Further details on electroluminescent displays can be found in [1629]. Other forms of radiation detected with scintillators are α-, β-, and γ-radiation, X-rays and neutrons [1630]. Different excitation conditions require different phosphors for optimal performance.

23.2.1 CIE Chromaticity Diagram

The CIE[4] procedure converts the spectral power distribution (SPD) of light from an object into a brightness parameter Y and two chromaticity coordinates x and y. The chromaticity coordinates map the color[5] with respect to hue and saturation on the two-dimensional CIE chromaticity diagram. The procedure for obtaining the chromaticity coordinates for a given colored object involves determination of its

[4]Commission Internationale de l'Éclairage. The color space can be described by different coordinate systems, and the three most widely used color systems, Munsell, Ostwald, and CIE, describe the color space with different parameters. The Munsell system uses hue, value, and chroma and the Ostwald system uses dominant wavelength, purity, and luminance. The more precise CIE system uses a parameter Y to measure brightness and parameters x and y to specify the chromaticity that covers the properties hue and saturation on a two-dimensional chromaticity diagram.

[5]This definition is motivated by the color vision of the eye. Two light sources will have the same color, even if they have different SPDs, when they evoke the same color impression to the human eye.

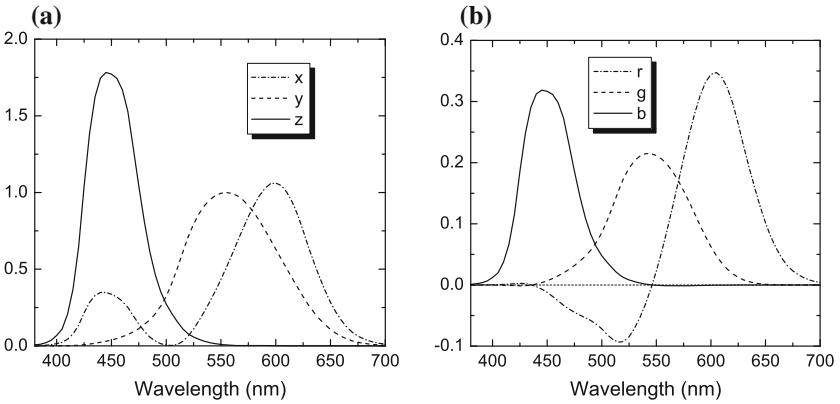


Fig. 23.2 (**a**) Color-matching functions $\bar{x}$, $\bar{y}$, and $\bar{z}$ for the calculation of the CIE chromaticity, (**b**) color-matching functions $\bar{r}$, $\bar{g}$, and $\bar{b}$ for the calculation of RGB values

spectral power distribution $P(\lambda)$ at each wavelength, multiplication by each of the three color-matching functions $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$ (Fig. 23.2a) and integration (or summation) of the three tristimulus values X, Y, Z

$$X = \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} P(\lambda)\,\bar{x}(\lambda)\,\mathrm{d}\lambda \tag{23.2a}$$

$$Y = \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} P(\lambda)\,\bar{y}(\lambda)\,\mathrm{d}\lambda \tag{23.2b}$$

$$Z = \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} P(\lambda)\,\bar{z}(\lambda)\,\mathrm{d}\lambda. \tag{23.2c}$$

Y gives the brightness. The tristimulus values are normalized to yield the chromaticity coordinates, e.g. $x = X/(X + Y + Z)$. x and y obtained in this way are the chromaticity coordinates. The third coordinate $z = 1 - x - y$ offers no additional information and is redundant. Therefore, the color is represented in a two-dimensional diagram, the CIE chromaticity diagram[6] as shown in Fig. 23.3a. White is represented by $x = y = z = 1/3$. In order to relate the differences between colors as perceived by the human eye more closely to the geometrical distance in the chart, a revision was made (Fig. 23.3b) with new coordinates

$$u' = 4x/(-2x + 12y + 3) \tag{23.3a}$$

$$v' = 9y/(-2x + 12y + 3). \tag{23.3b}$$

[6]The coloring of the chart is provided for an understanding of color relationships. CRT monitors and printed materials cannot reproduce the full gamut of the color spectrum as perceived in human vision. The color areas that are shown only depict rough categories and are not precise statements of color.

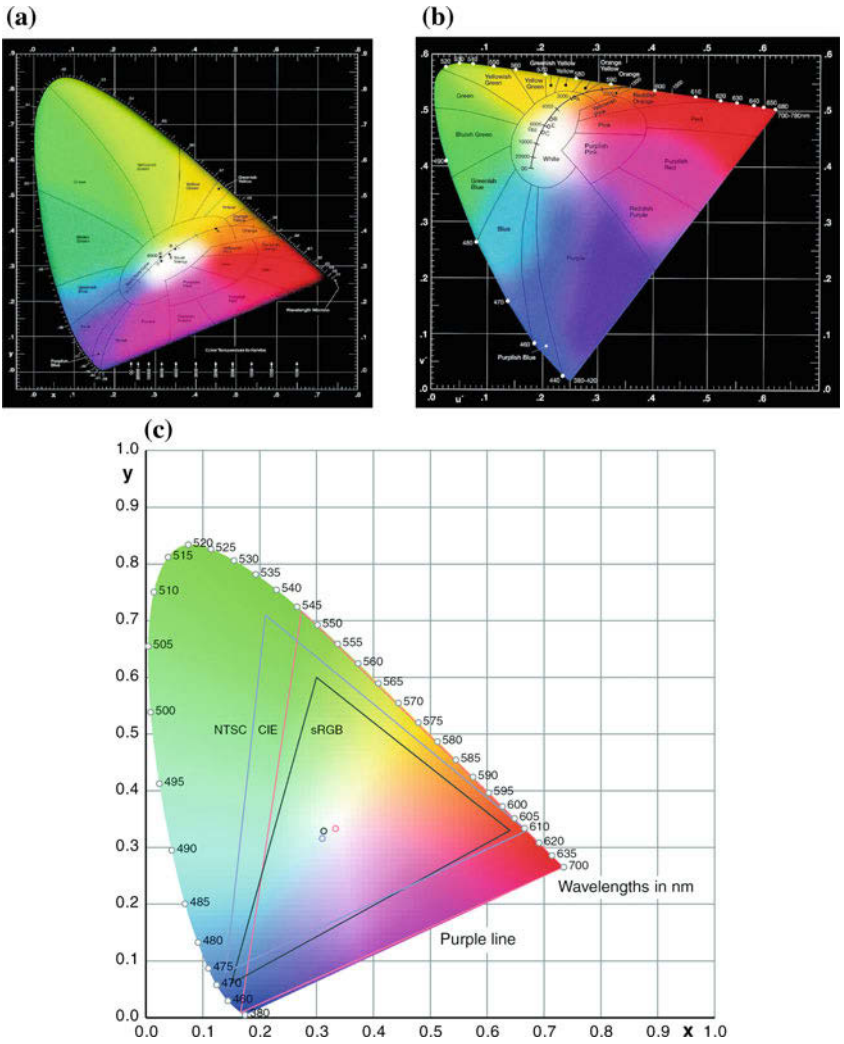


Fig. 23.3 CIE chromaticity diagram from 1931 (**a**) in the coordinates x and y and from 1976 (**b**) in the coordinates u' and v' (23.3b). The curved upper boundary is called the 'spectrum locus' and contains monochromatic colors, the straight line at the lower left is termed the 'purple boundary'. In the graph also the color of blackbody radiation is given, $T = 5440\,\mathrm{K}$ corresponds to $x = y = 1/3$. 'A', 'B', 'C', and 'E' are standard illuminants, 'D65' denotes daylight with color temperature $T = 6500\,\mathrm{K}$. (**c**) CIE chart with the color ranges of sRGB, CIE and NTSC. Part (**c**) adapted from [1631]

Table 23.2 Primaries and white points for sRGB, CIE and NTSC

Primary	Red		Green		Blue		White	
CIE	0.73467	0.26533	0.27376	0.71741	0.16658	0.00886	0.33333	0.33333
NTSC	0.6700	0.3300	0.2100	0.7100	0.1400	0.0800	0.3100	0.3160
sRGB	0.6400	0.3300	0.3000	0.6000	0.1500	0.0600	0.3127	0.3290

For CRTs the red-green-blue (RGB) color space is used.[7] The color matching functions for RGB values are shown in Fig. 23.2b. The RGB values are related to the XYZ values according to

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} 2.36461 & -0.89654 & -0.46807 \\ -0.51517 & 1.42641 & 0.08876 \\ 0.00520 & -0.01441 & 1.00920 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} . \tag{23.4}$$

The CIE RGB primaries from 1931 are at 700, 546.1, and 435.8 nm with the relative intensities 1.0, 4.5907, and 0.0601. A display device using three phosphors can only display colors in the triangular area of the CIE chart between the three chromaticity coordinates. For sRGB,[8] the 1931 CIE primaries and the NTSC[9] norm the coordinates are given in Table 23.2 and visualized in Fig. 23.3c. An optimal coverage of the CIE chart involves monochromatic sources (for laser TV or LED displays) at about 680, 520 and 440 nm.

23.2.2 Display Applications

The once ubiquitous amber-colored monochrome displays are mostly fabricated using ZnS:Mn [1629], having broad emission (540–680 nm) with its spectral peak at 585 nm ($x = 0.50$, $y = 0.50$) with an efficiency of 2–4 lm/W. In color television (and similar applications such as color computer monitors, tubes for aviation use, projection television) the image is reproduced by selective and time-multiplexed cathode excitation of three phosphors (blue, green and red) deposited on the internal face of the screen. The chromaticity coordinates of the standard CRT phosphors P-22B, P-22G and P-22R are given in Table 23.3. They cover about the color range

[7] RGB is an additive color system. However, printing devices use a subtractive color system. This means that the ink absorbs a particular color, and the visible impression stems from what is reflected (not absorbed). When inks are combined, they absorb a combination of colors, and hence the reflected colors are reduced, or subtracted. The subtractive primaries are cyan, magenta and yellow (CMY) and are related to RGB via $(C, M, Y) = (1 - R, 1 - G, 1 - B)$.

[8] Standard RGB color space as defined mainly by Hewlett-Packard and Microsoft, almost identical to PAL/SECAM European television phosphors.

[9] National television standard colors, US norm.

Table 23.3 CIE color coordinates, peak emission wavelength and decay time (10 %) of standard CRT phosphors

Phosphor	x	y	λ_p (nm)	Decay time
P-22B	0.148	0.062	440	~20 μs
P-22G	0.310	0.594	540	~60 μs
P-22R	0.661	0.332	625	1 ms

labeled 'sRGB' in Fig. 23.3c. For blue and green ZnS:Ag ($x = 0.157$, $y = 0.069$), ZnS:Ag,Cl, ZnS:Ag,Al and ZnS:Cu,Al ($x = 0.312$, $y = 0.597$), ZnS:Cu,Au,Al are used as phosphors, respectively. Y_2O_2S:Eu ($x = 0.624$, $y = 0.337$) activated with trivalent europium (Eu^{3+}) facilitated such a gain in the brilliance of red over ZnS:Ag (more than doubled it) that it has totally replaced it at about one fifth of the cost. For reproducible image quality, precise grain-size control (median size for CRT phosphors is about 8 μm), dispersion control and surface treatment are necessary. Flat-panel displays with their lower excitation voltage require different phosphors for optimal efficiency.

23.2.3 Radiation Detection

The most commonly used scintillation detector for alpha measurements is ZnS activated with silver, ZnS:Ag. This material is not very transparent to light and is usually prepared as a large number of crystals with sub-mm size attached with an adhesive to a flat piece of plastic or other material. The flat screen is optically coupled to a photomultiplier tube that is attached to associated electronics. The voltage and discriminator levels are selected so that the detector is sensitive to the rather large pulses from alpha interactions but insensitive to beta- or gamma-induced pulses. The alpha particles deposit all of their energies in a small thickness of material compared to beta and gamma radiations.

Scintillation detectors for beta radiation are often made from organic materials. In an organic scintillator, the light emission occurs as a result of fluorescence when a molecule relaxes from an excited level following excitation by energy absorption from ionizing radiation. Molecules such as anthracene, trans-stilbene, para-terphenyl, and phenyl oxazole derivatives are among the many organic species that have useful scintillation properties. The organic molecules are dissolved in organic solvents and used as liquid scintillators. A classic application is in the measurement of low-energy beta radiation from, e.g. tritium, ^{14}C, or ^{35}S. In such cases, the sample containing the radioactive beta emitter is dissolved in, or in some cases suspended in, the liquid scintillation solution. The emitted beta radiation transfers energy through the solvent to the scintillator molecule that emits light, subsequently detected by photomultiplier tubes. Organic scintillator molecules can also be dissolved in an organic monomer that can then be polymerized to produce a plastic scintillator in a wide variety of shapes and sizes. Very thin scintillators have been used for alpha detection, somewhat

Table 23.4 Emission peak wavelength and decay time of various scintillator materials

Material	λ_p (nm)	Decay time
Zn_2SiO_4:Mn	525	24 ms
ZnS:Cu	543	35–100 μs
$CdWO_4$	475	5 μs
CsI:Tl	540	1 μs
CsI:Na	425	630 ns
$Y_3Al_5O_{12}$:Ce	550	65 ns
Lu_2SiO_5:Ce	400	40 ns
$YAlO_3$:Ce	365	30 ns
ZnO:Ga	385	2 ns

thicker scintillators for beta detection. Large-volume plastic scintillators have been used in gamma detection, particularly for dose-related measurements.

Other inorganic crystalline scintillators, especially sodium iodide activated with thallium, NaI:Tl, have been used for gamma-ray energy measurements. Such detectors can be grown as large single crystals that have a reasonably high efficiency for absorbing all of the energy from incident gamma rays. There exists a rather large number of inorganic scintillators; some examples of these include cesium iodide activated with thallium, CsI:Tl, bismuth germanate, $Bi_4Ge_3O_{12}$, and barium fluoride, BaF_2. These are mostly used for gamma measurements but can also be prepared with thin windows and have been used for charged particle (e.g. alpha and beta) counting. A number of scintillator materials including tungstates like $CdWO_4$ has been reviewed in [1632].

In Table 23.4, the peak emission wavelength and the characteristic decay time are listed for a variety of scintillator materials. Direct semiconductors, although not offering the highest efficiency, are particularly useful for high time resolution in, e.g., time-of-flight measurements or fast scanning electron microscopy.

23.2.4 Luminescence Mechanisms

Self-Trapped Excitons

In a strongly ionic crystal, such as NaI, a hole becomes localized to an atomic site via the polaron effect. A spatially diffuse electron is attracted, and a self-trapped exciton is formed that can recombine radiatively.

Self-Activated Scintillator

In such material, the luminescent species is a constituent of the crystal. The emission involves an intraionic transition, e.g. 6p→6s in Bi^{3+} of $Bi_4Ge_3O_{12}$, or a charge-transfer transition in the case of $(WO_4)^{2-}$ in $CaWO_4$. At room temperature, nonradiative competing processes limit the efficiency.

Activator Ions

For dopant ions such as Eu^{2+} in YO_2S:Eu, Ce^{3+} in $YAlO_3$:Ce or Tl^+ in NaI:Tl, the hole and electron excited by the radiation are sequentially trapped by the same ion that then undergoes a radiative transition, in the case of Eu and Ce[10] 5d→4f, for Tl $^3P_{0,1} \rightarrow S_0$. CsI:Tl has one of the highest efficiencies of 64.8 photons/keV [1633].

Core–Valence Luminescence

In some materials, e.g., BaF_2, CsF, $BaLu_2F_8$ the energy gap between the valence band and the top core band is less than the fundamental band gap. A radiative transition occurs when an electron from the valence band fills a hole in the top core band that has been created by the radiation. The light yield is limited to about 2 photons/keV.

Semiconductor Recombination Processes

Free excitons or excitons bound to impurities can recombine radiatively. This process is most efficient at low temperatures. At room temperature, the emission is typically much weaker ($\gtrsim 10\times$) since excitons become unbound or dissociated. Highly doped n-type semiconductors, e.g. CdS:In, exhibit recombination between donor-band electrons and holes. ZnO:Ga has an efficiency of about 15 photons/keV and a fast response (with 2.4 photons/keV emitted in the first 100 ps). Luminescence can also stem from donor–acceptor pair transitions, e.g. in PbI_2 with an efficiency of 3 photons/keV at 10 K. Isoelectronic impurities such as nitrogen in GaP:N and tellurium in CdS:Te attract an electron and subsequently a hole. In ZnS:Ag and ZnS:Cu (conduction) band to trap recombination is dominant. In a codoping scheme like CdS:In,Te, In supplies electrons in an impurity band that can recombine with holes trapped at Te.

23.3 Light-Emitting Diodes

23.3.1 Introduction

Light-emitting diodes (LEDs) are semiconductor devices in which injected carriers recombine radiatively. The recombination process leading to light emission can be of intrinsic nature, i.e. band–band recombination, or extrinsic, e.g. impurity-bound excitons. Impurity-related luminescence can also be excited via impact excitation. For an extensive treatment of LEDs see [1634], for a review of the early field [1635] and for recent reviews [1636, 1637]. Mostly LEDs are pn-diodes although also some MIS-based devices have been reported [1638, 1639].

[10] This transition is dipole allowed for Ce and partially forbidden for Eu.

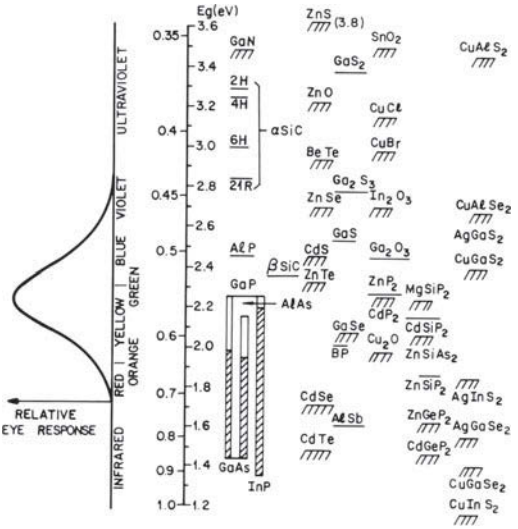


Fig. 23.4 Spectral coverage by various semiconductor materials. Reprinted with permission from [500], © 1981 Wiley

23.3.2 Spectral Ranges

Applications for LEDs can be sorted by the color of emission. In Fig. 23.4, the standard sensitivity $V(\lambda)$ of the human eye is shown (see Fig. 23.1a). In the visible spectral region (about 400–750 nm) the perceived brightness of the LED depends on the eye sensitivity. It is largest in the green (at 555 nm) and drops strongly towards the red and blue.

The most important spectral regions and applications are:

- infrared ($\lambda > 800$ nm): remote controls, optocouplers, low-cost data transmission, IR interface
- visible: indicator LED, lighting[11] (room, buildings, cars), white LED for broad spectrum
- ultraviolet ($\lambda < 400$ nm): pumping of phosphors for white LEDs, biotechnology

In Fig. 23.4, potentially useful semiconductors for the various spectral regions are shown. The semiconductors that are currently used for the various colors of the visible spectrum are

- red–yellow: Ga(As,P)/GaAs, now (Al,In,Ga)P/GaP
- yellow–green: GaAsP:N, GaP:N
- green–blue: SiC, now GaN, (In,Ga)N
- violet: GaN
- ultraviolet: (Al,Ga)N

[11] Penetration of white LEDs into the general lighting market could translate (globally) into cost savings of $\$ 10^{11}$ or a reduction of power generation capacity of 120 GW.

23.3.3 *Efficiencies*

External Quantum Efficiency

The *external* (or total) quantum efficiency η_{ext} is the number of photons emitted from the device per injected electron–hole pair. It is given by the product of the *internal* quantum efficiency η_{int} and the light extraction efficiency χ_{ex}:

$$\eta_{\text{ext}} = \chi_{\text{ex}}\, \eta_{\text{int}}. \tag{23.5}$$

In a commercial device another factor, the packaging efficiency may enter which accounts for photon loss due to packaging the LED dice into its housing.

Wall-Plug Efficiency

The wall-plug efficiency η_{w} is the power conversion ratio of the electrical power and the light output P_{out},

$$\eta_{\text{w}} = \frac{P_{\text{out}}}{I\,V} = \frac{\hbar\omega}{e\,V}\,\eta_{\text{ext}}. \tag{23.6}$$

At first it seems reasonable to assume that always $\eta_{\text{w}} < 1$. However, it has been reported that at small current and at elevated temperature the wall-plug efficiency is found larger than 100 % due to the electrical work pumping heat from the lattice to the photon field in a GaSb-based diode [1640]. The groundwork for such effect was laid in [1641], essentially predicting that $\hbar\omega > e\,V$ is possible.

Internal Quantum Efficiency

The internal quantum efficiency is the number of photons generated (inside the semiconductor) per injected electron–hole pair. High material quality, low defect density and low trap concentration are important for a large value of η_{int}. The recombination current in the pn-diode is given in (21.132).

Light Extraction Efficiency

The light extraction efficiency of the LED chip is ratio of the number of photons leaving the device and the total number of generated photons.[12] The geometry of the LED is of prime importance to optimize χ_{ex}. Due to the large index of refraction of the semiconductors ($n_{\text{s}} \sim 2.5$–3.5), light can leave the semiconductor only under a small angle θ_{c} from the surface normal due to total reflection (cf. (9.11) and see right part of Fig. 9.2). Against air ($n_1 \approx 1$) the critical angle is

$$\theta_{\text{c}} = \sin^{-1}(1/n_{\text{s}}). \tag{23.7}$$

The critical angle for total reflection is 16° for GaAs and 17° for GaP. Additionally, a portion of the photons that do not suffer total reflection is reflected back from the surface with the reflectance R given by (cf. (9.15))

[12] Note that the light extraction efficiency is also important for solar cells, cf. Sect. 22.4.3.

$$R = \left(\frac{n_s - 1}{n_s + 1}\right)^2 . \tag{23.8}$$

We note that the above formula is valid strictly for vertical incidence. For the GaAs/air interface, the surface reflectance (for normal incidence) is about 30 %. Thus, the light extraction efficiency for a LED is given by (1–R) and the critical angle by

$$\chi_{ex} \cong \frac{4\, n_1\, n_s}{(n_1 + n_s)^2}\,(1 - \cos\theta_c) \approx \frac{4\, n_s}{(n_s + 1)^2}\,(1 - \cos\theta_c)\,. \tag{23.9}$$

The latter approximation is valid when the outer medium is air. For GaAs, the light extraction efficiency is $0.7 \times 4\,\% \approx 2.7\,\%$. Thus, for this simple geometry, only a small fraction of generated photons can leave the device and contribute to the LED emission.

23.3.4 Device Design

In the following subsections the strategies that have allowed significant improvement of the extraction efficiency (Fig. 23.5) and thus LED performance are briefly discussed. The record by 2014 for light extraction efficiency is 89 % [1642], however not for a mass-produced device.

Nonplanar Surfaces

With curved surfaces, the problem of total reflection can be (partially) circumvented (Fig. 23.6). Spherically polished chips are feasible, but, very expensive. The epoxy seal of the standard LED case (Fig. 23.7a) and its shape play a similar role, however, with a smaller index of refraction than the semiconductor, and are important for the beam shape.

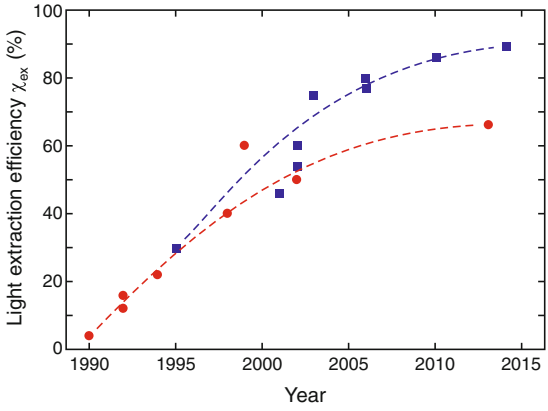


Fig. 23.5 Historic development of maximum light extraction efficiency for AlGaInP (*red circles*) and (In,Ga)N (*blue squares*) LEDs. *Dashed lines* are guides to the eyes. Adapted from [1637, 1643]

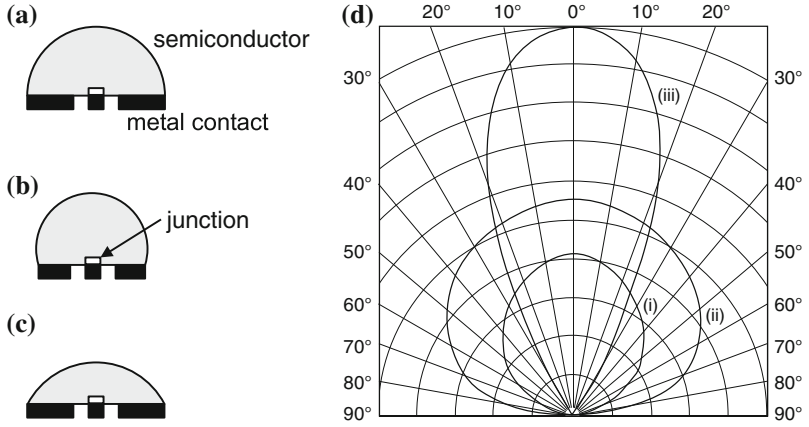


Fig. 23.6 Form of various LED casings with (**a**) hemispherical, (**b**) truncated sphere and (**c**) parabolic geometry. Adapted from [1644]. (**d**) Emission characteristics for rectangular (**i**), hemispheric (**ii**) and parabolic (**iii**) geometry. Adapted from [1645]

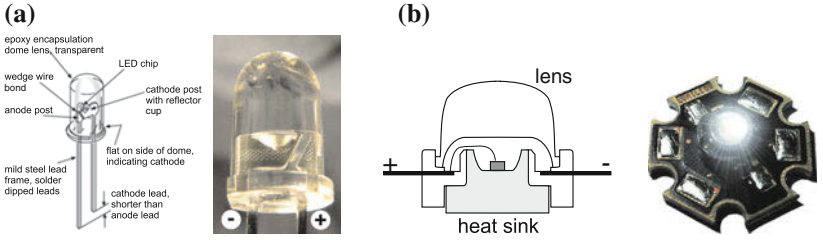


Fig. 23.7 (**a**) Standard LED casing (schematic drawing and macrophoto), (**b**) high-power mounting (schematic drawing and image of Luxeon® LED)

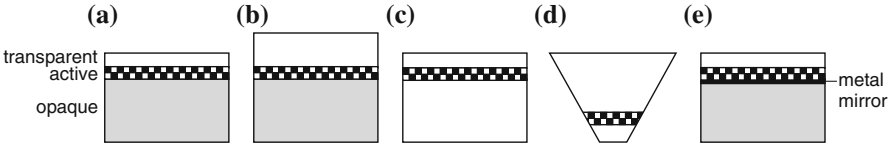


Fig. 23.8 (**a**) Standard LED layer sequence with opaque substrate (*grey*), active layer (*checkered*) and transparent top, (**b**) thick window design with thick top layer (50–70 μm). (**c**) Transparent substrate (by rebonding, see Fig. 23.9), (**d**) chip shaping (cf. also Fig. 23.11). (**e**) Thin-film LED with metal mirror (*black*) and rebonding (cf. also Fig. 23.12)

Thick-Window Chip Geometry

An increase in light extraction efficiency to about 10–12 % can be achieved if the top layer is fabricated with a much larger thickness (Fig. 23.8b) of 50–70 μm instead of a few μm. However, such approach is not scalable since larger device area would demand even larger thickness.

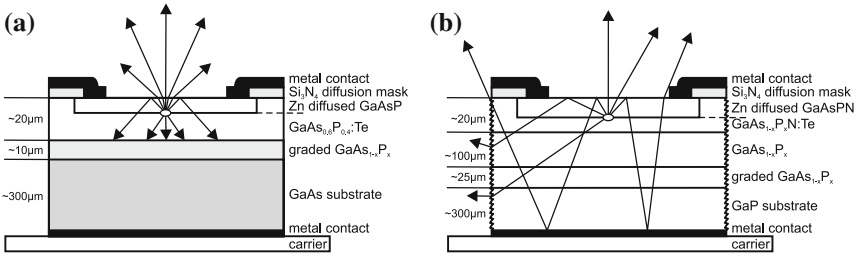


Fig. 23.9 Comparison of light paths in a GaAsP-based LED with (**a**) opaque (GaAs) and (**b**) transparent (GaP) substrate (side facets roughened). Adapted from [1646]

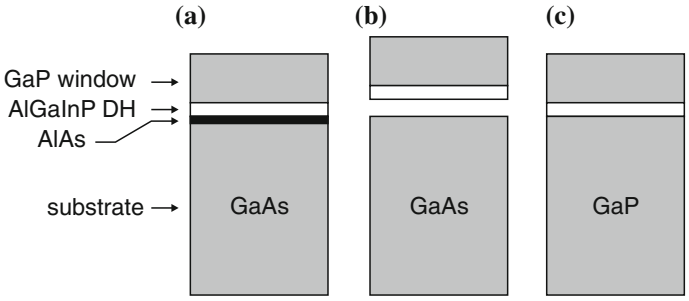


Fig. 23.10 Scheme of fabrication for red high brightness LED: (**a**) AlGaInP double heterostructure (DH) with GaP window on GaAs substrate (growth with MOCVD). (**b**) Lift-off using HF etch of sacrificial AlAs layer. (**c**) Wafer bonding on GaP (transparent for red light)

Transparent Substrate

Reflection of photons is not so detrimental if they are not lost later due to absorption in the substrate. In Fig. 23.8, the evolution of LED chip design is shown schematically. In Fig. 23.9, the light path is compared for opaque and transparent substrates. The latter provides higher light extraction efficiency due to the 'photon recycling' effect. Efficiencies of 20–25 % are possible. In Fig. 23.10, the technological steps are shown to fabricate a GaP LED with an AlGaInP active layer. The active layer is initially grown on GaAs due to lattice-match conditions.

Nonrectangular Chip Geometry

If the chip is made with an inverted structure and mounted on a mirror, a high light extraction efficiency (>50 %) can be achieved. Typical commercial designs are shown in Fig. 23.11.

The increase in quantum efficiency allows the devices to run on much higher output power. While initially LEDs delivered power only in the mW regime, now output power in the ∼1 W regime is possible (high brightness LEDs). The higher currents made a redesign of the LED mount towards better heat sinks necessary (Fig. 23.7b). While the standard case has a thermal resistance of 220 K/W

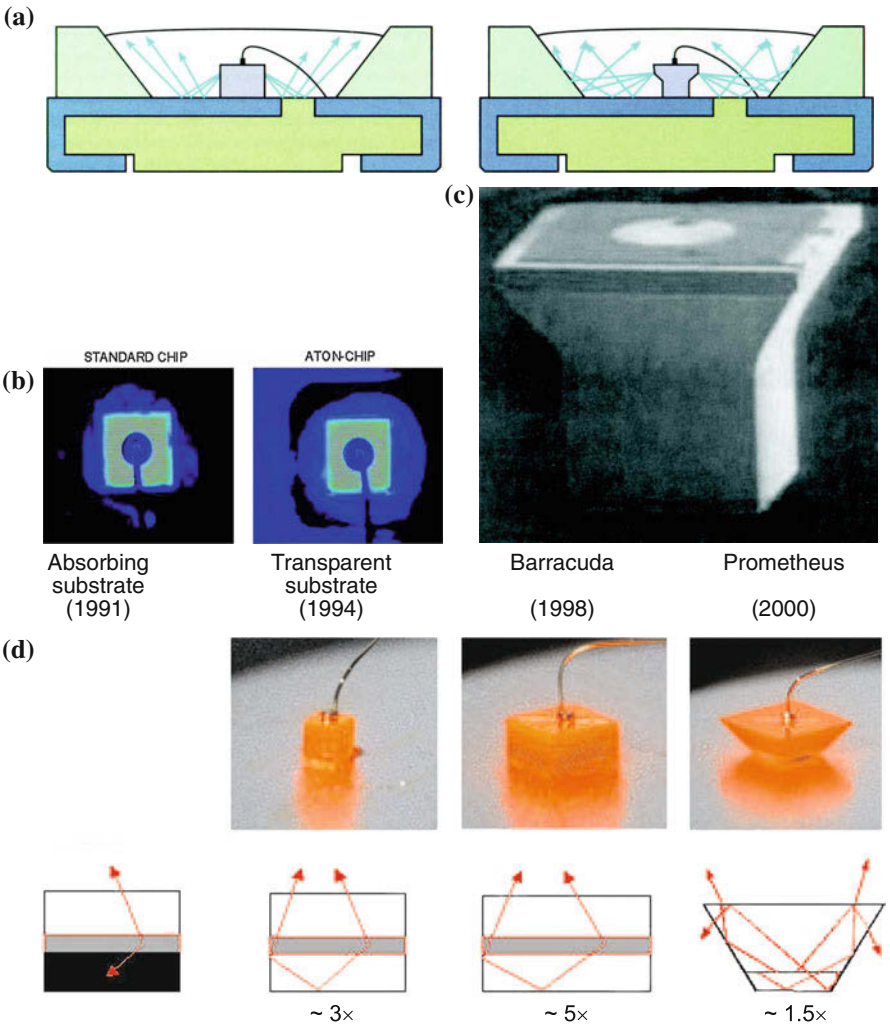


Fig. 23.11 Optimization of light exit by 3D design of the LED chip, (**a**) scheme, (**b**) emission pattern comparison and (**c**) SEM image of the ATON chip. Reprinted with permission from [1647]. (**d**) Development stages towards the truncated inverted pyramid (Prometheus) chip. From [1648]

(chip size $(0.25\,\mathrm{mm})^2$ for 0.05–0.1 W and 0.2–2 lm), the high-power case has 15 K/W (chip size $(0.5\,\mathrm{mm})^2$ for 0.5–2 W and 10–100 lm). An epoxy-free technique for encapsulation also enhances the color uniformity and maintains the brightness.

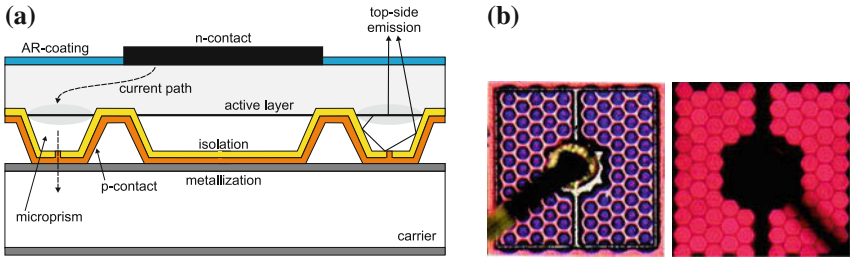


Fig. 23.12 (**a**) Scheme of thin-film flip-chip LED with microprisms. (**b**) image and luminescence image of thin film AlInGaP LED (chip length: 320 μm). Reprinted with permission from [1650]

Thin-Film LED

In the thin-film LED design [1649], as schematically shown in Fig. 23.12a, a metal mirror is evaporated on the LED layers. In a flip-chip design, subsequently the metal side is wafer bonded to another metallized substrate and the original substrate is removed. Additionally, the LED surface can be patterned (before bonding) into an (hexagonal) array of (hexagonal) microprism mesas with an insulating (e.g. silicon nitride) layer with openings in order to optimize the current path. The microprisms are optimized to allow efficient reflection of light towards the emitting surface. This technology is scalable to large areas without loss in efficiency.

In order to avoid bonding from the top which leads to partial shielding of the emitted radiation and is a mechanical process, contact schemes for flip-chip LEDs have been devised for contacting both the n- and the p-layer from the same side [1651]. In Fig. 23.13 a schematic cross-section and an emission image are shown. Here, the n-contact runs through a via hole with insulated side walls. In the emission image the array of via holes can be seen.

Bulk Flip-Chip

In the GaN material system a LED based on a triangular chip with 400 μm side length using a bulk substrate with 150 μm thickness has been presented that exhibits the so

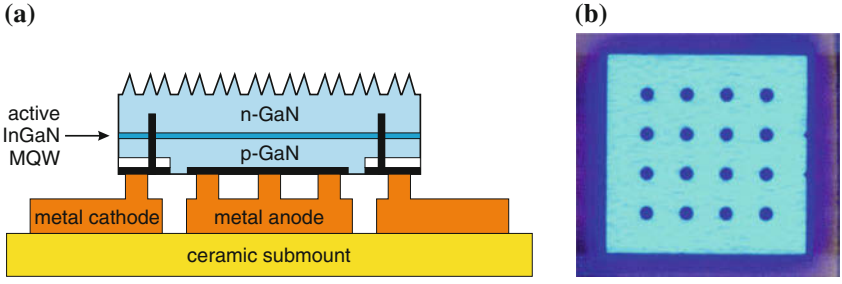


Fig. 23.13 (**a**) Scheme of thin-film flip-chip LED with rough surface and contacts from one side. (**b**) Emission image of $1 \times 1\,\mathrm{cm}^2$ chip. Adapted from [1651]

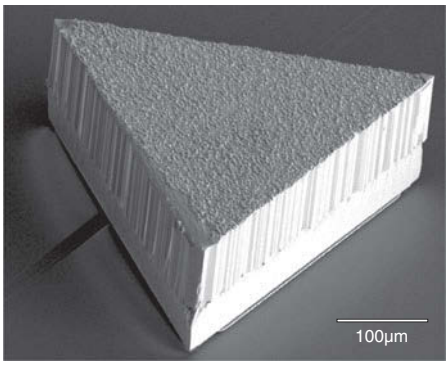


Fig. 23.14 Triangular GaN LED chip with high extraction efficiency. Both the p- and n-contact are on the bottom. Adapted from [1642]

far highest extraction efficiency of almost 90 % (Fig. 23.14) [1642, 1652]. The top and all side facets are rough.

23.3.4.1 Cascaded LEDs

Similar to the use of tunneling junctions in multi-junction solar cells (Fig. 22.64), the monolithic serial connection of several LED layers has been proposed [1653] (Fig. 23.15). For a given output power P at forward voltage V_F and current density j of a single junction LED, ideally a LED with N identical stacks and $N-1$ tunneling junctions provides the same output power P at the N-fold forward voltage, as reported for $N = 2, 3$ in [1654], and the current density j/N. Since the internal quantum efficiency of LEDs decreases with increasing current density (*droop*), such stacked LED design holds promise for increased wall-plug efficiency if the series resistance of the tunneling junctions is small. The quantum efficiency of such device is then larger than N times the efficiency of the single stack LED and thus way larger than 100% (similar to quantum cascade lasers, Sect. 23.4.16).

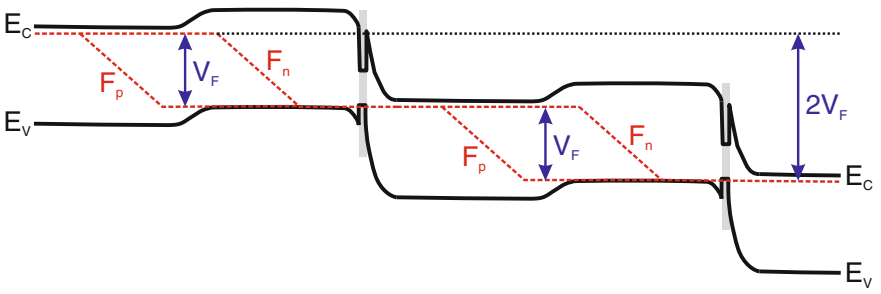


Fig. 23.15 Schematic band structure of cascaded LED. The *grey area* denotes an additional heterostructure for reduction of the series resistance of the tunnel junctions. Adapted from [1653]

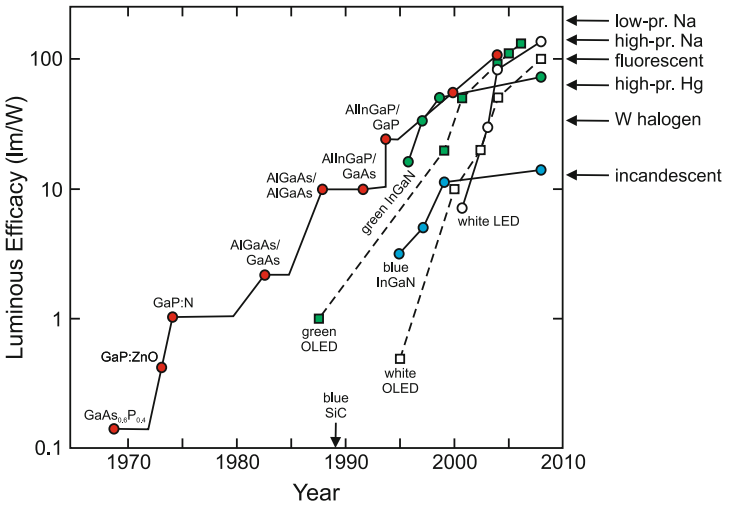


Fig. 23.16 Historic development of the luminous efficacy of semiconductor LEDs and OLEDs. Based on [1655] with the addition of data on OLEDs and recent data on LEDs. The arrows on the right indicate efficacy of various other light sources

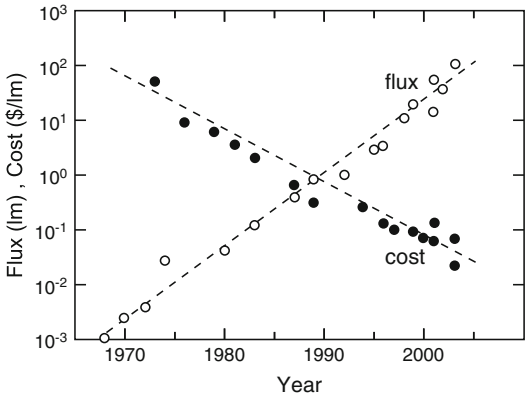


Fig. 23.17 Historic development of the flux (in lumen) and cost (in $/lm) for semiconductor LEDs. Data from [1648]

Historic Development

In Fig. 23.16, the historic development of the LED luminous efficacy (luminous flux per electrical input power) is shown for various material systems. While the luminosity has increased by a factor of 20 per decade in the last 40 years, the price has decreased by a factor of ten per decade (Fig. 23.17). Currently, there is a need for the development of efficient LEDs in the green spectral range since their luminosity is small compared to devices for the blue and red spectral regions (Fig. 23.18).

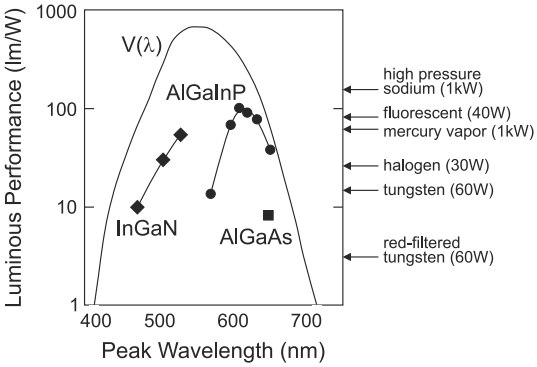


Fig. 23.18 Luminous performance of various LED materials in comparison with other light sources. Adapted from [1655]

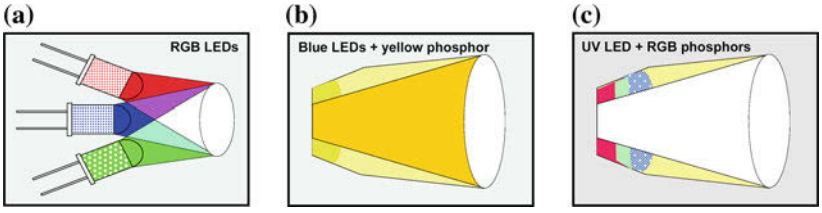


Fig. 23.19 Different strategies to generate white light with LEDs. (**a**) Additive mixing of R, G, and B LEDs, (**b**) blue LED and yellow phosphor, (**c**) UV LED (invisible) and R, G, and B phosphors. From [1648]

23.3.5 White LEDs

There are different possibilities to generate white light with an LED as shown schematically in Fig. 23.19. The highest color gamut and a tunable white point can be achieved by combining a red, a green and a blue LED (Fig. 23.19a). Using a blue LED and a yellow phosphor (Figs. 23.19b and 23.20a, b), a white spectrum can be achieved that is, however, not very close to a blackbody spectrum (Fig. 23.20c). A better color rendering can be obtained with the combination of two phosphors [1656]. With an UV LED that is itself invisible (and must be shielded so no UV radiation leaves the LED), phosphors with various colors can be pumped (Fig. 23.19c). The mix of phosphors determines the white point.

Using a blue-emitting LED based on (In,Ga)N material, phosphors (similar to those used in fluorescence lamps) can be pumped. Blue light is converted into green, yellow or red light such that the resulting total spectrum appears white to the human eye. Also, a broad range of other colors can be designed (color on demand), e.g. pink or particular corporate colors.

The color of a white LED depends on the operation conditions. In Fig. 23.21a the intensity versus dc driving current characteristic of a white LED is shown. In Fig. 23.21b the chromaticity coordinates are shown for various dc currents. A change of wavelength with forward voltage occurs for the blue-emitting (In,Ga)N material

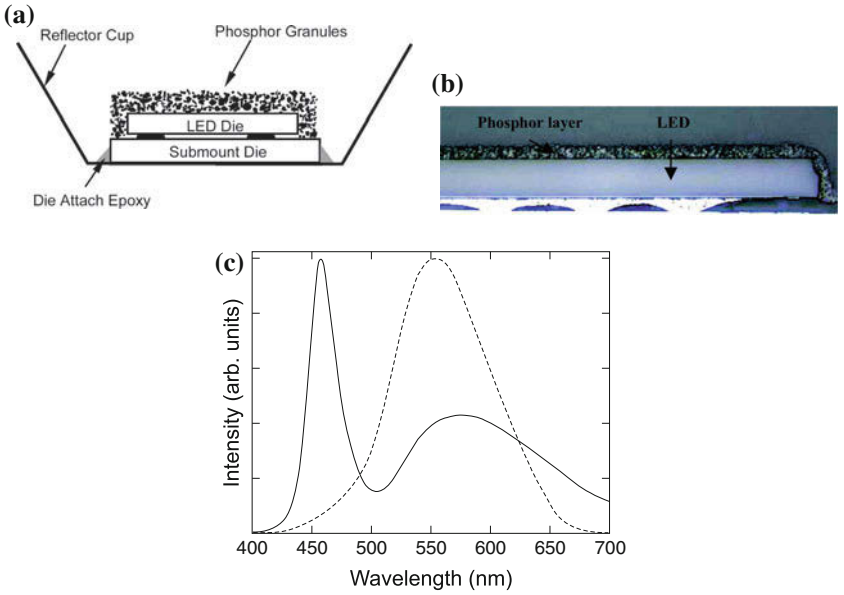


Fig. 23.20 (**a**) Scheme and (**b**) image of color conversion Luxeon® LED. From [1648]. (**c**) Spectrum (*solid line*) of white LED with blue LED pumping a yellow phosphor together with eye-sensitivity curve $V(\lambda)$ (*dashed line*). Adapted from [1650]

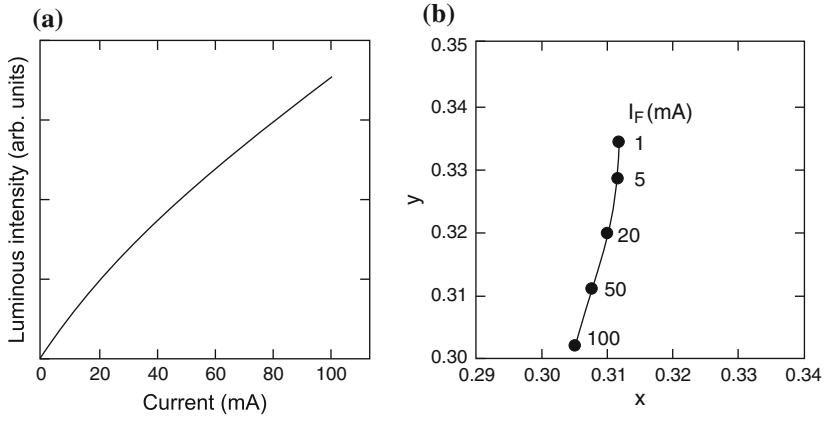


Fig. 23.21 (**a**) Luminous intensity of white LED (NSCW215) versus dc forward current. (**b**) CIE chromaticity coordinates for various dc driving conditions as labeled. Data taken from [1657]

due to filling of low-energy states (Fig. 23.22). In order to avoid this effect, the LED is driven with pulses of a fixed current amplitude and a repetition frequency that is high enough to provide a flicker-free image to the human eye, e.g. 100 Hz. The

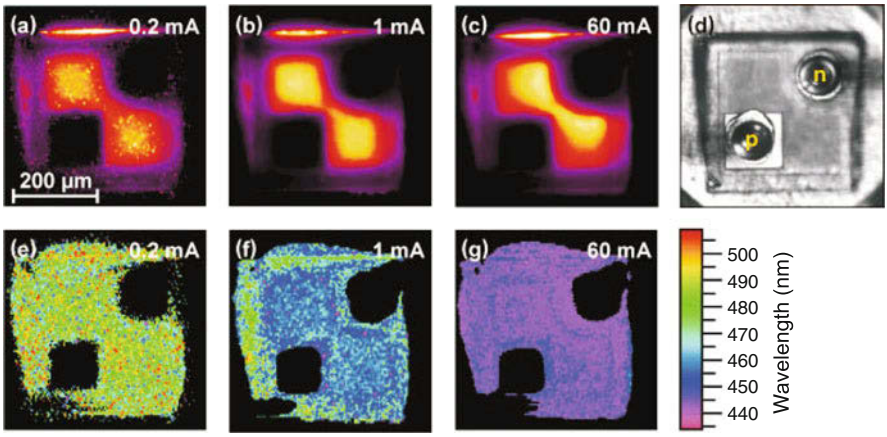


Fig. 23.22 (**a**–**c**) Electroluminescence intensity map of (In,Ga)N LED at different currents as labeled. (**d**) Optical image of the LED chip (*top view*). (**e**–**g**) Wavelength images (wavelength of spectral emission maximum) for different currents. Adapted from [1658]

intensity of the LED is modulated via the pulse width, i.e. between 0–10 ms in this case (PWM, pulsewidth modulation).

The direct light from white LEDs is visually appealing and cannot be distinguished from a blackbody source of matching temperature. But since the spectral power distribution of white LEDs is different from natural light, objects illuminated by such light source can appear in 'wrong' colors. For the spectrum of Fig. 23.20c, in particular green is reproduced poorly. A quantitative measure for the ability of a light source to reproduce the colors of an illuminated object faithfully compared with a natural (blackbody) light source is the *color rendering index* (CRI).

A major advantage of LEDs for display and lighting applications is their long lifetime compared to halogen (about 2000 h), xenon (10,000 h) or fluorescent (6000–10,000 h) lights. Philips Lumileds projects (for white LUXEON® K2 LEDs)

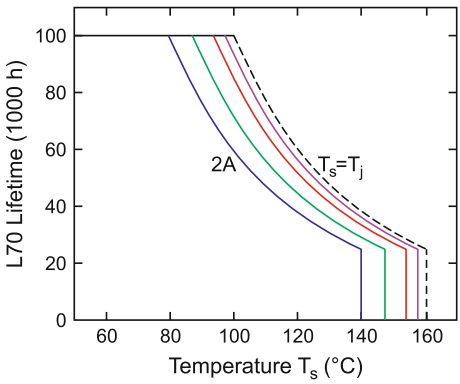


Fig. 23.23 Lifetime for 70 % lumen as a function of solder temperature T_s (for white Diamond Dragon® LED) for various driving currents (0.3, 0.7, 1.4 and 2.0 A, *solid lines* from right to left). The *dashed line* is for low driving current and $T_s = T_j$. Adapted from [1662]

70 % lumen maintenance at given current (1 A) at 50,000 h for junction temperature $T_j \leq 120°$ [1659]. Similar values are given by OSRAM [1662] for white high power LEDs (Fig. 23.23). A current problem is the decrease of efficiency with increasing current density termed *droop*, probably due to Auger recombination [1660, 1661].

23.3.6 Quantum Dot LEDs

Quantum dots are an interesting active medium for LEDs due to their spectroscopic properties (Sect. 14.3.4).

Ultranarrow Spectral Emission

A LED based on a single QD exhibits a rather unique spectrum consisting of a single spectral line, at least at low temperatures [1663], due to exciton recombination as shown in Fig. 23.24. Such device can deliver single photons on demand and be a photon source for quantum cryptographic communication. At higher current also biexciton recombination appears. In [1664, 1665] the triggered emission of photon pairs from cascade-like XX and X recombination in a single dot and their polarization entanglement is reported. Entanglement is related to degenerate X and XX emission energy [1666] (cmp. Fig. 14.33).

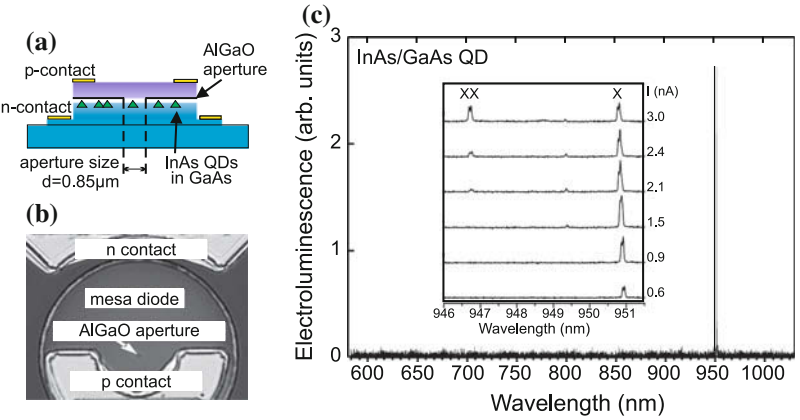


Fig. 23.24 (**a**) Schematic cross-section of QD LED. Current is fed to a single QD via an oxide aperture. (**b**) Plan-view SEM image of QD LED. **c** Electroluminescence (EL) spectrum ($T = 10\,\mathrm{K}$, $U = 1.65\,\mathrm{V}$, $I = 0.87\,\mathrm{nA}$) of single InGaAs/GaAs QD LED (diameter of oxide aperture $0.85\,\mu\mathrm{m}$, thickness 60 nm). The single line is due to (neutral) exciton recombination. The *inset* shows dependence of EL spectrum on injection current; at higher currents a second peak due to biexciton recombination (XX→X) appears. Adapted from [1667]

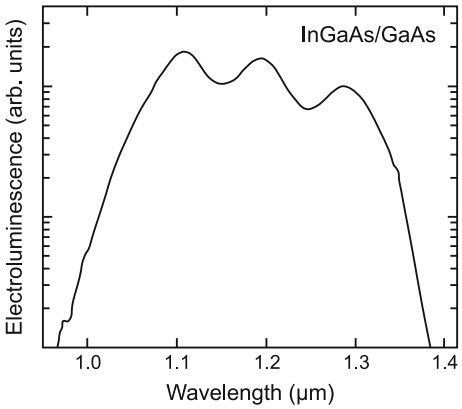


Fig. 23.25 Electroluminescence spectrum of a quantum dot LED designed for broad spectral range (at 5 kA/cm^2). Adapted from [1669]

Ultrabroad Spectral Emission

An LED based on the emission from a quantum dot ensemble exhibits a fairly broad spectrum because of inhomogeneous broadening due to size fluctuations of the quantum dots (cmp. Fig. 14.36). Additionally several ensembles of QDs with different mean emission wavelength can be incorporated in a device, e.g. in stacked layers [1668]. This way ultrabroad electroluminescence spectra can be realized (Fig. 23.25). Also emission on the ground and excited state can be used for broad spectral emission.

23.3.7 Organic LEDs

An organic light emitting diode (OLED) is made from organic semiconductors. The pioneering work was made by Tang and Van Slyke [1503, 1670]. Typical layer sequences are depicted in Fig. 21.60. The light emission occurs through the anode (and the transparent ITO layer) while the metal cathode is opaque. Two major configurations are possible, emission through the transparent substrate (glass) or top emission (Fig. 23.26).

The optimization of materials for the various functional layers is ongoing. The emission layer (EML) is optimized for efficient radiative recombination for the design wavelength or wavelength range. The highest efficacy of over 100 lm/W (Fig. 23.16) is achieved using phosphorescent materials (Sect. 17.6). The contacts are optimized for high carrier injection efficiency and the transport layers are optimized for high conductivity.

End of 2007 a transparent white OLED panel was introduced [1671] (Fig. 23.27a). Its transparency is 55 % and shall be improved in the future. A crucial point is

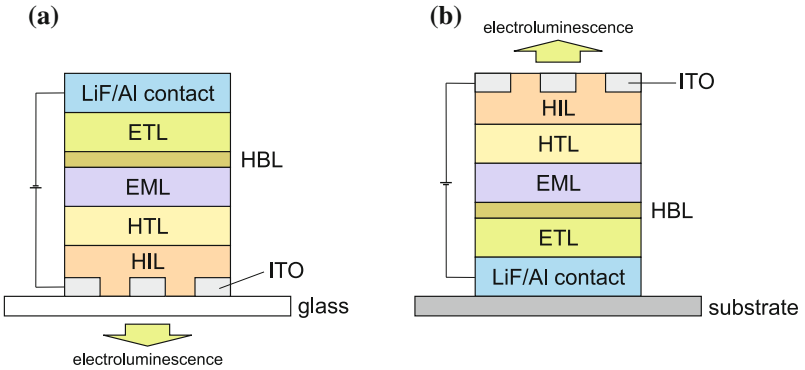


Fig. 23.26 Typical OLED design for (**a**) bottom and (**b**) top emission

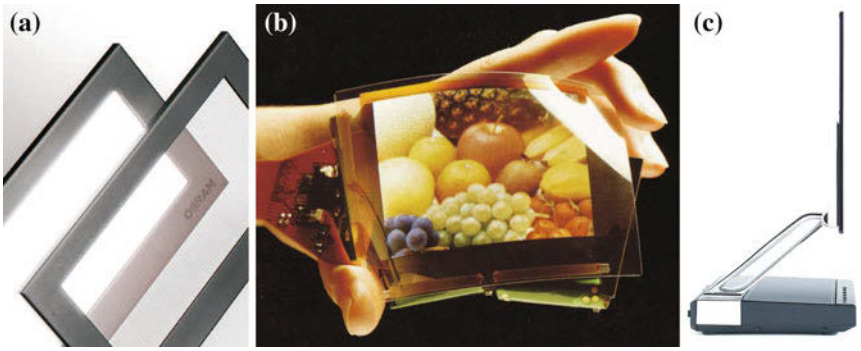


Fig. 23.27 (**a**) Transparent OLED panel. From [1671]. (**b**) Flexible OLED display. From [1672]. (**c**) 3 mm thin, 11 inch diagonal OLED TV. From [1673]

the protection of the organic films against moisture and air. The encapsulation by glass is very good. Flexible OLED panels with polymer substrate and encapsulation have been demonstrated (Fig. 23.27b). OLED technology is currently used for small displays in digital cameras and cellular phones. It enables very thin TV panels, only a few mm thick (Fig. 23.27c) entering the mass market in 2010. An increase of the lifetime from 30,000 to beyond 50,000 h is expected.

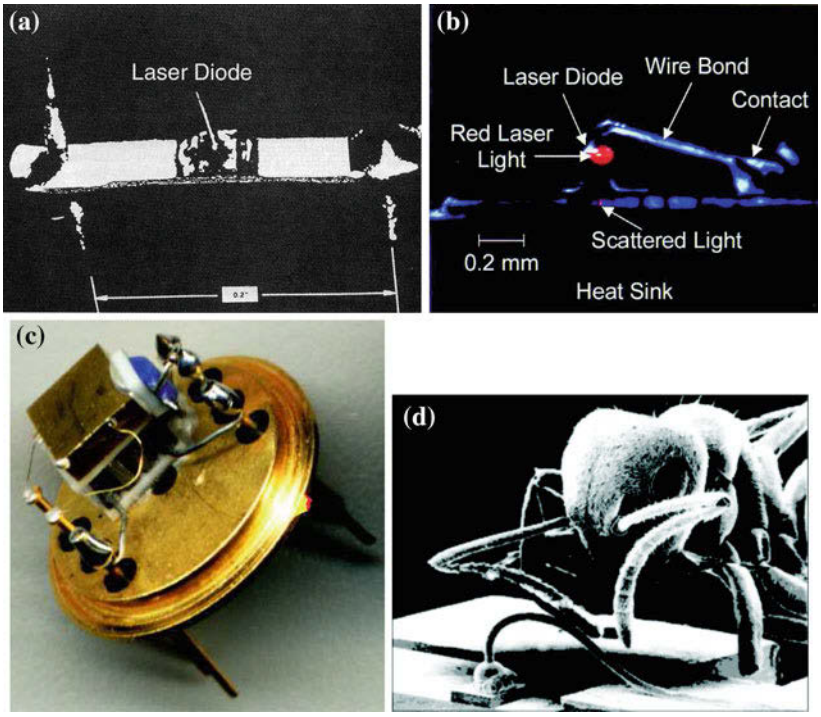


Fig. 23.28 Images of the first semiconductor lasers, 1962: (**a**) GaAs laser, Lincoln Laboratories and (**b**) GaInP laser, N. Holonyak and S.F. Bevacqua, Urbana Champaign. (**c**) Laser (at the end of gold bond wire) mounted on Peltier heat sink and a TO chip, Universität Leipzig. (**d**) Size comparison of an ant with a laser chip (underneath the bond wire)

23.4 Lasers

23.4.1 Introduction

Semiconductor lasers[13] [1674, 1675] contain a zone (mostly called the *active layer*) that has gain if sufficiently pumped and that overlaps with an optical wave. The wave bounces back and forth in an optical cavity that leads to optical feedback. The part of the wave that exits the semiconductor forms the laser beam. Some of the first semiconductor lasers and a mounting design are shown in Fig. 23.28.

[13]The term 'laser' is an acronym for 'light amplification by stimulated emission of radiation'. The amplification relies on stimulated emission, theoretically predicted by Einstein in 1917. The laser concept was first explored in the microwave wavelength region (1954, MASER using ammonia, Ch.H. Townes, Nobel prize 1964). The first optical laser (1958, US patent No. 2,929,922 awarded 1960, A.L. Schawlow, Ch.H. Townes) was the ruby laser developed in 1960 by Th. Maiman. A device is a laser when it emits stimulated light. This light must neither be monochromatic nor be emitted in a narrow, directed beam.

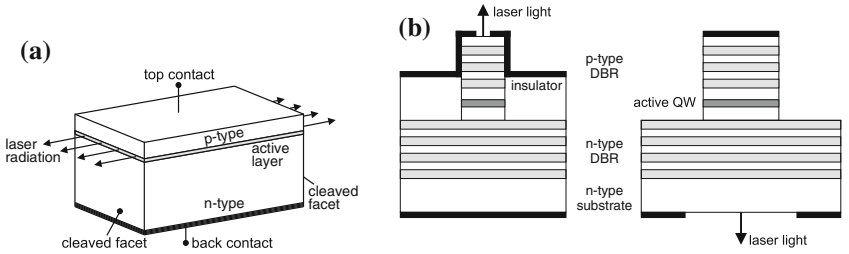


Fig. 23.29 (**a**) Schematic drawing of edge-emitting semiconductor laser. (**b**) Schematic drawings of vertical-cavity surface-emitting lasers with top emission (*left*) and emission through the substrate (*right*). *Black areas* are metal contacts

Generally, two main geometrical laser types, *edge* emitters (Fig. 23.29a) and *surface* emitters (Fig. 23.29b), are distinguished. The emission of the edge emitter exits through cleaved {110} side facets[14] (≈30 % reflectance), of which an opposite pair acts as a Fabry–Perot optical cavity. The surface emission is directed along (001), since this is the (standard) growth direction of the heterostructure sequence making up the laser. The mirrors in a vertical-cavity surface-emitting laser (VCSEL) are made from dielectric Bragg mirrors (cf. Sect. 19.1.4) with typically $R > 99.6\,\%$. Using antireflection coating on one facet, semiconductor lasers can be set up with an external cavity.[15] If both facets are antireflection coated, feedback is missing and the chip can be used as an optical amplifier (cf. Sect. 23.5).

Most lasers are pn-diodes and are then called laser diodes. They rely on the gain of interband transitions and the emission wavelength is determined and (more or less) given by the band gap of the semiconductor. The cascade laser [1676] (Sect. 23.4.16) is a unipolar structure with a superlattice as active layer. Here, the intersubband transitions (mostly in the conduction band but also in the valence band) carry the gain. The emission wavelength depends on the subband separation and lies typically in the far- and mid-infrared. Extensions to the THz regime and also to shorter wavelengths are possible. A third type of laser is the 'hot-hole' laser (Sect. 23.4.17), typically fabricated with p-doped Ge, which can be viewed as unipolar and functions only in a magnetic field; its emission is in the THz regime.

23.4.2 Applications

In Fig. 23.30, the revenue in the worldwide diode laser market is shown. The drop after 2000 is due to the burst of the 'internet bubble'. Nondiode laser (gas, ruby,

[14] Or etched facets in possibly any direction.

[15] Such external cavities can be used for manipulation of the laser properties such as wavelength tuning.

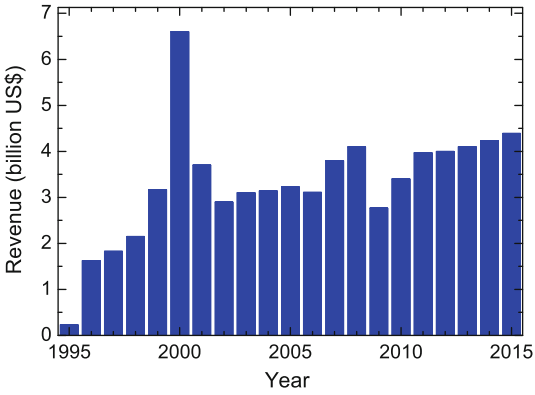


Fig. 23.30 Revenue in worldwide diode laser market. Based on numbers from [1677], data or 2015 estimated

excimer, …) revenue is currently stable at around 2 billion US$, thus semiconductor lasers account for the largest share of all laser types sold.

The following applications rely on semiconductor lasers:

- optical communication, mostly optical fiber based (senders), typically at 10 GBit/s, in special situations also 40 GBit/s data rate.
- optical information storage and retrieval (CD, DVD, BD[16]) with as short of a wavelength as possible, as shown in Fig. 23.31, currently 405 nm.
- pumping of solid-state lasers, typically 910 or 940 nm for pumping Nd:YAG.
- portable projectors, laser TV, entertainment.
- laser pointers, see Fig. 23.32. A red laser pointer simply uses the collimated red emission of a GaAs-based diode. In a green laser pointer, an infrared diode pumps a Nd:YAG or Nd:YVO_4 crystal. The emitted beam is then frequency doubled, typically with a $KTiOPO_4$ (KTP) crystal.
- medical instruments with a variety of wavelengths in ophthalmology, dermatology, cosmetics (hair removal, tattoo removal).
- various other uses, such as remote control, position detection, distance measurement, printing, scientific instrumentation.

The market for photonic devices is much more dynamic than the electronics market. An example is the rapid change of dominating laser applications. For diode lasers, the two most prominent applications are telecommunication (77 % market share in 2000, 25 % in 2003, 45 % in 2008) and optical data storage (17 % market share in 2000, 60 % in 2003, 44 % in 2008).

[16] 16 million 405 nm laser diodes were shipped in 2006–2008. 85 % of those are built into SONY's PS3, the rest into HD-DVD and other Blu-ray™disc (BD) players.

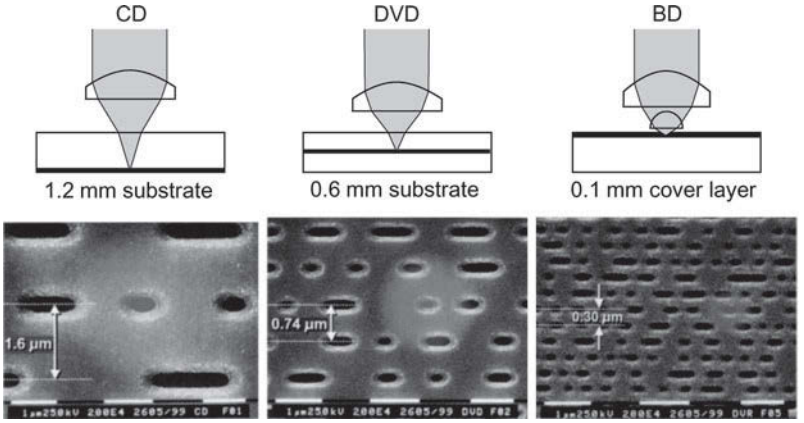


Fig. 23.31 Evolution of optical data storage technology, 'CD': compact disk (laser: 780 nm, pitch: 1.6 µm, capacity: 0.7 GB), 'DVD': digital versatile disk (laser: 635–650 nm, pitch: 0.74 µm, capacity: 4.7 GB for one layer), 'BD': 'Blu-ray' disk (laser: 405 nm, pitch: 0.32 µm, capacity: 27 GB for one layer)

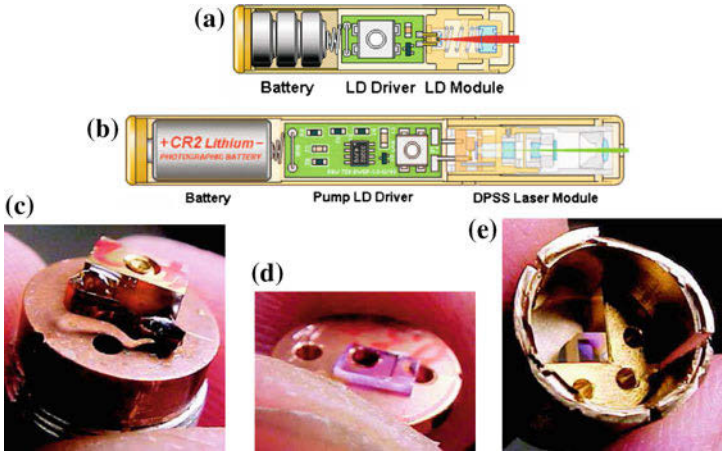


Fig. 23.32 (**a**) Scheme of red laser pointer, (**b**) scheme of green laser pointer. Parts of a green laser pointer: (**c**) pump laser diode, (**d**) YVO_4 crystal, (**e**) KTP doubler

23.4.3 Gain

Due to current injection,[17] a nonequilibrium carrier distribution is created. After fast thermalization processes (phonon scattering), it can mostly described by quasi-Fermi

[17] Or due to optical pumping. If electrical contacts are not available, the laser action can be invoked by supplying a high-intensity light beam, possibly in a stripe-like shape. For optically pumped semiconductor lasers see Sect. 23.4.15.

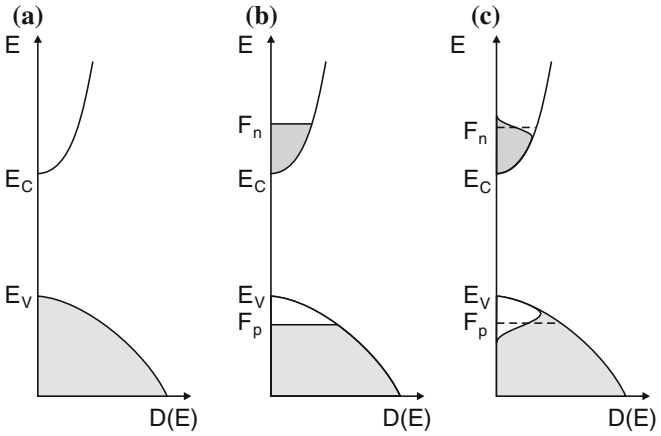


Fig. 23.33 Population (**a**) in thermodynamic equilibrium $T = 0\,\mathrm{K}$, (**b**) under inversion for $T = 0\,\mathrm{K}$, (**c**) under inversion for $T > 0\,\mathrm{K}$. *Shaded areas* are populated with electrons

levels. Sufficiently strong pumping leads to inversion, i.e. conduction-band states are more strongly populated with electrons than valence-band states (Fig. 23.33). In this case, the stimulated emission rate is stronger than the absorption rate (cf. Sect. 10.2.6). The thermodynamic laser condition (cf. (10.23)) requires the difference of the quasi-Fermi levels to be larger than the band gap.

$$F_\mathrm{n} - F_\mathrm{p} > E_\mathrm{g} \tag{23.10}$$

The gain is defined as the (frequency-dependent) coefficient $g(\hbar\omega)$ that describes the light intensity along a path L according to

$$I(L) = I(0)\,\exp(g\,L)\,. \tag{23.11}$$

The gain spectrum as a function of the photon energy $\hbar\omega$ is given for non-k-conserving recombination by (cf. (10.5) and (10.6))

$$g(\hbar\omega) = \int_0^{\hbar\omega - E_\mathrm{g}} D_\mathrm{e}(E) D_\mathrm{h}(E') \left[f_\mathrm{e}(E) f_\mathrm{h}(E') - (1 - f_\mathrm{e}(E))(1 - f_\mathrm{h}(E')) \right] \mathrm{d}E, \tag{23.12}$$

with $E' = \hbar\omega - E_\mathrm{g} - E$. The gain is positive for those photon energies for which light is amplified and negative for those that are absorbed. In Fig. 23.34a, the electron and hole concentrations are shown for GaAs as a function of the quasi-Fermi energies. In Fig. 23.34b, the difference of the quasi-Fermi energies is shown as a function of the carrier density (for neutrality $n = p$). The gain spectrum is shown in Fig. 23.34c

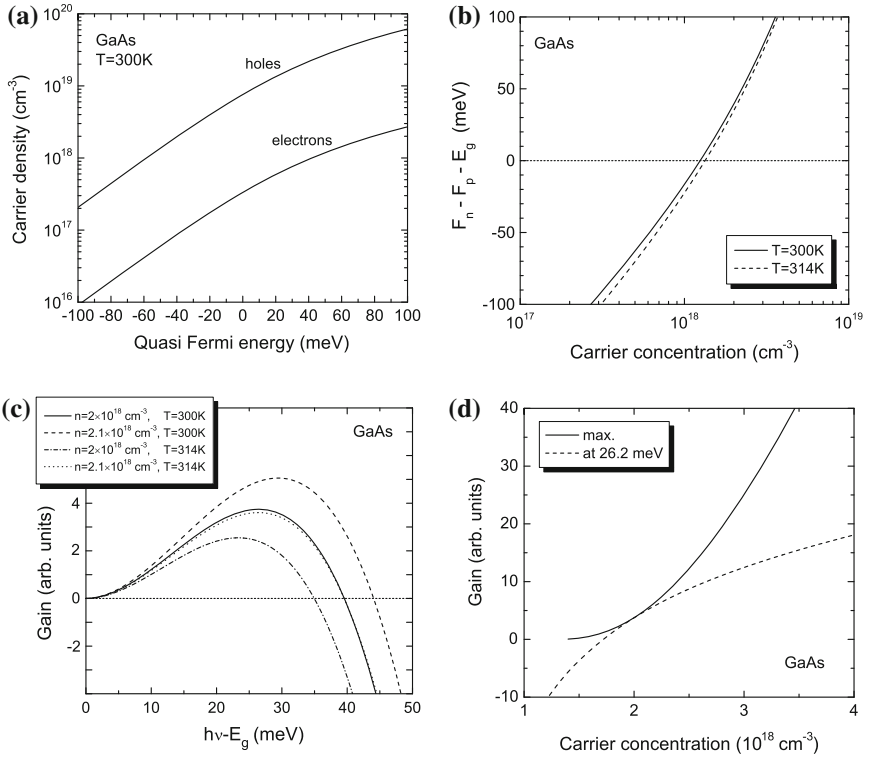


Fig. 23.34 Gain in the two-band model for GaAs. (**a**) Electron and hole concentrations at $T = 300\,\mathrm{K}$ as a function of the quasi-Fermi energies counted relative to the band edges, i.e. $F_\mathrm{n} - E_\mathrm{C}$ and $E_\mathrm{V} - F_\mathrm{p}$. (**b**) Difference of quasi-Fermi levels as a function of carrier concentration ($n = p$) for GaAs at two different temperatures. (**c**) Gain spectra according to (23.12) for $n = 2 \times 10^{18}$ and $T = 300\,\mathrm{K}$ (*solid line*), increased carrier density $n = 2.1 \times 10^{18}$ and $T = 300\,\mathrm{K}$ (*dashed line*), higher temperature $n = 2 \times 10^{18}$ and $T = 314\,\mathrm{K}$ (*dash-dotted line*) and same difference of the quasi-Fermi levels as for the *solid line*, $n = 2.1 \times 10^{18}$ and $T = 314\,\mathrm{K}$ (*dotted line*). (**d**) Maximum gain (*solid line*) and gain at a particular energy (*dashed line*, for photon energy $E_\mathrm{g} + 26.2\,\mathrm{meV}$ for which the gain is maximal for $n = 2 \times 10^{18}$ and $T = 300\,\mathrm{K}$, see *solid line* in part (**c**))

for a simple two-band model.[18] For a more elaborate discussion of such matters we refer to [1678]. In the case of inversion, the gain is positive for energies between E_g and $F_\mathrm{n} - F_\mathrm{p}$. At $\hbar\omega = F_\mathrm{n} - F_\mathrm{p}$, the gain is zero (transparency) and for larger energies negative (positive absorption coefficient). The agreement of experimental gain spectra of quantum wells with theoretical considerations, including carrier collision effects at the level of quantum kinetic theory in the Markovian limit, is very good (Fig. 23.35a) [1679].

[18] One electron and one hole band are considered; the heavy and light hole bands are taken into account via the mass according to (6.69).

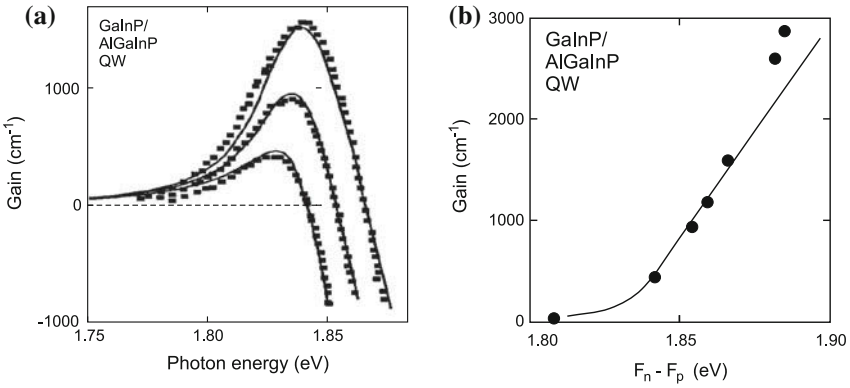


Fig. 23.35 (**a**) Gain spectra of a 6.8 nm thick $Ga_{0.41}In_{0.59}P/(Al_{0.5}Ga_{0.5})_{0.51}In_{0.49}P$ quantum well, experimental data (*symbols*) and theory (*lines*) for three different sheet carrier densities $n_{2D} = 2.2$, 2.7, and $3.2 \times 10^{12}\,\mathrm{cm}^{-2}$. (**b**) Maximum gain as a function of the separation of the quasi-Fermi levels, experimental data (*symbols*) and theory (*lines*). Adapted from [1679]

For a given fixed energy, the gain increases with increasing pumping and increasing carrier density n (Fig. 23.34d). For very small density, it is given as $g(n \to 0) = -\alpha$. The gain rises around transparency approximately linearly with the pumping intensity. At transparency carrier density n_{tr}, the gain is zero. Therefore, the relation $g(n)$ can be approximated as (linear gain model)

$$g(n) \cong \hat{\alpha}\,\frac{n - n_{\mathrm{tr}}}{n_{\mathrm{tr}}}. \tag{23.13}$$

For large carrier density, the gain saturates (at a value similar to α). The onset of positive gain is related to the separation of the quasi-Fermi levels being larger than the band gap (23.10), Fig. 23.35b. The gain in quantum dot lasers [1680] has been discussed in [1681].

23.4.4 Optical Mode

The light wave that is amplified must be guided in the laser. An optical cavity is needed to provide optical feedback such that the photons travel several times through the gain medium and contribute to amplification. We explain the light-wave management for the edge emitter first:

Vertical Mode Guiding

In the course of the historical development of the semiconductor laser, the most significant improvements (reduction of lasing threshold current) have been achieved through the improvement of the overlap of the optical wave with the gain medium, as

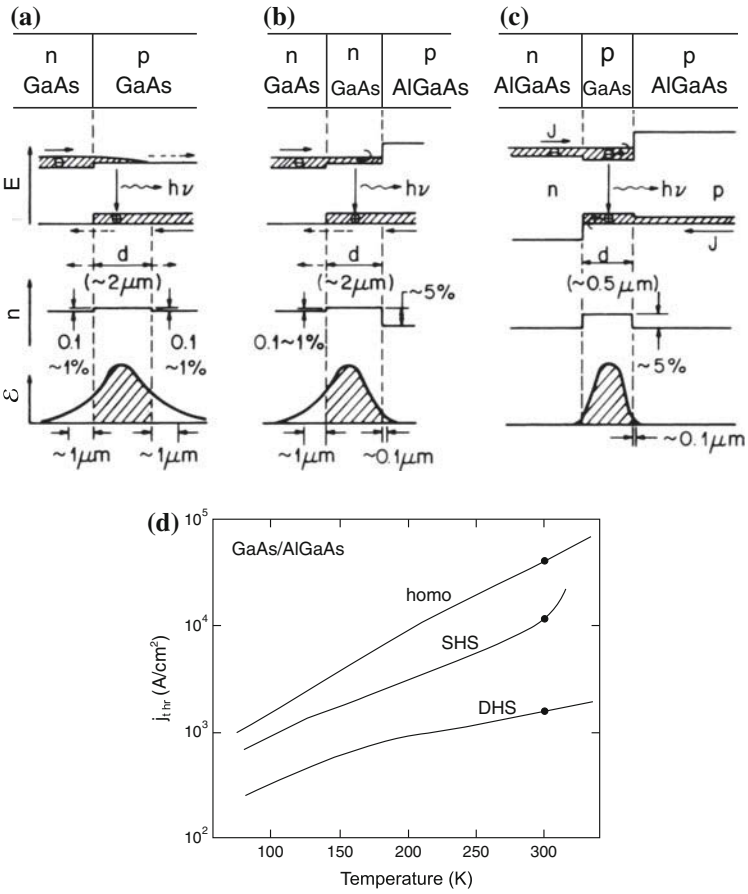


Fig. 23.36 Laser with (**a**) homojunction, (**b**) single heterostructure (SHS), (**c**) double heterostructure ('DHS'), (**d**) reduction of threshold current with design progress ('SHS': $d = 2\,\mu\text{m}$, 'DHS': $d = 0.5\,\mu\text{m}$). Adapted from [1682]

shown in Fig. 23.36. From homojunctions over the single heterojunction, eventually the double heterostructure (DHS) design could reduce the laser threshold current density to the $1\,\text{kA/cm}^2$ level.

The band diagram of a double heterostructure is shown in Fig. 23.37 for zero and forward bias. In the DHS, the optical mode is guided by total reflection within the low band gap center layer, which has a larger index of refraction than the outer, large band gap layer.[19] When the layer thickness is in the range of λ/n_r, the form of the optical mode must be determined from the (one-dimensional) wave equation (Helmholtz equation) for the electrical field E_z

[19] A smaller band gap coincides for many cases with a larger index of refraction.

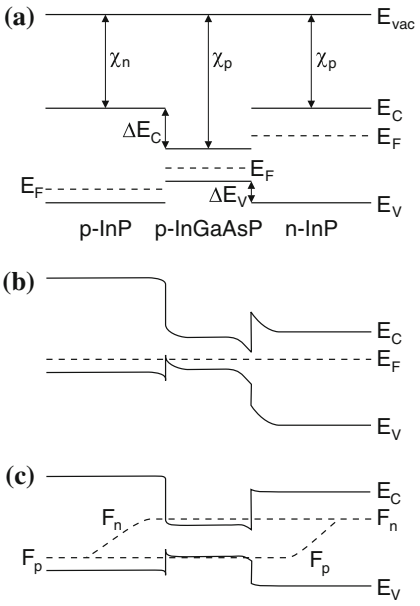


Fig. 23.37 Schematic band diagram of a pn double heterostructure (DHS) diode (InP/InGaAsP/InP) (**a**) before contact of the materials, (**b**) in thermodynamic equilibrium (zero bias, *dashed line* is Fermi level E_F = const.), (**c**) with forward bias close to flat-band conditions, *dashed lines* are quasi-Fermi levels

$$\frac{\partial^2 E_z}{\partial z^2} + \omega^2 \mu \epsilon(z) E_z = 0. \tag{23.14}$$

In Fig. 23.38a, the shape of the optical mode for GaAs/$Al_{0.3}Ga_{0.7}As$ DHS with different GaAs thickness is shown.

The optical confinement factor Γ is the part of the wave that has geometrical overlap with the gain medium, i.e. is subject to amplification. It is shown for GaAs/$Al_xGa_{1-x}As$ DHS with different GaAs thickness and different Al concentration in Fig. 23.38b. The modal gain g_{mod} that is responsible for light amplification in the cavity consists of the material gain g_{mat} due to inversion and the optical confinement factor.

$$g_{mod} = \Gamma g_{mat}. \tag{23.15}$$

In order to allow simultaneous optimization of the light mode and the carrier confinement, the separate confinement heterostructure (SCH) has been designed. Here, a single or multiple quantum well of a third material with even smaller band gap is the active medium (Fig. 23.39a, b, d). A single quantum well has an optical confinement factor of a few per cent only. It offers, however, efficient carrier capture and efficient radiative recombination. An increase in the carrier capture efficiency can be achieved using a graded index in the barrier (GRINSCH, Fig. 23.39c).

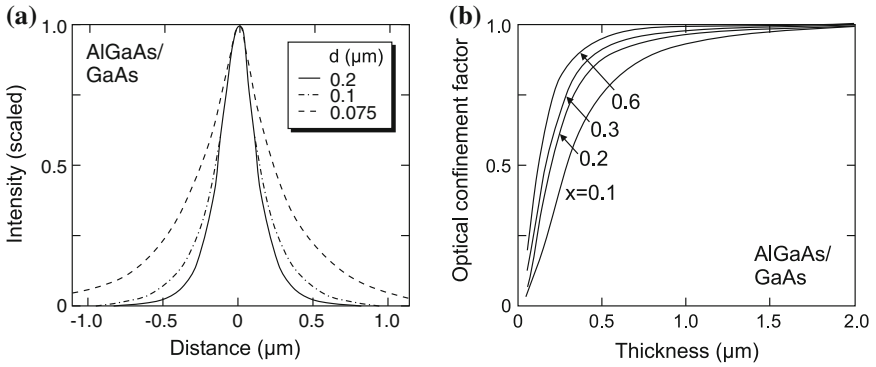


Fig. 23.38 (**a**) Optical mode (relative intensity) for various values of the thickness d of the active layer as labeled of a GaAs/$Al_{0.3}Ga_{0.7}As$ DHS laser, (**b**) Optical confinement factor Γ as a function of the thickness of the active layer and the Al concentration x of the barrier as labeled in a GaAs/$Al_xGa_{1-x}As$ DHS laser. Adapted from [1674]

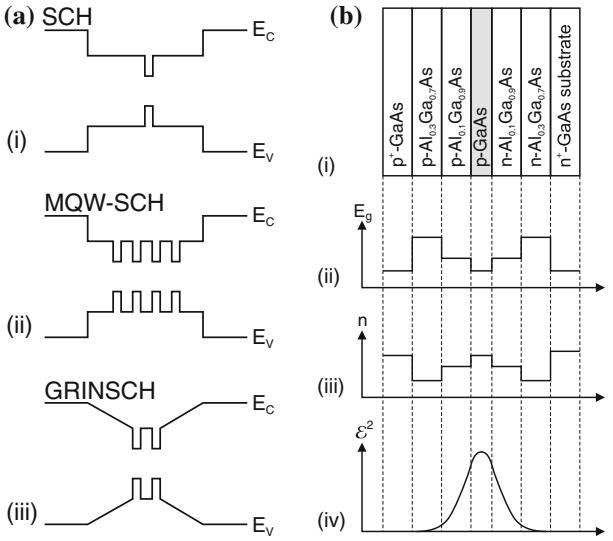


Fig. 23.39 (**a**) Various geometries of the active layer of a DHS laser with quantum wells as active medium, (**i**) single QW (separate confinement heterostructure, SCH), (**ii**) multiple QW SCH, and (**c**) GRINSCH (graded-index SCH) structure. (**b**) Layer sequence for a separate confinement heterostructure laser

The thin wave-guiding layer leads to large divergence of the laser beam along the vertical direction, typically about 90°. The strong confinement of light also limits the maximum achievable output power due to catastrophic optical damage (COD). Several ideas have been realized to overcome this problem and achieve much smaller divergence of about 18°. The waveguide can be designed to be very thick (large optical

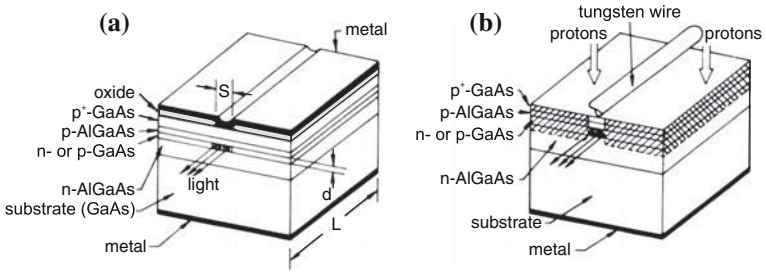


Fig. 23.40 Scheme of gain-guided lasers with stripe contact: (**a**) oxide stripe, (**b**) proton implanted with shadow mask from tungsten wire ($\sim 10\,\mu$m). Adapted from [1683]

cavity, LOC) that leads to an increase of threshold. Other schemes are insertion of a low-index layer into the confinement layer, insertion of a high-index layer into the cladding layer or the use of high-index quarter-wavelength reflecting layers [1684].

Lateral Mode Guiding

Lateral waveguiding can be achieved with gain guiding and index guiding (or a mixture of the two). In the gain-guiding scheme (Fig. 23.40), the current path that is defined by the stripe contact and the current spreading underneath it, defines the gain region and therefore the volume of amplification that guides the optical wave. Since a high carrier density reduces the index of refraction, a competing antiguiding effect can occur. For index guiding, the lateral light confinement is caused by a lateral increase of the index of refraction. This index modulation can be achieved by using a mesa-like contact stripe (Fig. 23.41a). A shallow mesa reaches down into the upper cladding, a deep mesa reaches down into or through the active layer. Possible problems with surface recombination can be avoided by regrowth of the structure (Fig. 23.41b) with a wide band gap material (compared to the active layer). Optimization of regrowth is targeted to achieve a well-defined surface for subsequent contact processing. A lateral pn-diode can be incorporated that avoids current spreading in the upper part of the structure.

Depending on the width of the lateral mode, it can be monomode or multimode (Fig. 23.42a, b). For laterally monomode lasers, the stripe width may only be a few μm. In particular for such lasers, the current spreading must be controlled. Problems can arise for wide stripe widths due to current filamentation and inhomogeneous laser emission from the facet. Since the optical mode is typically more confined in the growth direction than in the lateral direction, the far field is asymmetric (Fig. 23.43a, b). The vertical axis has the higher divergence and is called the fast axis. The lateral axis is called the slow axis. The asymmetric beam shape is detrimental when the laser needs to be coupled into an optical fiber or a symmetric beam profile is needed for subsequent optics. The beam can be made symmetric using special optic components such as anamorphic prisms (Fig. 23.43c) and graded-index lenses. The beam from a laterally monomode laser is diffraction limited and can therefore generally be refocused efficiently (beam quality $M^2 \gtrsim 1$).

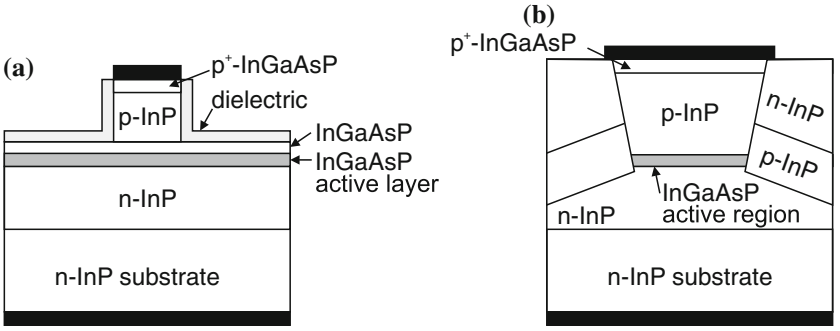


Fig. 23.41 Schematic cross section of index-guided lasers: (**a**) shallow ridge, (**b**) deep etch and regrowth. *Black areas* are metal contacts

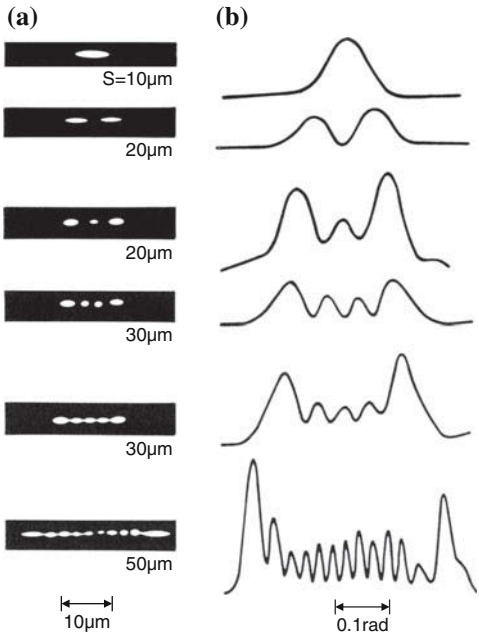


Fig. 23.42 Lateral near field (**a**) and far field (**b**) of lasers with various width S of the injection stripe as labeled. Adapted from [1685]

Longitudinal Modes

The spectral positions of the laser modes for a cavity with length L is given by the condition (cf. (19.34))

$$L = \frac{m\,\lambda}{2\,n_{\mathrm{r}}(\lambda)}, \tag{23.16}$$

where m is a natural number and $n(\lambda)$ is the dispersion of the index of refraction. The distance of neighboring modes is given by (for large m)

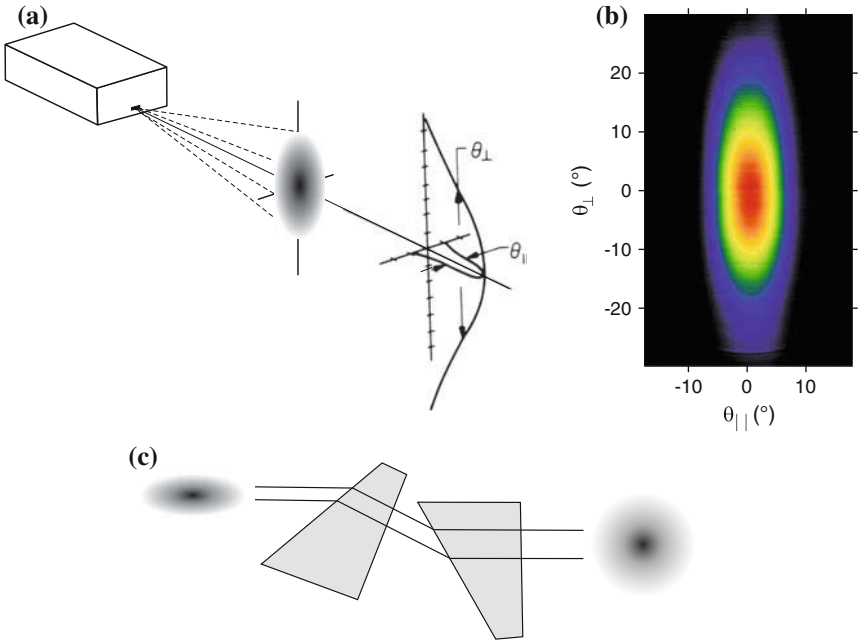


Fig. 23.43 (**a**) Schematics of the asymmetric far field of an edge emitter. Adapted from [500]. (**b**) Far-field distribution of violet GaN-based laser. Adapted from [1686]. (**c**) Correction of asymmetric far field with a pair of anamorphic prisms

$$\Delta\lambda = \frac{\lambda^2}{2\,n_\mathrm{r}\,L\,\left(1 - \frac{\lambda}{n_\mathrm{r}}\frac{\mathrm{d}n_\mathrm{r}}{\mathrm{d}\lambda}\right)}. \tag{23.17}$$

The dispersion $\mathrm{d}n_\mathrm{r}/\mathrm{d}\lambda$ can sometimes be neglected.

The facets of edge-emitting lasers are typically cleaved. Cleaving bears the danger of mechanical breakage and tends to have poor reproducibility, low yield and therefore high cost. Etched facets are another possibility to form the cavity mirror. The etch process, typically reactive ion dry etching, must yield sufficiently smooth surfaces to avoid scattering losses. A highly efficient distributed Bragg mirror (cf. Sect. 19.1.4) with only a few periods can be created by using the large index contrast between the semiconductor and air. As shown in Fig. 23.44, slabs can be etched that make a Bragg mirror with the air gaps [1687]. In this way, very short longitudinal cavities can be made ($\approx 10\,\mu\mathrm{m}$).

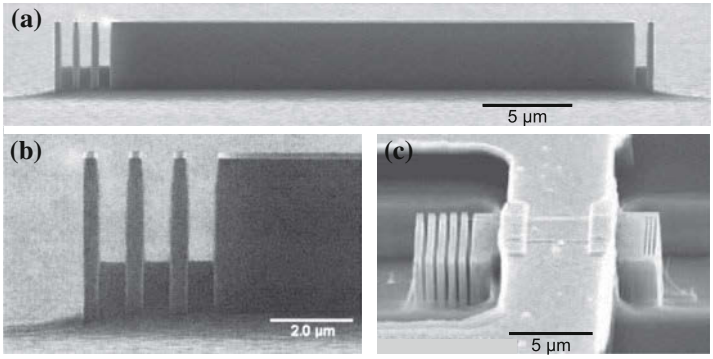


Fig. 23.44 SEM image of (**a**) an InP microlaser with third-order Bragg mirrors, (**b**) magnified view of the front facet with three slabs, (**c**) a 12 µm long microlaser with five third-order mirrors on the rear side and three first-order mirrors on the front side with top contact. From [1688]; part (**b**) reprinted with permission from [1687], © 2001 AIP

23.4.5 Loss Mechanisms

While the light travels through the cavity, it is not only amplified but it also suffers losses. The internal loss α_i and the mirror loss α_m contribute to the total loss α_tot

$$\alpha_\mathrm{tot} = \alpha_\mathrm{i} + \alpha_\mathrm{m}. \tag{23.18}$$

The internal loss is due to absorption in the cladding, scattering at waveguide inhomogeneities and possibly other processes. It can be written as

$$\alpha_\mathrm{i} = \alpha_0\, \Gamma + \alpha_\mathrm{g}\,(1 - \Gamma), \tag{23.19}$$

where α_0 is the loss coefficient in the active medium and α_g is the loss coefficient outside the active medium.

The mirror loss is due to the incomplete reflection of the optical wave at the laser facets. This condition is necessary, however, to observe a laser beam outside the cavity. R_1 and R_2 are the values of reflectance of the two facets, respectively. An as-cleaved facet has a reflectance of about 30 % (cf. (23.8)). Using dielectric layers on the facets, the reflectance can be increased (high reflection, HR-coating) or decreased (antireflection, AR-coating). One round-trip through the cavity of length L has the length $2L$. The intensity loss due to reflection at the facets is expressed via $\exp(-2\alpha_\mathrm{m} L) = R_1\, R_2$

$$\alpha_\mathrm{m} = \frac{1}{2L} \ln\left(\frac{1}{R_1\, R_2}\right). \tag{23.20}$$

If both mirrors have the same reflectance R, we have $\alpha_\mathrm{m} = -L^{-1} \ln R$. For $R = 0.3$ a 1 mm cavity has a loss of $12\,\mathrm{cm}^{-1}$. For the internal loss a typical value is $10\,\mathrm{cm}^{-1}$, very good waveguides go down to $1\text{–}2\,\mathrm{cm}^{-1}$.

Lasing is only possible if the gain overcomes all losses (at least for one wavelength), i.e.

$$g_\mathrm{mod} = g_\mathrm{mat}\, \Gamma \geq \alpha_\mathrm{tot}. \tag{23.21}$$

23.4.6 Threshold

When the laser reaches threshold, the (material) gain is pinned at the threshold value

$$g_\mathrm{thr} = \frac{\alpha_\mathrm{i} + \alpha_\mathrm{m}}{\Gamma}. \tag{23.22}$$

Since $g \propto n$, the carrier density is also pinned at its threshold value and does not increase further with increasing injection current. Instead, additional carriers are quickly converted into photons by stimulated emission. The threshold carrier density is (using the linear gain model, cf. (23.13))

$$n_\mathrm{thr} = n_\mathrm{tr} + \frac{\alpha_\mathrm{i} + \alpha_\mathrm{m}}{\hat{\alpha}\, \Gamma}. \tag{23.23}$$

For an active layer of thickness d, the threshold current density is

$$j_\mathrm{thr} \cong \frac{e\, d\, n_\mathrm{thr}}{\tau(n_\mathrm{thr})}, \tag{23.24}$$

where $\tau(n_\mathrm{thr})$ is the (minority) carrier lifetime at the threshold carrier density (cmp. Sect. 10.10) from

$$\tau(n) = \frac{1}{A + B\, n + C\, n^2}. \tag{23.25}$$

Using (23.23), we can write (for $R = R_1 = R_2$)

$$j_\mathrm{thr} = j_\mathrm{tr} + \frac{e\, d}{\tau\, \hat{\alpha}\, \Gamma} \left(\alpha_\mathrm{i} - \frac{1}{L} \ln R \right), \tag{23.26}$$

where the transparency current density is $j_\mathrm{tr} = e\, d\, n_\mathrm{tr}/\tau$. Thus, the plot of j_thr versus $1/L$ (or the optical loss) should be linear and can be extrapolated towards the transparency current density (cf. Fig. 23.45a).

Any additional increase of the current j leads to stimulated emission with the rate

$$r_\mathrm{st} = d\, v_\mathrm{g}\, g_\mathrm{thr}\, N_\mathrm{ph}, \tag{23.27}$$

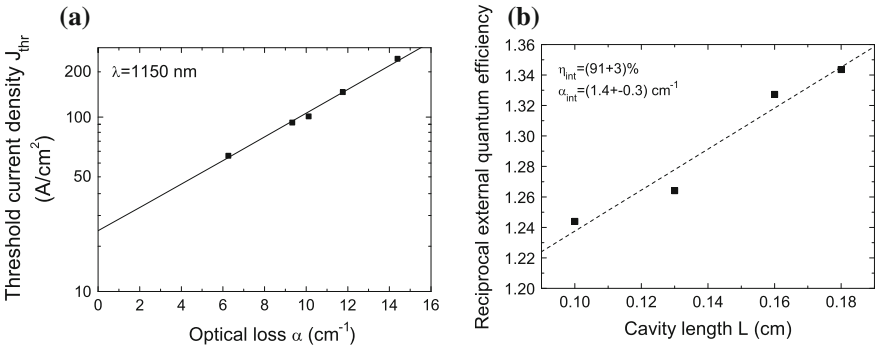


Fig. 23.45 (**a**) Threshold current density for (three-fold InGaAs/GaAs QD stack) laser ($\lambda = 1150\,\text{nm}$) at 10 °C with different cavity length versus the optical loss ($\propto 1/L$). The extrapolated transparency current density is $21.5 \pm 0.9\,\text{A/cm}^2$. (**b**) Inverse external quantum efficiency versus cavity length. The internal quantum efficiency determined from the plot is 91 % and the internal loss is $1.4\,\text{cm}^{-1}$

where v_g is the group velocity (mostly c_0/n_r) and N_{ph} is the photon density (per length) in the cavity. The photon density increases linearly beyond the threshold

$$N_{ph} = \frac{1}{e\,d\,v_g\,g_{thr}}\,(j - j_{thr})\,. \tag{23.28}$$

The photon lifetime

$$\frac{1}{\tau_{ph}} = v_g\,(\alpha_i + \alpha_m) = v_g\,\Gamma\,g_{thr} \tag{23.29}$$

is introduced that describes the loss rate of photons. $v_g\,\alpha_m$ describes the escape rate of photons from the cavity into the laser beam(s). Therefore,

$$N_{ph} = \frac{\tau_{ph}\,\Gamma}{e\,d}\,(j - j_{thr})\,. \tag{23.30}$$

Since the threshold depends on the carrier density, it is advantageous to reduce the active volume further and further. In this way, the same *amount* of injected carriers creates a larger carrier *density*. Figure 23.46 shows the historic development of laser threshold due to design improvements.

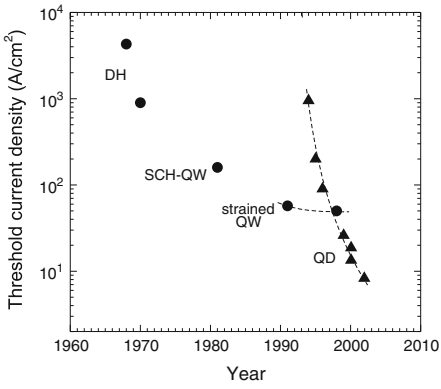


Fig. 23.46 Historic development of threshold current density (at room temperature, extrapolated for infinite cavity length and injection stripe width) for various laser designs, 'DH': double heterostructure, 'SCH–QW': separate confinement heterostructure with quantum wells. 'QD': quantum dots. *Dashed lines* are guides to the eye

23.4.7 Spontaneous Emission Factor

The spontaneous emission factor β is the fraction of spontaneous emission (emitted into all angles) that is emitted into laser modes. For Fabry–Perot lasers, β is typically in the order of 10^{-4}–10^{-5}. The design of a microcavity can increase β drastically by several orders of magnitude to $\approx$0.1 [1689] or above and thus reduce the threshold current. The photon number as a function of the pump current can be calculated from the laser rate equations and is depicted in Fig. 23.47. For $\beta = 1$, all emitted power goes into the laser mode regardless of whether emission is spontaneous or stimulated. The definition of threshold in such 'nonclassical' lasers with large β is discussed in detail in [1690].

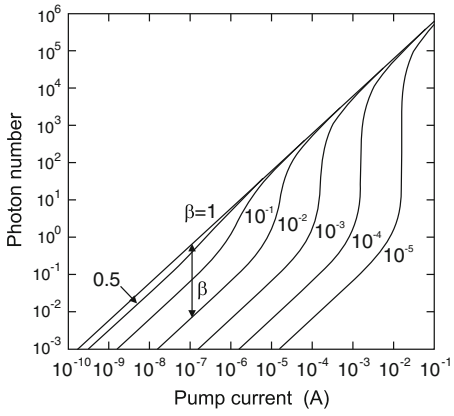


Fig. 23.47 Photon number versus pump current for a model laser. Adapted from [1691]

23.4.8 Output Power

The output power is given by the product of photon energy, the photon density in the cavity, the effective mode volume and the escape rate:

$$P_{\text{out}} = \hbar\omega\, N_{\text{ph}}\, \frac{L\, w\, d}{\Gamma}\, v_{\text{g}}\, \alpha_{\text{m}}. \tag{23.31}$$

Thus, it is given by

$$P_{\text{out}} = \hbar\omega\, v_{\text{g}}\, \alpha_{\text{m}}\, \frac{\tau_{\text{ph}}}{e}\, L\, w\, (j - j_{\text{thr}}) = \frac{\hbar\omega}{e}\, \frac{\alpha_{\text{m}}}{\alpha_{\text{m}} + \alpha_{\text{i}}}\, (I - I_{\text{thr}})\,. \tag{23.32}$$

To this equation, the factor η_{int} must be added. The internal quantum efficiency describes the efficiency of the conversion of electron–hole pairs into photons (cmp. (10.59)):

$$\eta_{\text{int}} = \frac{B\, n^2 + v_{\text{g}}\, g_{\text{thr}}\, N_{\text{ph}}}{A\, n + B\, n^2 + C\, n^3 + v_{\text{g}}\, g_{\text{thr}}\, N_{\text{ph}}}. \tag{23.33}$$

All in all, now (see Fig. 23.48a)

$$P_{\text{out}} = \frac{\hbar\omega}{e}\, \frac{\alpha_{\text{m}}}{\alpha_{\text{m}} + \alpha_{\text{i}}}\, \eta_{\text{int}}\, (I - I_{\text{thr}})\,. \tag{23.34}$$

The differential (or slope) quantum efficiency, also called the external quantum efficiency η_{ext}, is the slope of the P_{out} curve versus the current in the lasing regime. It is given by

$$\eta_{\text{ext}} = \frac{\mathrm{d}P_{\text{out}}/\mathrm{d}I}{\hbar\omega/e} = \eta_{\text{int}}\, \frac{\alpha_{\text{m}}}{\alpha_{\text{m}} + \alpha_{\text{i}}}. \tag{23.35}$$

The external quantum efficiency can also be written as

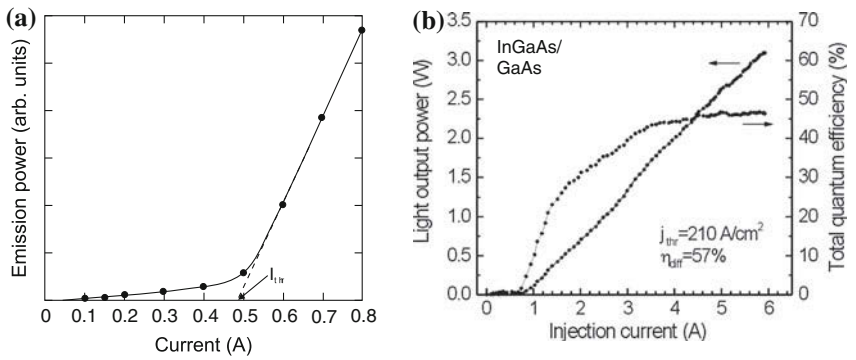


Fig. 23.48 (**a**) Typical P–I characteristic of a semiconductor laser. Adapted from [500]. (**b**) Output power and total quantum efficiency of a quantum dot laser (3 stacks of InGaAs/GaAs QDs, $L = 2\,\text{mm}$, $w = 200\,\mu\text{m}$, $\lambda = 1100\,\text{nm}$, $T = 293\,\text{K}$) versus injection current

$$\frac{1}{\eta_{\text{ext}}} = \frac{1}{\eta_{\text{int}}}\left(1 + \frac{\alpha_{\text{i}}}{\alpha_{\text{m}}}\right) = \frac{1}{\eta_{\text{int}}}\left[1 - 2\alpha_{\text{i}}\, L\, \ln\left(R_1 R_2\right)\right]. \tag{23.36}$$

Therefore, if η_{ext}^{-1} is plotted for similar lasers with different cavity length (see Fig. 23.45b), a straight line should arise from which the internal quantum efficiency (extrapolation to $L \to 0$) and the internal loss ($\propto$ slope) can be determined experimentally.

The threshold current for a given laser is determined from the P–I characteristic via extrapolation of the linear regime as shown in Fig. 23.48a. Record values for the threshold current density are often given for the limit $L \to \infty$. Due to current spreading, the threshold current density also depends on the width of the injection stripe. Record low thresholds are therefore often given for the limit $w \to \infty$.

The total quantum efficiency is given by

$$\eta_{\text{tot}} = \frac{P_{\text{out}}/I}{\hbar\omega/e}. \tag{23.37}$$

This quantity is shown in Fig. 23.48b for a laser as a function of the current. For a linear P–I lasing characteristic, the total quantum efficiency converges towards the external quantum efficiency for high currents because the low quantum efficiency subthreshold regime no longer plays any role. Another important figure of merit is the wall-plug efficiency η_{w} that describes the power conversion:

$$\eta_{\text{w}} = \frac{P_{\text{out}}}{U\, I}. \tag{23.38}$$

Additionally to the current balance discussed so far, typically a leakage current exists that flows without contributing to recombination or lasing. Carriers not captured into or escaping from the active layers can contribute to this current. The present record for wall-plug efficiency of high-power laser diodes is above 70 % [1692, 1693], employing careful control of band alignment (graded junctions, avoiding voltage barriers), optical losses, Joule heating, spontaneous emission and carrier leakage. It seems possible to achieve η_{w} of 80 %.

The P–I characteristic is not linear to arbitrary high currents. Generally, the output power will saturate or even decrease for increasing current. These effects can be due to increasing leakage current, increasing internal loss at high current or temperature effects, e.g. an increase of threshold with increasing temperature (cf. Sect. 23.4.9) and therefore a reduction of total efficiency. All nonradiative losses will eventually show up as heat in the laser that must be managed with a heat sink.

A radical effect is catastrophical optical damage (COD) at which the laser facet is irreversible (partially) destructed. Antioxidation or protective layers can increase the damage threshold to $>20\,\text{MW/cm}^2$. The record power from a single edge emitter is $\sim$12 W (200 µm stripe width). For a lateral monomode laser, cw power of about 1.2 W has been reached from a 1480 nm InGaAsP/InP double quantum-well lasers

with 3–5 μm stripes and 3 mm cavity length [1694]. About 500 mW can be coupled into a single-mode fiber [1695].

23.4.9 Temperature Dependence

The threshold of a laser typically increases with increasing temperature as shown in Fig. 23.49a. Empirically, in the vicinity of a temperature T_1 the threshold follows an exponential law (see Fig. 23.49b)

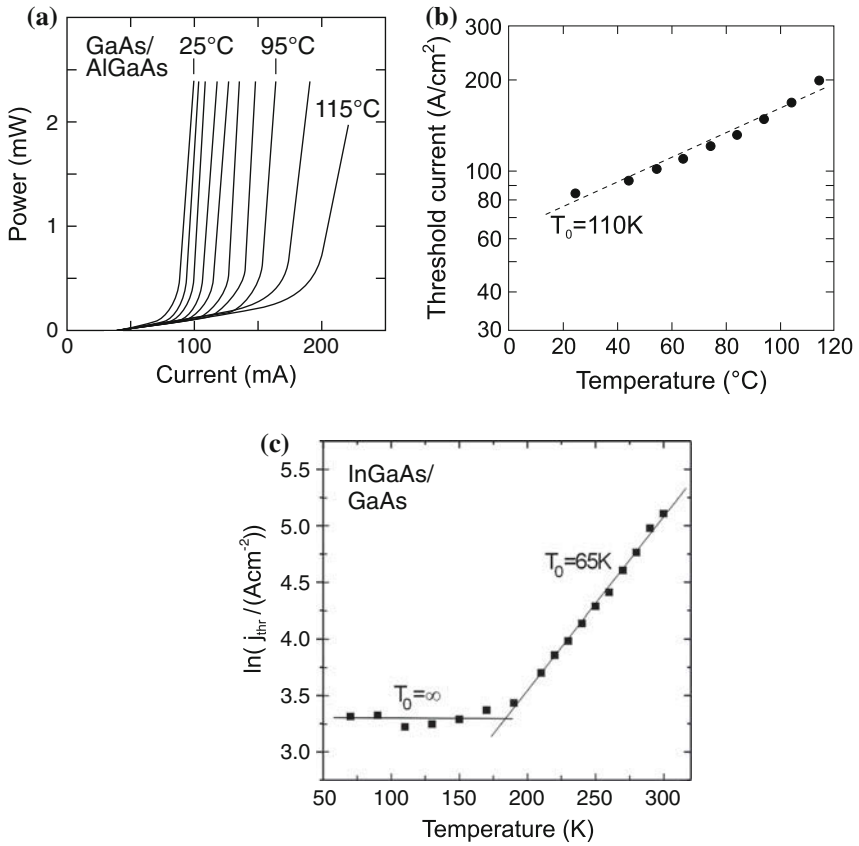


Fig. 23.49 (**a**) *P–I* characteristic (cw output power per mirror facet) of stripe-buried heterostructure laser at various temperatures of the heat sink between 25 °C and 115 °C in steps of 10 K. (**b**) Threshold current (in logarithmic scale) of this laser as a function of heat-sink temperature and exponential fit (*dashed line*) with $T_0 = 110$ K. Parts (**a**) and (**b**) adapted from [1696]. (**c**) Temperature dependence of the threshold current density of a quantum dot laser (3 stacks of InGaAs/GaAs QDs, $\lambda = 1150$ nm) with T_0 (*solid lines* are fits) given in the figure

$$j_{\mathrm{thr}}(T) = j_{\mathrm{thr}}(T_1)\,\exp\left(\frac{T - T_1}{T_0}\right) \propto \exp\left(\frac{T}{T_0}\right), \tag{23.39}$$

with T_0 being the so-called characteristic temperature.[20] It is the inverse logarithmic slope, $T_0^{-1} = \mathrm{d}\ln j_{\mathrm{thr}}/\mathrm{d}T$.

T_0 summarizes the temperature-dependent loss and the carrier redistribution in **k**-space due to the change of the Fermi distribution with temperature. With increasing temperature, populated states below the quasi-Fermi level become unpopulated and nonlasing states become populated. Therefore, the gain decreases with increasing temperature. This redistribution must be compensated by an increase of the quasi-Fermi energy, i.e. stronger pumping. This effect is present for (even ideal) bulk, quantum well and quantum wire lasers. Only for quantum dots with a δ-like density of states is the change of Fermi distribution irrelevant as long as excited states are energetically well separated from the (lasing) ground state. In Fig. 23.49c, the threshold of a quantum dot laser is indeed temperature independent ($T_0 = \infty$) as long as excited states are not thermally populated (for $T < 170\,\mathrm{K}$ for the present laser).

23.4.10 Mode Spectrum

In Fig. 23.50a, the mode spectrum of a typical edge-emitting laser is shown. Below threshold, the amplified spontaneous emission (ASE) spectrum exhibits a comb-like structure due to the Fabry–Perot modes. Above threshold, some modes grow much faster than others, possibly resulting in single longitudinal mode operation at high injection. The relative strength of the strongest side mode is expressed through the side-mode suppression ratio (SSR) in dB

$$SSR = 10\,\log\left(\frac{I_{\mathrm{mm}}}{I_{\mathrm{sm}}}\right), \tag{23.40}$$

where I_{mm} (I_{sm}) is the intensity of the maximum (strongest side) mode in the lasing spectrum.

As a tendency a DHS or QW semiconductor laser above threshold develops a narrow spectrum since the pump power is channeled into one or few modes with large gain. A quantum dot laser behaves differently when pumped largely above threshold. Since the gain of individual QDs in an inhomogeneously broadened ensemble (due to different QD sizes) is independent, a broad gain spectrum is present [1699]. The lasing spectrum takes a hat-like shape when homogeneous broadening is small compared to the inhomogeneous broadening [1698, 1700] (Fig. 23.51) as predicted theoretically [1699, 1701].

[20] Since T_0 has the dimension of a temperature difference, it can be expressed in °C or K. For the sake of unambiguity it should be given in K.

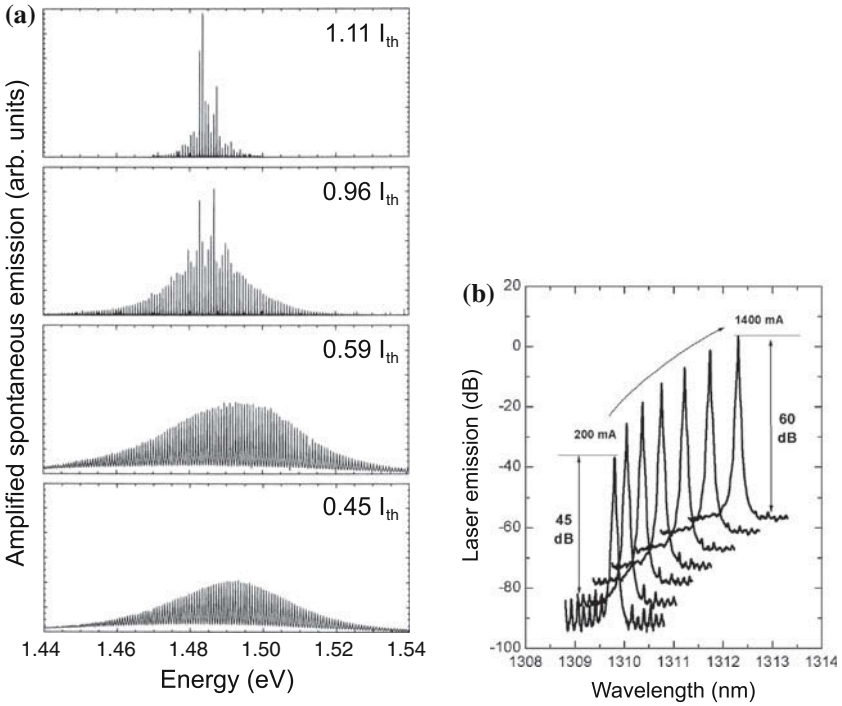


Fig. 23.50 (**a**) Mode spectra of a Fabry–Perot laser, under, at and above threshold ($I_{thr} = 13.5\,mA$). Adapted from [1047]. (**b**) Mode spectra of a cw DFB InGaAs/InP laser with 2 mm cavity length at various currents of 200, 400, ..., 1400 mA ($I_{thr} = 65\,mA$), SSR $> 40\,dB$ at $T = 293\,K$. Adapted from [1697], reproduced with permission from SPIE

23.4.11 Longitudinal Single-Mode Lasers

In order to achieve a high SSR or single longitudinal mode lasing, the feedback must offer a higher wavelength selectivity than a simple mirror. A preferential feedback for certain modes can be obtained using a periodic dielectric structure that 'fits' to a particular wavelength, similar to a Bragg mirror. The periodic modulation of the refractive index can be made within the cavity (distributed feedback, DFB, Fig. 23.52a) or at the mirror (distributed reflection, DBR, Fig. 23.52b). In this way, monochromatic lasers with SSR $\gg$30 dB are possible (Fig. 23.50b).

It is possible to couple several hundred mW optical power of a laterally and spectrally monomode laser into a monomode optical fiber [1702] (Fig. 23.53).

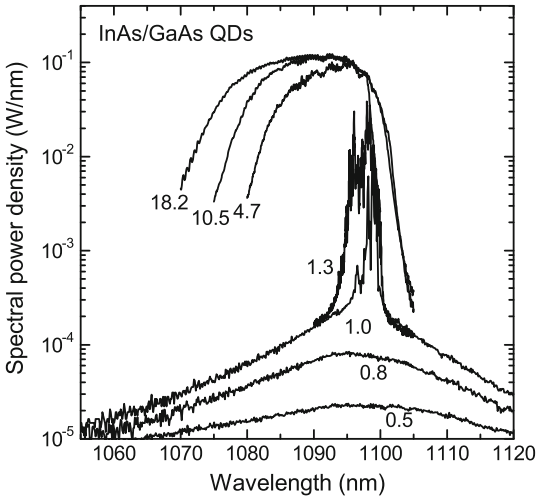


Fig. 23.51 Lasing spectra of quantum dot laser ($L = 1.2\,\text{mm}$, stripe width $w = 75\,\mu\text{m}$) at room temperature. The active medium is a three-fold stack of InGaAs/GaAs QDs. The current density as labeled is given in units of the threshold current density ($j_{\text{thr}} = 230\,\text{A/cm}^2$). Adapted from [1698]

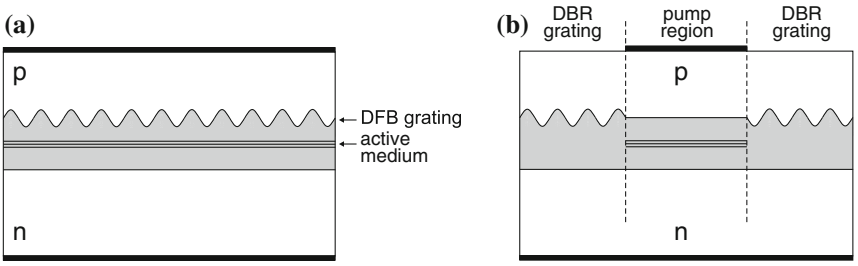


Fig. 23.52 Schematic drawing of (**a**) DFB (distributed feedback) and (**b**) DBR (distributed Bragg reflection) lasers. The active medium is schematically shown as a triple quantum well, the waveguide is shown as a *grey area*

23.4.12 Tunability

The tunability of the emission wavelength [1704] is important for several applications such as wavelength division multiplexing[21] and spectroscopy.

The simplest possibility to tune a laser is to vary its temperature and thus its band gap. This method is particularly used for lead salt lasers in the mid-infrared region,[22] as shown in Fig. 23.54.

[21] In order to make better use of the high bandwidth of the optical fiber several information channels with closely lying wavelengths are transmitted.

[22] Note the anomalous positive coefficient $\mathrm{d}E_{\mathrm{g}}/\mathrm{d}T$ as discussed in Sect. 6.7.

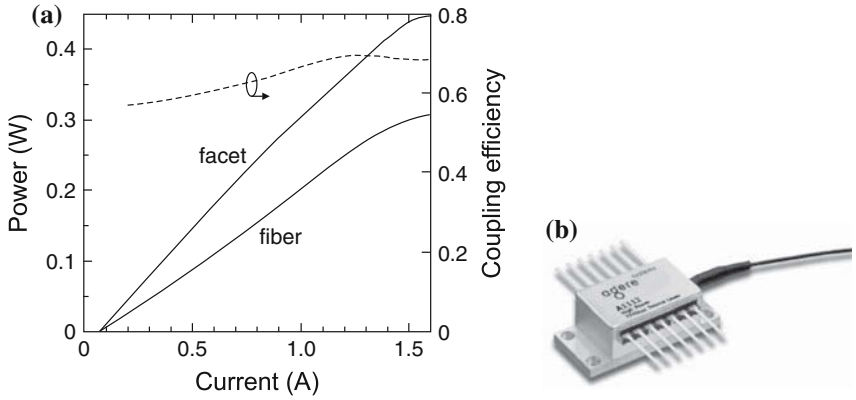


Fig. 23.53 (**a**) Output power of an InGaAsP/InP cw single-mode DFB laser at 1427 nm with 2 mm cavity length from the facet and coupled to a single-mode fiber versus driving current ($T = 293$ K). The *dashed* line represents the coupling efficiency to the fiber (right scale). Adapted from [1702]. (**b**) Package with pigtail of fiber coupled 1550 nm DFB laser with 40 mW output power in the fiber. From [1703]

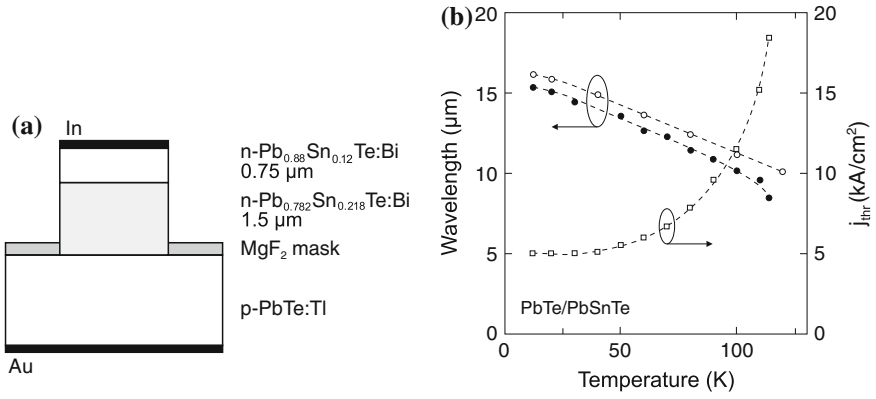


Fig. 23.54 (**a**) Schematic drawing of PbTe lead salt laser. (**b**) Tuning characteristics of such laser: Emission wavelength (*left scale*, *filled circles*: emission wavelength at cw threshold, *empty circles*: emission maximum under pulsed operation) and cw threshold current density (*right scale*) as a function of the heat-sink temperature. *Symbols* are experimental data, *dashed lines* are guides to the eye. Adapted from [1705]

For monomode lasers, mode hopping, i.e. the discontinuous shift of lasing wavelength (or gain maximum) from one mode to the next, poses a problem for continuous tuning, as shown in Fig. 23.55. The continuous shift of emission wavelength is due to the temperature dependence of the index of refraction and subsequently the longitudinal modes. The index of refraction increases with increasing temperature at typically $\sim 3 \times 10^{-4}\,\mathrm{K}^{-1}$. Generally, a red-shift is the consequence.

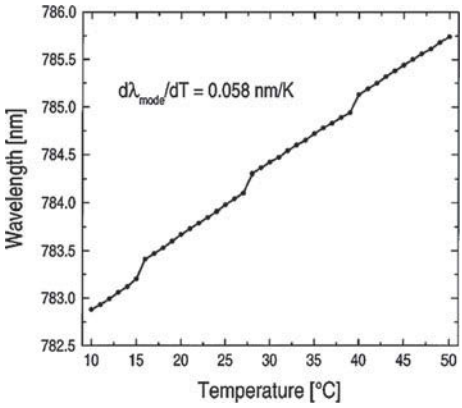


Fig. 23.55 Wavelength as a function of temperature (with mode hopping) for a GaAs-based DFB laser

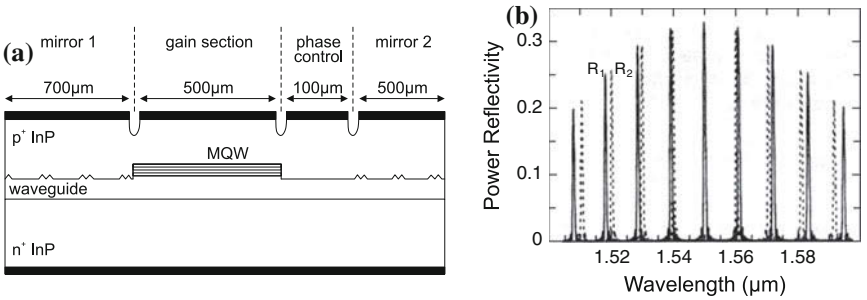


Fig. 23.56 (**a**) Schematic representation of SGDBR (sampled-grating DBR) laser with four sections. Adapted from [1706]. (**b**) Reflectance of two sampled gratings DBR mirrors

Another possibility to vary the index of refraction (and thus the optical path length) is a variation of the carrier density. The coefficient $\mathrm{d}n_\mathrm{r}/\mathrm{d}n$ is about $-10^{-20}\,\mathrm{cm}^3$. In a two-section laser, separate regions (with separately controlled currents) for gain and tuning are present. The regions are separated with deep-etched trenches to avoid crosstalk. The tuning range is limited to about 10 nm. For a mode-hopping free tuning, the control of the phase in the cavity is important and requires an additional section for the phase control. Such a three-section laser has separate regions (and current control) for the reflection, phase and amplification (or gain) region.

Using sampled gratings, the tuning range can be strongly increased to about 100 nm. A sampled grating is a nonperiodic lattice that has several ($\sim$10) reflection peaks. The laser structure has four sections (Fig. 23.56) with two mirrors that have slightly different sampled gratings. Via the carrier densities in the two mirror sections, different maxima can be brought to overlap (Vernier effect) and the position of the selected maximum can be tuned over a wide spectral range (Fig. 23.57).

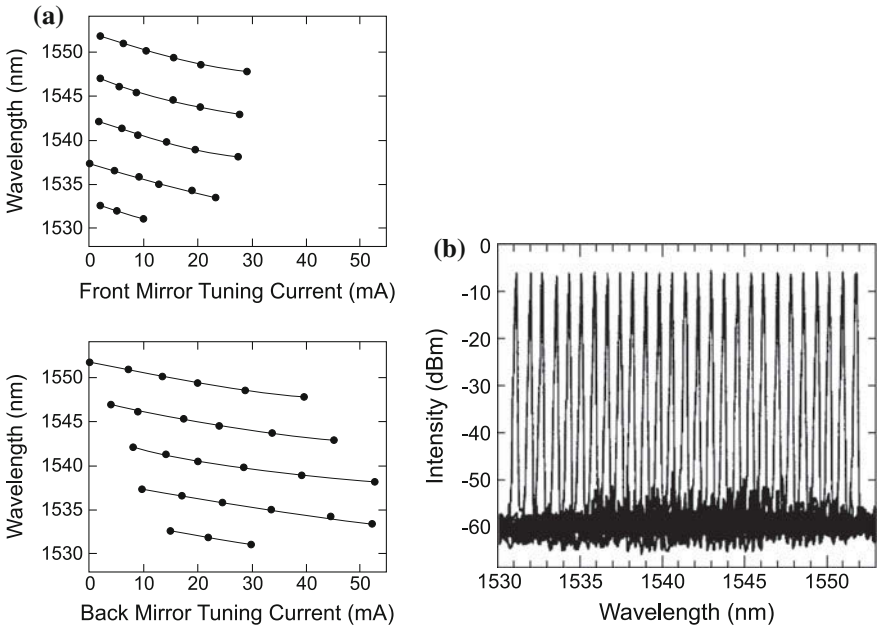


Fig. 23.57 (**a**) Tuning curves of two sampled gratings DBR mirrors for the front and back mirror current. (**b**) 27 wavelength channels (1531.12–1551.72 nm) with a channel separation of 1 nm. Adapted from [1707]

23.4.13 Dynamics and Modulation

For transmission of information in the time domain, the laser intensity must be modulated. This can be accomplished by *direct* modulation, i.e. modulation of the injection current, or *external* modulators, for example using the voltage-induced shift of the absorption spectrum due to QCSE (cf. Sect. 13.1.2). For direct modulation, small- and large-signal modulation are distinguished. A detailed treatment of laser dynamics and modulation can be found in [1708].

Laser Dynamics

The dynamical properties of the laser are described by the coupled rate equations for the carrier density n and the photon density (per volume) S:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{j}{e\,d} - \frac{n}{\tau(n)} - g(n, S)\, v_\mathrm{g}\, S \tag{23.41a}$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \Gamma\, g(n, S)\, v_\mathrm{g}\, S + \beta\, B\, n^2 - \frac{S}{\tau_\mathrm{ph}}. \tag{23.41b}$$

The first term in (23.41a) describes the electrical injection[23] ($=I/(e\,V)$), the second one the recombination (all channels) and the third one the carrier conversion to photons through stimulated emission via the gain g. In the photon dynamics (23.41b), the first term represents the gain term, the second one the photons in the laser mode from the recombination (spontaneous emission, cmp. Sect. 23.4.7) and the third term the photon loss (due to internal losses and the mirrors). The recombination rate $n/\tau(n)$ is typically given by (23.25). The photon lifetime τ_{ph} is given by (23.29). For the gain various models are used, a simple one being (cmp. (23.13)):

$$g(n,S) = g' \,\frac{n - n_{\mathrm{thr}}}{1 + \epsilon_S\, S}, \tag{23.42}$$

g' being the differential gain and ϵ_S being the gain compression coefficient due to spectral hole burning.

Large-Signal Modulation

If a current pulse is fed to the laser, the laser radiation is emitted with a short time delay, the so-called turn-on delay (TOD) time. This time is needed to build up the carrier density for inversion. The time dependence of the density is (neglecting the density dependence of the lifetime)

$$n(t) = \frac{I\,\tau}{e\,A\,d}\left[1 - \exp\left(-\frac{t}{\tau}\right)\right]. \tag{23.43}$$

The TOD time to reach the threshold density (using (23.24)) is

$$\tau_{\mathrm{TOD}} = \tau\,\ln\left(\frac{I}{I - I_{\mathrm{thr}}}\right). \tag{23.44}$$

We note that $\tau_{\mathrm{TOD}} > 0$ for $I > I_{\mathrm{thr}}$. Such a dependence is found experimentally (Fig. 23.58). The turn-on delay time decreases with increasing pump current but typically is at least 1 ns. In order to circumvent this limitation for more than about 1 GHz pulse repetition rate, the laser is biased slightly below threshold.

In Fig. 23.59a, the response (light emission) of a LED to a short current pulse is shown schematically. The monotonously decreasing transient (that is more or less exponential) corresponds to the carrier recombination dynamics. When a laser is excited with a steep (long) current pulse, the response exhibits so-called relaxation oscillations (RO) before the steady-state (cw) intensity level is reached (Fig. 23.59b).

In the laser, first the carrier density is built up. It surpasses the threshold density that leads to a build-up of the photon density. The laser pulse depletes the carrier density faster below threshold than the current can supply further carriers. Therefore, the laser intensity drops below the cw level. From the solution of the coupled rate equations for the electron and photon densities n and S (23.41a, b), the frequency of the relaxation oscillations is found as

[23] This term can be multiplied by the injection efficiency η_{inj} to account for leakage currents.

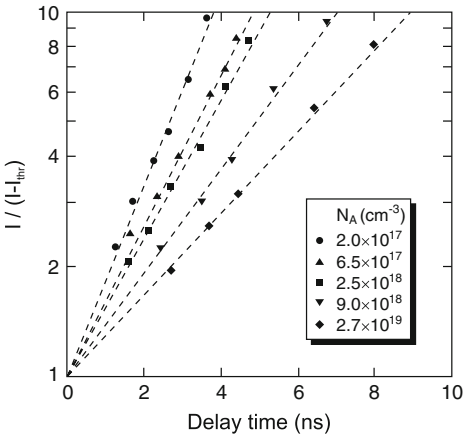


Fig. 23.58 Variation of turn-on delay time with the injected current for a laser at room temperature. Adapted from [1709]

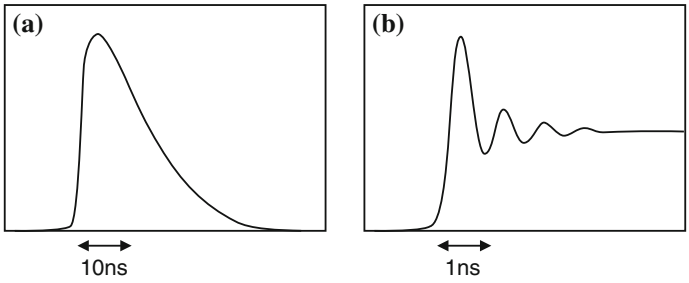


Fig. 23.59 Schematic response of (**a**) LED to current pulse and (**b**) of laser to current step

$$f_{\mathrm{RO}} = \frac{1}{2\pi}\left(\frac{v_{\mathrm{g}}\, g'\, S_0}{\tau_{\mathrm{ph}}}\right)^{1/2}, \tag{23.45}$$

where g' is the differential gain as defined in (23.42) and S_0 is the steady-state photon density per volume that is proportional to the laser power P. The dependence $f_{\mathrm{RO}} \propto S_0^{1/2}$ is also found experimentally (Fig. 23.60a). For higher power the relation $f_{\mathrm{RO}}^2 \propto S_0$ becomes non-linear (Fig. 23.60b) due to gain compression; in this case S_0 is replaced with $S_0/(1 + \epsilon_S\, S_0) \propto P/(1 + P/P_{\mathrm{sat}})$, with P_{sat} being the saturation power.

Pattern Response

For digital data transmission, the laser is driven with pulse sequences (bit patterns). The response to a random bit pattern is called an 'eye pattern' and is shown in Fig. 23.61. The pattern consists of traces of the type shown in Fig. 23.59b. A clear distinction with well-defined trigger thresholds between 'on'- and 'off'-states can only be made if the eye formed by the overlay of all possible traces remains open. From the eye patterns in Fig. 23.61, it can be seen that the RO overshoot can be suppressed by driving the laser with a dc offset current well above threshold.

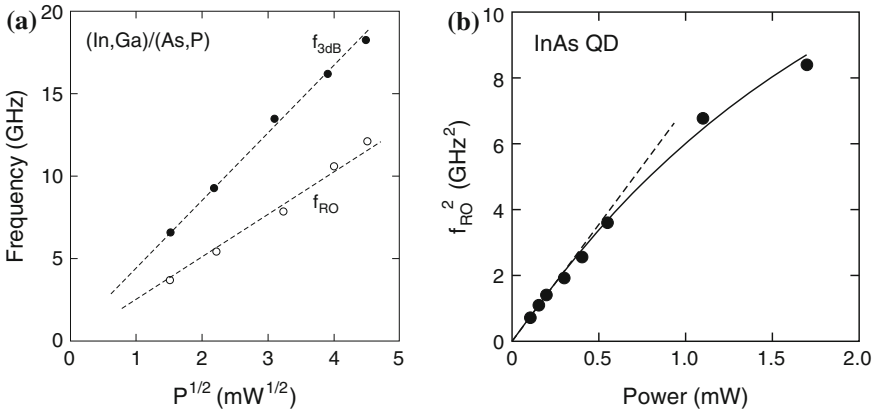


Fig. 23.60 (**a**) Dependence of small-signal 3 dB cutoff frequency $f_{3\mathrm{dB}}$ (*filled symbols*) and relaxation oscillation frequency f_{RO} (*empty symbols*) on the square root of the output power P for a DFB-laser. Adapted from [1710]. (**b**) Relaxation oscillation frequency f_{RO}^2 versus power for an InAs/GaAs QD laser. The *solid line* is a model including gain compression ($P_{\mathrm{sat}} = 3.3\,\mathrm{mW}$), the *dashed line* is the linear relation (23.45) for small power. Adapted from [1711]

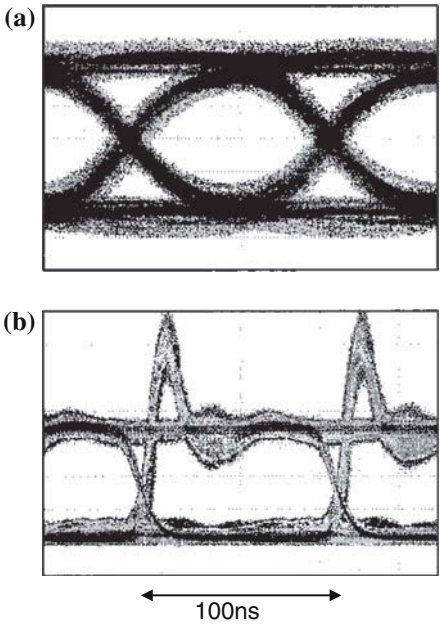


Fig. 23.61 Eye pattern of a single-mode VCSEL in response to a 10 Gb/s random bit pattern. The patterns are measured (**a**) with an offset current well above threshold and (**b**) with an offset current above but close to threshold. Adapted from [1712]

Small-Signal Modulation

In small-signal modulation, the injection current I is varied periodically by a small amount δI with $\delta I \ll I$ in the lasing regime. The current modulation leads to a

corresponding variation of the output intensity. The frequency response is limited by the differential gain and the gain compression coefficient. The frequency response shifts to higher frequency with increasing laser power as shown in Fig. 23.62b.

α Factor

Another important quantity is the α factor, also called the linewidth enhancement factor [1713, 1714]. Due to the coupling of amplitude and phase fluctuations in the laser, the linewidth Δf is larger than expected.

$$\Delta f = \frac{\hbar\omega\, v_\mathrm{g}\, R_\mathrm{spont}\, \ln R}{8\pi\, P_\mathrm{out}\, L}(1+\alpha^2). \tag{23.46}$$

The linewidth enhancement is described via $(1+\alpha^2)$ with

$$\alpha = \frac{\mathrm{d}n_\mathrm{r}/\mathrm{d}n}{\mathrm{d}\kappa/\mathrm{d}n} = -\frac{4\pi}{\lambda}\frac{\mathrm{d}n_\mathrm{r}/\mathrm{d}n}{g'}, \tag{23.47}$$

where κ denotes the imaginary part of the index of refraction (9.4). Typical values for α are between 1 and 10. The linewidth is inversely proportional to the output power (Fig. 23.63).

23.4.14 Surface-Emitting Lasers

Surface-emitting lasers emit their beam normal to the surface. They can be fabricated from horizontal (edge-) emitters by reflecting the beam with a suitable mirror into the surface direction. This technology requires tilted facets or micro-optical components but allows for high power per area. In Fig. 23.64, a schematic cross section of a horizontal-cavity surface-emitting laser (HCSEL) and the light emission from an

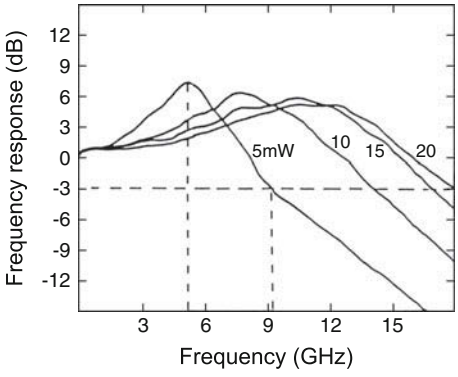


Fig. 23.62 Frequency response of a DFB-laser for various output powers as labeled. Adapted from [1710]

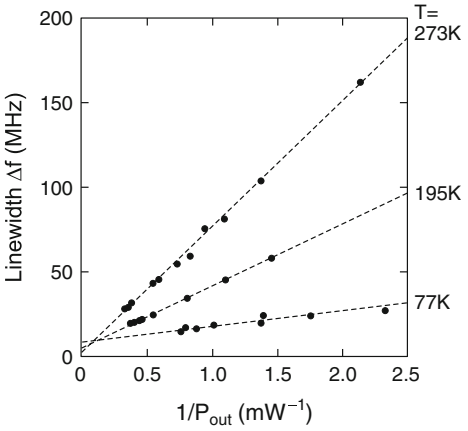


Fig. 23.63 Linewidth Δf of a cw GaAs/AlGaAs laser diode at various temperatures as a function of the inverse output power P_{out}^{-1}. At room temperature $\alpha \approx 5$. Adapted from [1715]

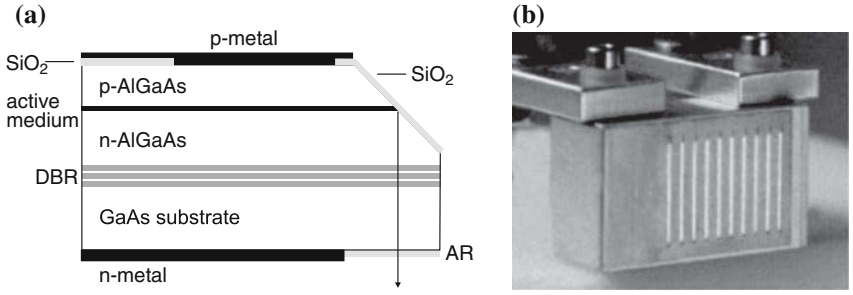


Fig. 23.64 (**a**) Principle of a surface-emitting laser. Light generated in the active region is internally reflected by the 45° angled mirror and directed through the substrate; 'AR': antireflection coating, 'DBR': epitaxial Bragg mirror. (**b**) Light emission from a 10×22 surface-emitting diode array. The light emission appears as stripes due to the broad beam divergence in the vertical direction. Part (**b**) reprinted with permission from [1716]

array of 220 such lasers are shown. The laser contains a 45° mirror that steers the light through the substrate and a Bragg mirror to provide the cavity mirror. The facet can also be fabricated such that the emission is to the top surface (Fig. 23.65). Another possibility to couple the beam out of a horizontal cavity is a surface grating.

Now, surface-emitting lasers with vertical-cavity (VCSEL), as shown in Fig. 23.29b, will be discussed. A detailed treatment can be found in [1718]. VCSELs are of increasing importance after many issues regarding their technology and fabrication have been solved. VCSEL fabrication is essentially a planar technology and VCSELs can be fabricated as arrays (Fig. 23.67). An on-wafer test of their properties is possible. They offer a symmetrical (or possibly a controlled asymmetrical) beam profile (Fig. 23.66) with possible polarization control or fixation.

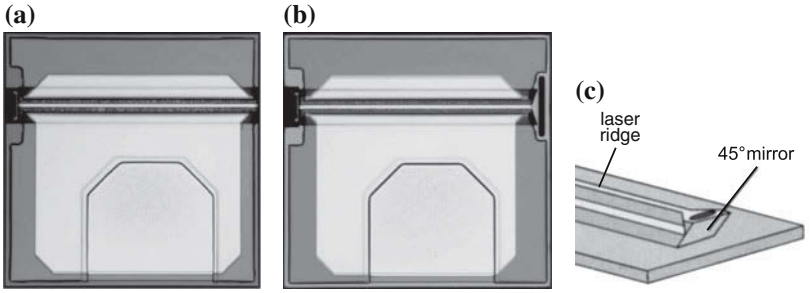


Fig. 23.65 (**a**) Horizontal Fabry–Perot cavity InP-based laser with 1310 nm emission length and 10 mW output power for modulation at 2.5 GB/s. The right facet is formed as DBR, emission is to the left. The trapezoid area in the center bottom of the image is the bond pad for the top contact. (**b**) Horizontal-cavity surface-emitting laser. Compared to (**a**), the right facet is replaced with a 45° mirror, leading to surface emission. (**c**) Schematic drawing of the tilted facet. Parts (**a**) and (**b**) from [1717]

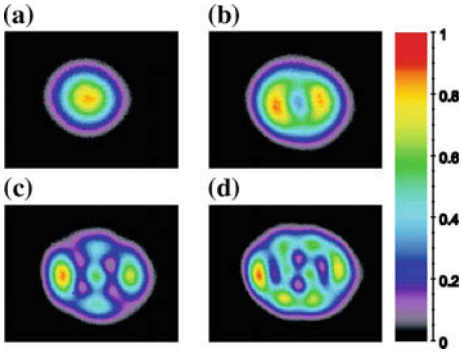


Fig. 23.66 In-plane near field of a VCSEL with 6 μm oxide aperture at various currents, (**a**) 3.0 mA, (**b**) 6.2 mA, (**c**) 14.7 mA, (**d**) 18 mA

The cavity is formed by two highly reflecting Bragg mirrors with a distance of $\lambda/2$ or $3\lambda/2$ forming a microcavity (see Sect. 19.1.7). A high index contrast can be obtained from GaAs/AlAs Bragg mirrors in which the AlAs layers have been selectively oxidized in a hot moist atmosphere. Pure semiconductor Bragg mirrors suffer typically from small index contrast and require many pairs. This poses a problem, e.g. for InP-based VCSELs. In Fig. 23.68, the distribution of light intensity along a $3\lambda/2$ cavity is shown. In the stop band of the mirrors, there is only one optical mode, the cavity mode, that can propagate along the vertical (z) direction.

The current path through the active layer can be defined with an oxide aperture. This aperture is fabricated by selective oxidation of an AlAs layer, leaving a circular opening in the center of the VCSEL pillar as shown in Fig. 23.69. The current can be injected through the mirrors if they are doped. Alternatively, the current can be directly fed to the active layer by so-called intracavity contacts.

The emission wavelength of a VCSEL can be shifted via a variation of temperature or pump power. Tuning of the VCSEL emission can also be accomplished by leaving an air gap between the cavity and the upper mirror [1722]. Applying a voltage to the

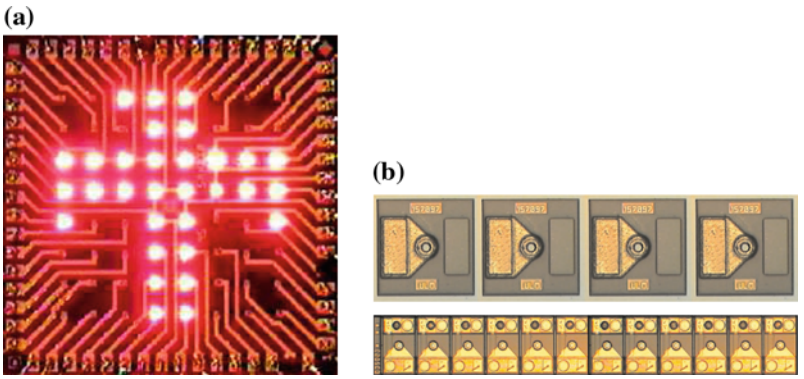


Fig. 23.67 (**a**), (**b**) VCSEL arrays. Part (**a**) reprinted from [1719] with permission, part (**b**) reprinted from [1720] with permission

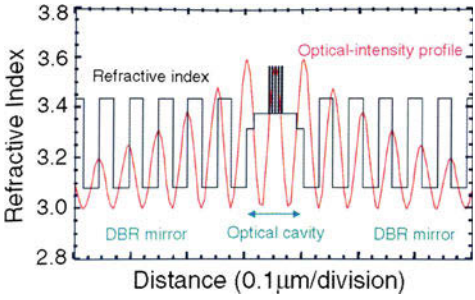


Fig. 23.68 Simulation of the longitudinal distribution of the optical field in a VCSEL structure. The active medium are five quantum wells in the center. Reproduced from [1721] by permission from the MRS Bulletin

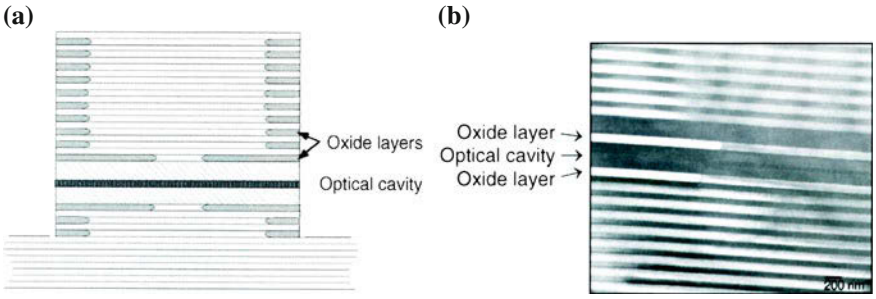


Fig. 23.69 (**a**) Schematic cross section of VCSEL with oxide aperture, (**b**) TEM image of cross section. Reproduced from [1721] by permission from the MRS Bulletin

lever arm with the top mirror, the width of the air gap can be varied. This variation leads to a shift of the cavity mode and therefore of the laser emission wavelength (Fig. 23.70). A VCSEL with air gap and particularly a high contrast Bragg mirror is achieved with InP/air as shown in Fig. 23.71.

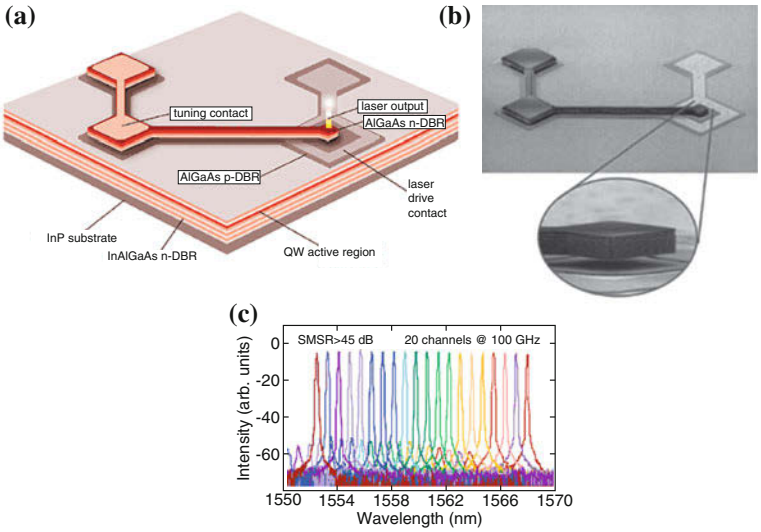


Fig. 23.70 (**a**) Schematic setup and (**b**) SEM image of VCSEL with air gap between active layer and top Bragg mirror, (**c**) spectra for different tuning conditions (via the width of the air gap). From [1723]

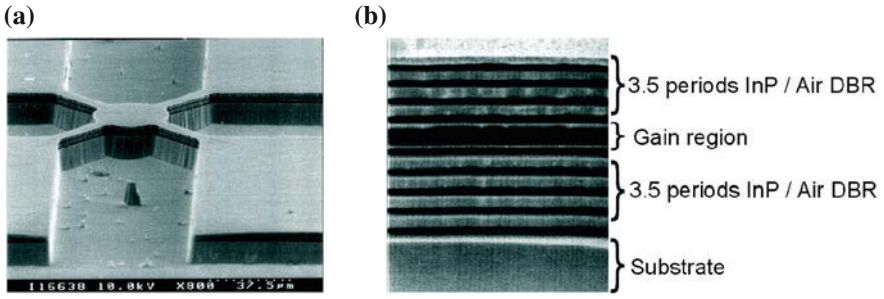


Fig. 23.71 (**a**) VCSEL with air gap and (**b**) Bragg mirror with high dielectric contrast InP/air interfaces. Reprinted with permission from [1724], © 2002 IEEE

23.4.15 Optically Pumped Semiconductor Lasers

An easy way to pump semiconductor lasers is optical pumping. This technique is similar to diode-pumped solid-state lasers (DPSS). A (semiconductor) pump diode illuminates a suitable semiconductor structure (Fig. 23.72). The resonator is built between the bottom Bragg mirror of the semiconductor and the output coupler. The semiconductor structure contains suitable absorption layers (barriers) that absorb the pump light and quantum wells that emit laser radiation. This radiation is intracavity frequency doubled. In order to reach, e.g., a 488 nm output laser beam, a standard 808 nm pump diode is employed. The InGaAs/GaAs quantum wells are

designed to emit at 976 nm. Other design wavelengths of the quantum wells allow for other output wavelengths. This technology allows compact lasers with little heat dissipation [1725]. The optically pumped semiconductor laser (OPSL) is also known as a semiconductor disc laser.

23.4.16 Quantum Cascade Lasers

In a quantum cascade laser (QCL), the gain stems from an intersubband transition. The concept was conceived in 1971 [1729, 1730] and realized in 1994 [1676]. In Fig. 23.73a, the schematic conduction-band structure at operation is shown. The injector supplies electrons into the active region. The electron is removed quickly from the lower level in order to allow inversion. The electron is then extracted into the next injector. The laser medium consists of several such units as shown in Fig. 23.73b. Since every unit can deliver a photon per electron (with efficiency η_1), the total quantum efficiency of N units $\eta = N\,\eta_1$ can be larger than 1.

The emission wavelength is in the far- or mid-infrared, depending only on the designed layer thicknesses and *not* on the band gap of the material (Fig. 23.73d). In the mid-infrared, room-temperature operation has been achieved while operation in the far-infrared requires cooling so far. Extensions to the THz-range and the infrared spectral region (telecommunication wavelengths of 1.3 and 1.55 µm) seem feasible. The cascade laser concept can also be combined with the DFB technology to create single-mode laser emission (Fig. 23.73d).

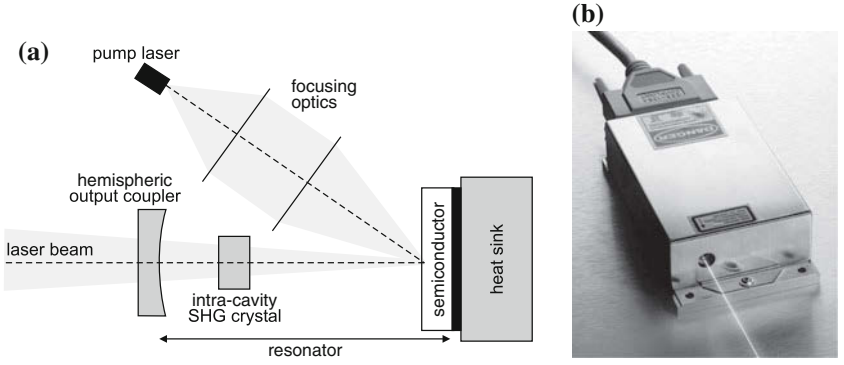


Fig. 23.72 (**a**) Schematic setup of optically pumped semiconductor laser (OPSL). The semiconductor chip consists of a Bragg mirror on the bottom, multiple quantum wells and an antireflection coating on the top. Adapted from [1725]. (**b**) OPSL source (488 nm, 20 mW, footprint: 125 × 70 mm^2). Reprinted with permission from [1726]

23.4.17 Hot-Hole Lasers

The hot-hole laser, which is mostly realized with p-doped Ge, is based on a population inversion between the light- and heavy-hole valence subbands. The laser operates with crossed electric and magnetic fields (Voigt configuration, typically $E = 0.5$–$3\,\mathrm{kV/cm}$, $B = 0.3$–$2\,\mathrm{T}$) and at cryogenic temperatures ($T = 4$–$40\,\mathrm{K}$) [1731–1733].

A significant scattering process of hot carriers is interaction with optical phonons, mainly optical phonon emission. This process has a threshold in carrier energy given by the optical phonon energy. For sufficiently high electric fields and at low temperature, hot carriers accelerate without acoustical phonon interaction (ballistic transport) along the crystallographic direction in which the electric field is applied. These hot carriers reach the optical phonon energy and lose all their energy due to emission

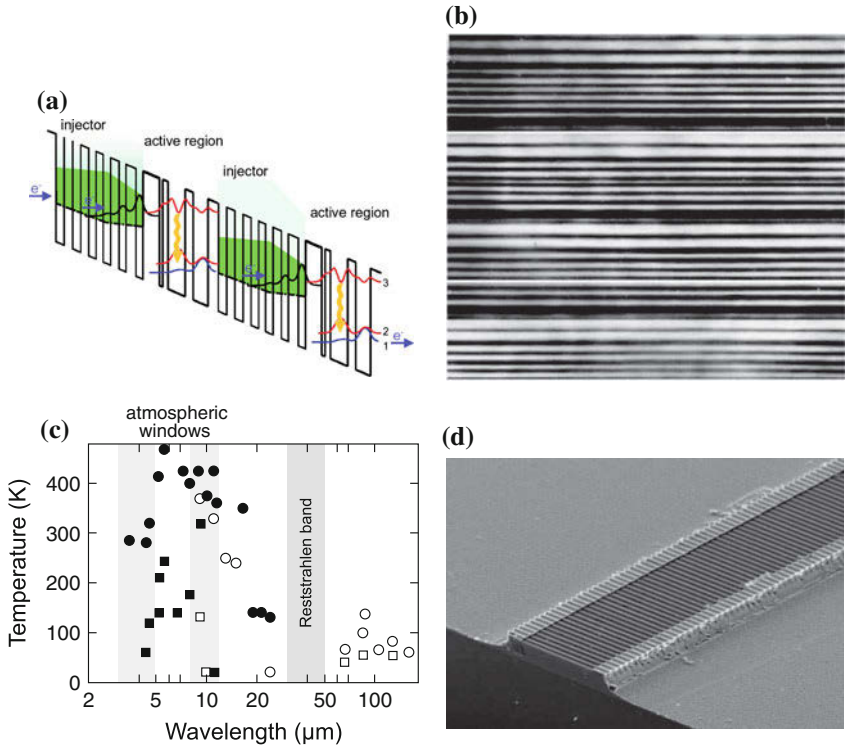


Fig. 23.73 (**a**) Schematic band diagram of quantum cascade laser. (**b**) Cross sectional TEM of cascade layer sequence. The periodicity of the vertical layer sequence is 45 nm. From [1727]. (**c**) Laser emission wavelengths and operation temperatures for various realized quantum cascade lasers (*squares*: cw, *circles*: pulsed operation, *solid symbols*: InP-, *empty symbols*: GaAs-based). Data from [1728]. (**d**) SEM image of a quantum cascade DFB laser (grating period: 1.6 μm). From [1728]

of an optical phonon. They accelerate again, repeating this directional motion in momentum space. This motion is called *streaming motion*.

For $|E/B|$ ratios of about 1.5 kV/cmT, the heavy holes are accelerated up to energies above the optical phonon energy (37 meV in germanium) and consequently are scattered strongly by these phonons. Under these conditions, a few per cent of the heavy holes are scattered into the light-hole band. The light holes remain at much lower energies and are accumulated at the bottom of the light-hole band below the optical phonon energy as sketched in Fig. 23.74. The continuous pumping of heavy holes into the light-hole band can lead to a population inversion. Consequently, laser radiation is emitted from optical (radiative) intervalence-band transitions (cf. Sect. 9.8.3). The emission wavelength is in the far-infrared around 100 μm. Typical p-Ge lasers span the frequency range 1–4 THz (300–70 μm) [1734] and deliver 1–10 W peak output power for 1 cm^3 typical active volume.

Since the applied electric field causes considerable heating, the temperature of the laser crystal rises quickly, within a few μs, up to 40 K. Then the laser action stops. Thus the duration of the electric-field excitation is limited to 1–5 μs (limiting the emission power) and the repetition frequency is only a few Hertz due to the necessary cooling. Research is underway towards high duty cycle (possibly cw) operation by using smaller volumes and planar vertical-cavities [1735, 1736].

23.5 Semiconductor Optical Amplifiers

If the facets of a laser cavity are antireflection coated, a laser gain medium can be used as a semiconductor optical amplifier (SOA). A textbook on this subject is [1737]. Optical feedback from facets can also be avoided using tilted facets [1738].

A tapered amplifier geometry, as shown in Fig. 23.75a, allows for laterally monomode input and a preservation of the lateral beam quality during the propagation of the optical wave through the gain medium. The active medium is an 8 nm compressively strained InGaAs quantum well. A typical taper angle is 5–10°.

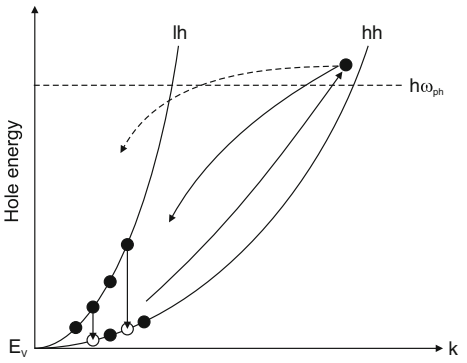


Fig. 23.74 Schematic cycle of hole motion in a hot-hole laser. *Filled* (*empty*) *circles* represent populated (unpopulated) hole states. The *solid lines* represent streaming motion of heavy hole, the *dashed line* represents scattering into the light hole band. *Arrows* denote radiative intervalence-band transitions

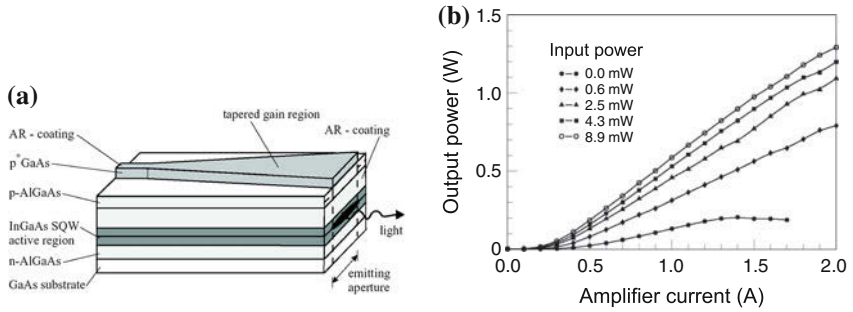


Fig. 23.75 (**a**) Schematic geometry of tapered semiconductor amplifier. (**b**) Optical output power versus amplifier current for various values of the optical input power, taper angle was 5°. For zero input power only spontaneous and amplified spontaneous emission is observed. Reprinted with permission from [1739]

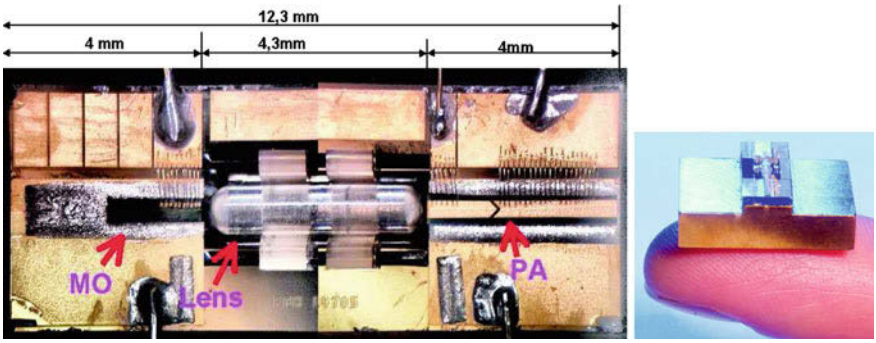


Fig. 23.76 Photographs of a MOPA arrangement of a laser (master oscillator, 'MO'), glass lens and tapered amplifier (power amplifier, 'PA') on a silicon micro-optical 'bench'. Reprinted with permission from [1740]

The input aperture is between 5 and 7 μm. The amplifier length is 2040 μm. More than 20 dB optical amplification can be obtained (Fig. 23.75b). The self-oscillation is suppressed for currents up to 2 A by AR facet coating of 10^{-4} in a 70 nm band. The wall-plug efficiency of the discussed amplifier is up to over 40 %. If such an amplifier is arranged together with a seed laser diode (master oscillator), the setup is called MOPA (master oscillator power amplifier), as shown in Fig. 23.76. A modulated input also leads to a modulated output.

Quantum dot arrays can be a useful gain medium in a SOA [1741, 1742] due to their fast gain dynamics [1743] and broad spectrum (Sect. 23.3.6).

Chapter 24
Transistors

Frequently, I have been asked if an experiment I have planned is pure or applied research; to me it is more important to know if the experiment will yield new and probably enduring knowledge about nature. If it is likely to yield such knowledge, it is, in my opinion, good fundamental research; and this is much more important than whether the motivation is purely esthetic satisfaction on the part of the experimenter on the one hand or the improvement of the stability of a high-power transistor on the other.

W.B. Shockley [1744, 1745]

Abstract The device functionalities of bipolar, heterobipolar and field effect transistors (JFET, MESFET and MOSFET) are explained. Within physical models for drift, diffusion and recombination given earlier in the book, the characteristics of these devices are derived. Remarks on integrated circuits, miniaturization and thin film transistors finish this chapter.

24.1 Introduction

Transistors[1] are the key elements for electronic circuits such as amplifiers, memories and microprocessors. Transistors can be realized in bipolar technology (bipolar junction transistor (BJT), Sect. 24.2) or as unipolar devices using the field effect (field-effect transistor (FET), Sect. 24.3) [500, 1746]. The equivalent in vacuum-tube technology to the transistor is the triode (Fig. 24.1a). Transistors can be optimized for their properties in analog circuits such as linearity and frequency response or their properties in digital circuits such as switching speed and power consumption. Transistors for microwave applications are discussed in [1526]. Early commercial models are shown in Fig. 24.2.

[1]The term 'transistor' was coined from the combination of 'transconductance' or 'transfer' and 'varistor' after initially such devices were termed 'semiconductor triodes'. The major breakthrough was achieved in 1947 when the first transistor was realized that showed gain (Figs. 1.9 and 24.1b).

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7_24

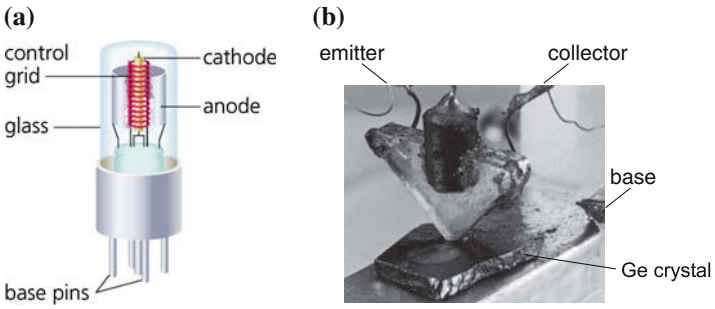


Fig. 24.1 (**a**) Schematic image of a vacuum triode. The electron current flows from the heated cathode to the anode when the latter is at a positive potential. The flow of electrons is controlled with the grid voltage. (**b**) Bell Laboratories' first (experimental) transistor, 1947

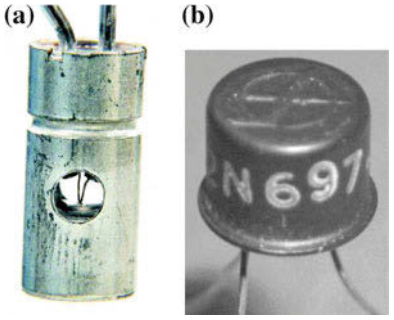


Fig. 24.2 (**a**) First commercial, developmental (point contact) transistor from BTL (Bell Telephone Laboratories) with access holes for adjustment of the whiskers pressing on a piece of Ge, diameter $7/32'' = 5\,\text{mm}$, 1948. (**b**) First high-performance silicon transistor (npn mesa technology), model 2N697 from Fairchild Semiconductor, 1958 (at \$200, in 1960 \$28.50). The product number is still in use (now \$0.95)

24.2 Bipolar Transistors

Bipolar transistors consist of a pnp or npn sequence (Fig. 24.3). The layers (or parts) are named emitter (highly doped), base (thin, highly doped) and collector (normal doping level). The transistor can be considered to consist of two diodes (emitter–base and base–collector) back to back. However, the important point is that the base is sufficiently thin (in relation to its minority carrier diffusion length) and carriers from the emitter (which are minority carriers in the base) can dominantly reach the collector by diffusion.

In Fig. 24.4, the three basic circuits with a transistor are shown. They are classified by the common contact for the input and output circuit. The space charges and band diagram for a pnp transistor in the base circuit configuration are depicted in Fig. 24.5. The emitter–base diode is switched in the forward direction to inject electrons into the

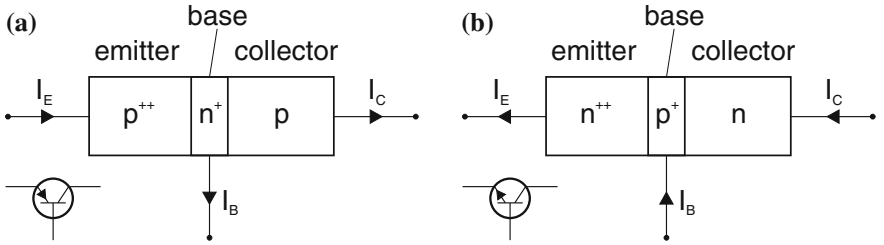


Fig. 24.3 Schematic structure and circuit symbol for (**a**) pnp and (**b**) npn transistors

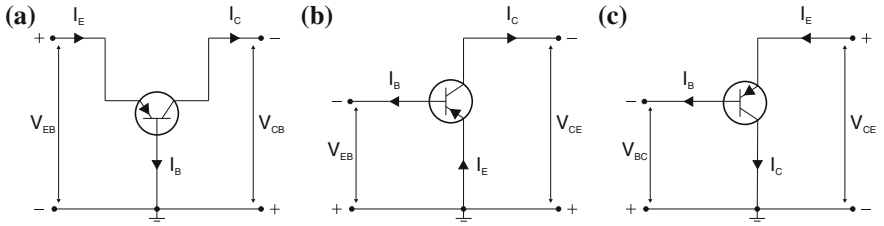


Fig. 24.4 Basic transistor circuits, named after the common contact: (**a**) Common base circuit, (**b**) common emitter circuit and (**c**) common collector circuit

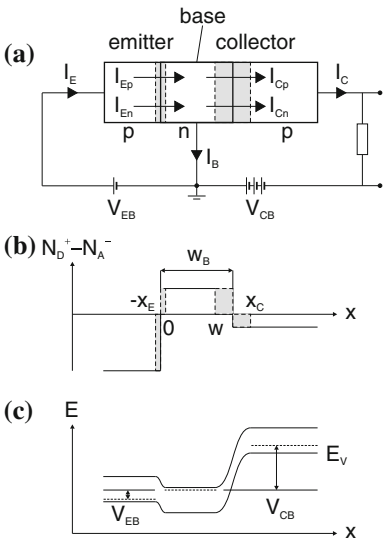


Fig. 24.5 pnp transistor in (**a**) base circuit. (**b**) Doping profile and space charges (abrupt approximation) and (**c**) band diagram for typical operation conditions

base. The base–collector diode is switched in the reverse direction. The electrons that diffuse through the base and reach the neutral region of the collector are transported by the high drift field away from the base.

24.2.1 Carrier Density and Currents

The modeling of transistors is a complex topic. We treat the transistor on the level of the abrupt junction. As an approximation, we assume that all voltages drop at the junctions. Series resistances, capacities and stray capacities and other parasitic impedances are neglected at this point.

The major result is that the emitter–base current from the forward-biased emitter–base diode will be transferred to the collector. The current flowing from the base contact is small compared to the collector current. This explains the most prominent property of the transistor, the current amplification.

For the neutral part of the base region of a pnp transistor, the stationary equations for diffusion and continuity are

$$0 = D_\mathrm{B}\frac{\partial^2 p}{\partial x^2} - \frac{p - p_\mathrm{B}}{\tau_\mathrm{B}} \tag{24.1a}$$

$$j_\mathrm{p} = -eD_\mathrm{B}\frac{\partial p}{\partial x} \tag{24.1b}$$

$$j_\mathrm{tot} = j_\mathrm{n} + j_\mathrm{p}, \tag{24.1c}$$

where p_B is the equilibrium minority carrier density in the base. From the discussion of the pn-diode, we know that at the boundary of the depletion layer the minority carrier density is increased by $\exp(eV/kT)$ (cf. (21.98a, b)). At the boundaries of the emitter–base diode (for geometry see Fig. 24.5a)

$$\delta p(0) = p(0) - p_\mathrm{B} = p_\mathrm{B}\left[\exp\left(\beta V_\mathrm{EB}\right) - 1\right] \tag{24.2a}$$

$$\delta n(-x_\mathrm{E}) = n(-x_\mathrm{E}) - n_\mathrm{E} = n_\mathrm{E}\left[\exp\left(\beta V_\mathrm{EB}\right) - 1\right], \tag{24.2b}$$

where n_E and p_B are the equilibrium minority-carrier densities in the emitter and base, respectively. Accordingly, at the boundaries of the base–collector diode we have

$$\delta p(w) = p(w) - p_\mathrm{B} = p_\mathrm{B}\left[\exp\left(\beta V_\mathrm{CB}\right) - 1\right] \tag{24.3a}$$

$$\delta n(x_\mathrm{C}) = n(x_\mathrm{C}) - n_\mathrm{C} = n_\mathrm{C}\left[\exp\left(\beta V_\mathrm{CB}\right) - 1\right]. \tag{24.3b}$$

These are the boundary conditions for the diffusion equations in the p-doped layers and in the neutral region of the n-doped base. For the p-layers (with infinitely long contacts), the solution is (similar to (21.124)) for $x < -x_\mathrm{E}$ and $x > -x_\mathrm{C}$, respectively

$$n(x) = n_\mathrm{E} + \delta n(-x_\mathrm{E})\,\exp\left(\frac{x + x_\mathrm{E}}{L_\mathrm{E}}\right) \tag{24.4a}$$

$$n(x) = n_\mathrm{C} + \delta n(x_\mathrm{C})\,\exp\left(-\frac{x - x_\mathrm{C}}{L_\mathrm{C}}\right). \tag{24.4b}$$

L_E and L_C are the minority carrier (electron) diffusion lengths in the emitter and collector, respectively. The solution for the hole density in the neutral region in the base ($0 < x < w$) is

$$p(x) = p_B + \left[\frac{\delta p(w) - \delta p(0)\exp(-w/L_B)}{2\sinh(w/L_B)}\right]\exp\left(\frac{x}{L_B}\right) - \left[\frac{\delta p(w) - \delta p(0)\exp(w/L_B)}{2\sinh(w/L_B)}\right]\exp\left(-\frac{x}{L_B}\right). \tag{24.5}$$

We shall denote the excess hole density at $x = 0$ and $x = w$ as $\delta p_E = \delta p(0)$ and $\delta p_C = \delta p(w)$, respectively. Typical ('normal') operation condition in the common base circuit is that $\delta p_C = 0$ (Fig. 24.8a). In the 'inverted' configuration, the role of emitter and collector are reversed and $\delta p_E = 0$. We can write (24.5) also as

$$p(x) = p_B + \delta p_E \frac{\sinh[(w-x)/L_B)]}{\sinh[w/L_B]} + \delta p_C \frac{\sinh[x/L_B)]}{\sinh[w/L_B]}. \tag{24.6}$$

If the base is thick, i.e. $w \to \infty$, or at least large compared to the diffusion length ($w/L_B \gg 1$), the carrier concentration is given by

$$p(x) = p_B + \delta p(0)\exp\left(-\frac{x}{L_B}\right) \tag{24.7}$$

and does not depend on the collector. In this case there is no transistor effect. A 'coupling' between emitter and collector currents that are given by the derivative $\partial p/\partial x$ at 0 and w, respectively, is only present for a sufficiently thin base.

From (24.6), the hole current densities at $x = 0$ and $x = w$ are given as[2]

$$j_{Ep} = j_p(0) = e\frac{D_B}{L_B}\left[\delta p_E \coth\left(\frac{w}{L_B}\right) - \delta p_C \operatorname{csch}\left(\frac{w}{L_B}\right)\right] \tag{24.8a}$$

$$j_{Cp} = j_p(w) = e\frac{D_B}{L_B}\left[\delta p_E \operatorname{csch}\left(\frac{w}{L_B}\right) - \delta p_C \coth\left(\frac{w}{L_B}\right)\right]. \tag{24.8b}$$

From (24.4a, b), the electron current densities at $x = -x_E$ and $x = x_C$ are given (with $\delta n_E = \delta n(-x_E)$ and $\delta n_C = \delta n(x_C)$) by

$$j_{En} = j_n(-x_E) = e\frac{D_E}{L_E}\delta n_E \tag{24.9a}$$

$$j_{Cn} = j_n(x_C) = -e\frac{D_C}{L_C}\delta n_C. \tag{24.9b}$$

[2] $\coth x \equiv \cosh x/\sinh x$, $\operatorname{csch} x \equiv 1/\sinh x$.

The emitter current density is (similar to (21.127))

$$\begin{aligned} j_\mathrm{E} &= j_\mathrm{p}(0) + j_\mathrm{n}(-x_\mathrm{E}) \\ &= e\,\frac{D_\mathrm{B}}{L_\mathrm{B}}\left[\delta p_\mathrm{E}\,\coth\left(\frac{w}{L_\mathrm{B}}\right) - \delta p_\mathrm{C}\,\mathrm{csch}\left(\frac{w}{L_\mathrm{B}}\right)\right] + e\,\frac{D_\mathrm{E}}{L_\mathrm{E}}\,\delta n_\mathrm{E}. \end{aligned} \tag{24.10}$$

The collector current density is given as

$$\begin{aligned} j_\mathrm{C} &= j_\mathrm{p}(w) + j_\mathrm{n}(x_\mathrm{C}) \\ &= e\,\frac{D_\mathrm{B}}{L_\mathrm{B}}\left[\delta p_\mathrm{E}\,\mathrm{csch}\left(\frac{w}{L_\mathrm{B}}\right) - \delta p_\mathrm{C}\,\coth\left(\frac{w}{L_\mathrm{B}}\right)\right] - e\,\frac{D_\mathrm{C}}{L_\mathrm{C}}\,\delta n_\mathrm{C}. \end{aligned} \tag{24.11}$$

In these equations, only the diffusion currents are considered. Additionally, the recombination currents in the depletion layers must be considered, in particular at small junction voltages.

24.2.2 Current Amplification

The emitter current consists of two parts, the hole current I_pE injected from the base and the electron current I_nE that flows from the emitter to the base (Fig. 24.5a). Similarly, the collector current is made up from the hole and electron currents I_pC and I_pC, respectively.

The total emitter current splits into the base and collector currents

$$I_\mathrm{E} = I_\mathrm{B} + I_\mathrm{C}. \tag{24.12}$$

The amplification (gain) in common base circuits

$$\alpha_0 = h_\mathrm{FB} = \frac{\partial I_\mathrm{C}}{\partial I_\mathrm{E}} = \frac{\partial I_\mathrm{pE}}{\partial I_\mathrm{E}}\,\frac{\partial I_\mathrm{pC}}{\partial I_\mathrm{pE}}\,\frac{\partial I_\mathrm{C}}{\partial I_\mathrm{pC}} = \gamma\,\alpha_\mathrm{T}\,M, \tag{24.13}$$

where γ is the emitter efficiency, α_T the base transport factor and M the collector multiplication factor. Since the collector is normally operated below the threshold for avalanche multiplication, $M = 1$.

The current amplification in the common emitter circuit is

$$\beta_0 = h_\mathrm{FE} = \frac{\partial I_\mathrm{C}}{\partial I_\mathrm{B}}. \tag{24.14}$$

Using (24.12), we find

$$\beta_0 = \frac{\partial I_\mathrm{E}}{\partial I_\mathrm{B}} - 1 = \frac{\partial I_\mathrm{E}}{\partial I_\mathrm{C}}\,\frac{\partial I_\mathrm{C}}{\partial I_\mathrm{B}} - 1 = \frac{1}{\alpha_0}\,\beta_0 - 1 = \frac{\alpha_0}{1-\alpha_0}. \tag{24.15}$$

Since α_0 is close to 1 for a well-designed transistor, β_0 is a large number, e.g. $\beta_0 = 99$ for $\alpha_0 = 0.99$.

The emitter efficiency is (A denotes the device area)

$$\gamma = \frac{A j_{\mathrm{Ep}}}{I_{\mathrm{E}}} = \left[1 + \frac{n_{\mathrm{E}}}{p_{\mathrm{B}}} \frac{D_{\mathrm{E}}}{D_{\mathrm{B}}} \frac{L_{\mathrm{B}}}{L_{\mathrm{E}}} \tanh\left(\frac{w}{L_{\mathrm{B}}}\right)\right]^{-1}. \tag{24.16}$$

The base transport factor, i.e. the ratio of minority carriers reaching the collector and the total number of injected minority carriers, is (for reverse bias $|\beta U_{\mathrm{CB}}| \gg kT$)

$$\begin{aligned} \alpha_{\mathrm{T}} = \frac{j_{\mathrm{Cp}}}{j_{\mathrm{Ep}}} &= \frac{\exp(\beta U_{\mathrm{EB}}) - 1 + \cosh w/L_{\mathrm{B}}}{1 + (\exp(\beta U_{\mathrm{EB}}) - 1)\ \cosh w/L_{\mathrm{B}}} \\ &\approx \frac{1}{\cosh\,(w/L_{\mathrm{B}})} \approx 1 - \frac{w^2}{2L_{\mathrm{B}}^2}. \end{aligned} \tag{24.17}$$

The first approximation is for $\beta U_{\mathrm{EB}} \gg 1$ (emitter diode injecting in forward direction), the second approximation is for $w \ll L_{\mathrm{B}}$. If the base length is a tenth of the diffusion length, the base transport factor is $\alpha_{\mathrm{T}} > 0.995$. M is also very close to 1; for reverse bias U_{CB} and $w \ll L_{\mathrm{B}}$ we find

$$M \approx 1 + \frac{w}{L_{\mathrm{C}}} \frac{D_{\mathrm{C}}}{D_{\mathrm{B}}} \frac{\delta n_{\mathrm{C}}}{\delta p_{\mathrm{C}} - \delta p_{\mathrm{E}}} \approx 1 + \frac{w}{L_{\mathrm{C}}} \frac{D_{\mathrm{C}}}{D_{\mathrm{B}}} \frac{n_{\mathrm{C}}}{p_{\mathrm{B}}} \exp(-\beta U_{\mathrm{EB}}). \tag{24.18}$$

Thus for $w \ll L_{\mathrm{B}}$, α_0 is dominated by γ and given as (approximating (24.16))

$$\alpha_0 \approx \gamma \approx 1 - \frac{w}{L_{\mathrm{E}}} \frac{n_{\mathrm{E}}}{p_{\mathrm{B}}} \frac{D_{\mathrm{E}}}{D_{\mathrm{B}}}. \tag{24.19}$$

The gain β_0 is then also determined by γ:

$$\beta_0 = h_{\mathrm{FE}} \approx \frac{\gamma}{1-\gamma} \approx \frac{1}{1-\gamma} \propto \frac{N_{\mathrm{E}}}{N_{\mathrm{B}}} \frac{L_{\mathrm{E}}}{w}, \tag{24.20}$$

N_{E} and N_{B} being the doping levels in the emitter and base, respectively. The base and collector current are shown in Fig. 24.6 as a function of the emitter–base voltage, i.e. the voltage at the injection diode. The collector current is close to the emitter–base diode current and displays a dependence $\propto \exp(eV_{\mathrm{EB}}/kT)$. The base current shows a similar slope but is orders of magnitude smaller in amplitude. For small forward voltages of the emitter–base diode, the current is typically dominated by a nonradiative recombination current that flows through the base contact and has an ideality factor (m in Fig. 24.6) close to 2.

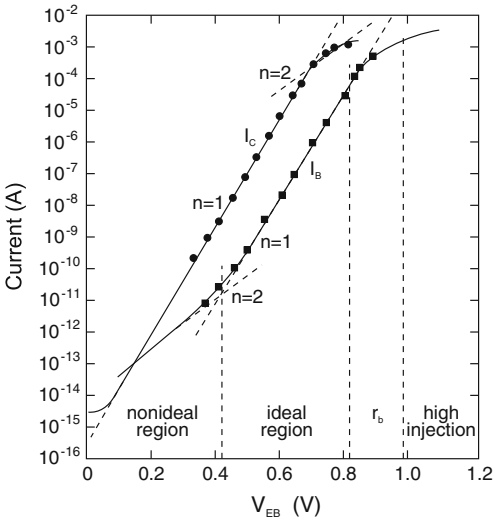


Fig. 24.6 Collector current I_C and base current I_B as a function of the emitter–base voltage V_{EB} (Gummel plot). Adapted from [1747]

24.2.3 Ebers–Moll Model

The Ebers–Moll model (Fig. 24.7) was developed in 1954 and is a relatively simple transistor model that needs, at its simplest level (Fig. 24.7a) just three parameters. It can (and must) be refined (Fig. 24.7b, c). The model considers two ideal diodes ('F' (forward) and 'R' (reverse)) back to back, each feeding a current source. The F diode represents the emitter–base diode and the R diode the collector–base diode. The currents are

$$I_F = I_{F0}\left[\exp\left(\beta V_{EB}\right) - 1\right] \tag{24.21a}$$
$$I_R = I_{R0}\left[\exp\left(\beta V_{CB}\right) - 1\right]. \tag{24.21b}$$

Using (24.8a, b)–(24.11), the emitter and collector currents are

$$I_E = \hat{a}_{11}\left[\exp\left(\beta V_{EB}\right) - 1\right] + \hat{a}_{12}\left[\exp\left(\beta V_{CB}\right) - 1\right] \tag{24.22a}$$
$$I_C = \hat{a}_{21}\left[\exp\left(\beta V_{EB}\right) - 1\right] + \hat{a}_{22}\left[\exp\left(\beta V_{CB}\right) - 1\right], \tag{24.22b}$$

with

$$\hat{a}_{11} = eA\left[p_B \frac{D_B}{L_B}\coth\left(\frac{w}{L_B}\right) + n_E \frac{D_E}{L_E}\right] \tag{24.23a}$$
$$\hat{a}_{12} = -eAp_B \frac{D_B}{L_B}\operatorname{csch}\left(\frac{w}{L_B}\right) \tag{24.23b}$$

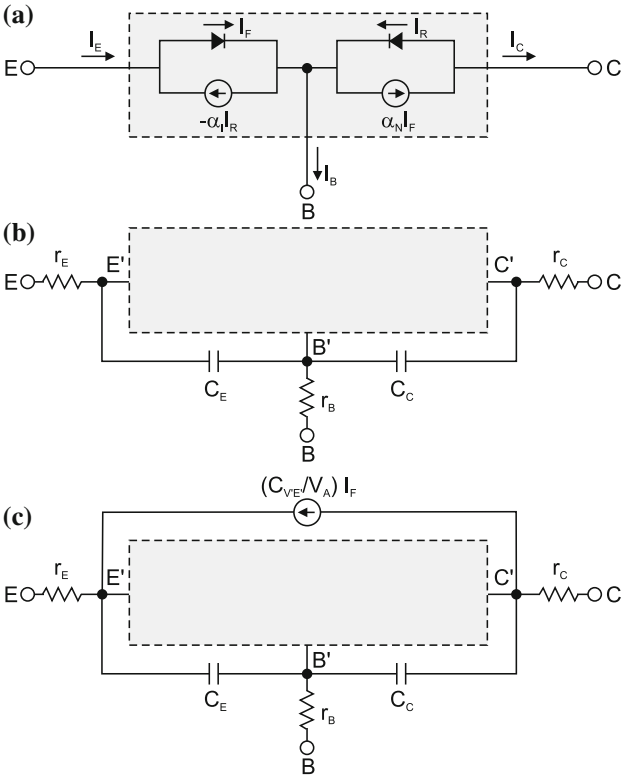


Fig. 24.7 Ebers–Moll model of a transistor, 'E': emitter, 'C': collector and 'B': base. Currents are shown for a pnp transistor. (**a**) Basic model (*grey area* in (**b**, **c**)), (**b**) model with series resistances and depletion-layer capacitances, (**c**) model additionally including the Early effect (V_A: Early voltage)

$$\hat{a}_{21} = eAp_B \frac{D_B}{L_B} \operatorname{csch}\left(\frac{w}{L_B}\right) = -\hat{a}_{12} \tag{24.23c}$$

$$\hat{a}_{22} = -eA\left[p_B \frac{D_B}{L_B} \coth\left(\frac{w}{L_B}\right) + n_C \frac{D_C}{L_C}\right]. \tag{24.23d}$$

The currents at the three contacts are

$$I_E = I_F - \alpha_I I_R \tag{24.24a}$$

$$I_C = \alpha_N I_F - I_R \tag{24.24b}$$

$$I_B = (1-\alpha_N) I_F + (1-\alpha_I) I_R. \tag{24.24c}$$

The last equation is obtained from (24.24a, b) using (24.12). By comparison with (24.21a, b) and (24.23a–d) we find

$$I_{F0} = \hat{a}_{11} \tag{24.25a}$$
$$I_{R0} = -\hat{a}_{22} \tag{24.25b}$$
$$\alpha_I = \hat{a}_{12}/I_{R0} \tag{24.25c}$$
$$\alpha_N = \hat{a}_{21}/I_{F0} = -\hat{a}_{12}/I_{F0} = -\alpha_I I_{R0}/I_{F0}. \tag{24.25d}$$

The constants α_N and α_I are the forward ('normal') ($\alpha_N = \alpha_0$ from (24.13)) and reverse ('inverted') gains in the common base circuit, respectively. Both constants are larger than zero. Typically, $\alpha_N \approx 0.98 \ldots 0.998 \lessapprox 1$ and $\alpha_I \approx 0.5 \ldots 0.9 < \alpha_N$. The model has three independent parameters, e.g. α_N, I_{F0} and I_{R0}. Equation (24.24a, b) can be rewritten as

$$I_E = \alpha_I\, I_C + (1 - \alpha_I\, \alpha_N)\, I_F \tag{24.26a}$$
$$I_C = \alpha_N\, I_E - (1 - \alpha_I\, \alpha_N)\, I_R. \tag{24.26b}$$

Under normal operation we have

$$I_E = I_F \tag{24.27a}$$
$$I_C = \alpha_N\, I_E. \tag{24.27b}$$

The model can be refined and made more realistic by including the effect of series resistances and depletion-layer capacitances, increasing the number of parameters to eight. The Early effect (see p. 794) can be included by adding a further current source. This level is the 'standard' Ebers–Moll model with a total of nine parameters. Further parameters can be added. However, as is always the case with simulations, there is a tradeoff between the simplicity of the model and to what detail a real situation is approximated.

24.2.4 Current–Voltage Characteristics

In Fig. 24.8, the hole density in the base (of a pnp transistor) is shown for various voltage conditions. In Fig. 24.9, the I–V characteristics of a bipolar transistor in common base and common collector circuit are shown. In the common base circuit (Fig. 24.9a), the collector current is practically equal to the emitter current and is almost independent of the collector–base voltage. From (24.26b), the dependence of the collector current on the collector–base voltage is given (within the Ebers–Moll model) as

$$I_C = \alpha_N I_E - (1 - \alpha_I \alpha_N) I_{R0} \left[\exp(\beta V_{CB}) - 1\right]. \tag{24.28}$$

V_{CB} is in the reverse direction. Therefore, the second term is zero for normal operating conditions. Since $\alpha_N \lessapprox 1$, the collector current is almost equal to the emitter current.

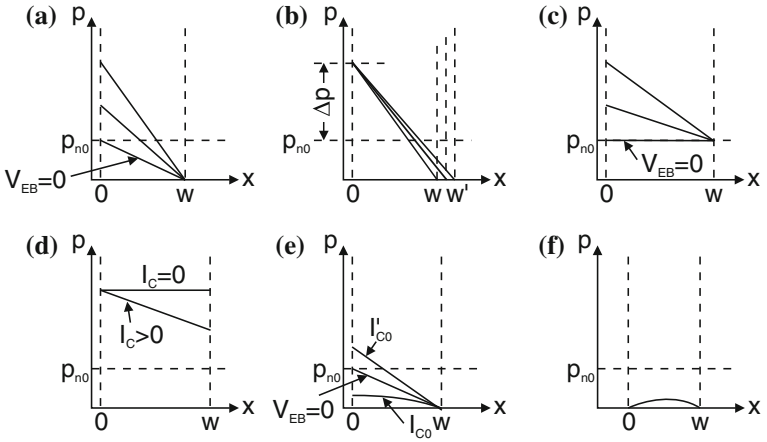


Fig. 24.8 Hole density (linear scale) in the base region (the neutral part of the base ranges from 0 to w) of a pnp transistor for various voltages. (**a**) normal voltages, V_{CB} = const. and various V_{EB} (in forward direction). (**b**) V_{EB} = const. and various values of V_{CB}. (**c**) Various values of $V_{EB} > 0$, $V_{CB} = 0$. (**d**) Both pn-junctions in forward direction. (**e**) Conditions for I_{C0} and I'_{C0}. (**f**) Both junctions in reverse direction. Adapted from [500]

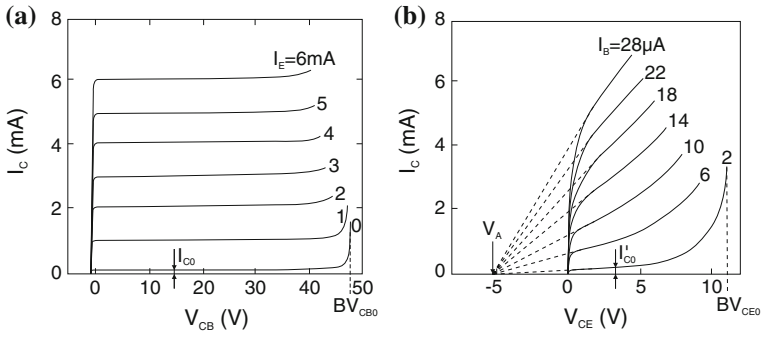


Fig. 24.9 Characteristics (I_C vs. V_{CB}) of a pnp transistor in (**a**) common base (CB) circuit (Fig. 24.4a) for various values of the emitter current as labeled. Adapted from [1748]. (**b**) Characteristics in common emitter (CE) circuit (Fig. 24.4b). Adapted from [1749]

Even at $V_{CB} = 0$ (the case of (Fig. 24.8c), holes are extracted from the base since $\partial p/\partial x|_{x=w} > 0$. A small forward voltage must be applied to the collector–base diode in order to make the current zero, i.e. $\partial p/\partial x|_{x=w} = 0$ (Fig. 24.8d). The collector saturation current I_{C0} is measured with an open emitter side. This current is smaller than the saturation current of the CB diode, since at the emitter side of the basis a vanishing gradient of the hole density is present (Fig. 24.8e). This reduces the gradient (and thus the current) at the collector side. The current I_{C0} is therefore smaller than the collector current for shorted emitter–base contact ($V_{EB} = 0$). At high collector voltage, the current increases rapidly at BV_{CB0} due to breakdown of the collector–base diode. It can also occur that the width of the neutral base region

w becomes zero (punch-through). In this case, the emitter and collector are short-circuited.

In the common emitter circuit (Fig. 24.9b), there is a high current amplification I_C/I_B. Note that the collector current is given in mA and the base current in μA. The current increases with increasing V_{CE} because the base width w decreases and β_0 increases. There is no saturation of the I–V characteristics (Early effect [1750]). Instead, the I–V curves look as if they start at a negative collector–emitter voltage, the so-called Early voltage V_A. In the linear regime, the characteristic can be approximated by

$$I_C = \left(1 + \frac{V_{CE}}{V_A}\right) \beta_0 \, I_B. \tag{24.29}$$

Here, β_0 is the current gain for $V_{CE} \approx 0$.

The physical reason for the increase of the collector current with increasing V_{CE} is the increasing reverse voltage at the collector–base diode that causes the so-called 'base-width modulation', as shown in Fig. 24.8b. The expansion of the CB depletion layer leads subsequently to a reduction of the neutral base width w. w will be smaller and smaller compared to the geometrical base width w_B. When w is reduced, the common base gain α_0 (24.19) becomes closer to 1 and the current gain increases. Therefore, the collector current increases with V_{CE} for a given (fixed) base current. The Early voltage is the coefficient of the increase of collector current with V_{CE},

$$\frac{\beta_0 \, I_B}{V_A} = \frac{\partial I_C}{\partial V_{CE}} = \frac{\partial I_C}{\partial V_{CB}} \frac{\partial V_{CB}}{\partial V_{CE}} \approx \frac{\partial I_C}{\partial V_{CB}}. \tag{24.30}$$

For constant base current, the emitter-base voltage is almost constant and the approximation in (24.30) holds. The dependence of the CB depletion layer width on the base side x_C^n on U_{CB} is given by (21.106a) for a pnp-transistor. Typical values for the Early voltage are 50–300 V. The modeling of the Early effect in the SPICE simulation program is discussed in [1751].

For small collector–emitter voltage, the current quickly drops to zero. V_{CE} is typically split in such a way that the emitter–base diode is well biased forward and the CB diode has a high reverse voltage. If V_{CE} drops below a certain value (≈1 V for silicon transistors), there is no longer any bias at the CB diode. A further reduction of V_{CE} biases the CB diode in the forward direction and quickly brings the collector current down to zero.

24.2.5 Basic Circuits

Common Base Circuit

In the common base configuration, there is no current amplification since the currents flowing through emitter and base are almost the same. However, there is voltage gain since the collector current causes a large voltage drop across the load resistor.

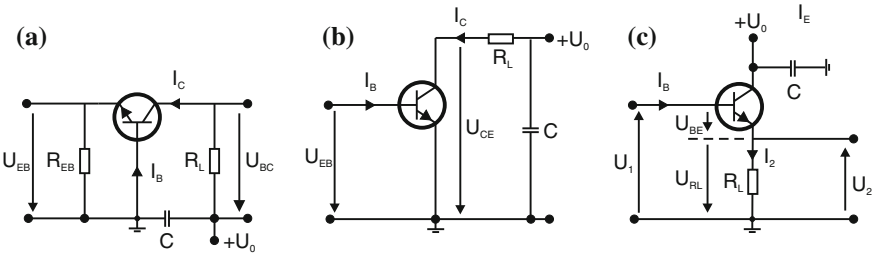


Fig. 24.10 (**a**) Common base, (**b**) common emitter and (**c**) common collector circuits with external loads

Common Emitter Circuit

The input resistance of the common emitter circuit (Fig. 24.10a) depends on the emitter–base diode and varies between a value of the order of 100 kΩ at small current and a few Ω at larger current and high V_{EB}. The voltage gain is

$$r_V = \frac{V_{CE}}{V_{EB}} = \frac{I_C}{V_{EB}} R_L, \tag{24.31}$$

where R_L is the load resistance in the output circuit (see Fig. 24.4b). The ratio $g_m = I_C / V_{EB}$ is called the forward transconductance. Also, the differential transconductance $g'_m = \partial I_C / \partial V_{EB}$ is used. The voltage gain of the common emitter circuit is typically 10^2–10^3. Since current *and* voltage are amplified, this circuit has the highest power gain.

If the input voltage V_{EB} (U_1 in Fig. 24.10a) is increased, the collector current rises. This increase causes an increase of the voltage drop across the load resistance R_L and a decrease of the output voltage U_2. Therefore, the phase of the input signal is reversed and the amplifier is inverting.

Common Collector Circuit

In Fig. 24.10c, the collector is connected to mass for alternating currents. Input and output current flow through the load resistance at which part of the input voltage drops. The input voltage is divided between the load resistor R_L and the emitter–base diode. At the transistor, the voltage $V_{BE} = V_1 - V_{RL}$ is applied. If the input voltage is increased, I_2 increases. This leads to a larger voltage drop at the load resistor and therefore to a decrease of V_{BE}, working against the original increase. The input resistance R_1 is large despite a small load resistance, $R_1 \approx \beta_0 R_L$. The input voltage is larger than V_{RL}, thus no voltage gain occurs (actually it is a little smaller than 1). The current amplification is $(\beta + 1)$. The output resistance R_2 is small, $R_2 = U_2 / I_2 = R_L \approx R_1 / \beta_0$. Therefore, this circuit is also called an impedance amplifier that allows high-impedance sources to be connected low-impedance loads.

Since an increase of the input voltage leads to an increase of the output voltage that is present at the emitter, this circuit is a direct amplifier and is also called an emitter follower.

24.2.6 High-Frequency Properties

Transistors for amplification of high-frequency signals are typically chosen as npn transistors since electrons, the minority carriers in the base, have higher mobility than holes. The active area and parasitic capacitance must be minimized. The emitter is formed in the shape of a stripe, nowadays in the 100 nm regime. The base width is in the 10 nm range. High p-doping of GaAs with low diffusion of the dopant is accomplished with carbon. Defects that would short emitter and collector at such thin base width must be avoided.

An important figure of merit is the cutoff frequency f_T for which h_{FE} is unity in the common emitter configuration. The cutoff frequency is related to the emitter–collector delay time τ_{EC} by

$$f_T = \frac{1}{2\pi\,\tau_{EC}}. \tag{24.32}$$

The delay time is determined by the charging time of the emitter–base depletion layer, the base capacitance, and the transport through the base–collector depletion layer. It is favorable if all times are short and similar. It does not help to minimize only one or two of the three processes since the longest time determines the transistor performance.

Another important figure of merit is the maximum frequency with which the transistor can oscillate in a feedback circuit with zero loss. This frequency is denoted by f_{max}. Approximately,

$$f_{max} \simeq \sqrt{\frac{f_T}{8\pi\,R_B\,C_{CB}}}, \tag{24.33}$$

where R_B is the base resistance and C_{CB} is the collector–base capacitance. f_{max} is larger than f_T, by a factor of the order of three.

24.2.7 Heterojunction Bipolar Transistors

In a heterojunction bipolar transistor (HBT), the emitter–base diode is formed with a heterostructure diode. The desired functionality is obtained when the emitter is made from the higher-bandgap material and the base from the lower-bandgap material. The

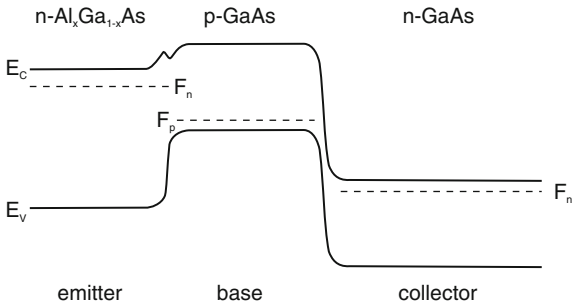


Fig. 24.11 Schematic band diagram of a heterojunction bipolar transistor

schematic band diagram is shown in Fig. 24.11 (see Fig. 21.57c for the emitter–base diode).

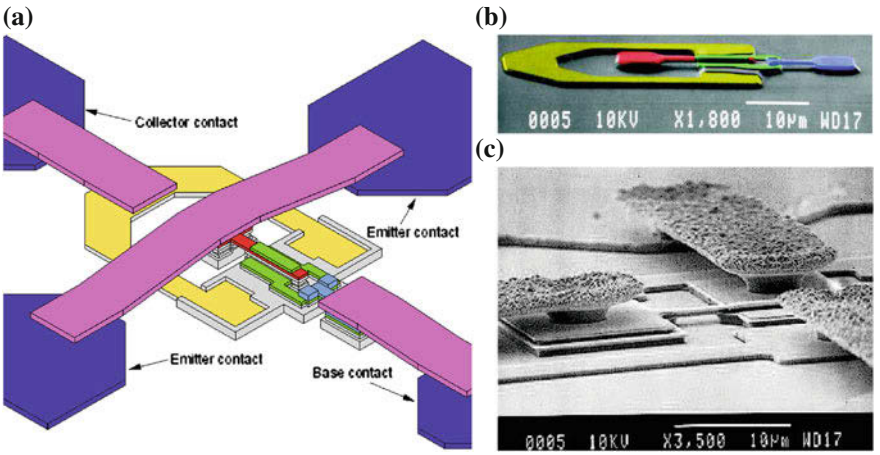
The higher discontinuity in the valence band, compared to a homojunction with the base material, provides a higher barrier for hole transport from the base to the emitter. Thus, the emitter efficiency is increased. Another advantage is the possibility for higher doping of the base without loss of emitter efficiency. This reduces the base series resistance and leads to better high-frequency behavior due to higher current gain and a smaller RC time constant. Also, operation at higher temperature is possible when the emitter has a larger band gap. Current InP/InGaAs-based HBTs have cutoff frequencies beyond 30 GHz, SiGe-HBTs beyond 80 GHz. The high-frequency performance is influenced by the velocity-overshoot effect (cf. Sect. 8.4.3) [1752].

In Fig. 24.12, an InAlAs/InGaAs HBT is shown [1754]. The cutoff frequency is 90 GHz. For the layer design, a fairly thick collector with low doping was chosen. This design allows a broad depletion layer with fairly small maximum electric field and thus a high breakdown voltage of $B V_{CE0} > 8.5$ V. The base is not too thin (80 instead of maybe 60 nm) to reduce the series resistance. A graded region between emitter and base was chosen to avoid a spike occurring in the conduction band (Fig. 21.57b) and keep the turn-on voltage low.

24.2.8 Light-Emitting Transistors

The base current has two components. One is the recombination current in the neutral region of the emitter; this current can be suppressed in the HBT. The other is the recombination in the base region itself.[3] If quantum wells are introduced into the

[3] Also, a recombination current in the emitter–base depletion region is possible. However, since in normal operating conditions this diode is forward biased, the depletion layer is short and the associated recombination current is small, cf. p. 638.



(d)

Material	d (nm)	type	function
(In,Ga)As	75	n^+	emitter cap
(In,Al)As	45	N^+	emitter cap
(In,Al)As	100	N	emitter
(In,Al,Ga)As	45	N	grade
(In,Ga)As	80	p^+	base
(In,Ga)As	700	n^-	collector
(In,Ga)As	430	n^+	subcollector
InP substrate			

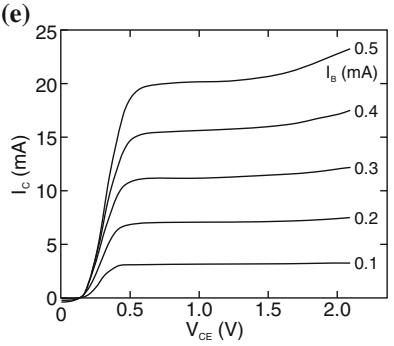


Fig. 24.12 (**a**) Schematic layout of a high-frequency HBT and SEM images (**b**) without and (**c**) with contacts. (**d**) Epitaxial layer sequence and (**e**) static performance data. Parts (**a**, **b**) from [1753], parts (**d**, **e**) from [1754]

base region, this recombination can occur radiatively between electrons and holes captured into the quantum well (Fig. 24.13). The spectrum exhibits two peaks from the QWs and the GaAs barrier.

24.3 Field-Effect Transistors

Next to the bipolar transistors, the field-effect transistors (FET) are another large class of transistors. FETs were conceptualized first but due to technological difficulties with semiconductor surfaces, realized second. The principle is fairly simple: A current flows through a channel from source to drain. The current is varied via the channel conductivity upon the change of the gate voltage. The gate needs to make a nonohmic contact to the semiconductor. Since the conductivity in the channel is a

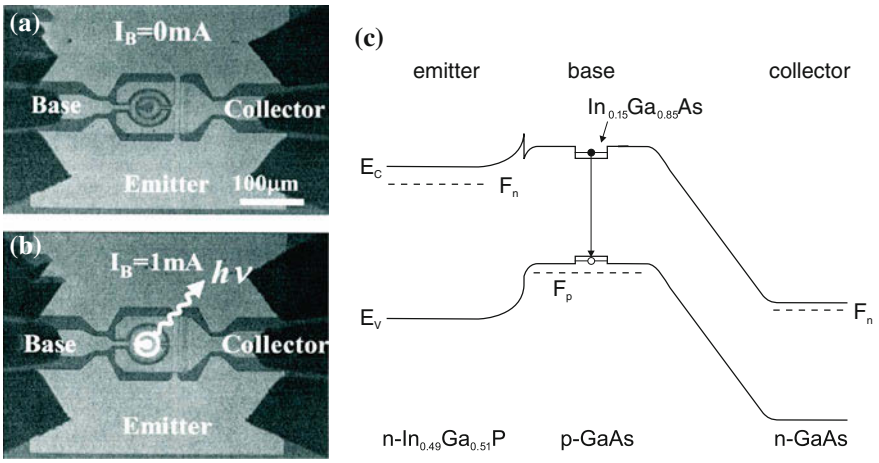


Fig. 24.13 Microscopic image of an InGaP/GaAs HBT with two 5-nm InGaAs/GaAs QWs in the 30-nm wide base at (**a**) zero base current and (**b**) at 1 mA base current in the common emitter configuration with Si CCD image of light emission. (**c**) Schematic band diagram of a HBT with a single InGaAs/GaAs quantum well in the base. Parts (**a**, **b**) from [1755], part (**c**) adapted from [1756]

property related to the majority charge carriers, FETs are called unipolar transistors. FETs feature a higher input impedance than bipolar transistors, a good linearity, and a negative temperature coefficient and thus a more homogeneous temperature distribution. According to the structure of the gate diode we distinguish JFETs, MESFETs and MOSFETs, as discussed in the following.

In the junction FET (JFET), the variation of channel conductivity is accomplished via the extension of the depletion layer of the pn-junction formed by the gate and the channel material (Fig. 24.14a). The JFET was analyzed by Schottky in 1952 [98] and realized by Dacey and Ross in 1953 [99].

In a MESFET, a metal–semiconductor diode (Schottky diode) is used as rectifying contact instead of a pn-diode. Otherwise, the principle is the same as that of the JFET. After the proposal by Mead in 1966 [123], the first (epitaxial) GaAs MESFET was realized by Hooper and Lehrer in 1967 [126]. The MESFET offers some advantages, such as the fabrication of the metal gate at lower temperature than necessary for the (diffusion or epitaxy of the) pn-diode, lower resistance, good thermal contact. The JFET can be made with a heterostructure gate to improve the frequency response.

In a MISFET, the gate diode is a metal–insulator–semiconductor diode (Fig. 24.14b). If the insulator is an oxide, the related FET is a MOSFET. When the gate is put at a positive voltage (for a p-channel), an inversion layer is formed close to the insulator–semiconductor interface. This layer is an n-conducting channel allowing conduction between the two oppositely biased pn-diodes. It can carry a high current. The MOSFET was theoretically envisioned early by Lilienfeld in 1925 [48] and realized only in 1960 by Kahng and Atalla [111].

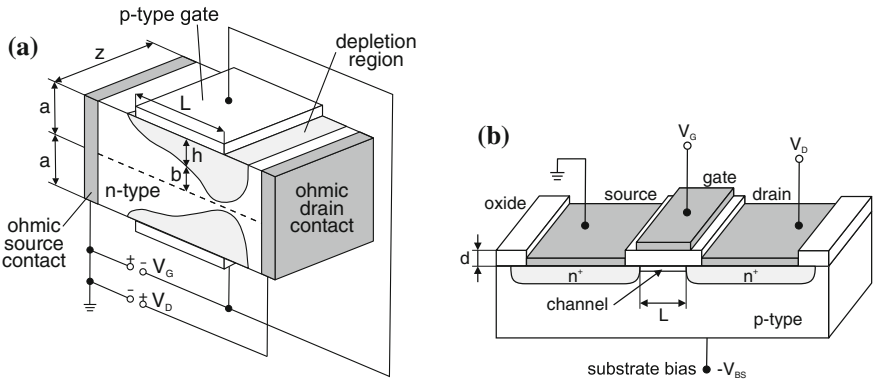


Fig. 24.14 (**a**) Shockley's model of a JFET. The *dashed line* represents the middle of the symmetric channel of total thickness $2a$. The *light grey area* is the depletion layer with thickness h. The gate length is L. The *dark grey* areas are ohmic metal contacts. Based on [99]. (**b**) Scheme of a MOSFET with channel length L and oxide thickness d. The *dark grey* areas are ohmic metal contacts. Adapted from [500]

FETs come in 'n' and 'p' flavors, depending on the conductivity type of the channel. For high-frequency applications, typically an n-channel is used due to the higher mobility or drift velocity. In CMOS (complementary MOS) technology, both n-FETs and p-FETs are integrated in high density, allowing the effective realization of logic gates with minimized power consumption.

24.4 JFET and MESFET

24.4.1 General Principle

The principal characteristic of a JFET is shown in Fig. 24.15. At $V_D = 0$ and $V_G = 0$, the transistor is in thermodynamic equilibrium and there are no net currents. Underneath the gate diode, a depletion layer is present. If for zero gate voltage the source–drain voltage is applied to the channel, the current increases linearly. The positive voltage at the drain contact causes the expansion of the depletion layer of the (reversely biased) gate–drain pn-diode. When the two (the upper and the lower) depletion regions meet (pinch-off), the current saturates. The respective source–drain voltage is denoted as $V_{D,sat}$. For high gate–drain (reverse) voltage V_D, breakdown occurs with a strong increase of the source–drain current. A variation of the gate voltage V_G leads to a variation of the source–drain current. A reverse voltage leads to a reduction of the saturation current and saturation at lower source–drain voltage. For a certain gate voltage V_P, the pinch-off voltage, no current can flow in the channel any longer since pinch-off exists even for $V_D = 0$.

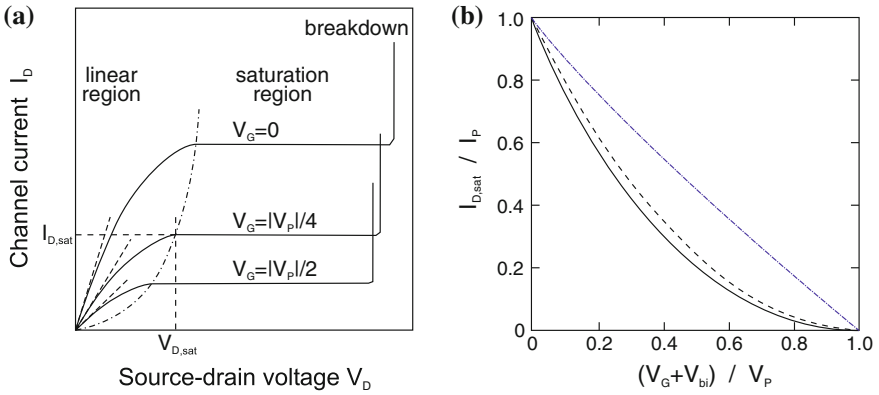


Fig. 24.15 (**a**) Principal characteristics of a JFET. The channel current I_D is shown as a function of the source–drain voltage V_D for three different values of the (absolute value of the) gate voltage V_G. The saturation values $V_{D,sat}$ and $I_{D,sat}$ are indicated for one curve. The intersections with the *dash-dotted* line yield the saturation voltage. Adapted from [500]. (**b**) Transfer behavior of a JFET for two different carrier distributions, homogeneous (*solid* line) and δ-like (*dashed* line). The *blue, dash-dotted* line is $\sqrt{I_D/I_P}$ versus the gate voltage. After [1756, 1757]

24.4.2 Static Characteristics

Here, we will calculate the general static behavior outlined in the previous section. We assume a long channel ($L \gg a$), the abrupt approximation for the depletion layer, the gradual channel approximation, i.e. the depletion layer depth changes slowly along x, and a field-independent, constant mobility. In this case, the two-dimensional Poisson equation for the potential distribution V can be used by solving it along the y direction (channel depth) for all x-positions (adiabatic approximation),

$$\frac{\partial^2 V}{\partial y^2} = -\frac{\rho(y)}{\epsilon_s}. \tag{24.34}$$

The geometry is shown in the inset of Fig. 24.15b.

The depth h of the depletion layer in the abrupt approximation is given by (cf. (21.107), reverse voltages are counted as positive here)

$$h = \sqrt{\frac{2\epsilon_s}{eN_D}\,(V_{bi} + V_G + V(x))}. \tag{24.35}$$

Here, we have assumed homogeneous doping, i.e. N_D does not depend on y (or x). The built-in voltage (for a p^+n gate diode) is given by $V_{bi} = \beta^{-1} \ln(N_D/n_i)$ (21.97a). The voltage V is the applied source–drain voltage in relation to the source. The depth of the depletion layer at $x = 0$ (source) and $x = L$ (drain) is given by

$$y_1 = h(0) = \sqrt{\frac{2\,\epsilon_\mathrm{s}}{e\,N_\mathrm{D}}\,(V_\mathrm{bi} + V_\mathrm{G})} \tag{24.36a}$$

$$y_2 = h(L) = \sqrt{\frac{2\,\epsilon_\mathrm{s}}{e\,N_\mathrm{D}}\,(V_\mathrm{bi} + V_\mathrm{G} + V_\mathrm{D})}. \tag{24.36b}$$

The maximum value of h is a. Therefore, the pinch-off voltage V_P, at which $V_\mathrm{P} = V_\mathrm{bi} + V_\mathrm{G} + V_\mathrm{D}$ is such that $h = a$, is given by

$$V_\mathrm{P} = \frac{e\,N_\mathrm{D}\,a^2}{2\,\epsilon_\mathrm{s}}. \tag{24.37}$$

The (drift) current density along x is given by (cf. (8.53a))

$$j_x = -e\,N_\mathrm{D}\,\mu_\mathrm{n}\,E_x = e\,N_\mathrm{D}\,\mu_\mathrm{n}\,\frac{\partial V}{\partial x} \tag{24.38}$$

for the neutral part of the semiconductor. Therefore, the current in the upper half of the channel is given by

$$I_\mathrm{D} = e\,N_\mathrm{D}\,\mu_\mathrm{n}\,\frac{\partial V(x)}{\partial x}\,Z\,[a - h(x)]\,, \tag{24.39}$$

where Z is the width of the channel (Fig. 24.14a). Although it seems that I_D depends on x, it is of course constant along the channel due to Kirchhoff's law.[4] Using the triviality $\int_0^L I_\mathrm{D}\,\mathrm{d}x = L\,I_\mathrm{D}$ and $\frac{\partial V}{\partial x} = \frac{\partial V}{\partial h}\,\frac{\partial h}{\partial x}$ with $\frac{\partial V}{\partial h} = eN_\mathrm{D}h/\epsilon_\mathrm{s}$ from (24.35), we find from (24.39)

$$I_\mathrm{D} = \frac{e^2\,\mu_\mathrm{n}\,N_\mathrm{D}^2\,Z\,a^3}{6\,\epsilon_\mathrm{s}\,L}\left[\frac{3}{a^2}\,\left(y_2^2 - y_1^2\right) - \frac{2}{a^3}\,\left(y_1^3 - y_2^3\right)\right]. \tag{24.40}$$

[4] We neglect recombination, in particular since the current is a majority-carrier current.

This equation can also be written, using (24.37) and

$$I_{\mathrm{P}} = \frac{e^2\,\mu_{\mathrm{n}}\,N_{\mathrm{D}}^2\,Z\,a^3}{6\,\epsilon_{\mathrm{s}}\,L}, \tag{24.41}$$

as

$$I_{\mathrm{D}} = I_{\mathrm{P}}\left[\frac{3V_{\mathrm{D}}}{V_{\mathrm{P}}} - 2\,\frac{(V_{\mathrm{bi}}+V_{\mathrm{G}}+V_{\mathrm{D}})^{3/2} - (V_{\mathrm{bi}}+V_{\mathrm{G}})^{3/2}}{V_{\mathrm{P}}^{3/2}}\right]. \tag{24.42}$$

The saturation current is reached for $y_2 = a$ or $V_{\mathrm{bi}} + V_{\mathrm{G}} + V_{\mathrm{D}} = V_{\mathrm{P}}$ and is given by

$$I_{\mathrm{D,sat}} = I_{\mathrm{P}}\left[1 - 3\,\frac{V_{\mathrm{bi}}+V_{\mathrm{G}}}{V_{\mathrm{P}}} + 2\left(\frac{V_{\mathrm{bi}}+V_{\mathrm{G}}}{V_{\mathrm{P}}}\right)^{3/2}\right]. \tag{24.43}$$

The dependence of the saturation current on $(V_{\mathrm{G}} + V_{\mathrm{bi}})/V_{\mathrm{P}}$ is depicted in Fig. 24.15b. For the threshold (gate) voltage of

$$V_{\mathrm{T}} = V_{\mathrm{P}} - V_{\mathrm{bi}}, \tag{24.44}$$

the saturation current is zero since then $V_{\mathrm{D}} = 0$.[5] Around the threshold voltage, the drain saturation current is given in lowest order of V_{G} as

$$I_{\mathrm{D,sat}} \approx \frac{3\,I_{\mathrm{P}}}{4}\left(\frac{V_{\mathrm{G}} - V_{\mathrm{T}}}{V_{\mathrm{P}}}\right)^2. \tag{24.45}$$

Thus, in order to experimentally determine the threshold voltage, $\sqrt{I_{\mathrm{D}}}$ is plotted versus the gate voltage and extrapolated to $I_{\mathrm{D}} = 0$ (dash-dotted line in Figs. 24.15 and 24.16).

The source–drain voltage at the saturation point decreases with decreasing saturation current, shown as dashed parabola-like line in Fig. 24.15a.

If the charge-carrier distribution differs from the homogeneous distribution assumed so far, a change of transistor properties arises, as shown in Fig. 24.15b for a δ-like carrier distribution. The I–V characteristic is slightly less curved, but not linear. A linear characteristic is only achievable in the drift velocity saturation regime (cf. Sect. 24.4.4).

For high source–drain voltage $V_{\mathrm{D}} > V_{\mathrm{P}} - V_{\mathrm{bi}} - V_{\mathrm{G}}$, the current remains essentially at its saturation value. For very high source–drain voltage, breakdown in the gate–drain diode can occur, when the maximum voltage, which is given by $V_{\mathrm{G}} + V_{\mathrm{D}}$ at the end of the channel, is equal to the breakdown voltage V_{B}.

[5] The threshold voltage can also be obtained from the condition $g_{\mathrm{D0}} = 0$ (cf. (24.49)).

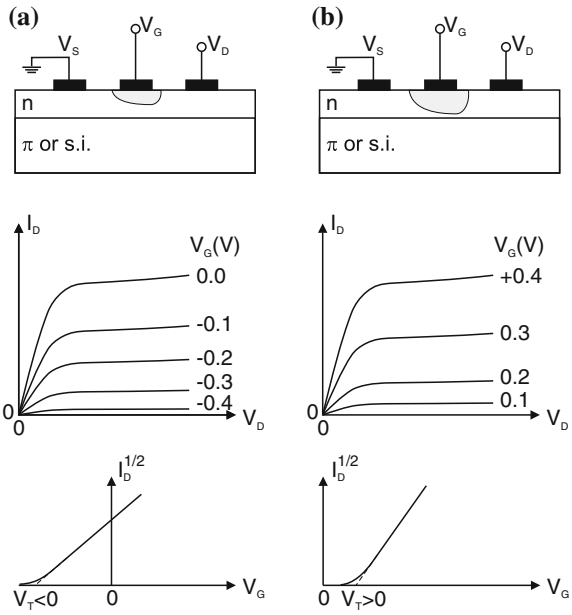


Fig. 24.16 Scheme (*top*), I_D versus V_D (*center*) and $I_D^{1/2}$ versus V_G (*bottom*) I–V characteristics for (**a**) normally on (depletion) and (**b**) normally off (accumulation) n-type JFET. Adapted from [500]

The forward transconductance g_m and the drain transconductance g_D are given by

$$g_m = \frac{\partial I_D}{\partial V_G} = g_{max} \left[\sqrt{\frac{V_{bi} + V_G}{V_P}} - \sqrt{\frac{V_{bi} + V_G + V_D}{V_P}} \right] \tag{24.46}$$

$$g_D = \frac{\partial I_D}{\partial V_D} = g_{max} \left[1 - \sqrt{\frac{V_{bi} + V_G + V_D}{V_P}} \right], \tag{24.47}$$

where

$$g_{max} = \frac{3\, I_P}{V_P} = \frac{e\, N_D\, \mu\, Z\, a}{L}. \tag{24.48}$$

The drain transconductance for $V_D \to 0$ (linear regime, dashed straight lines in Fig. 24.15a) is given by

$$g_{D0} = g_{max} \left[1 - \sqrt{\frac{V_{bi} + V_G}{V_P}} \right] = g_{m,sat}, \tag{24.49}$$

which is equal[6] to the forward transconductance in the saturation regime $g_{m,sat} = \partial I_{D,sat} / \partial V_G$.

[6]Technically, here $g_{D0} = -g_{m,sat}$, however, we had counted V_G positive for the reverse direction.

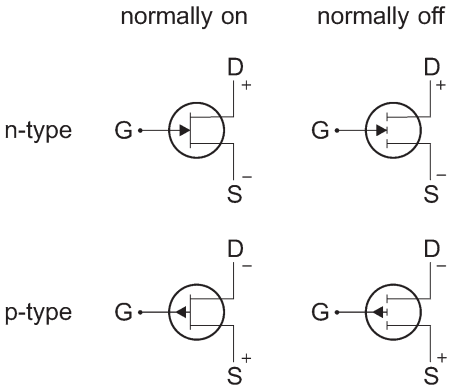


Fig. 24.17 Circuit symbols for various types of FETs

24.4.3 Normally on and Normally Off FETs

The JFET discussed so far had an n-conductive channel and was conductive at $V_G = 0$. It is termed an 'n-type, normally on' (or depletion) FET. If the channel is p-conductive, the FET is called 'p-type'. A FET that has a nonconductive channel at $V_G = 0$ is called 'normally off' (or accumulation) FET. In this case, the built-in voltage must be large enough to cause pinch-off. For a positive gate voltage (in the forward direction of the gate–drain diode), current begins to flow. The I–V characteristics of the four FET-types are depicted in Fig. 24.16. The circuit symbols for the four different FET types are shown in Fig. 24.17.

24.4.4 Field-Dependent Mobility

So far, we have considered FETs with long channels ($L \gg a$). This situation is often not the case, in particular for high-integration or high-frequency applications. For short channels, the I–V characteristics exhibit changes. The theory needs to be modified to take into account, among other effects, the electric-field dependence of the mobility (Fig. 8.10) that was discussed in Sect. 8.4.1.

Drift-Velocity Saturation

A material without negative differential mobility, such as Si or Ge, can be described with a drift-velocity model

$$v_d = \mu E \frac{1}{1 + \mu E / v_s}. \tag{24.50}$$

In this model, μ denotes the low-field (ohmic) mobility and v_s the drift-saturation velocity reached for $E \gg v_s/\mu$. The fraction in (24.50) describes the drift-velocity saturation.

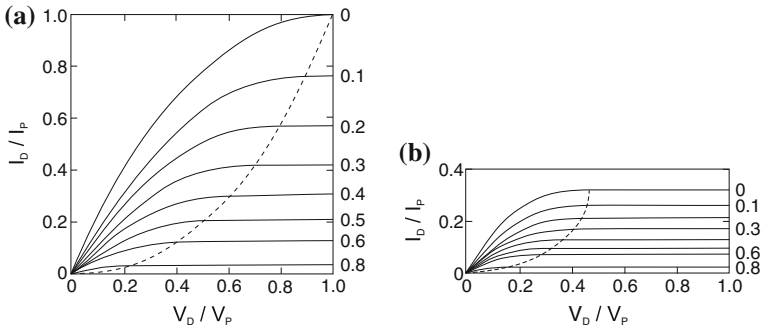


Fig. 24.18 I–V characteristic (**a**) without consideration of drift saturation ($z = 0$) and (**b**) with drift saturation ($z = 3$) for various values of $(V_G + V_{bi})/V_P = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.8$ as indicated at the right side. The intersections of the *dashed line* and the *solid lines* indicate the beginning of saturation. Adapted from [1758]

By inserting (24.50) into (24.39), we obtain (for a n-channel)

$$I_D = -e\, N_D\, \mu_n\, E(x)\, \frac{1}{1 + \mu\, E(x)/v_s}\, [a - h(x)]\, Z, \tag{24.51}$$

and after a short calculation the drain current is given by (cf. (24.42))

$$I_D = I_P \left(1 + \frac{\mu\, V_G}{v_s\, L}\right)^{-1} \left[\frac{3V_D}{V_P} - 2\, \frac{(V_{bi} + V_G + V_D)^{3/2} - (V_{bi} + V_G)^{3/2}}{V_P^{3/2}}\right]. \tag{24.52}$$

The factor $1/(1 + z)$ with $z = \mu V_G / v_s L$ reduces the channel current due to the drift saturation effect. The effect of the parameter z is depicted in Fig. 24.18 in comparison to $z = 0$, i.e. without the drift saturation effect (or $v_s \to \infty$). The forward conductance $g_{m,sat}$ decreases with increasing z, as shown in Fig. 24.19.

Two-Region Model

In order to model the GaAs drift velocity versus field characteristic, a two-region model is used. In the front region of the channel (region I), the field is small enough and a constant mobility μ is used. In the back region of the channel (region II) is the high-field region where a constant drift velocity v_s is used. With increasing source–drain voltage, the region II (I) increases (decreases) in length. The relative width of region II is also increased with decreasing channel length.

Saturated-Drift Model

Here, the drift velocity is taken everywhere as v_s, i.e. complete drift saturation. This is a good approximation for short channels (high fields) that are in current saturation.

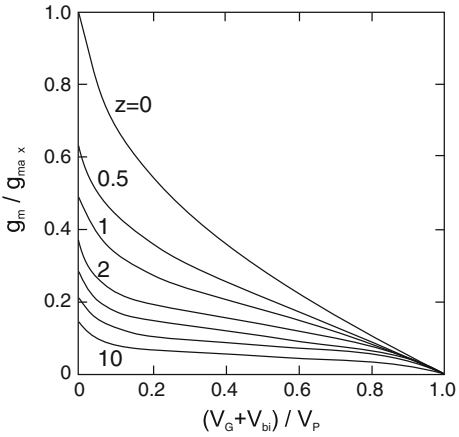


Fig. 24.19 Decrease of (saturated) forward conductance with gate voltage (according to (24.49)) and parametric dependence on z for $z = 0$, 0.5, 1, 2, 3, 5 and 10. Adapted from [1758]

In this case, the current is given by

$$I_D = -e\, N_D\, v_s\, [a - h(x)]\, Z. \tag{24.53}$$

Equation (24.53) is valid for homogeneous doping. For other doping profiles, the current is given by

$$I_D = v_s\, Z \int_h^a \rho(y)\, dy. \tag{24.54}$$

The forward conductance is given by

$$g_m = \frac{v_s\, Z\, \epsilon_s}{h(V_G)}. \tag{24.55}$$

The transistor is more linear if the depletion-layer depth only weakly depends on the gate voltage. This can be accomplished with a doping profile with increasing doping with depth. An increase with a power law and a step-wise or exponential increase lead to a more linear $I(V)$-dependence. In the limit of δ-like doping, a linear $I_{D,sat}$ versus V_G relation develops. Indeed, FETs with graded or stepped doping profiles exhibit improved linearity and are used for analog circuits.

Nonequilibrium Velocity

Below the electric field for which the drift velocity in GaAs peaks, the carriers can be considered to be in equilibrium. If the field is higher, velocity overshoot (Sect. 24.4.4 and Fig. 8.13) occurs. The carriers have a higher velocity (and ballistic transport) before they relax to the lower equilibrium (or steady-state) velocity after intervalley scattering. This effect will shorten the transit time in short-channel FETs.

24.4.5 High-Frequency Properties

Two factors limit the high-frequency performance of a FET: The transit time and the RC time constant. The transit time t_r is the time that the carrier needs to go from source to drain. For the case of constant mobility (long channel) and constant drift velocity (short channel), the transit time is given by (24.56a and b), respectively.

$$t_r = \frac{L}{\mu E} \approx \frac{L^2}{\mu V_G} \tag{24.56a}$$

$$t_r = \frac{L}{v_s}. \tag{24.56b}$$

For a 1-μm long gate in a GaAs FET, the transit time is of the order of 10 ps. This time is typically small compared to the RC time constant due to the capacitance C_{GS} and transconductance. The cutoff frequency is given by

$$f_T = \frac{g_m}{2\pi C_{GS}}. \tag{24.57}$$

24.5 MOSFETs

The MOSFET has four terminals. In Fig. 24.14b, two n-type regions (source and drain) are within a p-type substrate. The n-type channel (length L) forms underneath a MIS diode. A forth electrode sets the substrate bias. The source electrode is considered to be at zero potential. The important parameters are the substrate doping N_A, the insulator thickness d and the depth r_j of the n-type regions. Around the MOSFET structure is an oxide to insulate the transistor from neighboring devices.

24.5.1 Operation Principle

When there is no applied gate voltage, only the saturation current of the pn-diode(s) between source and drain flows. In thermodynamic equilibrium (Fig. 24.20c), the necessary surface potential for inversion at the MIS diode is $\Psi_s^{inv} \approx 2\Psi_B$. If there is a finite drain voltage, a current flows and there is no longer equilibrium. In this case, the quasi-Fermi level of the electrons (or generally of the minority carriers) is lowered and a higher gate voltage is needed to create inversion. The situation at the drain is depicted in Fig. 24.21.

In nonequilibrium, the depletion layer width is a function of the drain voltage V_D. In order to reach strong inversion at the drain, the surface potential must be at least $\Psi_s^{inv} \approx V_D + 2\Psi_B$.

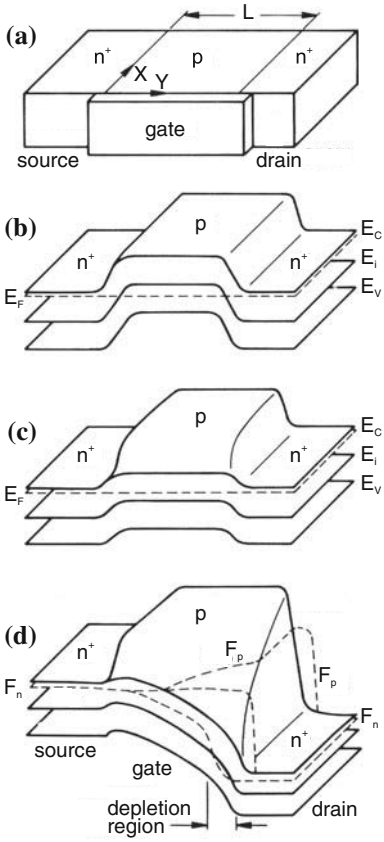


Fig. 24.20 (**a**) Schematic geometry of a MOSFET and its band diagram for (**b**) flat-band conditions for zero gate voltage (and $V_D = 0$), (**c**) thermodynamic equilibrium with reverse gate voltage (weak inversion, still $V_D = 0$) and (**d**) nonequilibrium with nonzero drain voltage and gate voltages (with most of the channel being inverted, the depletion region is indicated). Adapted from [1759]

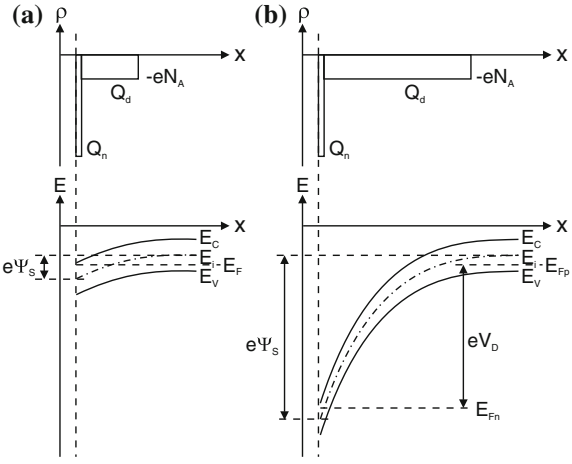


Fig. 24.21 Charge-carrier distribution (*top*) and band diagram (*bottom*) at the inverted p-region of a MOSFET for (**a**) thermodynamic equilibrium ($V_D = 0$) and (**b**) nonequilibrium at drain

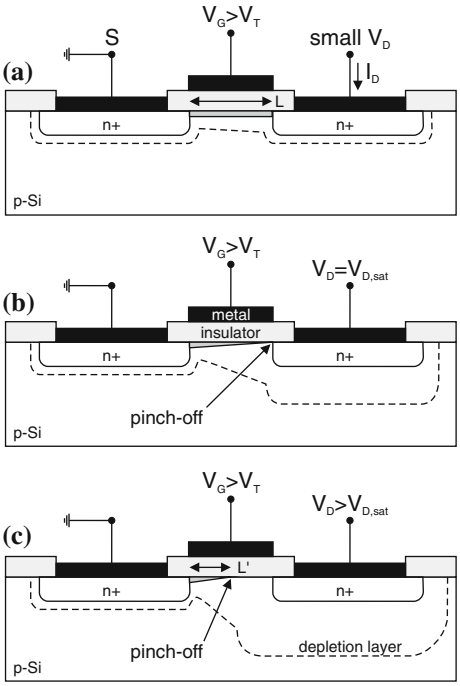


Fig. 24.22 (**a**) MOSFET with inverted channel (*dark grey*) of length L for small source–drain voltage V_D in linear regime, (**b**) at the start of saturation at pinch-off, (**c**) in the saturation regime with reduced channel length L'. The pinch-off point is denoted by an arrow in (**b**) and (**c**). The *dashed lines* denote the extension of the depletion region. Adapted from [500]

If the gate voltage is such that an inversion channel is present from source to drain, a current will flow for a small drain voltage (Fig. 24.22a). Initially, the current will increase linearly with V_D, depending on the conductivity of the channel. With increasing drain voltage, the quasi-Fermi level of the electrons is lowered until, finally at $V_D = V_{D,sat}$, the inversion channel depth becomes zero (pinch-off at the point denoted with an arrow in Fig. 24.22b). The current at this condition is denoted as $I_{D,sat}$. For a further increase of V_D, the pinch-off point moves closer to the source contact and the channel length (inverted region) is shortened (arrow in Fig. 24.22c). The voltage at the pinch-off point remains $V_{D,sat}$ and thus the current in the channel remains constant at $I_{D,sat}$.

24.5.2 Current–Voltage Characteristics

We assume now that the potential $V(y)$ varies along the channel from $V = 0$ at $y = 0$ to $V = V_D$ at $y = L$. In the gradual-channel approximation, the voltage drop V_i across the oxide is

$$V_i(y) = V_G - \Psi_s(y), \tag{24.58}$$

where Ψ_s is the surface potential in the semiconductor (see Fig. 21.33). The total charge induced in the semiconductor (per unit area) is, using (21.88), given by

$$Q_s(y) = -\left[V_G - \Psi_s(y)\right] C_i, \tag{24.59}$$

with C_i being the insulator capacitance (per unit area), as given in (21.89).

The inversion surface potential can be approximated by $\Psi_s(y) \approx 2\Psi_B + V(y)$ (see Fig. 24.21). With (21.93) the depletion-layer charge is

$$Q_d(y) = -e\, N_A\, w_m = -\sqrt{2\,\epsilon_s\, e\, N_A\, [2\Psi_B + V(y)]}, \tag{24.60}$$

such that, using (24.59), the inversion layer charge is

$$\begin{aligned} Q_n(y) &= Q_s(y) - Q_d(y) \\ &= -\left[V_G - V(y) - 2\Psi_B\right] C_i + \sqrt{2\,\epsilon_s\, e\, N_A\, [2\Psi_B + V(y)]}. \end{aligned} \tag{24.61}$$

For the calculation of the drain current, we consider the increase of channel resistance $\mathrm{d}R(y)$ along a line element $\mathrm{d}y$ of the channel. The integral of the conductivity over the cross section A of the channel (width Z) is

$$\iint_A \sigma(x, z)\mathrm{d}x\, \mathrm{d}z = -e\, \mu_n \iint_A n(x, z)\, \mathrm{d}x\, \mathrm{d}z = Z\, \mu_n\, |Q_n(y)|. \tag{24.62}$$

Therefore,

$$\mathrm{d}R(y) = \mathrm{d}y\, \frac{1}{Z\, \mu_n\, |Q_n(y)|}. \tag{24.63}$$

Here we have assumed that the mobility is constant along the channel, i.e. not field dependent. The change of voltage across the line element $\mathrm{d}x$ is

$$\mathrm{d}V(y) = I_D\, \mathrm{d}R = \frac{I_D\, \mathrm{d}y}{Z\, \mu_n\, |Q_n(y)|}. \tag{24.64}$$

We note the drain current is independent of x. Using (24.61) and performing the integral of (24.64) from $V(y = 0) = 0$ to $V(y = L) = V_D$, we find

$$\begin{aligned} I_D = \mu_n\, C_i\, \frac{Z}{L} \Bigg\{ \left(V_G - 2\Psi_B - \frac{V_D}{2}\right) - \frac{2}{3}\, \frac{\sqrt{2e\epsilon_s N_A}}{C_i} \left[(V_D + 2\Psi_B)^{3/2}\right. \\ \left. - (2\Psi_B)^{3/2}\right] \Bigg\}. \end{aligned} \tag{24.65}$$

This characteristic is depicted in Fig. 24.23a. In the linear regime (small drain voltage, $V_D \ll (V_G - V_T)$), the drain current is given by

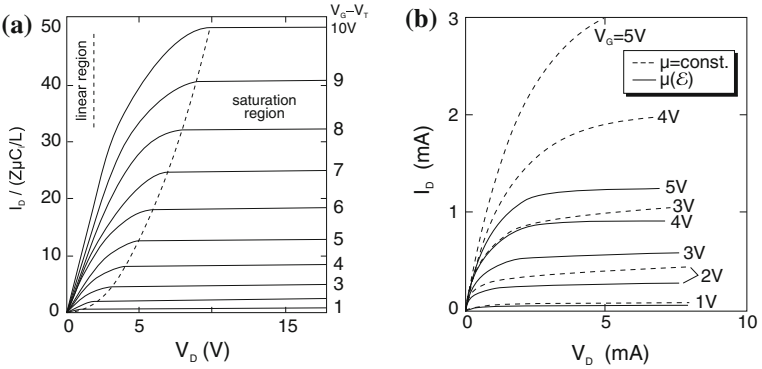


Fig. 24.23 (**a**) Idealized I–V characteristics for a MOSFET with constant mobility. The *dashed line* visualizes the drain (saturation) voltage for which the current is equal to $I_{D,sat}$. The *solid lines* are for various values of the gate voltage $V_G - V_T = 1$–10 V. Adapted from [500] (**b**) I–V characteristics taking into account the effect of field-dependent mobility (*solid lines*) in comparison to the constant-mobility model (*dashed lines*) for various gate voltages as labeled. Adapted from [1760]

$$I_D \cong \mu_n C_i \frac{Z}{L} (V_G - V_T) V_D. \tag{24.66}$$

The threshold voltage V_T, i.e. the gate voltage for which the channel is opened and a current can flow, is given for small drain voltage (linear regime) by

$$V_T = 2\Psi_B + \frac{\sqrt{4 e \epsilon_s N_A \Psi_B}}{C_i}. \tag{24.67}$$

The transconductances in the linear regime are easily obtained as

$$g_m = \mu_n C_i \frac{Z}{L} V_D \tag{24.68a}$$

$$g_D = \mu_n C_i \frac{Z}{L} (V_G - V_T). \tag{24.68b}$$

The saturation current (for constant mobility) is approximately

$$I_{D,sat} \cong \mu_n C_i \frac{m Z}{L} (V_G - V_T)^2, \tag{24.69}$$

where m depends on the doping concentration and is about 0.5 for low doping. For low p-doping of the substrate, the threshold voltage in (24.69) for the saturation regime is also given by (24.67). At higher doping, the threshold voltage becomes dependent on the gate voltage. C_i denotes the insulator capacitance

$$C_i = \epsilon_i / d_i. \tag{24.70}$$

The forward transconductance in the saturation regime is

$$g_{\mathrm{m,sat}} = \mu_{\mathrm{n}}\, C_{\mathrm{i}}\, \frac{2\,m\,Z}{L}\, (V_{\mathrm{G}} - V_{\mathrm{T}})\,. \tag{24.71}$$

For constant drift velocity (Fig. 24.23b for field-dependent mobility), the saturation current is given by

$$I_{\mathrm{D,sat}} = Z\, C_{\mathrm{i}}\, v_{\mathrm{s}}\, (V_{\mathrm{G}} - V_{\mathrm{T}})\,, \tag{24.72}$$

and the forward transconductance in the saturation regime is

$$g_{\mathrm{m,sat}} = Z\, C_{\mathrm{i}}\, v_{\mathrm{s}}. \tag{24.73}$$

We note that the transistor properties depend on and can be separated into the geometry factor (Z/L) and the material properties $(\mu\, C_{\mathrm{i}} = \mu\, \epsilon_{\mathrm{i}}/d_{\mathrm{i}})$.

The threshold voltage can be changed by the substrate bias V_{BS} as $(\beta = e/kT)$

$$\Delta V_{\mathrm{T}} = \frac{a}{\sqrt{\beta}} \left(\sqrt{2\Psi_{\mathrm{B}} + V_{\mathrm{BS}}} - \sqrt{2\Psi_{\mathrm{B}}} \right), \tag{24.74}$$

with (L_{D} being the Debye length (cf. 21.81b))

$$a = 2\, \frac{\epsilon_{\mathrm{s}}}{\epsilon_{\mathrm{i}}}\, \frac{d}{L_{\mathrm{D}}}. \tag{24.75}$$

Experimental data are shown in Fig. 24.24. For a Si/SiO_2 gate diode, $a = 1$ for, e.g. $d_{\mathrm{i}} = 10\,\mathrm{nm}$ and $N_{\mathrm{A}} = 10^{16}\,\mathrm{cm}^{-3}$. For gate voltages below V_{T}, the current is given by the diffusion current, similar to a npn transistor. This regime is important for low-voltage, low-power conditions. The related drain current is termed the *subthreshold* current and is given by

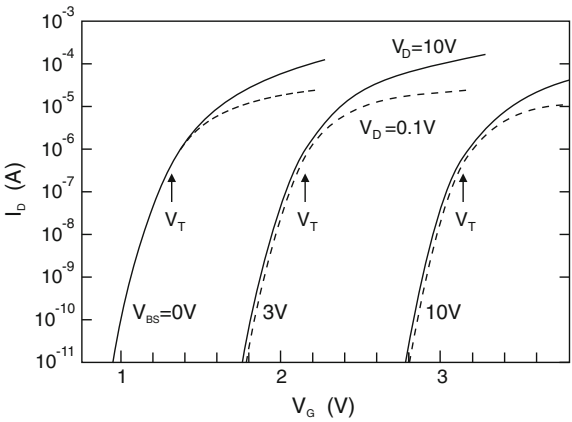


Fig. 24.24 Experimental subthreshold I–V characteristic of a MOSFET device with long channel (15.5 μm). *Solid lines* for $V_{\mathrm{D}} = 10\,\mathrm{V}$, *dashed lines* for $V_{\mathrm{D}} = 0.1\,\mathrm{V}$. Adapted from [1761]

$$I_D = \mu_n \frac{Z\, a\, C_i\, n_i^2}{2\, L\, \beta^2\, N_A^2} \left[1 - \exp\left(-\beta V_D\right)\right] \exp\left(-\beta \Psi_s\right)\, (\beta \Psi_s)^{-1/2}\,. \tag{24.76}$$

The drain current therefore increases exponentially with V_G, as shown in Fig. 24.24. V_G is ly proportional to Ψ_B:

$$\Psi_s = (V_G - V_{FB}) - \frac{a^2}{2\beta}\left(\sqrt{1 + \frac{4}{a^2}\,(\beta V_G - \beta V_{FB} - 1)} - 1\right), \tag{24.77}$$

where V_{FB} is the flat-band voltage of the gate MIS diode. The drain current is independent of V_D for $V_D \gtrsim 3kT/e$.

24.5.3 MOSFET Types

MOSFETs can have an n-type channel (on a p-substrate) or a p-channel (on an n-type substrate). So far, we have discussed the normally off MOSFET. If there is a conductive channel even without a gate voltage, the MOSFET is normally on. Here, a negative gate voltage must be applied to close the channel. Therefore, similar to the JFET, a total of four different types of MOSFET exist, see Fig. 24.25.

24.5.4 Complementary MOS

Complementary metal–oxide–semiconductor technology (CMOS) is the dominating technology for highly integrated circuits. In such devices, MOSFETs with n-channel (NMOS) and p-channel (PMOS) are used on the same chip. The basic structure of logic circuits, the inverter, can be realized with a pair of NMOS and PMOS transistors, as shown in Fig. 24.26a with two normally off transistors. The load capacitor represents the capacitance of the following elements.

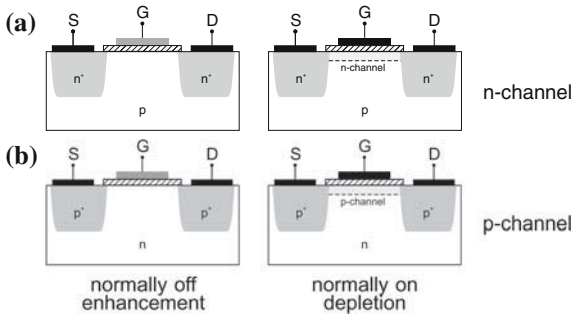


Fig. 24.25 The four MOSFET types. (**a**) Enhancement and (**b**) depletion type with n-channel (*top row*) and p-channel (*bottom row*)

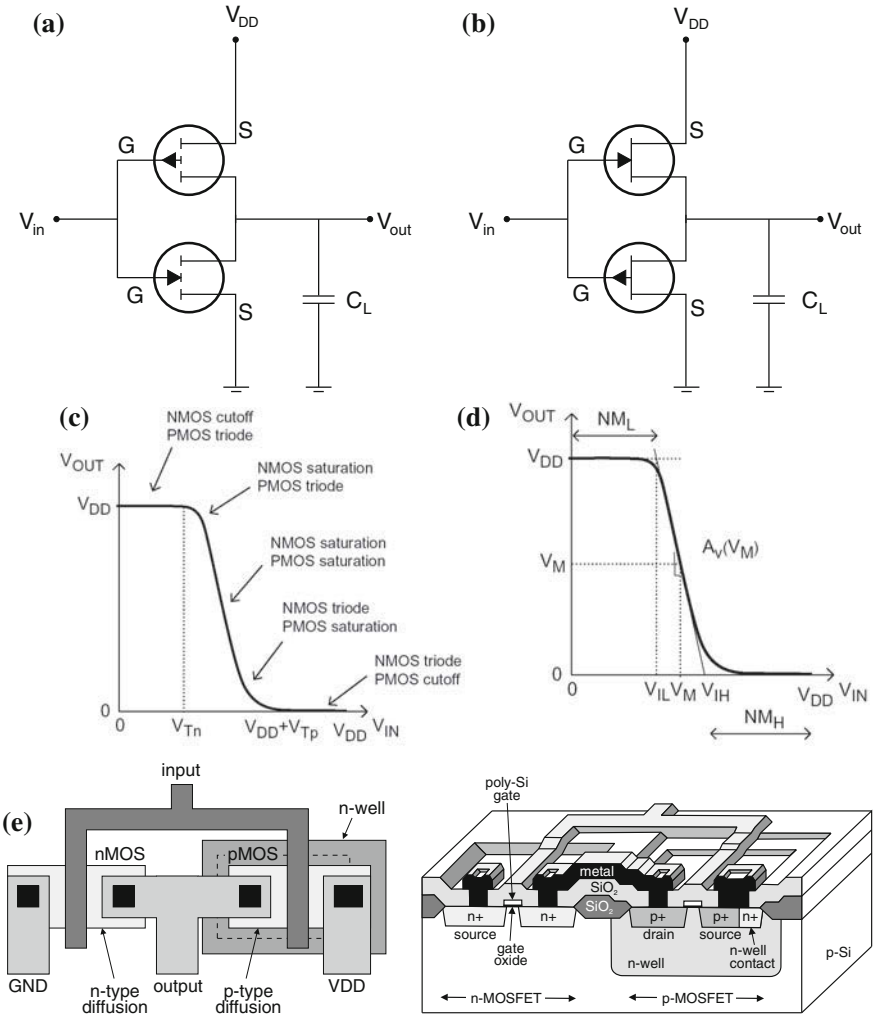


Fig. 24.26 Circuit diagram of (**a**) inverter with n-type (bottom) and p-type (normally off, enhancement mode) FETs and (**b**) inverter with p-type (bottom) and n-type (normally on, depletion mode) FETs. (**c**) Inverter characteristic with the transistor thresholds indicated, (**d**) inverter characteristic with middle voltage V_M indicated. $NM_{L,H}$ denotes the low- and high-noise margins, respectively, i.e. the voltage by which the input voltage can fluctuate without leading to switching. (**e**) Composite layout (*left panel*) and cross-sectional view (*right panel*) of CMOS inverter. Part (**e**) adapted from [1762]

If the input voltage is $V_{in} = 0$, the NMOS transistor is nonconductive ('off'). The (positive) voltage V_{DD} is at the PMOS transistor source, thus the gate is negative in relation to the source and the transistor is conductive ('on') since $-V_{DD} = V_{Gp} < V_{Tp} < 0$ (see Fig. 24.25). The current flows through the capacitor that becomes charged to $V_{out} = V_{DD}$. The current then subsides, since V_D at the PMOS becomes

zero. If the input voltage is set to V_{DD}, the NMOS transistor has a positive gate–source voltage larger than the threshold $V_{Tn} < V_{Gn} = V_{DD}$ and becomes conductive. The charge from the capacitor flows over the NMOS to ground. The PMOS transistor has zero gate–source voltage and is in the 'off' state. In this case, the voltage V_{DD} drops entirely across the PMOS and the capacitor is uncharged with $V_{out} = 0$.

In both its logic states, the CMOS inverter does not consume power. No current[7] flows in either of the two steady states since one of the two transistors is in both cases in the 'off' state. Current flows only during the switching operation. Therefore, the CMOS scheme allows for low power consumption.

The middle voltage for which $V_{in} = V_{out}$ can be calculated from the MOSFET characteristics. Both are, for this condition, in saturation and the currents are given by (cf. (24.69))

$$I_{Dn} = \mu_n C_{ox} \frac{Z_n}{2L_n} (V_M - V_{Tn})^2 \tag{24.78a}$$

$$I_{Dp} = \mu_p C_{ox} \frac{Z_p}{2L_p} \left(V_{DD} - V_M - V_{Tp}\right)^2 . \tag{24.78b}$$

With $\gamma = \frac{Z_p}{Z_n} \frac{L_n}{L_p} \frac{\mu_p}{(-\mu_n)}$, we find from $I_{Dn} = -I_{Dp}$,

$$V_M = \frac{V_{Tn} + \gamma \left(V_{DD} + V_{Tp}\right)}{1 + \gamma} . \tag{24.79}$$

As gate material, often polycrystalline silicon (poly-Si) is used (cf. Fig. 21.28). It is used instead of metals because its work function matches that of silicon closely. Also, poly-Si is more resistant to temperature. Despite its high doping, the resistance of poly-Si is two orders of magnitude larger than that of metals. Since it is easily oxidized, it cannot be used with high-k oxide dielectrics.[8]

For optimized ohmic contacts on the n- and p-Si, different metals are used to create a small barrier height (Fig. 21.23a) and low contact resistance (cf. Sect. 21.2.6). Figure 24.27 visualizes the band edges of silicon in relation to the work functions of various metals (see Table 21.1). For example, the work function of titanium matches the electron affinity of n-Si closely. However, a direct deposition of Ti on Si results in a Schottky barrier of 0.5 eV [1415]. A surface passivation with a group-VI element such as Se can help reduce this value to 0.19 eV [1763].

In the latest generation of CMOS ICs the PMOS (NMOS) device has a built-in compressive (tensile) channel strain for modifying the effective mass (cf. Sect. 6.10.2), both allowing higher drive current due to higher mobility. A detailed treatment can be found in [1764].

---

[7] Except for the subthreshold current and other leakage currents. These need to be reduced further since the dissipated power limits chip performance (speed and device density) and battery lifetime in handheld applications.

[8] The term 'high-k dielectric' means a dielectric material with large dielectric constant ϵ.

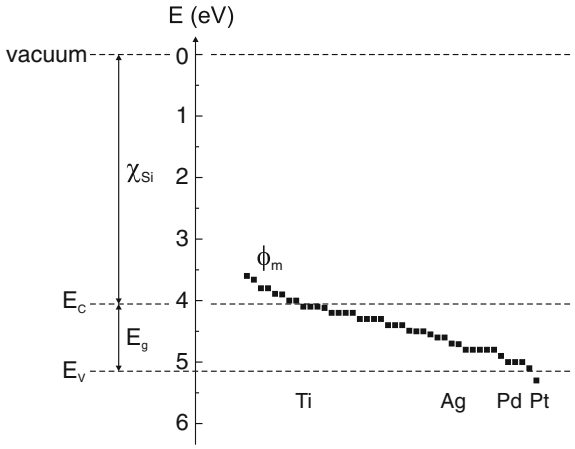


Fig. 24.27 Silicon band edges in relation to different metals and their work functions

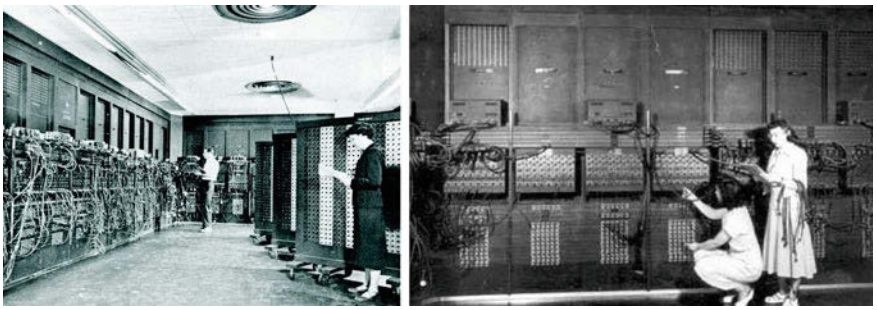


Fig. 24.28 ENIAC, the first electronic computer (J.P. Eckert, J.W. Mauchly, 1944/5). The images show only a small part of the 18 000 vacuum tubes

24.5.5 Large-Scale Integration

Historic Development

Compared to the first computers on the basis of vacuum tubes (triodes), e.g. ENIAC (Fig. 24.28), today's devices are extremely miniaturized and need many orders of magnitude less power per operation. ENIAC needed 174 kW of power. A comparable computing power was reached in 1971 with the few cm^2 large Intel 4004 microprocessor (Fig. 24.29b) consuming only several Watts with 2300 transistors. In 2004 about 42 million transistors were integrated in the Pentium 4 microprocessor (Fig. 24.30). Also, memory chips started to become highly integrated (Fig. 24.29a).

The development of electronic circuit integration is empirically described by Moore's 'law' [1765] that has been valid since the 1970s. According to this law,

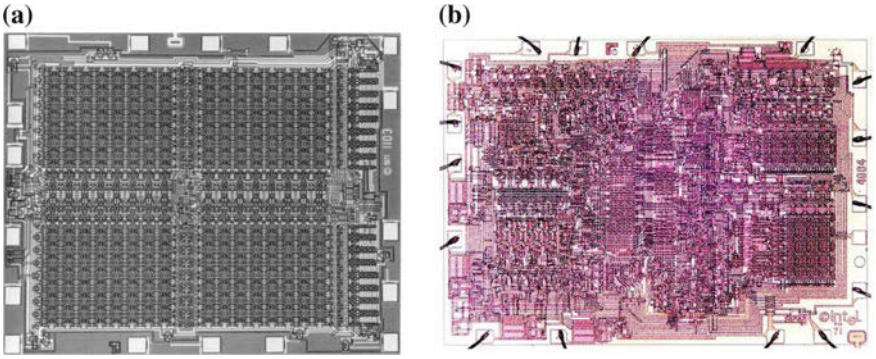


Fig. 24.29 (**a**) Intel^{TM} 1103 1 KByte (1024 memory cells) dynamic random access memory (RAM), arranged in four grids with 32 rows and columns (1970), chip size: $2.9 \times 3.5\,\text{mm}^2$. (**b**) Intel^{TM} 4004 microprocessor (1971), chip size: $2.8 \times 3.8\,\text{mm}^2$, circuit lines: $10\,\mu\text{m}$, 2,300 MOS transistors, clock speed: 108 kHz

the number of transistors doubles every 20 months (Fig. 24.31a). At the same time, the performance has been improved by an increase of the clock speed (Fig. 24.31b).[9]

Interconnects

Moore's second law says that the cost of production also doubles for each new chip generation and is currently (2004) in the multi-billion US$ range. Most of the cost saved by integration is due to efficient *wiring* (interconnects) of the components, in 2004 (65 nm node) in eight layers above the active elements (transistors and capacitors) (Fig. 24.33), in 2008 (45 nm node) in eleven layers. Plane-view images of the first three layers of the interconnects are shown in Fig. 24.34. The Cu interconnects are fabricated with the so-called damascene process [1766–1768]. Barrier layers (e.g. TaN or TiN) are required to avoid out-diffusion of Cu into the silicon or other parts of the circuit. Three effects limit the conductivity: The interconnect metal line width and height approaches the mean free path of carriers ($d_{\text{Cu}} \approx 40\,\text{nm}$) [633, 1769], grain boundary scattering can limit mobility since grain size is reduced for thinner lines, and the (high resistivity) barrier reduces space for the conductive part of the metal line. In Fig. 24.32 the increase of the resistivity of copper with reduced dimension is shown as a function of film thickness t and for a 100 nm-film as a function of line width w. In a simplified approach, the line resistivity ρ_{line} is given as [1769]

$$\frac{\rho_{\text{line}}}{\rho_0} = 1 + \frac{3}{8}(1-p)\left(\frac{d}{t} + \frac{d}{w}\right), \tag{24.80}$$

ρ_0 denoting the bulk resistivity ($1.7\,\mu\Omega\,\text{cm}$ for Cu), d being the mean free path (8.7) and p being the electron scattering parameter ($p = 0$ for diffuse scattering).

[9] After year 2003 data for maximum clock rate are not for highest integration density processors.

Fig. 24.30 The Intel^{TM} Pentium 4 microprocessor (2000), circuit lines: 0.18 µm, 42 million transistors, clock speed: 1.5 GHz

In order to achieve the best high frequency performance the material between the metal interconnects should have low dielectric constant ('low-k' dielectric). Alternative materials to the standard SiO_2 ($\epsilon_r \approx 4.1$) are investigated such as SiOF (≈3.8), SiCOH (≈3.0), porous materials (≈2.5) and air gaps [1770].

CMOS Scaling

Using planar technologies, LSI (large-scale integration), VLSI (very large-scale integration), ULSI (ultra large-scale integration) and further generations of devices have

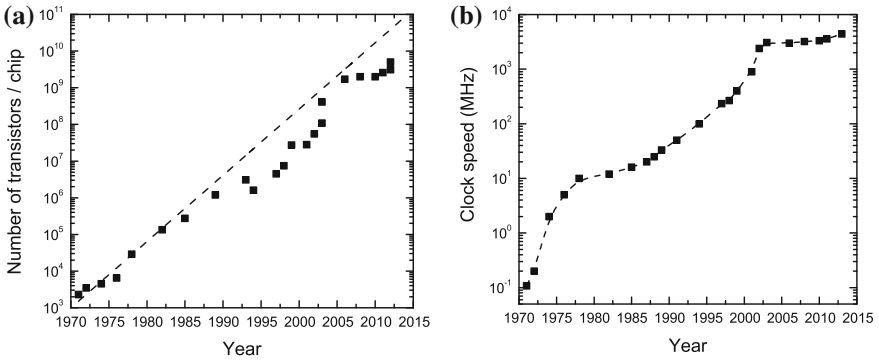


Fig. 24.31 (**a**) Moore's law on the exponential increase of transistors per chip (for Intel^{TM} processor chips). *Dashed line* corresponds to doubling in 20 months. (**b**) Historical increase of maximum clock speed, *dashed line* is guide to the eye. Note the almost constant rate of 10 MHz from the mid-1970s to the mid-1980s and another plateau developing after 2000

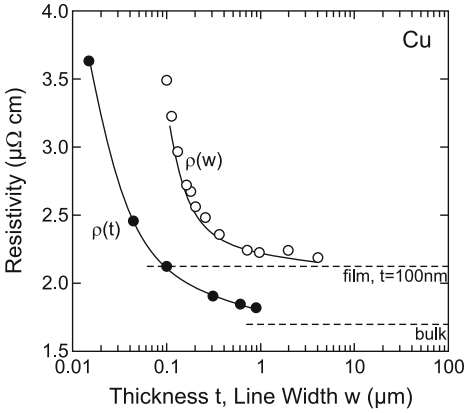


Fig. 24.32 Resistivity of copper at room temperature for various film thickness (*solid circles*), and for a 100 nm-film as a function of line width w (*empty circles*). *Solid lines* are theoretical dependence according to (24.80). The *dashed lines* indicate the limits for bulk material ($t \to \infty$) and for large line thickness ($d = 100\,\text{nm}$, $w \to \infty$). Adapted from [1769]

been conceived, driven by high-density electronic memory devices. Subsequently also logic devices are produced with reduced device size.

The increase of the number of transistors per area requires the scaling of their geometrical properties. This impacts many other properties of the transistor and their scaling needs to be considered as well. From a general perspective, the physical properties scale while the thermal energy kT remains constant for room-temperature electronics.

If channel width Z and channel length L of a transistor are scaled down by a factor of $s > 1$, $Z' = Z/s$ and $L' = L/s$, the area obviously scales as $A' = A/s^2$. In subsequent transistor generations $s = \sqrt{2}$, i.e. doubling of the number of devices per area. In order to maintain the aspect ratio of the device also the oxide thickness (d_i) is scaled, $t'_{ox} = t_{ox}/s$ ('classical scaling').

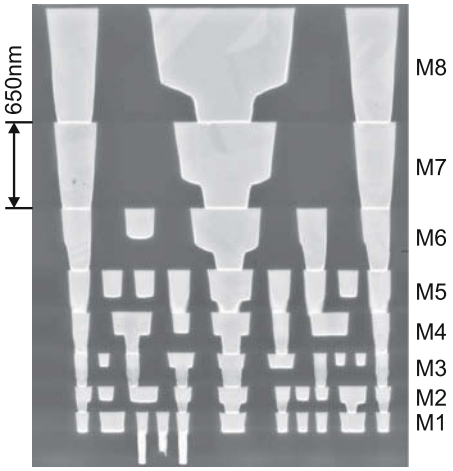


Fig. 24.33 Cross section through a logic chip (65 nm technology, 35 nm gate length) with eight layers of dual damascene Cu interconnects (M1–M8) with low-k carbon-doped oxide ($\epsilon_r = 2.9$) inter-level dielectric above the active elements. Adapted from [1771]

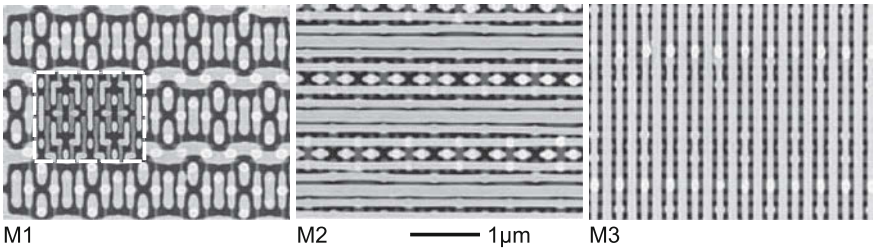


Fig. 24.34 Plane-view of the first three interconnect layers of a 45 nm node SRAM array (Intel® Xeon®). In the image of the M1 layer the gate layer metal connects are shown in the *inset*, framed with a *white dashed line*. Adapted from [1772]

The ultimate design criteria are maximum temperature and maximum power loss. The maximum temperature needs to be obeyed, the worst case usually taken as 100 °C. The power loss per area, e.g. heating, needs to stay constant at an appropriate maximum level around 200 kW/m^2 (Fig. 24.35b) unless higher (and more expensive) efforts on cooling are made. At the same time the device performance needs to be maintained if not improved, e.g. for lower power dissipation in battery operated devices. Very important is the reduction of operation voltage V_{DD} in order to keep electric fields and power consumption small enough (Fig. 24.35). The power consumption in stand-by mode P_{off} depends on V_{DD} and the subthreshold (off) current

$$P_{off} = W_{tot}\, V_{DD}\, I_{off}, \tag{24.81}$$

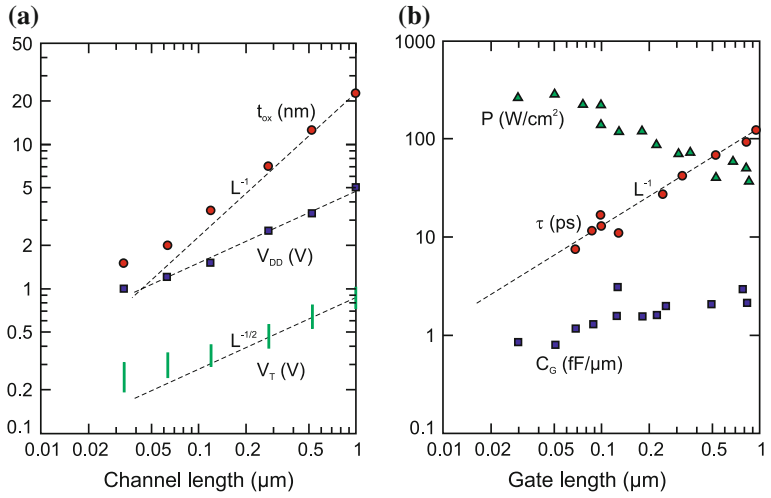


Fig. 24.35 Scaling of MOSFET parameters gate oxide thickness t_{ox}, power supply voltage V_{DD} (across source–drain), threshold voltage V_T, total power loss per area P, gate capacitance per channel width C_G and inverter delay τ, the time required to propagate a transition through a single inverter driving a second, identical inverter, commonly used as a means of gauging the speed of CMOS transistors. Data for (**a**) from [1773] and for (**b**) selected from [1774]

where W_{tot} is the total width of the turned-off devices and I_{off} is the average off-current per device per width. The latter increases exponentially with reduced threshold voltage V_T,

$$I_{off} = I_0 \exp\left(-\frac{e\,V_T}{n\,kT}\right), \tag{24.82}$$

with ideality factor $n \approx 1.2$ and $I_0 \approx 1$–$10\,\mu A/\mu m$ [1773]. A well-functioning MOSFET requires a ratio of V_T/V_{DD} of < 0.3.

The power consumption in active mode P_{ac} depends also on the clock speed (frequency f) that increases with higher integration due to shorter gate length,

$$P_{ac} = C_{sw}\,V_{DD}^2\,f, \tag{24.83}$$

where C_{sw} is the total node capacitance being charged and decharged in a clock cycle.

Historically the oxide thickness has been reduced less than the channel length [1773] (Fig. 24.35a) leading to increased local fields. The reduction of the physical gate oxide thickness is limited due to gate leakage through tunneling [1775]. While for a gate voltage of 1.5 V and oxide thickness $t_{ox} = 3.6\,nm$ the leakage current is only about $10^{-8}\,A/cm^2$, it is about $1\,A/cm^2$ for $t_{ox} = 2.0\,nm$ and about $10^4\,A/cm^2$ for $t_{ox} = 1.0\,nm$. Obviously variations of oxide thickness are more harmful at small average thickness. 1.2 nm physical SiO_2 thickness has been used in the 90 nm (gate length) logic node.

The technological solution for further reduction of oxide thickness is the use of geometrically thicker layers, to suppress tunneling, with higher dielectric constant ('high-k dielectrics'), e.g. HfO_2 [1776], to maintain reasonable gate capacitance per gate width

$$C_G = \frac{\epsilon_{ox}}{t_{ox}} L, \tag{24.84}$$

(cmp. (24.70)) at a value of about 1.0–1.5 fF/μm (Fig. 24.35b). For the 45 nm technology node a 0.7-fold reduction in electrical oxide thickness was achieved while reducing gate leakage 1000× for the PMOS and 25× for the NMOS transistors [1777].

Materials

The electronics industry is based on silicon as the material for transistors. However, many other materials are incorporated in the technology. Traditionally silicon dioxide gate oxide is used, silicon nitride for insulation layers and polysilicon for gate contacts. For wiring aluminum has been used. Silicides were introduced as contact materials around 1986.

Progress was made with copper interconnects (IBM, 1997), replacing aluminum. The better electrical and heat conductivity could previously not be used since Cu is a deep level in Si (cf. Fig. 7.6). The key to success was an improved barrier technology based on amorphous TaN- or TiN-based barrier layers to prevent the diffusion of Cu into the silicon and dielectric layers. The first chip from series production, incorporating the Cu technology, was the PowerPC 750 (400 MHz) in 1998. Since 2000 high-k, i.e. large ϵ_r, Hf-containing gate dielectrics are used (Fig. 24.36). HfO_2 has a dielectric constant of 25–30. 45 nm node technology probably uses HfZrO, HfSiO or HfSiON [1778] gate dielectrics with $k \sim 12$ and an electrical thickness of $t_{ox}\,\epsilon_{SiO_2}/\epsilon_{ox} = 1.0\,\text{nm}$.

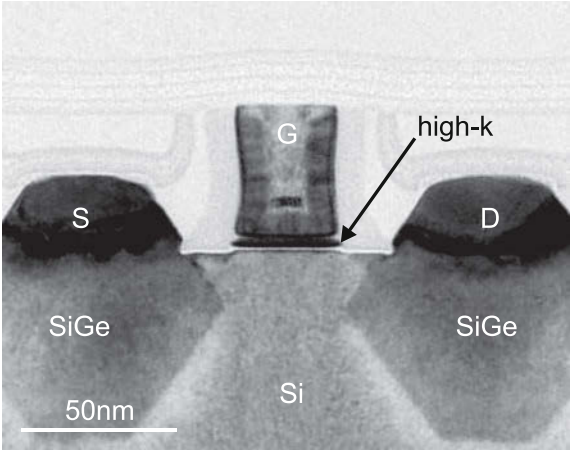


Fig. 24.36 Cross-section TEM image of 45 nm node PMOS transistor with high-k Hf-containing gate oxide (*dark*) above a thin SiO_2 layer (*white*). The role of the stressor SiGe pockets is explained in Fig. 24.35. Adapted from [1779]

Germanium is reintroduced into mainstream semiconductor technology via SiGe stressors in the source and drain for PMOS. Uniaxial compressive strain in the channel region leads to 30 % increased saturation current [1780] mostly due to reduced effective masses [658, 1781, 1782] (Sect. 8.3.11) for 90 nm transistors. Similarly, uniaxial tensile strain in NMOS, introduced by SiN caps or more recently tensile trench contacts [1777], allows for 10 % higher saturation current [1780] (Fig. 24.37). The enhanced electron mobility is due to strain-induced splitting of the X-valley and change of electron mass [1783]. Further improvement to 18 % (NMOS) and

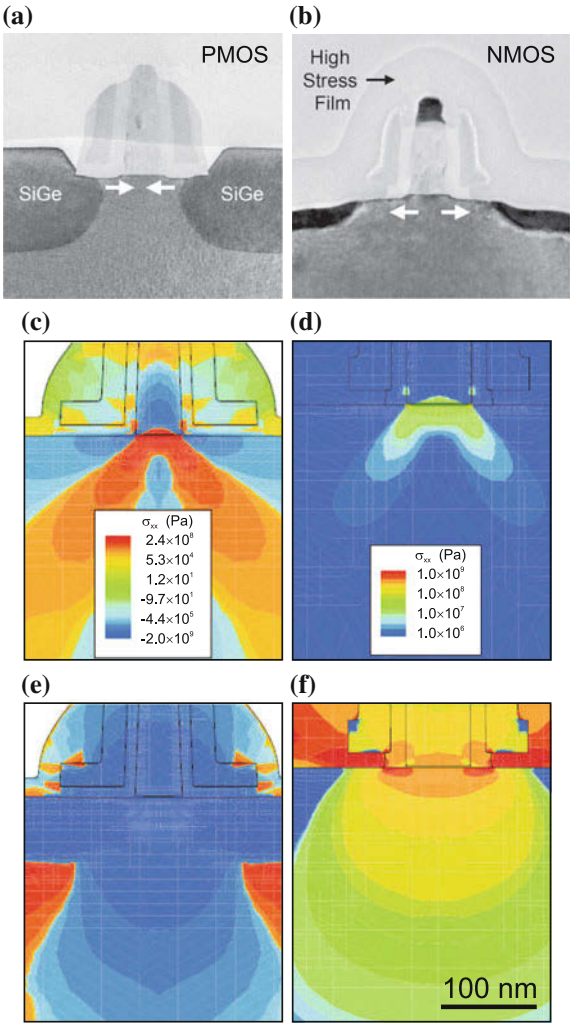


Fig. 24.37 Cross-section TEM images of strained (**a**) PMOS and (**b**) NMOS transistors. Adapted from [1784]. (**c**–**f**) Modelling of strain distribution: PMOS without (**c**) and with (**e**) $Si_{0.83}Ge_{0.17}$ pockets, NMOS without (**d**) and with (**f**) tensile cap layer. Adapted from [1785]

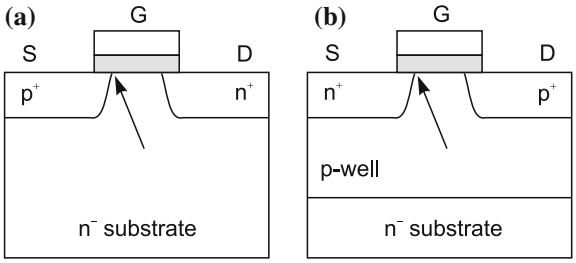


Fig. 24.38 Schematic of a (**a**) n-type and (**b**) p-type tunneling FET (TFET). D is reversely biased, i.e. positive for NTFET and negative for PTFET. The *grey areas* represent the gate oxide, the *arrows* denote the spatial position of tunneling (surface tunneling junction) for sufficient (NTFET: positive, PTFET: negative) gate voltage

50 % (PMOS) increase in $I_{D,sat}$ compared to unstrained Si have been made in 65 nm transistors [1771].

The end of the miniaturization has been theoretically predicted many times and for various feature sizes. Today, only fundamental limits such as the size of an atom seems to limit circuit design.[10] Such limits (and the effects in nanostructures in the few-nm regime) will be reached beyond 2010, projected at about 2020. Up to then, it is probable that at least a few companies will follow the road map for further miniaturization, as laid out by the Semiconductor Industry Association[11] (SIA).

24.5.6 Tunneling FETs

A decisive parameter for FET performance is low leakage current. With shrinking device dimensions it increases rapidly for conventional FET design. A novel type of FET has thus been conceptualized, the tunneling FET (TFET) [1786]. It is a lateral p-i-n diode with a MOS gate (Fig. 24.38). The leakage current is minimized due to the reverse biased p-i-n structure. A low leakage current (per gate width) of less than 10^{-14} A/μm has been realized [1787, 1788]. The channel current is due to band-to-band tunneling as in an Esaki diode (Sect. 21.5.9) and can be controlled by the gate voltage [1789]. The surface tunneling junction is close to the source electrode. The use of germanium instead of silicon allows further performance enhancements [1790].

[10] Only commercial profit, rather than testing physical limits, drives the miniaturization. Insufficient economic advantages or low yield of further chip generations possibly can limit or slow down large-scale integration.

[11] www.semichips.org.

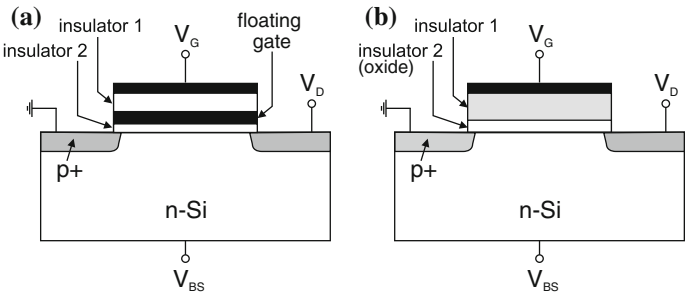


Fig. 24.39 MOSFET with (**a**) floating gate and (**b**) MIOS structure

24.5.7 Nonvolatile Memories

Floating Gate Memories

When the gate electrode of a MOSFET is modified in such a way that a (semi-)permanent charge can be stored in the gate, a nonvolatile electronic memory can be fabricated. In the floating-gate structure (Fig. 24.39a), an insulator–metal–insulator structure is used where charge is stored in the metal and cannot escape through the insulating barriers. The 'metal' is often realized by poly-Si. In the MIOS structure (Fig. 24.39b), the insulator–oxide interface is charged. The charge can be removed by UV light (EPROM, erasable programmable read-only memory) or by a sufficient voltage across the oxide at which the charge carriers can tunnel out (Fowler–Nordheim tunneling) (EEPROM, E^2PROM, electrically erasable programmable read-only memory).

Nowadays, a special type of EEPROM is used for the so-called *flash* memories. The stored gate charge causes a change in the MOSFET threshold voltage and is designed to switch between the on and off state. The storage time of the charge can be of the order of 100 years. Since tunneling limits the charge retention, the oxide must be sufficiently thick. In Fig. 24.40 a cross section of a 4 Gb, 73 nm SLC (single-level cell) flash memory is shown. The lower insulator (tunneling oxide at the channel) consist of 7.2 nm SiO_2, the upper insulator (insulator 1 in Fig. 24.39a) is a 18 nm thick oxide/nitride/oxide (ONO) stack. The floating gate has a $90 \times 90\,\text{nm}^2$ footprint, is about 86 nm high and consists of two polysilicon layers.

In a SLC memory the floating gate has two states, a certain charge value and the erased state. In a MLC (multi-level cell) the gate can store several charge states which can be sensed as different logic states, e.g. $2^2 = 4$ states. This increases the storage density, lowering cost per bit, but also increases the complexity. Typical endurance of SLC is at least 10^6 program–erase cycles. SLC cells so far have about ten times higher endurance (possible number of read–write cycles) and lower power consumption than MLC. Generally SLC flash memory is considered industrial grade and MLC flash is considered consumer grade. Recently also triple level cells (TLC),

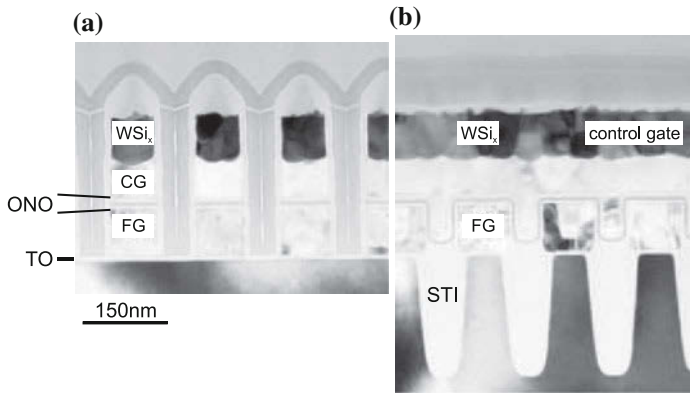


Fig. 24.40 Cross sections (**a**) perpendicular and (**b**) parallel to the control gate line of a 4 Gb, 73 nm SLC flash memory (Samsung K9F4G08U0M). 'CG' denotes the control gate, 'FG' the floating gate, 'TO' the tunneling oxide, 'ONO' the oxide/nitride/oxide insulator stack, and 'STI' the shallow trench insulation. Adapted from [1772]

storing 3 bit (8 states) are commercialized, however the increased storage density comes at high cost of reliability [1791].

The ultimate limit, explored currently, is to use a single electron charge to cause such an effect in the single-electron transistor (SET).

Future Concepts

Memory concepts beyond the storage of free charges include information storage via

- the static polarization in a ferroelectric material (either crystalline or polymer) (FeRAM [1792], Fig. 24.41a) which can be switched by an electric field.
- the phase change between amorphous and polycrystalline phases in a chalcogenide layer (typically GeSb [1793] or $Ge_2Sb_2Te_5$, GST [1794, 1795] with an $\alpha \leftrightarrow$ c transition, Fig. 24.42) upon local heating (similar to a rewritable DVD) and the related change is resistivity (PCM, phase change memory).
- the storage of magnetization direction (MRAM [1796, 1797]) and subsequent resistance change of a magneto-tunneling junction (MTJ) whose resistance depends on the relative magnetization (parallel or perpendicular) of two magnetic layers separated by a thin tunneling insulator (Fig. 24.41b). The largest TMR (tunnel-magnetoresistance) effect has been achieved with MgO as insulator [1797]. The magnetization of the bottom magnetic layer of the MTJ is fixed. The magnetization directions $\pm 45°$ are written into the free layer with the magnetic fields of two perpendicular high current wires in two subsequent back-end interconnect layers sandwiching the MTJ.
- resistance change based on solid electrolytes (PMC, programmable metallization cell memory). The lowering of the resistance is attained by the reduction of ions in a fairly high resistivity electrolyte (e.g. from the system Cu,Ag–Ge–Se,S,O [1798, 1799] or oxides [1800]) to form a conducting bridge between the electrodes. The

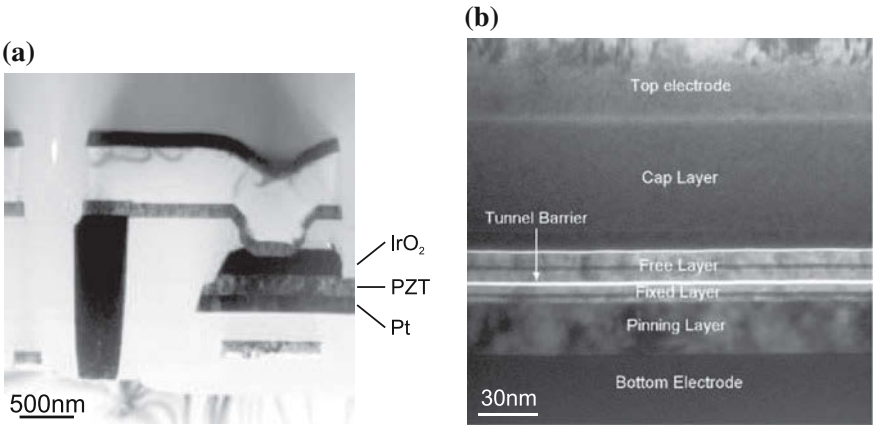


Fig. 24.41 (**a**) Cross-section TEM image of a cell from a Ramtron 4 Mb FeRAM. The information is stored in the electric polarization of a polycrystalline $Pb(Ti_xZr_{1-x})O_3$ (PZT) island, contacted on the bottom and top with platinum and iridium oxide, respectively. Adapted from [1772]. (**b**) Cross-section TEM of the magnetic tunneling junction from a Freescale 4.2 Mb MRAM, located between the M4 and M5 interconnect layers. The magnetization of the free layer can be switched, that of the fixed layer remains constant. Adapted from [1772]

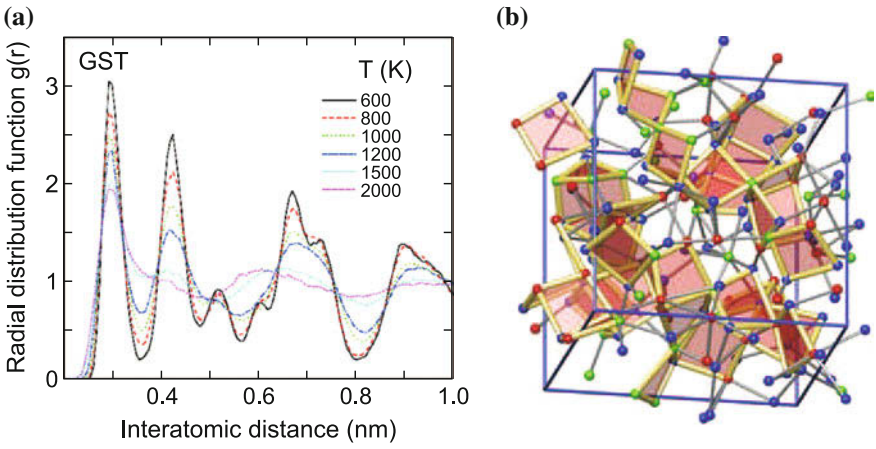


Fig. 24.42 (**a**) Radial distribution function of ions in $Ge_2Sb_2Te_5$ (GST) for various temperatures (cmp. Fig. 3.14b). Adapted from [756]. (**b**) Atom arrangement in the amorphous phase of GST with square units highlighted that nucleate crystallization. Adapted from [1795]

resistance is returned to the high value via the application of a reverse bias that results in the breaking of the conducting pathway.

- resistance change in transition metal oxides such as perovskites, e.g. $SrTiO_3$:Cr [1801, 1802] or NiO:Ti (RRAM). Electrical pulses of opposite polarity switch the resistance reversibly between a high- and a low-resistance state. Oxygen-vacancy

drift modulates the valence of the mixed-valence transition-metal ion (e.g. Ti^{3+}–Ti^{4+}) and thus the conducting state [1803].
- a molecular configuration change (e.g. redox reaction) between crossed wire lines (molecular electronics [1804–1806]).

24.5.8 Heterojunction FETs

Several types of field-effect transistors have been devised that use heterojunctions (HJFET).

HIGFET

As conducting channel, the two-dimensional electron gas at an undoped heterointerface is used. Such a transistor is called a heterojunction insulating gate FET (HIGFET). With forward or backward gate voltage, an electron or hole gas can be created (channel enhancement mode), as visualized in Fig. 24.43. Thus, a complementary logic can be realized. However, the p-channel suffers from low hole mobility.

HEMT

If the top wide-bandgap layer is n-doped, a modulation-doped FET (MODFET) is made (see Sect. 12.3.4). This structure is also called a HEMT (high electron mobility transistor) or TEGFET (two-dimensional electron gas FET) (Fig. 24.44). A thin undoped AlGaAs spacer layer is introduced between the doped AlGaAs and the

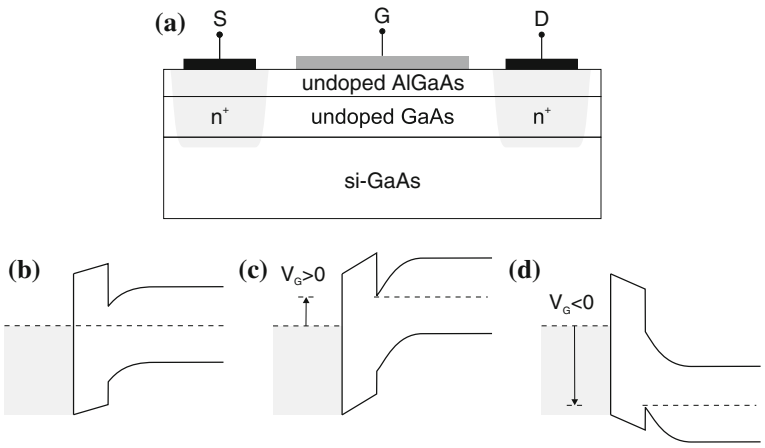


Fig. 24.43 (**a**) Scheme of a HIGFET structure with metal gate and undoped AlGaAs/GaAs heterointerface on semi-insulating GaAs. The source and drain contacts are n-doped such that this structure can be used as an n-HIGFET (see part (**c**)). (**b**) Band diagram for zero gate voltage. (**c**) Band diagram for positive gate voltage and n-channel, (**d**) for negative gate voltage and p-channel

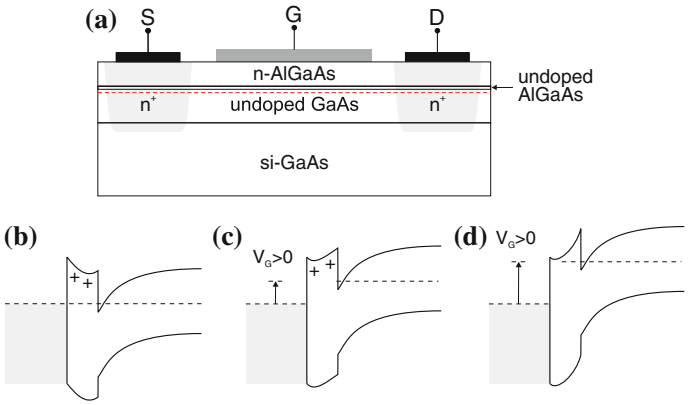


Fig. 24.44 (**a**) Scheme of a HEMT structure with n-AlGaAs/GaAs heterointerface on semi-insulating GaAs. The source and drain contacts are n-doped such that this structure can be used as an n-channel (normally-on) HEMT. The *horizontal dashed line* represents schematically the position of the 2DEG at the heterointerface on the GaAs side. (**b**) Band diagram at zero gate voltage. (**c**) Band diagram at positive gate voltage, increase of channel carrier concentration. (**d**) Band diagram at even larger positive gate voltage, formation of conducting channel in the AlGaAs layer

undoped GaAs to reduce impurity scattering from carriers that tunnel into the barrier. With increasing gate voltage, a parallel conduction channel in the AlGaAs is opened. The natural idea would be to increase the Al fraction in the AlGaAs to increase the quantum-well barrier height. Unfortunately, the barrier height is limited to 160 meV for an aluminum concentration of about 20 %. For Al content higher than about 22 %, the DX center (cf. Sect. 7.7.6) forms a deep level such that the apparent ionization energy increases drastically and no shallow donors can be used for modulation doping. An improvement for the barrier conduction problem is the use of δ-doping [1807], i.e. the introduction of a highly doped thin (mono-)layer (Fig. 24.45), which results in higher channel carrier concentration.

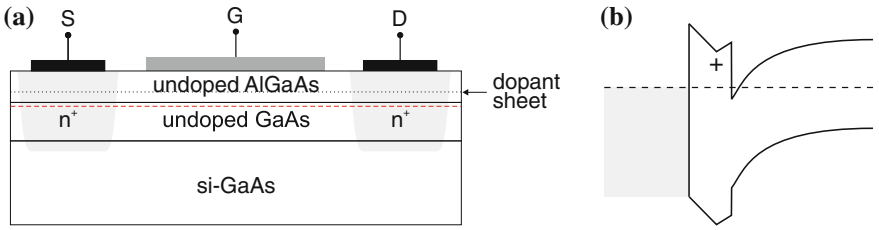


Fig. 24.45 (**a**) Scheme of a δ-doped HEMT structure with AlGaAs/GaAs heterointerface on semi-insulating GaAs. The source and drain contacts are n-doped such that this structure can be used as an n-channel HEMT. The *horizontal dashed line* represents schematically the position of the 2DEG in the GaAs layer. (**b**) Band diagram at zero gate voltage

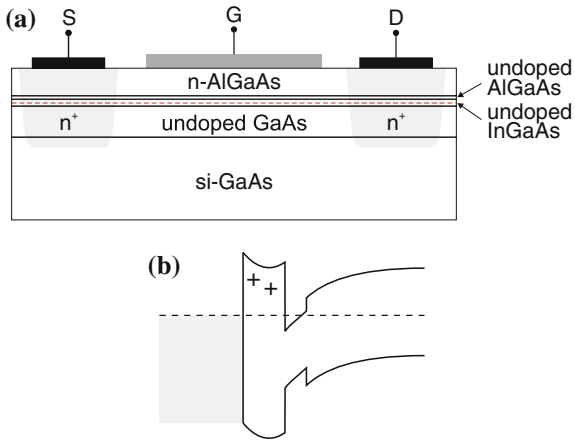


Fig. 24.46 (**a**) Scheme of a PHEMT structure with n-AlGaAs/InGaAs heterointerface on semi-insulating GaAs. The source and drain contacts are n-doped such that this structure can be used as an n-channel HEMT. The *horizontal dashed line* represents schematically the position of the 2DEG in the InGaAs layer. (**b**) Band diagram at zero gate voltage

Pseudomorphic HEMTs

Instead of increasing the height of the barrier, the depth of the well can be increased by using a low-bandgap material. On GaAs substrate, InGaAs is used (Fig. 24.46). However, strain is introduced in this case and the InGaAs layer thickness is limited by the onset of dislocation formation (cf. Sect. 5.4.1) (which reduces the channel mobility and the device reliability). For $In_{0.15}Ga_{0.85}As$ (thickness about 10–20 nm), a total barrier height of about 400 meV can be obtained. A barrier height of 500 meV can be reached with an InAlAs/InGaAs structure on InP (Fig. 24.47). The InAlAs does not suffer from the problem related to DX centers. The channel indium concentration is typically 50 %. The mobility increases with increasing indium concentration. This InP-based HEMT structure is widely used in satellite receivers for its excellent high-speed and low-noise properties in the 100–500 GHz range and beyond.

However, the InP technology is economically less favorable than GaAs due to smaller available substrate size and higher cost (2001: 4″ InP substrate: $1000, 6″ GaAs substrate: $450).

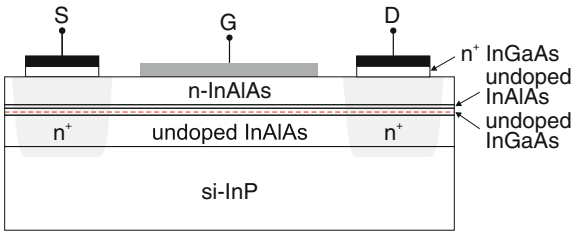


Fig. 24.47 Scheme of a PHEMT structure with n-AlInAs/InGaAs/InAlAs structure on semi-insulating InP. The source and drain contacts (with a highly doped InGaAs contact layer) are an n-doped such that this structure can be used as an n-channel HEMT. The *horizontal dashed line* represents schematically the position of the 2DEG in the InGaAs layer

Metamorphic HEMTs

A unification of the InAlAs/InGaAs structure with the best figure of merit and the GaAs substrate is achieved with the metamorphic HEMT (MHEMT). Here, a relaxed buffer is used to bring the in-plane lattice constant from that of GaAs to about that of InP. It is key that the defects occurring are confined to the relaxed buffer and do not enter the active device structure (see Fig. 24.48). The relaxed buffer is typically about 1 μm thick. It can be grown, e.g. with a graded $In_x(Ga,Al)_{1-x}As$ layer with x = 0–42 % or with a stepped structure with piecewise constant indium concentration in each layer. It is important that a smooth interface of the channel is achieved in order to avoid additional scattering mechanisms. For high-frequency operation, the fabrication of a small gate length is important, as shown in Fig. 24.49 for a 70-nm gate of a f_T = 293 GHz, f_{max} = 337 GHz transistor [1808]. SiGe channels, providing higher mobility than pure Si, can be fabricated using graded or stepped SiGe buffer layers on Si substrate. With such Si-based MHEMTs frequencies up to 100 GHz can be achieved.

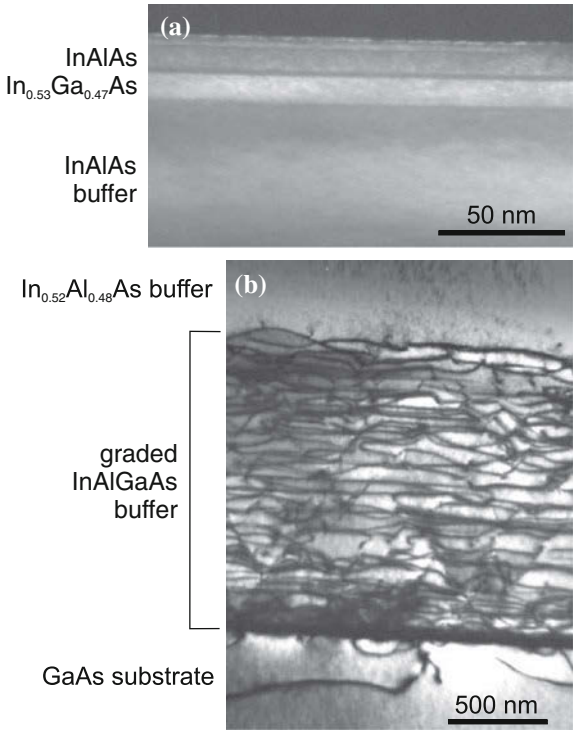


Fig. 24.48 Cross-sectional TEM image of an InAlAs/InGaAs MHEMT: (**a**) Active layer with rms surface roughness of 2.0 nm (from AFM), (**b**) graded InGaAlAs buffer layer (1.5 μm) on GaAs substrate. Adapted from [1809]

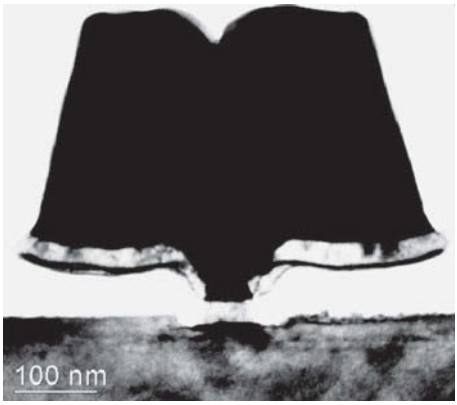


Fig. 24.49 Cross-sectional TEM image of the 70-nm gate of an InAlAs MHEMT on GaAs substrate. From [1808]

24.6 Thin-Film Transistors

Thin-film transistors (TFTs) are field-effect transistors with a channel formed as thin film on insulating substrate. A detailed treatment is available in [1810]. TFTs are typically fabricated as large-area arrays from thin layers of polycrystalline or amorphous silicon [1811] or organic semiconductors [1812–1815] on cheap substrates such as glass. Their most prominent use is driving pixels in active-matrix displays such as electroluminescence (EL) displays or twisted nematic liquid crystal displays (LCD) [1816]. Various gates and gate geometries have been reported as depicted in Fig. 24.50.

24.6.1 Annealing of Amorphous Silicon

Since the mobility in polycrystalline silicon is much higher (up to several hundred cm^2/Vs depending on grain size, see Sect. 8.3.8) than in amorphous silicon ($<1\,cm^2/Vs$), such material is much more desirable as channel in TFTs. However, it

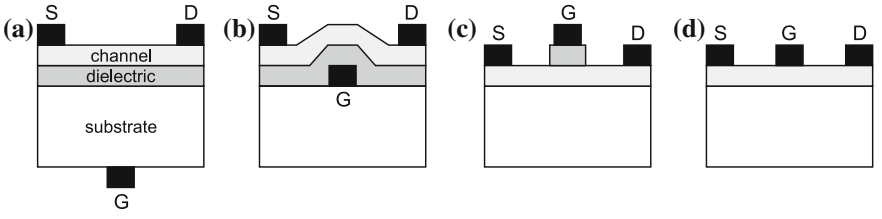


Fig. 24.50 Schematic geometries of TFTs: (**a**–**c**) MISFETs, (**d**) MESFET with (**a**, **b**) bottom gate and (**c**, **d**) top gate. Semiconductor channel layer (*light grey*), insulating dielectric (*dark grey*) and metals (*black*)

requires high deposition temperatures. In order to achieve polycrystalline silicon with large grain size from amorphous silicon films that can be deposited at low temperature (down to room temperature) several schemes have been developed, the most important being thermal annealing and (excimer) laser annealing (ELA). Crystallization occurs by thermally activated nucleation and growth processes [1817]. Polycrystalline layers will small grain size can be made amorphous with implantation of Si (self-implantation) and a subsequent optimized (re-)crystallization processes.

In laser annealing energy is locally introduced during short pulses (several 10 ns or even fs); subsequent material change occurs on a sub-μs time scale [1818]. Laser induced crystallization enables the use of inexpensive low-temperature substrates, such as plastic or glass, since it involves the ultrafast melting and resolidification of the near-surface region of the sample, and minimal heating of the substrate. Local processing is also possible using laser crystallization.

In Fig. 24.51 the effect of thermal annealing of amorphous silicon is shown. The annealing time necessary to convert the amorphous phase completely to polycrystalline, e.g. 10 h at 640 °C, depends largely on temperature as detailed in [1819, 1820] (Fig. 24.51a) with a large activation energy of 3.9 eV. Also the final grain size is temperature dependent (Fig. 24.51b).

The introduction of certain metals like Pd [1821], Al [1822], Au [1823] or Ni [1824] induces crystallization and allows for much lower annealing temperatures. Pd and Ni create silicides that play an important role for the grain nucleation or growth front. Au and Al are solved in the bulk but have a similar effect. For example, using Pd complete crystallization of a 150 nm thick a-Si film deposited at 480 °C can be achieved by thermal annealing after 10 h at only 500 °C [1825] (using metal-induced lateral crystallization, MILC).

24.6.2 TFT Devices

A schematic cross section of an amorphous silicon-based TFT is shown in Fig. 24.52a. Carriers in amorphous silicon have a low mobility typically less than 1 cm^2/Vs [1826, 1827]. As-grown polycrystalline silicon has a mobility of typically less than 10 cm^2/Vs. With the use of laser irradiation or thermal annealing, amorphous or small-grain polycrystalline silicon layers can be recrystallized, increasing the mobility up to several 100 cm^2/Vs, improving transistor performance [1826, 1828, 1829]. However, for display applications a mobility of 10 cm^2/Vs is large enough.

The main optimization criteria for thin-film transistors are high on-off ratio, long-term stability, good uniformity and reproducibility, and low cost. Recently, flexible (on polymer substrate) and transparent TFTs (TFET, transparent FET), e.g. with polycrystalline ZnO or GaInZnO (GIZO) channel (Fig. 24.52b), are being investigated for advanced applications such as all-transparent electronics and displays [1398, 1828–1832]. A compilation of recent results on transparent semiconducting oxide (TSO) channel FETs can be found in [1834]. In Fig. 24.53 performance data for

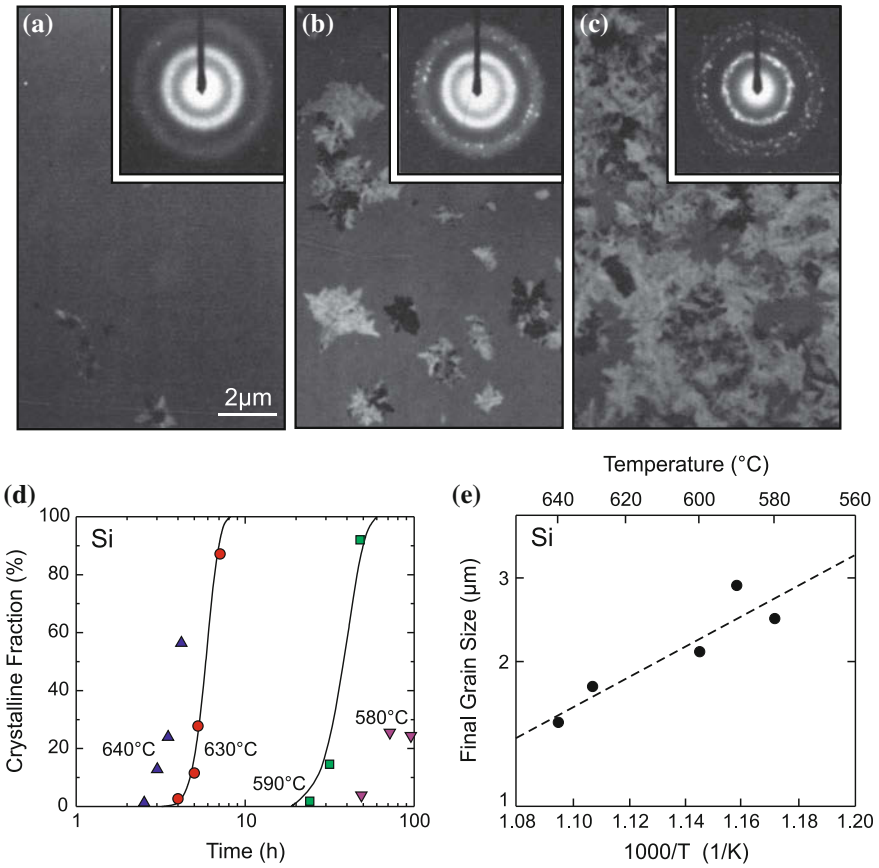


Fig. 24.51 Thermal annealing of 100 nm thick film of amorphous silicon (fabricated from LPCVD and amorphized by 100 keV Si^+ implantation with a dose of $5 \times 10^{15}\,cm^{-2}$). TEM images and diffraction patterns (*insets*) for amorphized Si after (**a**) 4 h, (**b**) 5.25 h and (**c**) 7.1 h annealing at $T = 630\,°C$. The crystalline fractions are 2, 28 and 87 %, respectively. (**d**) Crystalline fraction as a function of annealing time for various annealing temperatures as labeled. *Symbols* are experimental data, *solid lines* depict theory considering grain nucleation and growth. (**e**) Final grain size for various annealing temperature. *Dashed line* is exponential with a slope of 0.6 eV. Adapted from [1819]

various TSO channel FETs are visualized. In Fig. 24.54 a transparent inverter based on ZnO-MESFETs is depicted [1835].

24.6.3 OFETs

Organic field effect transistors (OFETs) [1837–1840] are transistors for which at least the channel consists of an organic material. Most work is done on thin film

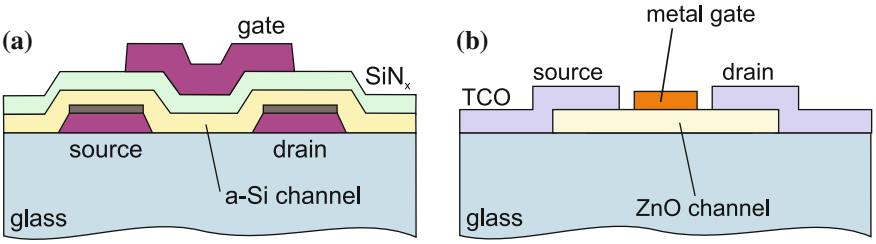


Fig. 24.52 (**a**) Schematic cross section of a top-gate amorphous silicon (a-Si) thin-film transistor (MISFET) on glass substrate. (**b**) Schematic cross section of a transparent ZnO thin-film transistor (MESFET)

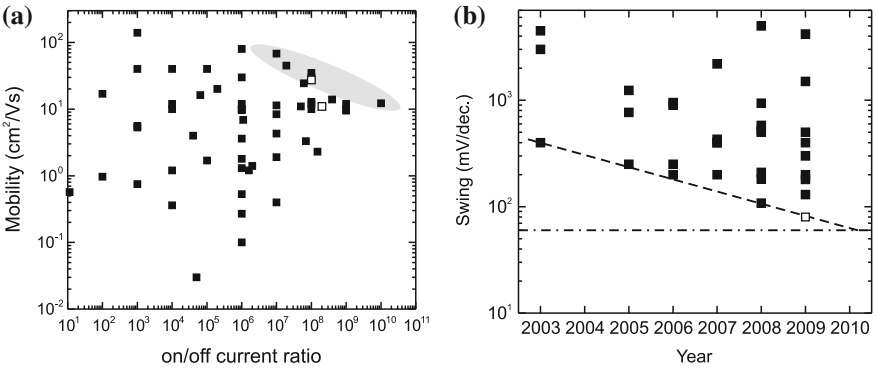


Fig. 24.53 (**a**) Field effect mobility and on/off current ratio for oxide channel transistors. *Filled squares* represent MISFET transistors; *open squares* are for MESFETs from [1833]. The *shaded area* indicates best performance. (**b**) Voltage swing for MISFETs (*filled squares*, subthreshold voltage swing) and MESFET (*empty square*, above turn-on voltage from [1833]) with TSO channels. The *dashed line* is guide to the eyes for the trend of best performance. The *dash-dotted line* indicates the thermodynamic limit of about 60 meV/decade for the swing [1836]. Adapted from [1834]

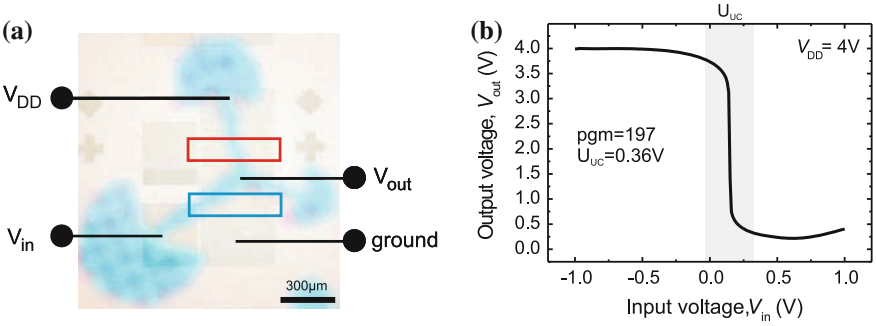


Fig. 24.54 (**a**) Optical image of transparent MESFET inverter based on ZnO. The *two rectangles* indicate the two gates. (**b**) Transfer characteristic for supply voltage $V_{DD} = 4\,V$

transistors, although some work on OFETs using bulk organic semiconductors has been reported [1841, 1842]. Organic materials are also used for the insulator and the contact materials. Often organic and flexible substrates are used. Applications are in low cost electronics, e.g. for driving display pixels or RFID tags (typically operating at 13.56 or 900 MHz). Processes like spin-on and printing can be used. Due to their larger chemical stability against oxidation, mostly p-type channel materials are used. The highest mobilities are reached for pentacene (6 cm^2/Vs) and sexithiophene (1 cm^2/Vs); n-type organic semiconductors exhibit field mobility below 0.1 cm^2/Vs [1838].

Appendix A
Tensors

A.1 Introduction

A physical quantity $T_{ij\ldots m}$ with a total of k indices that is independent of translations of the coordinate system and transforms with respect to all indices like a vector is called a tensor of rank k.

Often, Einstein's sum convention is used; a sum is built over indices with the same symbol. For example, $x_i' = D_{ij}\, x_j$ shall be read as $x_i' = \sum_{j=1}^{3} D_{ij}\, x_j$.

A.2 Rotation of Coordinate System

A rotation of the coordinate system is a transformation $\mathbf{x} \rightarrow \mathbf{x}'$ that is written in components as

$$x_i' = D_{ij}\, x_j. \tag{A.1}$$

$\mathbf{D}$ is called the rotation matrix. The inverse of the rotation matrix is $\mathbf{D}^{-1}$ with

$$D_{kl}^{-1} = D_{lk}, \tag{A.2}$$

i.e. it is the transposed matrix $\mathbf{D}$. The inverse transformation is $x_j = D_{ij}\, x_i'$. Thus,

$$D_{ij}\, D_{kj} = \delta_{ij}. \tag{A.3}$$

A simple example is the azimuthal rotation around the z-axis by an angle ϕ (in the mathematically positive direction)

$$\mathbf{D} = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{A.4}$$

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

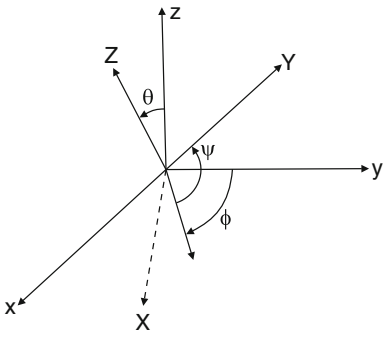


Fig. A.1 Rotation of a coordinate system (x, y, z) by the Euler angles (ϕ, θ, ψ) into the system (X, Y, Z)

For the description of an arbitrary rotation $(x, y, z) \rightarrow (X, Y, Z)$, generally three angles are necessary. Typically, the Euler angles (ϕ, θ, ψ) are used (Fig. A.1). First, the system is rotated by ϕ around the z-axis. The y-axis becomes the u-axis. Then, the system is tilted by θ around the u-axis and the z-axis becomes the Z-axis. Finally, the system is rotated by ψ around the Z-axis.

The matrix for the general rotation by the Euler angles is

$$\begin{pmatrix} \cos\psi\cos\theta\cos\phi - \sin\psi\sin\phi & -\sin\phi\cos\theta\cos\psi - \cos\phi\sin\psi & \sin\theta\cos\psi \\ \cos\phi\cos\theta\sin\psi + \sin\phi\cos\psi & -\sin\psi\cos\theta\sin\phi + \cos\psi\cos\phi & \sin\theta\sin\psi \\ -\cos\phi\sin\theta & \sin\phi\sin\theta & \cos\theta \end{pmatrix}. \quad \text{(A.5)}$$

A.3 Rank-*n* Tensors

Rank-0 Tensors

A tensor of rank 0 is also called a scalar. For example, the length $v_1^2 + v_2^2 + v_3^2$ of the vector $\mathbf{v} = (v_1, v_2, v_3)$ is a scalar since it is invariant under rotation of the coordinate system. However, 'scalar' is not equivalent to 'number' since, e.g. the number $v_1^2 + v_2^2$ is not rotationally invariant.

Rank-1 Tensors

A tensor of rank 1 is a vector. It transforms under coordinate rotation $\mathbf{D}$ as

$$v_i' = D_{ij}\, v_j. \quad \text{(A.6)}$$

Rank-2 Tensors

A tensor of rank 2 is also called a dyade and is a 3×3 matrix $\mathbf{T}$ that transforms under coordinate rotation as

$$T_{ij}' = D_{ik}\, D_{jl}\, T_{kl}. \quad \text{(A.7)}$$

The physical meaning is the following: Two vectors $\mathbf{s}$ and $\mathbf{r}$ shall be related to each other via $s_i = T_{ij}\, r_j$. This could be, e.g., the current $\mathbf{j}$ and the electric field $\mathbf{E}$ that are connected via the tensor of conductivity $\boldsymbol{\sigma}$, i.e. $j_i = \sigma_{ij}\, E_j$.

Such an equation only makes physical sense if it is also valid in a (any) rotated coordination system. The tensor $\mathbf{T}'$ in the rotated coordinate system must fulfill $s'_i = T'_{ij}\, r'_j$. This implies the transformation law (A.7). $s'_k = D_{ki}\, s_i = D_{ki}\, T_{ij} r_j$ and also $s'_k = T'_{km}\, r'_m = T'_{km}\, D_{mj}\, r_j$. Thus, $T'_{km}\, D_{mj} = D_{ki}\, T_{ij}$ since the previous relations are valid for arbitrary $\mathbf{r}$. Multiplication by D_{lj} yields $T'_{km}\, D_{mj}\, D_{lj} = T'_{km}\, \delta_{ml} = T'_{kl} = D_{ki}\, D_{lj}\, T_{ij}$.

The trace of a rank-2 tensor is defined as $\mathrm{tr}\,\mathbf{T} = T_{ii} = T_{11} + T_{22} + T_{33}$. It is a scalar, i.e. invariant under coordinate rotation, since $T'_{kk} = D_{ki} D_{kj} T_{ij} = \delta_{ij} T_{ij} = T_{ii}$.

A rank-2 tensor can be separated into a symmetric part $\mathbf{T}^S$ and an antisymmetric part $\mathbf{T}^A$, i.e. $T^S_{ji} = T^S_{ij}$ and $T^A_{ji} = -T^A_{ij}$ with

$$\mathbf{T} = \mathbf{T}^S + \mathbf{T}^A \tag{A.8a}$$

$$T^S_{ij} = \frac{T_{ij} + T_{ji}}{2} \tag{A.8b}$$

$$T^A_{ij} = \frac{T_{ij} - T_{ji}}{2}. \tag{A.8c}$$

A rank-2 tensor can be separated into an isotropic (spherical) part $\mathbf{T}^I$ and a deviatoric part $\mathbf{T}^D$. The isotropic part is invariant under coordinate rotation.

$$\mathbf{T} = \mathbf{T}^I + \mathbf{T}^D \tag{A.9a}$$

$$T^I_{ij} = \delta_{ij} \frac{\mathrm{tr}\,\mathbf{T}}{3} \tag{A.9b}$$

$$T^D_{ij} = T_{ij} - \delta_{ij} \frac{\mathrm{tr}\,\mathbf{T}}{3}. \tag{A.9c}$$

The trace of $\mathbf{T}$ is the same as that of $\mathbf{T}^I$. The trace of $\mathbf{T}^D$ is zero.

Rank-3 Tensors

A tensor of rank 3 transforms according to

$$T'_{ijk} = D_{il}\, D_{jm}\, D_{kn} T_{lmn}. \tag{A.10}$$

An example is the tensor $\mathbf{e}$ of piezoelectric constants that relates the rank-2 tensor of the strains with the polarization vector $\mathbf{P}$, i.e. $P_i = e_{ijk}\, \epsilon_{jk}$.

Rank-4 Tensors

A tensor of rank 4 transforms according to

$$T'_{ijkl} = D_{im}\, D_{jn}\, D_{ko}\, D_{lp}\, T_{mnop}. \tag{A.11}$$

An example is the tensor $\mathbf{C}$ of elastic constants that relates the rank 2 tensors $\boldsymbol{\epsilon}$ and $\boldsymbol{\sigma}$ of the elastic strains and stresses, i.e. $\sigma_{ij} = C_{ijkl}\, \epsilon_{kl}$.

Appendix B
Point and Space Groups

(See Tables B.1, B.2 and B.3.)

Table B.1 The 10 two-dimensional point groups in group and full and abbreviated international notation

Group	Notation		N_{sg}	Symmetry elements
	Full international	Abbrev.		
C_1	1	1	1	C_1
D_1	1m	m	3	C_1, m
C_2	2	2	1	C_2
D_2	2mm	mm	4	C_2, 2m
C_3	3	3	1	C_3
D_3	3m	3m	2	C_3, 3m
C_4	4	4	1	C_4
D_4	4mm	4m	2	C_4, 2m
C_6	6	6	1	C_6
D_6	6mm	6m	1	C_6, 6m

N_{sg} denotes the number of space groups

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

Table B.2 The 32 point groups in Schönfließ and international notation

System	Class		N_{sg}	Symmetry elements
	international	Schönfließ		
Triclinic	1	C_1	1	E
	$\bar{1}$	C_i	1	$E\ i$
Monoclinic	m	C_s	3	$E\ \sigma_h$
	2	C_2	4	$E\ C_2$
	$2/m$	C_{2h}	6	$E\ C_2\ i\ \sigma_h$
Orthorhombic	$2mm$	C_{2v}	9	$E\ C_2\ \sigma_v'\ \sigma_v''$
	222	D_2	22	$E\ C_2\ C_2'\ C_2''$
	mmm	D_{2h}	28	$E\ C_2\ C_2'\ C_2''\ i\ \sigma_h\ \sigma_v'\ \sigma_v''$
Tetragonal	4	C_4	6	$E\ 2C_4\ C_2$
	$\bar{4}$	S_4	2	$E\ 2S_4\ C_2$
	$4/m$	C_{4h}	6	$E\ 2C_4\ C_2\ i\ 2S_4\ \sigma_h$
	$4mm$	C_{4v}	10	$E\ 2C_4\ C_2\ 2\sigma_v'\ 2\sigma_d$
	$\bar{4}2m$	D_{2d}	12	$E\ C_2\ C_2'\ C_2''\ 2\sigma_d\ 2S_4$
	422	D_4	12	$E\ 2C_4\ C_2\ 2C_2'\ 2C_2''$
	$4/mmm$	D_{4h}	20	$E\ 2C_4\ C_2\ 2C_2'\ 2C_2''\ i\ 2S_4\ \sigma_h\ 2\sigma_v'\ 2\sigma_h$
Trigonal (rhombohedral)	3	C_3	4	$E\ 2C_3$
	$\bar{3}$	S_6	2	$E\ 2C_3\ i\ 2S_6$
	$3m$	C_{3v}	7	$E\ 2C_3\ 3\sigma_v$
	32	D_3	6	$E\ 2C_3\ 3C_2$
	$\bar{3}m$	D_{3d}	6	$E\ 2C_3\ 3C_2\ i\ 2S_6\ 3\sigma_d$
Hexagonal	$\bar{6}$	C_{3h}	6	$E\ 2C_3\ \sigma_h\ 2S_3$
	6	C_6	1	$E\ 2C_6\ 2C_3\ C_2$
	$6/m$	C_{6h}	2	$E\ 2C_6\ 2C_3\ C_2\ i\ 2S_3\ 2S_6\ \sigma_h$
	$\bar{6}m2$	D_{3h}	6	$E\ 2C_3\ 3C_2\ \sigma_h\ 2S_3\ 3\sigma_v$
	$6mm$	C_{6v}	4	$E\ 2C_6\ 2C_3\ C_2\ 3\sigma_v\ 3\sigma_d$
	622	D_6	4	$E\ 2C_6\ 2C_3\ C_2\ 3C_2'\ 3C_2''$
	$6/mmm$	D_{6h}	4	$E\ 2C_6\ 2C_3\ C_2\ 3C_2'\ 3C_2''\ i\ 2S_3\ \ 2S_6\ \sigma_h\ 3\sigma_d\ 3\sigma_v$
Cubic	23	T	5	$E\ 4C_3\ 4C_3^2\ 3C_2$
	$m3$	T_h	7	$E\ 4C_3\ 4C_3^2\ 3C_2\ i\ 8S_6\ 3\sigma_h$
	$\bar{4}3m$	T_d	8	$E\ 8C_3\ 3C_2\ 6\sigma_d\ 6S_4$
	432	O	6	$E\ 8C_3\ 3C_2\ 6C_2'\ 6C_4$
	$m3m$	O_h	10	$E\ 8C_3\ 3C_2\ 6C_2\ 6C_4\ i\ 8S_6\ 3\sigma_h\ 6\sigma_d\ 6S_4$

N_{sg} denotes the number of space groups

Table B.3 Space group numbers and corresponding space group symbols in standard international notation

1	P1	2	$P\bar{1}$	3	P2	4	$P2_1$	5	C2
6	Pm	7	Pc	8	Cm	9	Cc	10	P2/m
11	$P2_1/m$	12	C2/m	13	P2/c	14	$P2_1/c$	15	C2/c
16	P222	17	$P222_1$	18	$P2_12_12$	19	$P2_12_12_1$	20	$C222_1$
21	C222	22	F222	23	I222	24	$I2_12_12_1$	25	Pmm2
26	$Pmc2_1$	27	Pcc2	28	Pma2	29	$Pca2_1$	30	Pnc2
31	$Pmn2_1$	32	Pba2	33	$Pna2_1$	34	Pnn2	35	Cmm2
36	$Cmc2_1$	37	Ccc2	38	Amm2	39	Abm2	40	Ama2
41	Aba2	42	4_2	43	Fmm2	44	Imm2	45	Iba2
46	Ima2	47	Pmmm	48	Pnnn	49	Pccm	50	Pban
51	Pmma	52	Pnna	53	Pmna	54	Pcca	55	Pbam
56	Pccn	57	Pbcm	58	Pnnm	59	Pmmn	60	Pbcn
61	Pbca	62	Pnma	63	Cmcm	64	Cmca	65	Cmmm
66	Cccm	67	Cmma	68	Ccca	69	Fmmm	70	Fddd
71	Immm	72	Ibam	73	Ibca	74	Imma	75	P4
76	$P4_1$	77	$P4_2$	78	$P4_3$	79	I4	80	$I4_1$
81	$P\bar{4}$	82	$I\bar{4}$	83	P4/m	84	$P4_2/m$	85	P4/n
86	$P4_2/n$	87	I4/m	88	$I4_1/a$	89	P422	90	$P42_12$
91	P4122	92	$P412_12$	93	$P4_222$	94	$P4_22_12$	95	$P4_322$
96	$P4_32_12$	97	I422	98	$I4_122$	99	P4mm	100	P4bm
101	$P4_2cm$	102	$P4_2nm$	103	P4cc	104	P4nc	105	$P4_2mc$
106	$P4_2bc$	107	I4mm	108	I4cm	109	$I4_1md$	110	$I4_1cd$
111	$P\bar{4}2m$	112	$P\bar{4}2c$	113	$P\bar{4}2_1m$	114	$P\,\bar{4}2_1c$	115	$P\bar{4}m2$
116	$P\bar{4}c2$	117	$P\bar{4}b2$	118	$P\bar{4}n2$	119	$I\bar{4}m2$	120	$I\bar{4}c2$
121	$I\bar{4}2m$	122	$I\bar{4}2d$	123	P4/mmm	124	P4/mcc	125	P4/nbm
126	P4/nnc	127	P4/mbm	128	P4/mnc	129	P4/nmm	130	P4/ncc
131	$P4_2/mmc$	132	$P4_2/mcm$	133	$P4_2/nbc$	134	$P4_2/nnm$	135	$P4_2/mbc$
136	$P4_2/mnm$	137	$P4_2/nmc$	138	$P4_2/ncm$	139	I4/mmm	140	I4/mcm
141	$I4_1/amd$	142	$I4_1/acd$	143	P3	144	$P3_1$	145	$P3_2$
146	R3	147	$P\bar{3}$	148	$R\bar{3}$	149	P312	150	P321
151	$P3_112$	152	$P3_121$	153	$P3_212$	154	$P3_221$	155	R32
156	P3m1	157	P31m	158	P3c1	159	P31c	160	R3m
161	R3c	162	$P\bar{3}1m$	163	$P\bar{3}1c$	164	$P\bar{3}m1$	165	$P\bar{3}c1$
166	$R\bar{3}m$	167	$R\bar{3}c$	168	P6	169	$P6_1$	170	$P6_5$
171	$P6_2$	172	$P6_4$	173	$P6_3$	174	$P\bar{6}$	175	P6/m
176	$P6_3/m$	177	P622	178	$P6_122$	179	$P6_522$	180	$P6_222$
181	$P6_422$	182	$P6_322$	183	P6mm	184	P6cc	185	$P6_3cm$
186	$P6_3mc$	187	$P\bar{6}m2$	188	$P\bar{6}c2$	189	$P\bar{6}2m$	190	$P\bar{6}2c$

(continued)

Table B.3 (continued)

191	P6/mmm	192	P6/mcc	193	$P6_3/mcm$	194	$P6_3/mmc$	195	P23
196	F23	197	I23	198	$P2_13$	199	$I2_13$	200	$Pm\bar{3}$
201	$Pn\bar{3}$	202	$Fm\bar{3}$	203	$Fd\bar{3}$	204	$Im\bar{3}$	205	$Pa\bar{3}$
206	$Ia\bar{3}$	207	P432	208	$P4_232$	209	F432	210	$F4_132$
211	I432	212	$P4_332$	213	$P4_132$	214	$I4_132$	215	$P\bar{4}3m$
216	$F\bar{4}3m$	217	$I\bar{4}3m$	218	$P\bar{4}3n$	219	$F\bar{4}$ 3c	220	$I\bar{4}3d$
221	$Pm\bar{3}m$	222	$Pn\bar{3}n$	223	$Pm\bar{3}n$	224	$Pn\bar{3}m$	225	$Fm\bar{3}m$
226	$Fm\bar{3}c$	227	$Fd\bar{3}m$	228	$Fd\bar{3}c$	229	$Im\bar{3}m$	230	$Ia\bar{3}d$

Appendix C
Kramers–Kronig Relations

The Kramers–Kronig relations (KKR) are relations between the real and imaginary part of the dielectric function. They are of a general nature and are based on the properties of a complex, analytical response function $f(\omega) = f_1(\omega) + \mathrm{i}f_2(\omega)$ fulfilling the following conditions[1]:

- The poles of $f(\omega)$ are below the real axis.
- The integral of $f(\omega)/\omega$ along a semicircle with infinite radius in the upper half of the complex plane vanishes.
- The function $f_1(\omega)$ is even and the function $f_2(\omega)$ is odd for real values of the argument.

The integral of $f(s)/(s-\omega)\mathrm{d}s$ along the real axis and an infinite semicircle in the upper half of the complex plane is zero because the path is a closed line. The integral along a semicircle above the pole at $s = \omega$ yields $-\pi\mathrm{i}f(\omega)$, the integral over the infinite semicircle is zero. Therefore the value of $f(\omega)$ is given by[2]

$$f(\omega) = \frac{1}{\pi\mathrm{i}} \Pr\int_{-\infty}^{\infty} \frac{f(s)}{s-\omega}\mathrm{d}s. \tag{C.1}$$

Equating the real and imaginary parts of (C.1) yields for the real part

$$f_1(\omega) = \frac{1}{\pi} \Pr\int_{-\infty}^{\infty} \frac{f_2(s)}{s-\omega}\mathrm{d}s. \tag{C.2}$$

Splitting the integral into two parts $\int_0^\infty$ and $\int_{-\infty}^0$, going from s to $-s$ in the latter and using $f_2(-\omega) = -f_2(\omega)$ and $\frac{1}{s-\omega} + \frac{1}{s+\omega} = \frac{2s}{s^2-\omega^2}$ yields (C.3a)

[1] The requirements for the function f to which the KKR apply can be interpreted as that the function must represent the Fourier transform of a linear and causal physical process.

[2] The Cauchy principal value Pr of the integral is the limit for $\delta \to 0$ of the sum of the integrals over $-\infty < s < \omega - \delta$ and $\omega + \delta < s < \infty$.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

$$f_1(\omega) = \frac{2}{\pi} \Pr \int_0^\infty \frac{s f_2(s)}{s^2 - \omega^2} \mathrm{d}s \tag{C.3a}$$

$$f_2(\omega) = -\frac{2}{\pi} \Pr \int_0^\infty \frac{f_1(s)}{s^2 - \omega^2} \mathrm{d}s. \tag{C.3b}$$

In a similar way, (C.3b) is obtained. These two relations are the Kramers–Kronig relations [1839, 1840]. They are most often applied to the dielectric function ϵ. In this case, they apply to the susceptibility, i.e. $f(\omega) = \chi(\omega) = \epsilon(\omega)/\epsilon_0 - 1$. The susceptibility can be interpreted as the Fourier transform of the time-dependent polarization in the semiconductor after an infinitely short pulsed electric field, i.e. the impulse response of the polarization. For the dielectric function $\epsilon = \epsilon_1 + \mathrm{i}\epsilon_2$, the following KKR relations hold:

$$\epsilon_1(\omega) = \epsilon_0 + \frac{2}{\pi} \Pr \int_0^\infty \frac{s\, \epsilon_2(s)}{s^2 - \omega^2} \mathrm{d}s \tag{C.4a}$$

$$\epsilon_2(\omega) = -\frac{2\omega}{\pi} \Pr \int_0^\infty \frac{\epsilon_1(s) - \epsilon_0}{s^2 - \omega^2} \mathrm{d}s. \tag{C.4b}$$

The static dielectric constant is thus given by

$$\epsilon(0) = \epsilon_0 + \frac{2}{\pi} \Pr \int_0^\infty \frac{\epsilon_2(s)}{s} \mathrm{d}s. \tag{C.5}$$

The integral does not diverge since ϵ_2 is an odd function and zero at $\omega = 0$. Generally the jth momentum M_j of the imaginary part of the dielectric function is

$$M_j = \int_0^\infty \epsilon_2(\omega)\, \omega^j \, \mathrm{d}\omega. \tag{C.6}$$

Thus, $M_{-1} = \pi[\epsilon(0) - \epsilon_0]/2$.

Other KKRs are, e.g., the relation between the index of refraction n_r and the absorption coefficient α:

$$n_\mathrm{r}(\lambda) = \frac{1}{\pi} \Pr \int_0^\infty \frac{\alpha(s)}{1 - s^2/\lambda^2} \mathrm{d}s. \tag{C.7}$$

If the imaginary (real) part of the dielectric function is known (for all frequencies), the real (imaginary) part can be calculated via the KKR. If the dependence is not known for the entire frequency range, assumptions about the dielectric function in the unknown spectral regions must be made that limits the reliability of the transformation.

Appendix D
Oscillator Strength

The response of an oscillator to an electric field **E** is formulated with the dielectric function. The resulting polarization **P** is related to the electric field via

$$\mathbf{P} = \epsilon_0 \, \chi \, \mathbf{E}, \tag{D.1}$$

with χ being the electric susceptibility, and the displacement field **D** is given by

$$\mathbf{D} = \epsilon_0 \, \mathbf{E} + \mathbf{P} = \epsilon_0 \, \epsilon \, \mathbf{E}. \tag{D.2}$$

Thus the (relative) dielectric constant is

$$\epsilon = 1 + \chi. \tag{D.3}$$

We assume a harmonic oscillator model for an electron, i.e. an equation of motion for the amplitude $x = x_0 \exp(\mathrm{i}\omega t)$

$$m \, \ddot{x} = -C \, x. \tag{D.4}$$

The resonance frequency is $\omega_0^2 = C/m$. The presence of a harmonic electric field E of frequency ω and amplitude E_0 adds a force eE. Thus,

$$-\, m \, \omega^2 \, x = -m \, \omega_0^2 \, x + e \, E. \tag{D.5}$$

The polarization ex_0 is given by

$$e x_0 = \frac{e^2}{m} \, \frac{1}{\omega_0^2 - \omega^2} \, E_0 = \frac{e^2}{m \, \omega_0^2} \, \frac{1}{1 - \omega^2/\omega_0^2} \, E_0. \tag{D.6}$$

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

The pre-factor is called the (dimensionless) *oscillator strength* and will be denoted as

$$f = \frac{e^2}{\epsilon_0 m \omega_0^2} \tag{D.7}$$

in the following. The frequency-dependent dielectric function of the resonance is thus

$$\epsilon(\omega) = 1 + \frac{f}{1 - \omega^2/\omega_0^2}. \tag{D.8}$$

In the low-frequency limit, the dielectric function is $\epsilon(0) = 1 + f$, in the high-frequency limit $\epsilon(\infty) = 1$. The oscillator strength is the difference of ϵ for frequencies below and above the resonance.

For all systems, the high-frequency limit of ϵ is 1. This means that $\chi = 0$, i.e. there are no more oscillators to be polarized. The low-frequency limit includes all possible oscillators. If there are further oscillators between frequencies well above ω_0 and $\omega \to \infty$, these are summarized as the high-frequency dielectric constant $\epsilon_\infty > 1$. Equation (D.8) then reads

$$\epsilon(\omega) = \epsilon(\infty) + \frac{\hat{f}}{1 - \omega^2/\omega_0^2}. \tag{D.9}$$

The limit $\epsilon \to \epsilon(\infty)$ is only valid for frequencies above ω_0 but smaller than the next resonance(s) at higher frequencies.[3] Another common form is to include the background dielectric constant via

$$\epsilon(\omega) = \epsilon(\infty)\left[1 + \frac{f}{1 - \omega^2/\omega_0^2}\right]. \tag{D.10}$$

Obviously, $f = \hat{f}/\epsilon(\infty)$, making the two forms equivalent.

In order to discuss the lineshape, not only for ϵ but also for the index of refraction $n^* = n_\mathrm{r} + \mathrm{i}\kappa = \sqrt{\epsilon}$, we introduce damping to our calculation by adding a term $-m\Gamma\dot{x}$ to the left side of (D.5). This term is something like a 'friction' and would cause the oscillation amplitude to decay exponentially with a time constant $\tau = 2/\Gamma$ without external stimulus. The dielectric constant is

$$\epsilon(\omega) = \epsilon(\infty)\left[1 + \frac{f}{1 - (\omega^2 + \mathrm{i}\,\omega\Gamma)/\omega_0^2}\right] = \epsilon' + \mathrm{i}\epsilon''. \tag{D.11}$$

The real and imaginary part fulfill the Kramers–Kronig relations (C.3a) and (C.3b). For the oscillator strength, the regimes of large oscillator strength ($f \sim 1$) and small oscillator strength ($f \ll 1$) are distinguished. For the damping, two regimes

[3] For ω going to infinite values (beyond the X-ray regime), ϵ always goes towards one.

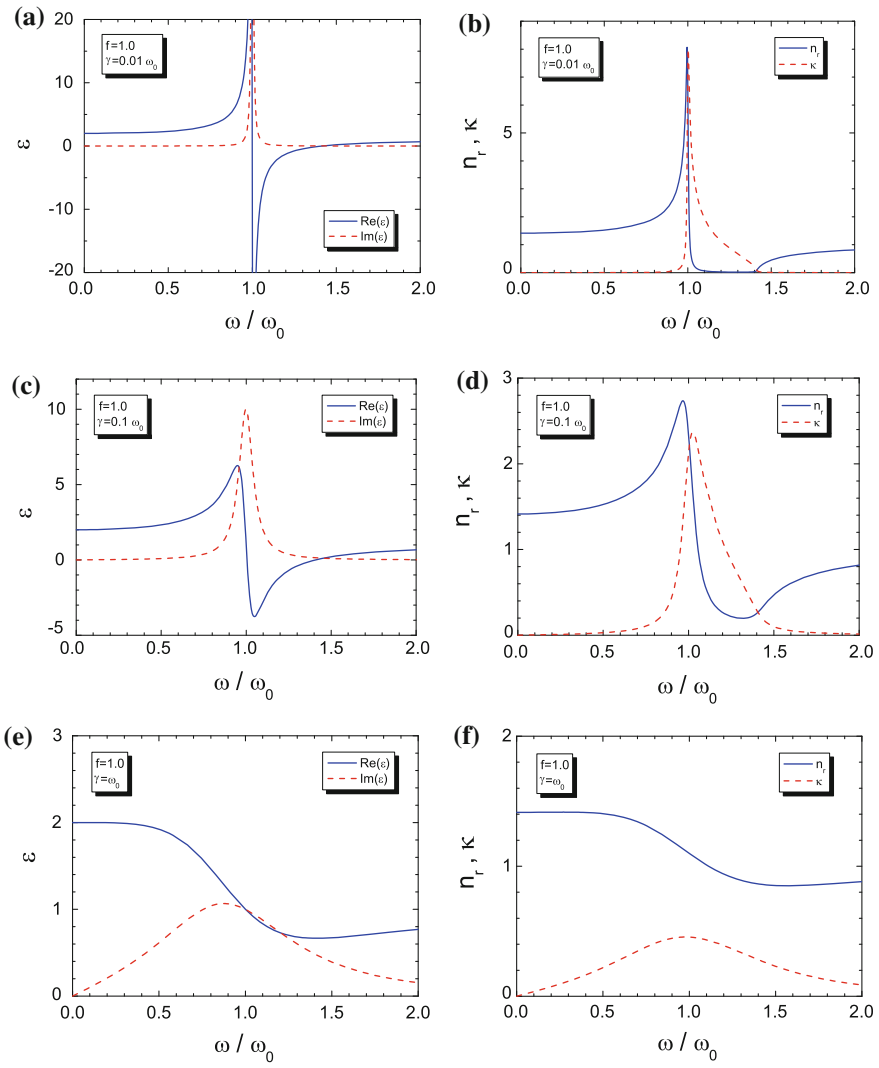


Fig. D.1 Real (*solid lines*) and imaginary (*dashed lines*) parts of the dielectric constant (**a**, **c**, **e**) and index of refraction (**b**, **d**, **f**) (D.11) for oscillator strength $f = 1$ and various values of damping: (**a**, **b**) $\Gamma = 10^{-2}\omega_0$, (**c**, **d**) $\Gamma = 10^{-1}\omega_0$, and (**e**, **f**) $\Gamma = \omega_0$

should be distinguished: Small damping ($\Gamma \ll \omega_0$) and strong damping ($\Gamma \gtrsim \omega_0$). Typical lineshapes are shown in Figs. D.1 and D.2.

For small oscillator strength, i.e. $f \ll 1$, the index of refraction $n^* = \sqrt{\epsilon} = n_\mathrm{r} + \mathrm{i}\kappa$ is given by ($n_\infty = \sqrt{\epsilon(\infty)}$)

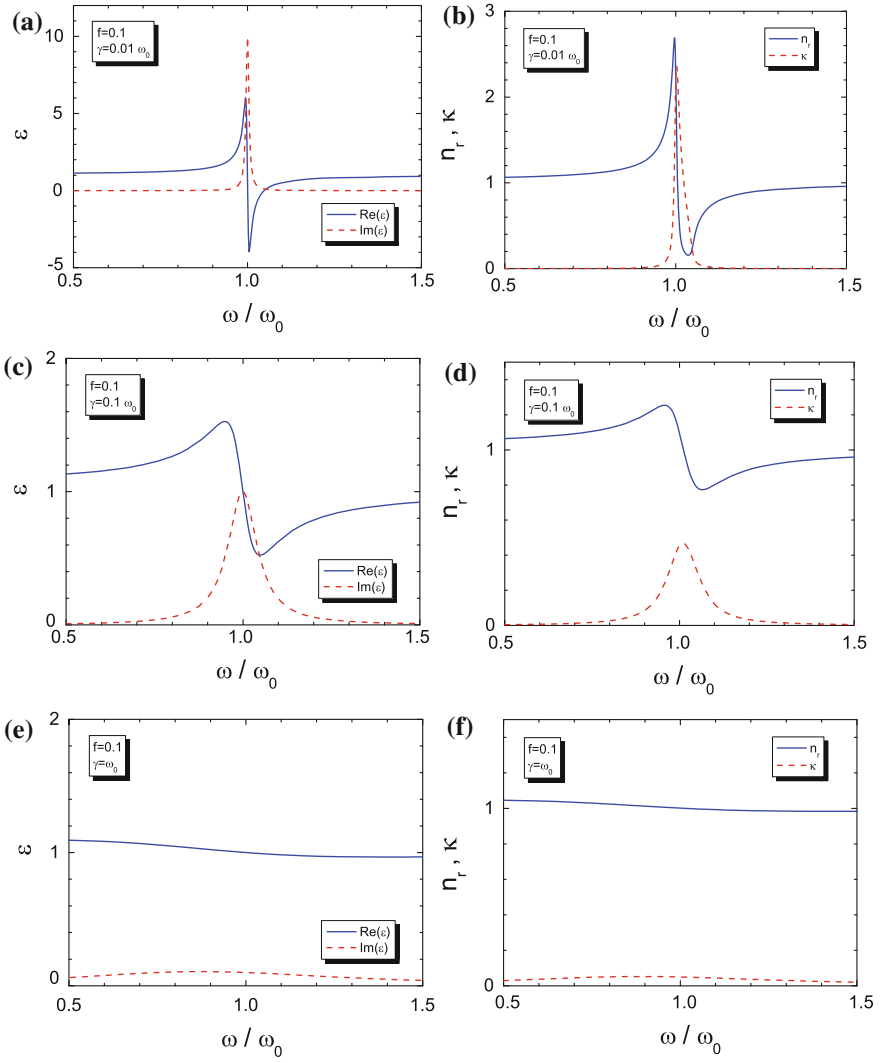


Fig. D.2 Real (*solid lines*) and imaginary (*dashed lines*) parts of the dielectric constant (**a**, **c**, **e**) and index of refraction (**b**, **d**, **f**) (D.11) for oscillator strength $f = 10^{-1}$ and various values of damping: (**a**, **b**) $\Gamma = 10^{-2}\omega_0$, (**c**, **d**) $\Gamma = 10^{-1}\omega_0$, and (**e**, **f**) $\Gamma = \omega_0$

$$n_{\mathrm{r}} = n_\infty \left[1 + \frac{f}{2}\,\frac{\omega_0^2\,(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + \Gamma^2\,\omega^2}\right] \tag{D.12a}$$

$$\kappa = n_\infty\,\frac{f}{2}\,\frac{\Gamma\,\omega_0\,(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + \Gamma^2\,\omega^2}. \tag{D.12b}$$

For small detuning from the resonance frequency, i.e. $\omega = \omega_0 + \delta\omega$ with $|\delta\omega|/\omega_0 \ll 1$, the index of refraction is given by

$$n_\mathrm{r} = n_\infty \left[1 - \frac{f}{4} \frac{\omega_0\,\delta\omega}{(\delta\omega)^2 + \Gamma^2/4}\right] \tag{D.13a}$$

$$\kappa = n_\infty \frac{f}{4} \frac{\omega_0\,\Gamma/2}{(\delta\omega)^2 + \Gamma^2/4}. \tag{D.13b}$$

The maximum absorption is given as

$$\alpha_\mathrm{m} = 2\,\frac{\omega_0}{c}\,\kappa(\omega_0) = f\,\frac{\omega_0^2}{\Gamma}\,\frac{n_\infty}{c}. \tag{D.14}$$

For zero damping, the dielectric function has a zero at

$$\omega_0' = \omega_0\,\sqrt{1+f} \approx \omega_0\,\left(1 + \frac{f}{2}\right). \tag{D.15}$$

The latter approximation is valid for $f \ll 1$. In the region between ω_0 and ω_0', the real part of the index of refraction is very small (for the physically unrealistic case of $\Gamma \equiv 0$ it is exactly zero since $\epsilon < 0$). The reflectance (for vertical incidence $R = [(1-n^*)/(1+n^*)]^2$) in this region (width: $f\omega_0/2$) is thus very high. For larger damping (and small oscillator strength), this effect is washed out.

The frequency $\omega_{\epsilon'',\mathrm{max}}$ of the maximum of the imaginary part of ϵ'' of the dielectric function ($\hat{\Gamma} = \Gamma/\omega_0$) is

$$\omega_{\epsilon'',\mathrm{max}}^2 = \omega_0^2\,\frac{2 - \hat{\Gamma}^2 + \sqrt{16 - 4\hat{\Gamma}^2 + \hat{\Gamma}^4}}{6} \approx \omega_0^2\left[1 - \left(\frac{\Gamma}{2\omega_0}\right)^2\right]. \tag{D.16}$$

The approximation is valid for small damping $\Gamma \ll \omega_0$. In this case, the detuned frequency of the maximum is close to ω_0 (Fig. D.3). The frequency position of the maximum of $\tan\delta = \epsilon''/\epsilon'$ is

$$\omega_{\tan\delta,\mathrm{max}}^2 = \omega_0^2\,\frac{2 + f - \hat{\Gamma}^2 + \Lambda^2}{6} \tag{D.17}$$

$$\Lambda^2 = \sqrt{12\,(1+f) + \left(2 + f - \hat{\Gamma}^2\right)^2}.$$

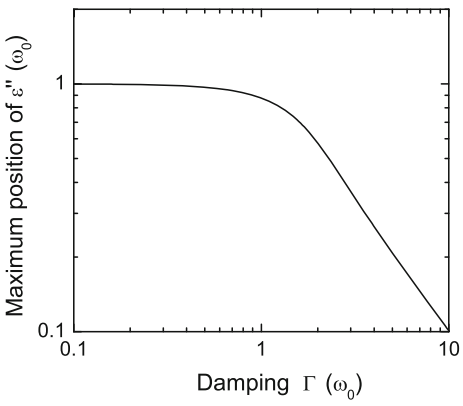


Fig. D.3 Frequency position of the maximum of ϵ'' as a function of the damping

The value of $\tan\delta$ at its maximum is (Λ has the same meaning as in (D.17))

$$(\tan\delta)_{\max} = \frac{-3\sqrt{\frac{3}{2}} f\,\hat{\Gamma}\,\sqrt{2+f-\hat{\Gamma}^2+\Lambda^2}}{-8-8f+f^2-4\,\hat{\Gamma}^2-2f\,\hat{\Gamma}^2+\hat{\Gamma}^4+\left(2+f-\hat{\Gamma}^2\right)\Lambda^2}. \quad \text{(D.18)}$$

Appendix E
Quantum Statistics

E.1 Introduction

Bosons are particles with integer spin $s = n$, fermions are particles with spin $s = n + 1/2$ with n being an integer including zero. The fundamental quantum-mechanical property of the wavefunction of a system with N such particles is that under exchange of any two particles, the wavefunction is symmetric in the case of bosons and antisymmetric in the case of fermions. For two particles, these conditions read

$$\Psi(q_1, q_2) = \Psi(q_2, q_1) \tag{E.1a}$$
$$\Psi(q_1, q_2) = -\Psi(q_2, q_1), \tag{E.1b}$$

where (E.1a) holds for bosons and (E.1b) holds for fermions. The variables q_i denote the coordinates and spin of the ith particle. The Pauli principle allows bosons to populate the same single particle state with an arbitrary number of particles (at least more than one). For fermions, the exclusion principle holds that each single particle state can only be populated once.

E.2 Partition Sum

We consider a gas of N identical particles in a volume V in equilibrium at a temperature T. The possible quantum-mechanical states of a particle is denoted as r. The energy of a particle in the state r is ϵ_r, the number of particles in the state r is n_r.

For vanishing interaction of the particles, the total energy of the gas in the state R (with n_r particles in the state r) is

$$E_R = \sum_r n_r \, \epsilon_r. \tag{E.2}$$

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

The sum runs over all possible states r. The total number of particles imposes the condition

$$N = \sum_r n_r. \tag{E.3}$$

In order to calculate the thermodynamic potentials, the partition sum Z needs to be calculated

$$Z = \sum_R \exp(-\beta E_R), \tag{E.4}$$

with $\beta = 1/(kT)$. The sum runs over all possible microscopic states R of the gas, i.e. all combinations of the n_r that fulfill (E.3). The probability P_S to find the system in a particular state S is given by (canonical ensemble)

$$P_S = \frac{\exp(-\beta E_S)}{Z}. \tag{E.5}$$

The average number $\bar{n}_s$ of particles in a state s is given by

$$\bar{n}_s = \frac{\sum_R n_s \exp(-\beta E_R)}{Z} = -\frac{1}{\beta Z}\frac{\partial Z}{\partial \epsilon_\mathrm{s}} = -\frac{1}{\beta}\frac{\partial \ln Z}{\partial \epsilon_\mathrm{s}}. \tag{E.6}$$

We note that the average deviation $\overline{(\Delta n_s)^2} = \overline{n_s^2 - \bar{n}_s^2} = \overline{n_s^2} - \bar{n}_s^2$ is given by

$$\overline{(\Delta n_s)^2} = \frac{1}{\beta^2}\frac{\partial^2 \ln Z}{\partial \epsilon_\mathrm{s}^2} = -\frac{1}{\beta}\frac{\partial \bar{n}_s}{\partial \epsilon_\mathrm{s}}. \tag{E.7}$$

In the Bose–Einstein statistics (for bosons), the particles are fundamentally indistinguishable. Thus, a set of $(n_1, n_2, \ldots)$ uniquely describes the system. In the case of fermions, for each state n_r is either 0 or 1. In both cases, (E.3) needs to be fulfilled.

E.3 Photon Statistics

This case is the Bose–Einstein statistics (cf. (E.24)) with undefined particle number. We rewrite (E.6) as

$$\bar{n}_s = \frac{\sum_{n_s} n_s \exp(-\beta n_s \epsilon_\mathrm{s}) \sum^{(s)}_{n_1,n_2,\ldots} \exp(-\beta(n_1\epsilon_1 + n_2\epsilon_2 + \ldots))}{\sum_{n_s} \exp(-\beta n_s \epsilon_\mathrm{s}) \sum^{(s)}_{n_1,n_2,\ldots} \exp(-\beta(n_1\epsilon_1 + n_2\epsilon_2 + \ldots))}, \tag{E.8}$$

where $\sum^{(s)}$ denotes a summation that does not include the index s. In the case of photons, the values n_r can take any value (integers including zero) without restriction and therefore the sums $\sum^{(s)}$ in the numerator and denominator of (E.8) are identical. After some calculation we find

$$\bar{n}_s = -\frac{1}{\beta}\frac{\partial}{\partial \epsilon_s} \ln\left(\sum_{n_s=0}^{\infty} \exp(-\beta n_s \epsilon_s)\right). \tag{E.9}$$

The argument of the logarithm is a geometrical series with the limit $[1- \exp(-\beta\epsilon_s)]^{-1}$. This leads to the so-called Planck distribution

$$\bar{n}_s = \frac{1}{\exp(\beta\epsilon_s) - 1}. \tag{E.10}$$

E.4 Fermi–Dirac Statistics

Now, the particle number is fixed to N. For the sum $\sum^{(s)}$ from (E.6), we introduce the term $Z_S(M)$

$$Z_s(M) = \sum_{n_1,n_2,\ldots}^{(s)} \exp(-\beta(n_1\epsilon_1 + n_2\epsilon_2 + \ldots)), \tag{E.11}$$

when M particles are to be distributed over all states except s ($\sum_r^{(s)} n_r = M$). M is either $N-1$ if $n_s = 1$ and N if $n_s = 0$. Using Z_s, we can write

$$\bar{n}_s = \frac{1}{\frac{Z_s(N)}{Z_s(N-1)} \exp(\beta\epsilon_s) + 1}. \tag{E.12}$$

We evaluate $Z_s(N-1)$

$$\ln Z_s(N - \Delta N) = \ln Z_s(N) - \frac{\partial \ln Z_s}{\partial N}|_N \, \Delta N, \tag{E.13}$$

or

$$Z_s(N - \Delta N) = Z_s(N) \exp(-\gamma_s \Delta N), \tag{E.14}$$

with

$$\gamma_s = \frac{\partial \ln Z_s}{\partial N}. \tag{E.15}$$

Since Z_s runs over many states, the derivative is approximately equal to

$$\gamma = \frac{\partial \ln Z}{\partial N}, \tag{E.16}$$

as will be shown below. Thus, we obtained so far

$$\bar{n}_s = \frac{1}{\exp(\gamma + \beta\epsilon_s) + 1}. \tag{E.17}$$

Equation (E.3) holds also for the average values $\bar{n}_s$, i.e.

$$N = \sum_r \bar{n}_r = \sum_r \frac{1}{\exp(\gamma + \beta\epsilon_s) + 1}, \tag{E.18}$$

from which the value of γ can be calculated. Given that the free energy is given as $F = -kT \ln Z$, we find that

$$\gamma = -\frac{1}{kT}\frac{\partial F}{\partial N} = -\beta\mu \,, \tag{E.19}$$

where μ is the chemical potential by definition. Therefore, the distribution function for the Fermi–Dirac statistics (also called the Fermi function) is

$$\bar{n}_s = \frac{1}{\exp(\beta(\epsilon_s - \mu)) + 1}. \tag{E.20}$$

Now, we briefly revisit the approximation $\gamma = \gamma_s$. Exactly, γ fulfills

$$\gamma = \gamma_s - n_s \frac{\partial \gamma}{\partial N}. \tag{E.21}$$

Thus, the approximation is valid if $n_s \frac{\partial \gamma}{\partial N} \ll \gamma$. Since $n_s < 1$, this means that the chemical potential does not change significantly upon addition of another particle.

The Fermi–Dirac distribution function (E.20) for electrons is typically written as

$$f_e(E) = \frac{1}{\exp\left(\frac{E - E_F}{kT}\right) + 1}, \tag{E.22}$$

where k (or k_B) denotes the Boltzmann constant, T is the temperature, and E_F is the Fermi level, which is called the chemical potential μ in thermodynamics. The Fermi distribution is shown in Fig. E.1 for various parameters. The distribution function gives the probability that a state at energy E is populated in thermodynamic equilibrium. For $E = E_F$ the population is $1/2$ for all temperatures. At (the unrealistic case of) $T = 0$, the function makes a step from 1 (for $E < E_F$) to 0.

The high-energy tail of the Fermi distribution, i.e. for $E - E_F \gg kT$, can be approximated by the Boltzmann distribution:

$$f_e(E) \cong \exp\left(-\frac{E - E_F}{kT}\right). \tag{E.23}$$

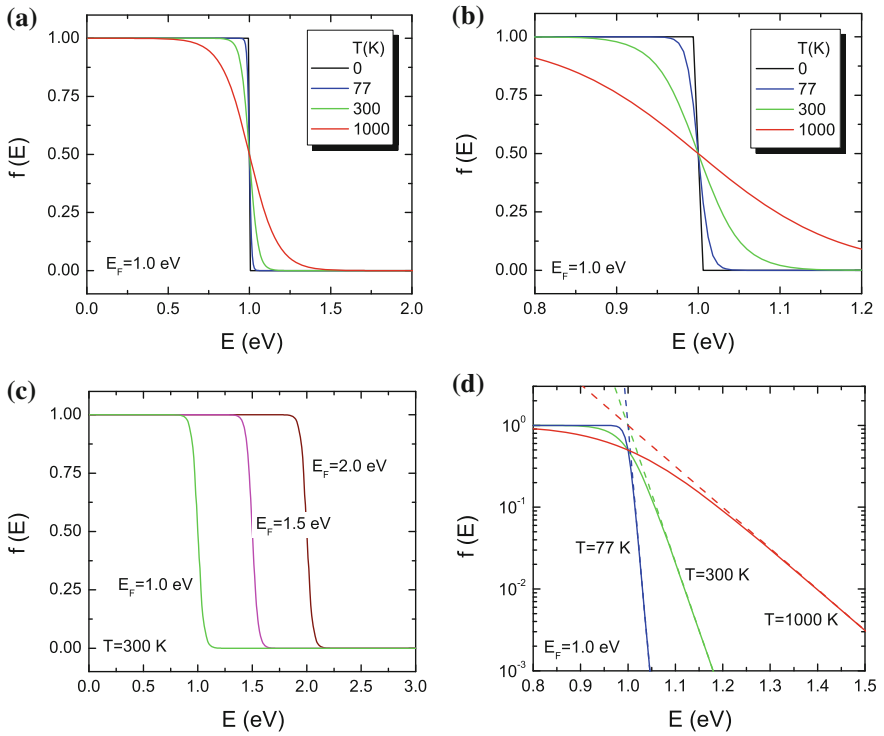


Fig. E.1 Fermi function for (**a**, **b**) different temperatures (for $E_{\mathrm{F}} = 1.0\,\mathrm{eV}$) and (**c**) for different chemical potentials (for $T = 300\,\mathrm{K}$). (**d**) Fermi function (*solid lines*) compared with Boltzmann approximation (*dashed lines*) for various temperatures and $E_{\mathrm{F}} = 1.0\,\mathrm{eV}$ on semilogarithmic plot

E.5 Bose–Einstein Distribution

Executing (E.8) with the approximation $\gamma = \gamma_s$, the Bose–Einstein distribution is found to be

$$\bar{n}_s = \frac{1}{\exp(\beta(\epsilon_{\mathrm{s}} - \mu)) - 1}. \tag{E.24}$$

Appendix F
Kronig–Penney Model

The Kronig–Penney model [64] is a simple, one-dimensional analytically solvable model that visualizes the effect of the periodic potential on the dispersion relation of electrons, i.e. the formation of a band structure.

A one-dimensional periodic hard-wall potential of finite height is assumed (Fig. F.1a). The well width is a, the barrier width b and thus the period $P = a + b$. The potential is zero in the well (regions $(0, a) + nP$) and $+U_0$ in the barrier. The Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{\partial^2 \Psi}{\partial x^2} + U(x)\,\Psi(x) = E\,\Psi(x) \tag{F.1}$$

has to be solved. The solutions for a single hard-wall potential well are well known. In the well, they have oscillatory character, i.e. $\Psi \propto \exp(\mathrm{i}kx)$ with real k. In the barrier, they have exponential character, i.e. $\Psi \propto \exp(kx)$ with real k. Thus we chose

$$\Psi(x) = A\exp(\mathrm{i}Kx) + B\exp(-\mathrm{i}Kx) \tag{F.2a}$$

$$\Psi(x) = C\exp(\kappa x) + D\exp(-\kappa x)\,. \tag{F.2b}$$

The wavefunction from (F.2a) is for the well between 0 and a with $E = \hbar^2K^2/2m$. The wavefunction from (F.2b) is for the barrier between a and $a + b$ with $U_0 - E = \hbar^2\kappa^2/2m$. From the periodicity and Bloch's theorem the wavefunction at $x = -b$ must have the form $\Psi(-b) = \exp(-\mathrm{i}kP)\,\Psi(a)$, i.e. between the two wavefunctions is only a phase factor. The wavevector k of the Bloch function (plane-wave part of the solution) is a new quantity and must be carefully distinguished from K and κ.

Both K and κ are real numbers. As boundary conditions, the continuity of Ψ and Ψ' are used.[4] At $x = 0$ and $x = a$ this yields

[4]Generally, Ψ'/m should be continuous, however, in the present example the mass is assumed constant throughout the structure.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

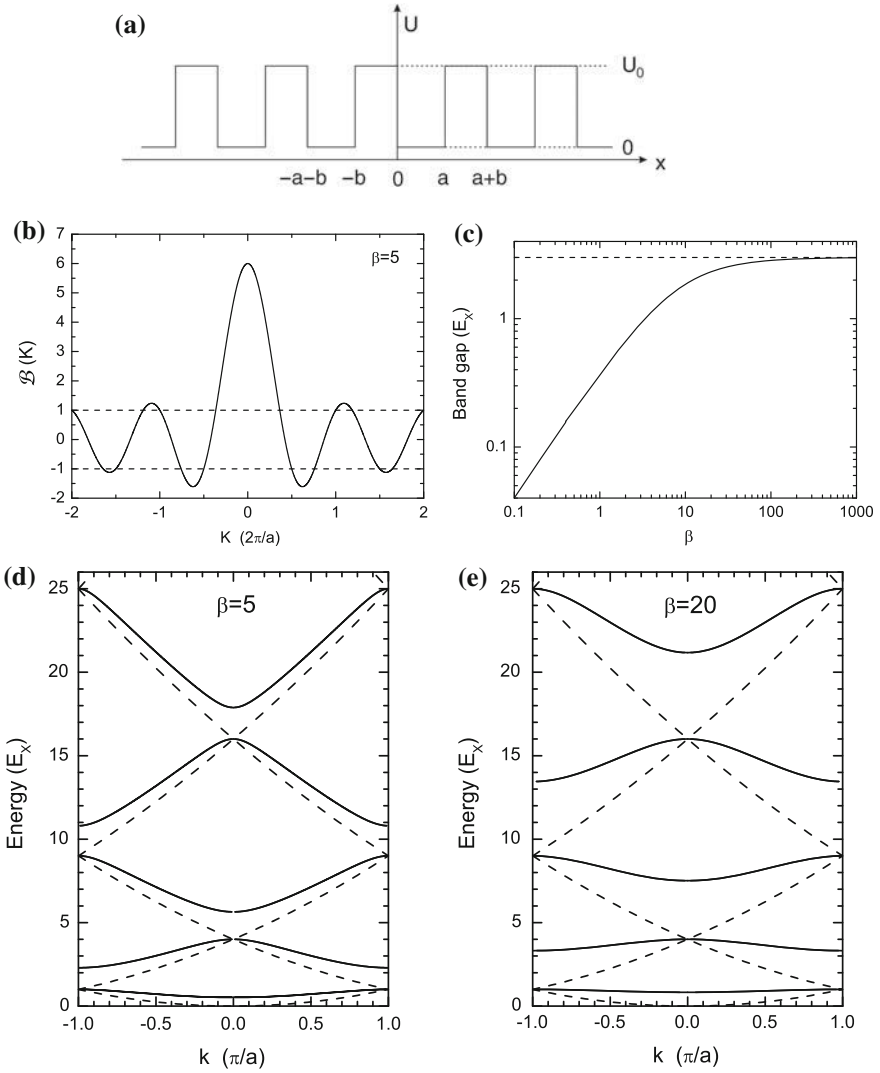


Fig. F.1 (**a**) One-dimensional periodic hard-wall potential (Kronig–Penney model). (**b**) Transcendental function $\mathcal{B}(K)$ from (F.5) for $\beta = 5$. The *dashed lines* indicate the $[-1, 1]$ interval for which solutions exist for (F.5). (**c**) Band gap between first and second subband (in units of $E_X = \hbar^2\pi^2/(2ma^2)$) as a function of β. For smaller β the band gap is $\propto \beta$. For thick barriers ($\beta \to \infty$) the band gap saturates towards $3E_X$ as expected for uncoupled wells. (**d**, **e**) The resulting energy dispersion (in units of E_X) as a function of the superlattice wavevector k for (**d**) $\beta = 5$ and (**e**) $\beta = 20$ in (F.5). The *dashed lines* are the free-electron dispersion (for $\beta = 0$) (see Fig. 6.2a)

$$A + B = C + D \tag{F.3a}$$
$$\mathrm{i}KA - \mathrm{i}KB = \kappa C - \kappa D \tag{F.3b}$$
$$A\exp(\mathrm{i}Ka) + B\exp(-\mathrm{i}Ka) = C\exp(\kappa a) + D\exp(-\kappa a) \tag{F.3c}$$
$$\mathrm{i}KA\exp(\mathrm{i}Ka) - \mathrm{i}KB\exp(-\mathrm{i}Ka) = \kappa C\exp(\kappa a) - \kappa D\exp(-\kappa a). \tag{F.3d}$$

The continuity of Ψ and Ψ' at $x = -b$ is used in the left sides of (F.3c, d).

A nontrivial solution arises only if the determinant of the coefficient matrix is zero. This leads (after some tedious algebra) to

$$\cos(kP) = \left[\frac{\kappa^2 - K^2}{2\kappa K}\right] \sinh(\kappa b)\ \sin(Ka) + \cosh(\kappa b)\ \cos(Ka). \tag{F.4}$$

Further simplification can be reached by letting the barrier thickness $b \to 0$ and $U_0 \to \infty$. Then $P \to a$. The limit, however, is performed in such a way that the barrier 'strength' $U_0 b \propto \kappa^2 b$ remains constant and finite. Equation (F.4) then reads (for $\kappa b \to 0$: $\sinh(\kappa b) \to \kappa b$ and $\coth(\kappa b) \to 1$):

$$\cos(ka) = \beta \frac{\sin(Ka)}{Ka} + \cos(Ka) = \mathcal{B}(K). \tag{F.5}$$

The coupling strength $\beta = \kappa^2 ba/2$ represents the strength of the barrier. Equation (F.5) only has a solution if the right side is in the interval $[-1, 1]$ (Fig. F.1b). The function $\sin(x)/x$ oscillates with decreasing amplitude such that for sufficiently high values of Ka a solution can always be found. The resulting dispersion is shown in Fig. F.1c. The dispersion is different from the free-electron dispersion and has several separated bands. The band gaps are related to the K values, i.e. energies for which (F.5) cannot be fulfilled. At the zone boundary ($k = \pi/a$) the bands are split and the tangent is horizontal ($\mathrm{d}E/\mathrm{d}k = 0$). The form of the dispersion is similar to the arccos function.

For large coupling between the potential wells (small β, $\beta \lesssim 1$) the band gap E_{12} between the first and the second subband at the X-point is $E_{12} = 4\beta/\pi^2 E_\mathrm{X}$ with $E_\mathrm{X} = \hbar^2\pi^2/(2ma^2)$. In this case, the width of the subbands is wide. For small coupling (large β) the band gap E_{12} converges towards $3E_\mathrm{X}$ as expected for decoupled potential wells with energy levels $E_n = E_\mathrm{X} n^2$ and the width of the bands is small.

Appendix G
The k · p Perturbation Theory

The solutions of the Schrödinger equation (cf. Sect. 6.2.1)

$$H\,\Psi_{n\mathbf{k}}(\mathbf{r}) = \left(-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r})\right)\Psi_{n\mathbf{k}}(\mathbf{r}) = E_n(\mathbf{k})\,\Psi_{\mathrm{n}\mathbf{k}}(\mathbf{r}), \tag{G.1}$$

with a lattice periodic potential U, i.e. $U(\mathbf{r}) = U(\mathbf{r}+\mathbf{R})$ for direct lattice vectors $\mathbf{R}$, are Bloch waves of the form

$$\Psi_{n\mathbf{k}}(\mathbf{r}) = \exp(\mathrm{i}\,\mathbf{k}\,\mathbf{r})\,u_{n\mathbf{k}}(\mathbf{r}), \tag{G.2}$$

with the lattice periodic Bloch function $u_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r}+\mathbf{R})$.

Inserting the Bloch wave into (G.1), the following equation is obtained for the periodic Bloch function:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r}) + \frac{\hbar}{m}\mathbf{k}\cdot\mathbf{p}\right)u_{n\mathbf{k}}(\mathbf{r}) = \left(E_n(\mathbf{k}) - \frac{\hbar^2k^2}{2m}\right)u_{\mathrm{n}\mathbf{k}}(\mathbf{r}). \tag{G.3}$$

For simplicity, we assume a band edge $E_n(0)$ at $\mathbf{k} = 0$. In its vicinity, the $\mathbf{k}\cdot\mathbf{p}$ term can be treated as a perturbation. The dispersion for a nondegenerate band[5] is given up to second order in k

$$E_n(\mathbf{k}) = E_n(0) + \sum_{i,j=1}^{3}\left(\frac{\hbar^2}{2m}\,\delta_{ij} + \frac{\hbar^2}{m}\sum_{l\neq n}\frac{p^i_{nl}\,p^j_{ln}}{E_n(0) - E_l(0)}\right)k_i\,k_j, \tag{G.4}$$

with l running over other, so-called *remote* bands. The momentum matrix element is given by $p^i_{nl} = \langle u_{n0}|p_i|u_{l0}\rangle$ (cf. (6.35)). The coefficients in front of the quadratic

[5] Apart from the spin degeneracy.

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

terms are the components of the dimensionless inverse effective-mass tensor (cf. (6.39))

$$\left(\frac{m}{m^*}\right)_{ij} = \delta_{ij} + \frac{2}{m}\sum_{l\neq n}\frac{p^i_{nl}\,p^j_{ln}}{E_n(0) - E_l(0)}. \tag{G.5}$$

For degenerate bands, the $p^i_{nn'}$ vanish when n and n' belong to the degenerate set and also the first-order correction is zero. In the Löwdin perturbation theory [1841], the bands are separated into the close-by degenerate or nearly degenerate bands and the remote bands. The effect of the remote bands is taken into account by an effective perturbation

$$\mathbf{k}\cdot\mathbf{p} + \mathbf{k}\cdot\mathbf{p}\sum_{l\neq n}\frac{|l\rangle\langle l|}{E_n(0) - E_l(0)}\,\mathbf{k}\cdot\mathbf{p}, \tag{G.6}$$

with the index l running over all bands not being in the degenerate set. The dispersion relation is obtained by diagonalization of the Hamiltonian (G.3) in the degenerate basis but with the perturbation given by (G.6).

The spin-orbit interaction [1244] adds an additional term

$$H_{\mathrm{so}} = \frac{\hbar}{4\,m^2\,c^2}\,(\boldsymbol{\sigma}\times\nabla U)\;\mathbf{p} \tag{G.7}$$

to the Hamiltonian, where $\boldsymbol{\sigma}$ are the Pauli spin matrices and c the vacuum speed of light. In the Schrödinger equation for the Bloch functions two new terms arise:

$$\begin{aligned}\Bigg(-\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r}) + \frac{\hbar}{4m^2c^2}\,(\boldsymbol{\sigma}\times\nabla U)\;\mathbf{p}\\ + \frac{\hbar}{m}\mathbf{k}\left[\mathbf{p} + \frac{\hbar}{4m^2c^2}\,(\boldsymbol{\sigma}\times\nabla U)\right]\Bigg)u_{n\mathbf{k}}(\mathbf{r}) = \left(E_n(\mathbf{k}) - \frac{\hbar^2k^2}{2m}\right)u_{n\mathbf{k}}(\mathbf{r}).\end{aligned} \tag{G.8}$$

The linear term in $\mathbf{k}$ is again treated as a perturbation. The first spin-orbit term in (G.8) is lattice periodic, thus the solutions at $\mathbf{k} = 0$ are still periodic Bloch functions, however, different ones from previously. If the band edge is not degenerate, the momentum operator in (G.3) is simply replaced by

$$\boldsymbol{\pi} = \mathbf{p} + \frac{\hbar}{4\,m^2\,c^2}\,(\boldsymbol{\sigma}\times\nabla U)\,, \tag{G.9}$$

and the band edge is still parabolic. For a degenerate band edge, the effect can be more profound, in particular it can lead to the lifting of a degeneracy.

In the 8-band Kane model [441], four bands (lowest conduction band, heavy, light and split-off hole band) are treated explicitly and the others through Löwdin perturbation theory. The basis is chosen to be diagonal in the spin-orbit interaction leaving the spin-orbit interaction Δ_0 as parameter. The band- edge Bloch functions are denoted as $|i\uparrow\rangle$, where the index $i = s, x, y, z$ labels the symmetry of the different

Table G.1 Basis set that diagonalizes the spin-orbit interaction

$\lvert J, m_j\rangle$	Wavefunction	Symmetry
$\lvert\frac{1}{2}, \frac{1}{2}\rangle$	$\mathrm{i}\lvert s\uparrow\rangle$	Γ_6
$\lvert\frac{1}{2}, -\frac{1}{2}\rangle$	$\mathrm{i}\lvert s\downarrow\rangle$	Γ_6
$\lvert\frac{3}{2}, \frac{3}{2}\rangle$	$\frac{1}{\sqrt{2}}\lvert(x+\mathrm{i}y)\uparrow\rangle$	Γ_8
$\lvert\frac{3}{2}, \frac{1}{2}\rangle$	$\frac{1}{\sqrt{6}}\lvert(x+\mathrm{i}y)\downarrow\rangle - \sqrt{\frac{2}{3}}\lvert z\uparrow\rangle$	Γ_8
$\lvert\frac{3}{2}, -\frac{1}{2}\rangle$	$-\frac{1}{\sqrt{6}}\lvert(x-\mathrm{i}y)\uparrow\rangle - \sqrt{\frac{2}{3}}\lvert z\downarrow\rangle$	Γ_8
$\lvert\frac{3}{2}, -\frac{3}{2}\rangle$	$\frac{1}{\sqrt{2}}\lvert(x-\mathrm{i}y)\uparrow\rangle$	Γ_8
$\lvert\frac{1}{2}, \frac{1}{2}\rangle$	$\frac{1}{\sqrt{3}}\lvert(x+\mathrm{i}y)\downarrow\rangle + \sqrt{\frac{1}{3}}\lvert z\uparrow\rangle$	Γ_7
$\lvert\frac{1}{2}, -\frac{1}{2}\rangle$	$-\frac{1}{\sqrt{3}}\lvert(x-\mathrm{i}y)\uparrow\rangle + \sqrt{\frac{1}{3}}\lvert z\downarrow\rangle$	Γ_7

bands. The linear combinations that diagonalize the spin-orbit interaction are given in Table G.1. The band gap and the spin-orbit interaction are given by

$$E_{\mathrm{g}} = E_{\Gamma_6} - E_{\Gamma_8} \tag{G.10a}$$

$$\Delta_0 = E_{\Gamma_8} - E_{\Gamma_7}. \tag{G.10b}$$

The Hamiltonian in the basis states of Table G.1 is given by

$$\begin{bmatrix}
k^2+E_{\mathrm{g}} & 0 & \sqrt{2}Pk_+ & -\sqrt{\frac{2}{3}}Pk_z & -\sqrt{\frac{2}{3}}Pk_- & 0 & \sqrt{\frac{1}{3}}Pk_z & -\sqrt{\frac{4}{3}}Pk_- \\
0 & k^2+E_{\mathrm{g}} & 0 & \sqrt{\frac{2}{3}}Pk_+ & -\sqrt{\frac{2}{3}}Pk_z & \sqrt{2}Pk_- & \sqrt{\frac{4}{3}}Pk_+ & \sqrt{\frac{1}{3}}Pk_z \\
\sqrt{2}Pk_- & 0 & k^2 & 0 & 0 & 0 & 0 & 0 \\
-\sqrt{\frac{2}{3}}Pk_z & \sqrt{\frac{2}{3}}Pk_- & 0 & k^2 & 0 & 0 & 0 & 0 \\
-\sqrt{\frac{2}{3}}Pk_+ & -\sqrt{\frac{2}{3}}Pk_z & 0 & 0 & k^2 & 0 & 0 & 0 \\
0 & \sqrt{2}Pk_+ & 0 & 0 & 0 & k^2 & 0 & 0 \\
\sqrt{\frac{1}{3}}Pk_z & \sqrt{\frac{4}{3}}Pk_- & 0 & 0 & 0 & 0 & k^2-\Delta_0 & 0 \\
-\sqrt{\frac{4}{3}}Pk_+ & \sqrt{\frac{1}{3}}Pk_z & 0 & 0 & 0 & 0 & 0 & k^2-\Delta_0
\end{bmatrix} \tag{G.11}$$

with the energy measured from the valence-band edge in units of $\hbar^2/(2m)$ and

$$\frac{1}{2}\,\mathrm{i}\,\hbar P = \langle s|\pi_x|x\rangle = \langle s|\pi_y|y\rangle = \langle s|\pi_z|z\rangle \tag{G.12a}$$

$$k_\pm = k_x \pm \mathrm{i}k_y. \tag{G.12b}$$

The inclusion of remote bands renormalizes the above Hamiltonian to

$$\begin{bmatrix}
Dk^2+E_\mathrm{g} & 0 & \sqrt{2}Pk_+ & -\sqrt{\frac{2}{3}}Pk_z & -\sqrt{\frac{2}{3}}Pk_- & 0 & \sqrt{\frac{1}{3}}Pk_z & -\sqrt{\frac{4}{3}}Pk_- \\
0 & Dk^2+E_\mathrm{g} & 0 & \sqrt{\frac{2}{3}}Pk_+ & -\sqrt{\frac{2}{3}}Pk_z & \sqrt{2}Pk_- & \sqrt{\frac{4}{3}}Pk_+ & \sqrt{\frac{1}{3}}Pk_z \\
\sqrt{2}Pk_- & 0 & H_\mathrm{h} & R & S & 0 & \frac{\mathrm{i}}{\sqrt{2}}R & -\mathrm{i}\sqrt{2}S \\
-\sqrt{\frac{2}{3}}Pk_z & \sqrt{\frac{2}{3}}Pk_- & R^* & H_\mathrm{l} & 0 & S & \frac{H_\mathrm{h}-H_\mathrm{l}}{\sqrt{2}\mathrm{i}} & \mathrm{i}\sqrt{\frac{3}{2}}R \\
-\sqrt{\frac{2}{3}}Pk_+ & -\sqrt{\frac{2}{3}}Pk_z & S^* & 0 & H_\mathrm{l} & -R & -\mathrm{i}\sqrt{\frac{3}{2}}R^* & \frac{H_\mathrm{h}-H_\mathrm{l}}{\sqrt{2}\mathrm{i}} \\
0 & \sqrt{2}Pk_+ & 0 & S^* & -R^* & H_\mathrm{h} & -\mathrm{i}\sqrt{2}S^* & -\frac{\mathrm{i}}{\sqrt{2}}R^* \\
\sqrt{\frac{1}{3}}Pk_z & \sqrt{\frac{4}{3}}Pk_- & -\frac{\mathrm{i}}{\sqrt{2}}R^* & -\frac{H_\mathrm{h}-H_\mathrm{l}}{\sqrt{2}\mathrm{i}} & \mathrm{i}\sqrt{\frac{3}{2}}R & \mathrm{i}\sqrt{2}S & \frac{H_\mathrm{h}+H_\mathrm{l}}{\sqrt{2}}-\Delta_0 & 0 \\
-\sqrt{\frac{4}{3}}Pk_+ & \sqrt{\frac{1}{3}}Pk_z & \mathrm{i}\sqrt{2}S^* & -\mathrm{i}\sqrt{\frac{3}{2}}R^* & -\frac{H_\mathrm{h}-H_\mathrm{l}}{\sqrt{2}\mathrm{i}} & \frac{\mathrm{i}}{\sqrt{2}}R & 0 & \frac{H_\mathrm{h}+H_\mathrm{l}}{\sqrt{2}}-\Delta_0
\end{bmatrix} \tag{G.13}$$

with

$$D = 1 + \frac{2}{m}\sum_{l\neq n}\frac{|\langle s|\pi_x|l\rangle|^2}{E_\mathrm{g}-E_l(0)} \tag{G.14a}$$

$$\gamma_1' = \left[1 + \frac{2}{m}\sum_{l\neq n}\frac{|p_{xl}^x|^2}{E_n(0)-E_l(0)}\right] - \frac{2P^2}{3E_\mathrm{g}} \tag{G.14b}$$

$$\gamma_2' = \left[1 + \frac{2}{m}\sum_{l\neq n}\frac{|p_{xl}^y|^2}{E_n(0)-E_l(0)}\right] - \frac{P^2}{3E_\mathrm{g}} \tag{G.14c}$$

$$\gamma_3' = \left[\frac{2}{m}\sum_{l\neq n}\frac{p_{xl}^x p_{ly}^y + p_{xl}^y p_{ly}^x}{E_n(0)-E_l(0)}\right] - \frac{P^2}{3E_\mathrm{g}} \tag{G.14d}$$

$$H_\mathrm{h} = (\gamma_1'+\gamma_2')(k_x^2+k_y^2) + (\gamma_1'-2\gamma_2')k_z^2 \tag{G.14e}$$

$$H_\mathrm{l} = (\gamma_1'-\gamma_2')(k_x^2+k_y^2) + (\gamma_1'+2\gamma_2')k_z^2 \tag{G.14f}$$

$$R = -2\sqrt{3}\gamma_3' k_- k_z \tag{G.14g}$$

$$S = \sqrt{3}\gamma_2'(k_x^2-k_y^2) + 2\sqrt{3}\gamma_3'\mathrm{i}k_x k_y. \tag{G.14h}$$

The Hamiltonian in the presence of inhomogeneous strain is given in [463]. The hole bands decouple from the conduction band for $E_\mathrm{g}\to\infty$ (six-band model [1082]). The heavy and light holes can be treated separately for $\Delta_0\to\infty$ (Luttinger Hamiltonian). For the Γ_8 states, the Hamiltonian is then given by

$$\begin{bmatrix}
H_\mathrm{h} & R & S & 0 \\
R^* & H_\mathrm{l} & 0 & S \\
S^* & 0 & H_\mathrm{l} & -R \\
0 & S^* & -R^* & H_\mathrm{h}
\end{bmatrix}. \tag{G.15}$$

Appendix H
Effective-Mass Theory

The effective-mass theory or approximation (EMA), also termed the envelope function approximation, is widely used for calculating the electronic properties of carriers in potentials in an otherwise periodic crystal. The strength of the method is that the complexities of the periodic potential are hidden in the effective-mass tensor m^*_{ij}. The effective-mass theory is a useful approximation for the treatment of shallow impurities (Sect. 7.5) or quantum wells (Sect. 12.3.2) with a potential that is slowly varying with respect to the scale of the lattice constant.

For the lattice-periodic potential, the Schrödinger equation

$$H_0\,\Psi_{n\mathbf{k}} = E_n(\mathbf{k})\,\Psi_{n\mathbf{k}} \tag{H.1}$$

is solved by the Bloch wave $\Psi_{n\mathbf{k}}$. With a perturbing potential V, the Schrödinger equation reads

$$(H_0 + V)\;\Psi_{n\mathbf{k}} = E_n(\mathbf{k})\,\Psi_{n\mathbf{k}}. \tag{H.2}$$

According to Wannier's theorem [1842], the solution is approximated by the solution of the

$$(E_n(-\mathrm{i}\nabla) + V)\,\Phi_n = E\,\Phi_n. \tag{H.3}$$

The dispersion relation is expanded to second order as described in Appendix G. The function Φ_n is termed the *envelope function* since it varies slowly compared to the lattice constant and the exact wavefunction is approximated (in lowest order) by

$$\Psi(\mathbf{r}) = \Phi_n(\mathbf{r})\;\exp(\mathrm{i}\,\mathbf{k}\,\mathbf{r})\,u_{n0}(\mathbf{r}). \tag{H.4}$$

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

Appendix I
Boltzmann Transport Theory

I.1 Boltzmann Transport Equation

The Boltzmann treatment of transport in semiconductors goes beyond the relaxation time approximation (cmp. Sect.sec:cond) and contains this approach as its simplest approximation. The distribution function of carriers $f(\mathbf{r}, \mathbf{p}, t)$ is considered with regard to their momentum $\mathbf{p} = (p_x, p_y, p_z)$, their position $(r) = (x, y, z)$ and time t. Via the dispersion relation(s) the momentum distribution also determines the energy distribution.

In thermodynamical equilibrium, the distribution function shall be termed $f_0(\mathbf{p})$. In a homogeneous semiconductor it should be independent of $\mathbf{r}$, not depend explicitly on time and the momentum distribution be such that the resulting energy distribution should match the Fermi-Dirac distribution.

In non-equilibrium, the flow of electrons and heat is determined by the external forces $\mathbf{F}$ (electrical and magnetic fields) and the scattering of charge carriers via various processes (termed here collisions). In a (non-equilibrium) steady-state situation with constant forces, the distribution function f is constant in time; thus in a given time interval δt the change δf is zero,

$$\frac{\delta f}{\delta t} = 0. \tag{I.1}$$

Within the time interval δt the momenta change as $\mathbf{p} \to \mathbf{p} + \mathbf{F}\,\delta t$ and the coordinates as $\mathbf{r} \to \mathbf{r} + \mathbf{p}/m^*\,\delta t$. We assume here for simplicity an isotropic mass and also the particle energy given by $E = \mathbf{p}^2/(2\,m^*)$. The condition (I.1) written in partial derivatives reads

$$\left(\frac{\partial}{\partial t} + \frac{1}{m^*}\,\mathbf{p}\cdot\nabla_{\mathbf{r}} + \mathbf{F}\cdot\nabla_{\mathbf{p}}\right) f(\mathbf{p}, \mathbf{r}, t) = 0. \tag{I.2}$$

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

The force may be taken as the Lorentz force. So far no collisions have been considered. Without giving an explicit form for the microscopic details of the collisions, the change of the distribution function due to collisions is written as

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}}. \tag{I.3}$$

Assuming that only two-particle collisions play a role, sample boundaries play no role and that position and velocity of particles are uncorrelated, the collision term can be written as

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \iiint [f(\mathbf{p}', \mathbf{r}, t)\, P(\mathbf{p}', \mathbf{p}) - f(\mathbf{p}, \mathbf{r}, t)\, P(\mathbf{p}, \mathbf{p}')]\, \mathrm{d}\mathbf{p}'. \tag{I.4}$$

with $P(\mathbf{p}, \mathbf{p}')$ being the transition probability per time that a momentum $\mathbf{p}$ is changed into $\mathbf{p}'$ by collisions. The collision integral must be calculated explicitly using microscopic and eventually quantum mechanical models. This leads now to the Boltzmann transport equation

$$\left(\frac{\partial}{\partial t} + \frac{1}{m^*}\mathbf{p} \cdot \nabla_{\mathbf{r}} + \mathbf{F} \cdot \nabla_{\mathbf{p}}\right) f(\mathbf{p}, \mathbf{r}, t) = \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}. \tag{I.5}$$

Under certain circumstances, the collision term can be effectively written as (for a homogeneous semiconductor and homogeneous fields, neglecting the spatial dependence of f)

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{f(\mathbf{p}) - f_0}{\tau(\mathbf{p})}. \tag{I.6}$$

Compared to the relaxation time approximation, the major difference on the level of (I.6) here is the consideration of the momentum (and energy) dependence of the distribution function and the relaxation time.

I.2 Conductivity

In thermodynamical equilibrium the number of electronic states per unit volume associated with an element $\mathrm{d}\mathbf{p} = \mathrm{d}p_x\, \mathrm{d}p_y\, \mathrm{d}p_z$, including spin degeneracy of 2 is

$$\frac{2}{h^3} f_0(\mathbf{p})\, \mathrm{d}\mathbf{p}. \tag{I.7}$$

In the presence of an electric field $\mathbf{E}$, which we assume here in x-direction, a steady-state current will arise and the number of electronic states changes to

$$\frac{2}{h^3} f(\mathbf{p})\, \mathrm{d}\mathbf{p}, \tag{I.8}$$

making the (electron) current density (along x-direction)

$$j_x = -\frac{2\,e}{h^3} \iiint v_x \,[f(\mathbf{p}) - f_0(\mathbf{p})]\,\mathrm{d}\mathbf{p}. \tag{I.9}$$

This is a generalization of (8.4). The Boltzmann transport equation (I.5) with (I.6) simplifies to

$$-\frac{f(\mathbf{p}) - f_0}{\tau(\mathbf{p})} = -e\,E_x\,\frac{\partial f}{\partial p_x} \approx -e\,E_x\,\frac{\partial f_0}{\partial p_x}. \tag{I.10}$$

The last approximation is valid for small fields and makes j_x proportional to E_x (ohmic regime). The derivative with respect to p_x is converted to a derivative with respect to energy, yielding

$$\frac{f(\mathbf{p}) - f_0}{\tau(\mathbf{p})} = e\,v_x\,E_x\,\frac{\partial f_0}{\partial E}. \tag{I.11}$$

We note that for the Fermi–Dirac distribution (E.22) $f_0(E)$:

$$\frac{\partial f_0}{\partial E} = -\frac{1}{kT} f_0\,[1 - f_0]\,, \tag{I.12}$$

and in the case of a non-degenerate semiconductor (Boltzmann approximation), the right-hand side simplyfies to

$$\frac{\partial f_0}{\partial E} \approx -\frac{1}{kT} f_0 = -\frac{1}{kT}\,\exp\left(-\frac{E - E_\mathrm{F}}{kT}\right), \tag{I.13}$$

Now the current density is given as

$$j_x = -\frac{2\,e^2}{h^3}\,E_x \iiint v_x^2\,\tau(\mathbf{p})\,\frac{\partial f_0}{\partial E}\,\mathrm{d}\mathbf{p}. \tag{I.14}$$

If we assume that τ depends only on the momentum and not its direction,[6] and replace v_x^2 by $v^2/3$ assuming isotropy, the integral reads[7]

$$j_x = -\frac{8\pi\,e^2}{3\,h^3}\,E_x \int_0^\infty v^2\,\tau(p)\,\frac{\partial f_0}{\partial E}\,p^2\,\mathrm{d}p. \tag{I.15}$$

The quantity $8\pi\,p^2\mathrm{d}p\,f_0/h^3$ (cmp. I.7) denotes the number $\mathrm{d}n$ of electrons with momentum in the range $\mathrm{d}p$. Thus the integral can also be written as (in Boltzmann approximation)

[6] This might be incorrect e.g. for piezoelectric scattering.

[7] Using $\mathrm{d}\mathbf{p} = 4\pi\,p^2\,\mathrm{d}p$.

$$j_x = \frac{e^2}{3\,kT}\,E_x \int_0^\infty v^2\,\tau\,\mathrm{d}n. \tag{I.16}$$

Denoting the average of a quantity a over the electron distribution with $\langle a\rangle$ according to

$$\langle a\rangle = \frac{\int a\,\mathrm{d}n}{n}, \tag{I.17}$$

the equation (I.16) can be written as

$$j_x = \frac{n\,e^2}{3\,kT}\,E_x\,\langle v^2\tau\rangle. \tag{I.18}$$

Using $m^*\langle v^2\rangle = 3\,kT$, we thus have obtained

$$\sigma = \frac{n\,e^2}{m^*}\,\frac{\langle v^2\tau\rangle}{\langle v^2\rangle}, \tag{I.19}$$

and with $\sigma = n\,(-e)\,\mu$ (for electrons), the mobility

$$\mu = -\frac{e}{m^*}\,\frac{\langle v^2\tau\rangle}{\langle v^2\rangle}. \tag{I.20}$$

For degenerate semiconductors, similar as for metals, the derivative of f_0 in (I.15) has a significant value only in the few-kT vicinity of the Fermi level. In an approximation we can evaluate the integral by replacing $E^{3/2}$ and τ by their values at the Fermi level and find[8] (using (6.66))

$$\sigma = \frac{j_x}{E_x} = \frac{n\,e^2\,\tau_\mathrm{F}}{m^*}. \tag{I.21}$$

Starting again with (I.15), using the density of states (6.67) in the form (per volume)

$$D(E) = m^*\,\frac{8\,\pi}{h^3}\,\sqrt{2\,m^*E}, \tag{I.22}$$

and $\mathrm{d}p/\mathrm{d}E = \sqrt{2m^*/E}$ we write

$$j_x = -\frac{e^2}{3}\,E_x \int_0^\infty D(E)\,v^2\,\tau(E)\,\frac{\partial f_0}{\partial E}\,\mathrm{d}E. \tag{I.23}$$

[8] $\int_0^\infty \frac{\partial f_0}{\partial E}\,\mathrm{d}E = -1 + [1 + \exp(E_\mathrm{F}/kT)]^{-1} \approx -1$ for $E_\mathrm{F} \gg kT$.

Using an energy-dependent mobility, in the spirit of (I.20) defined as

$$\mu(E) = -e\,\frac{v^2\,\tau(E)}{3\,kT}, \tag{I.24}$$

the conductivity can be written in a generalized form integrating over single electron states [1843] (neglecting correlation effects):

$$\begin{aligned}\sigma &= e\int D(E)\,\mu(E)\,kT\,\frac{\partial f_0}{\partial E}\,\mathrm{d}E \\ &= -e\int D(E)\,\mu(E) f_0(E)\,[1-f_0(E)]\,\mathrm{d}E.\end{aligned} \tag{I.25}$$

I.3 Hall Effect

Treating the Hall effect with the Boltzmann transport equation and making the assumptions of isotropy, one obtains (cmp. (13.12) and (13.21))

$$R_\mathrm{H} = \frac{1}{q\,n}\,\frac{\langle v^2\tau^2\rangle\,\langle v^2\rangle}{\langle v^2\tau\rangle^2}. \tag{I.26}$$

The Hall mobility determined from the Hall coefficient is

$$\mu_\mathrm{H} = \sigma\,R_\mathrm{H} = \frac{e}{m^*}\,\frac{\langle v^2\tau^2\rangle}{\langle v^2\tau\rangle^2}, \tag{I.27}$$

and thus different from the field mobility (I.20).

I.4 Thermopower

The electronic energy transported per electron is $E - E_\mathrm{F}$. Writing (I.25) as $\sigma = \int \sigma(E)\,\mathrm{d}E$, the weighing factor for electrons at energy E contributing to conduction is $\sigma(E)\,\mathrm{d}E/\sigma$. Therefore the Seebeck coefficient (thermopower) can be written [721]

$$S = -\frac{k}{e}\int\left(\frac{E-E_\mathrm{F}}{kT}\right)\frac{\sigma(E)}{\sigma}\,\mathrm{d}E. \tag{I.28}$$

or

$$S = -\frac{k}{e}\,\frac{\int D(E)\,\mu(E)\,[(E-E_\mathrm{F})/kT]f\,(1-f)\,\mathrm{d}E}{\int D(E)\,\mu(E) f\,(1-f)\,\mathrm{d}E}. \tag{I.29}$$

For band conduction the thermopower is obtained by integrating (I.29) for electrons (S_n) and holes (S_p) (using the Boltzmann approximation) as [721]

$$S_\mathrm{n} = -\frac{k}{e}\left(\frac{E_\mathrm{C} - E_\mathrm{F}}{kT} + A_\mathrm{C}\right) \tag{I.30a}$$

$$S_\mathrm{p} = \frac{k}{e}\left(\frac{E_\mathrm{F} - E_\mathrm{V}}{kT} + A_\mathrm{V}\right), \tag{I.30b}$$

where A_i are constants depending on the energy dependence of the density of states and the mobility,

$$A_\mathrm{C} = \frac{\int_0^\infty (E'/kT)\,\sigma(E')\,\mathrm{d}E'}{\int_0^\infty \sigma(E')\,\mathrm{d}E'}, E' = E - E_\mathrm{C} \tag{I.31a}$$

$$A_\mathrm{V} = \frac{\int_{-\infty}^0 (E'/kT)\,\sigma(E')\,\mathrm{d}E'}{\int_{-\infty}^0 \sigma(E')\,\mathrm{d}E'}, E' = E_\mathrm{V} - E. \tag{I.31b}$$

If the product of the density of states and the mobility $D\mu$ depends on the energy like E^γ, the constant is $A = 1 + \gamma$ (for $\gamma > -1$). For a parabolic band ($D \propto E^{1/2}$) and acoustic deformation potential scattering $\mu \propto E^{-1/2}$ (Sect. 8.3.4), A=1; for moderate ionized impurity scattering $\mu \propto E^{3/2}$ (Sect. 8.3.3) and $A = 3$.

For two-band conduction, when electrons *and* holes contribute to transport,

$$S = \frac{S_\mathrm{n}\,\sigma_\mathrm{n} + S_\mathrm{p}\,\sigma_\mathrm{p}}{\sigma_\mathrm{n} + \sigma_\mathrm{p}}. \tag{I.32}$$

At low temperatures the interaction of the phonon flow with the current via electron-phonon scattering (phonon-drag effect) leads to an increase of thermopower [723, 1844–1846].

Appendix J
Noise

Noise is a general phenomenon effecting every measurement process and the performance of semiconductor devices [1847–1852]. Eventually, always a signal-to-noise ratio is measured instead of a 'signal'. Electrical noise fundamentally limits the sensitivity and resolution of communication, navigation, measurement, and other electronic systems [1850].

Behind the fluctuating signal stand microscopic classical and quantum mechanical processes that inherently contain randomness. From the physical standpoint, seemingly constant physical quantities even in thermodynamical equilibrium such as the free carrier density or the density of carriers on a trap are subject to fluctuations, e.g. leading to generation-recombination noise. Also the random motion of carriers, in equilibrium without net charge transport, leads to fluctuations, e.g. thermal noise on a resistor.

In this appendix necessary definitions, some mathematical basics and simple physical examples regarding noise are given.

J.1 Fluctuating Signals

The noisy signal under consideration can be 'analog', for example in the case of a fluctuating current, voltage or power or it can be 'digital' for example a photon count rate.

Let $A(t)$ be an analog signal that fluctuates in time. Even under constant experimental conditions, it will fluctuate due to possibly many reasons, at least due to thermal fluctuations. We note that another, identically set-up experiment will have another signal $B(t)$. The time average (of first moment) of the signal within a time interval $2T$ (symmetric around $t = 0$) is defined as

$$\langle A \rangle_T = \frac{1}{2\,T} \int_{-T}^{T} A(t)\,\mathrm{d}t. \tag{J.1}$$

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

The time average of the signal $\langle A\rangle$ in general is the limit for large times,

$$\langle A\rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} A(t)\,\mathrm{d}t. \tag{J.2}$$

Two identical experiments will (should) share the same limits, i.e. $\langle A\rangle = \langle B\rangle$. The fluctuation or noise of A is defined as $a(t)$ via

$$a(t) = A(t) - \langle A\rangle, \tag{J.3}$$

thus evidently $\langle a\rangle = 0$. For an identical but different experiment, $a(t) \neq b(t)$ as stated before.

The *variance* σ^2 (or second moment) of the signal is the average of the squared fluctuation,

$$\sigma^2 = \langle a^2\rangle = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} a(t)^2\,\mathrm{d}t = \langle A^2\rangle - \langle A\rangle^2. \tag{J.4}$$

The *effective value* of the noise quantity a is the square root of the variance, also termed the 'root mean square' (or rms-value),

$$\sigma = \langle a^2\rangle^{1/2} = \sqrt{\langle A^2\rangle - \langle A\rangle^2}. \tag{J.5}$$

The quantity σ^2 is a measure of the noise power, where $\langle A\rangle^2$ is a measure of the dc power.[9]

In a measurement procedure, the noise of a signal can be reduced by integrating or averaging over time; however, the time for a specific measurement is always finite and maybe constricted by many conditions. Given a fixed (finite) averaging time of T_0, the measured signals A_{T_0} in a series of such subsequent identical measurements will still exhibit a fluctuation. How large this remaining fluctuation is depends on the choice of T_0 and the noise spectrum discussed below.

In the case of a digital signal, e.g. the count rate of a photomultiplier or from a scintillator, the signal consists of (integer) numbers $N(t_i)$ aggregated at times t_i, $i = 0, 1, \ldots, m$. The average is then defined as

$$\langle N\rangle = \lim_{m\to\infty} \frac{1}{m} \sum_{i=0}^{m} N(t_i). \tag{J.6}$$

[9]Imagine a fluctuating current $I(t) = \langle I\rangle + i(t)$ leading to Joule heating ($\propto I^2$) at a resistor. Comparing the heating from I and $\langle I\rangle = \langle I\rangle$ (from a low noise current source) can yield the noise power. Also $\langle i^2\rangle$ could be determined by first compensating I with $\langle I\rangle$ (from a low noise current source) and then measuring the temperature increase at the resistor.

The definition of the variance and rms are analog to this definition. A well known result for photon counting, based on the Poisson statistics of classical light is $\sigma^2 = \langle N \rangle = \bar{N}$.

J.2 Correlations

If a measurable quantity is subject to two fluctuating quantities $a_1(t)$ and $a_2(t)$, the time average of $a_1 + a_2$ is

$$\langle (a_1 + a_2)^2 \rangle = \langle a_1^2 \rangle + \langle a_2^2 \rangle + 2\langle a_1\, a_2 \rangle. \tag{J.7}$$

The third term is the decisive one; the correlation coefficient of noise quantities a_1 and a_2 is defined as

$$c_{12} = \frac{\langle a_1\, a_2 \rangle}{\sqrt{a_1^2\, a_2^2}} = \frac{\langle a_1\, a_2 \rangle}{\sigma_1\, \sigma_2}. \tag{J.8}$$

If the two noise quantities are independent of each other they are termed *uncorrelated* and $c_{12} = 0$. In the following it will become clear that this is a necessary but not sufficient condition for two noise sources to be uncorrelated. In the case $c_{12} = 0$, the noise powers of the two processes are simply added,

$$\langle (a_1 + a_2)^2 \rangle = \langle a_1^2 \rangle + \langle a_2^2 \rangle. \tag{J.9}$$

This concept can be generalized to several noise sources.

A more general concept to determine correlation of two functions a_1 and a_2 is the *cross correlation* function, defined by

$$\rho_{12}(\tau) = \langle a_1(t)\, a_2(t + \tau) \rangle, \tag{J.10}$$

which is the average of function a_1 and time-shifted function a_2. Often $t = 0$ is used when the nature of the fluctuations does not change with time. An important time-shift is $\tau = 0$, and it follows that

$$c_{12} = \frac{\rho_{12}(0)}{\sigma_1\, \sigma_2}, \tag{J.11}$$

Two noise quantities a_1 and a_2 are uncorrelated if $\rho_{12}(\tau) = 0$ holds for all times τ; thus $c_{12} = 0$ is a special but important case.[10]

[10] A simply example of correlated noise sources with $c_{12} = 0$ are the voltages at a resistor and a capacitance in series; the fact that they are 90° out of phase makes $c_{12} = 0$ although the fluctuations of the voltages, due to fluctuations of the driving current, are obviously correlated.

If a_1 and a_2 are the same function, i.e. $a = a_1 = a_2$, (J.10) becomes the *auto correlation* function,

$$\rho(\tau) = \langle a(t)\, a(t+\tau)\rangle, \tag{J.12}$$

In stationary processes the auto correlation function must be symmetric with regard to τ,

$$\rho(\tau) = \rho(-\tau). \tag{J.13}$$

The value at $\tau = 0$ is

$$\rho(0) = \langle a^2\rangle = \sigma^2. \tag{J.14}$$

Typically, $\rho(\tau \to \infty) = 0$ in a statistic (non-repetitive) process. For uncorrelated processes, the auto correlation function of the sum, is the sum of the individual auto correlation functions.

J.3 Noise spectrum

Since the function $a(t)$ is not known, the noise spectrum cannot be calculated from its Fourier transform. However, this is also unnecessary since we are not interested in the Fourier transform of a itself but rather the spectral power density for a given frequency $W(\nu)$, with

$$\int_0^\infty W(\nu)\,\mathrm{d}\nu = \langle a^2\rangle. \tag{J.15}$$

Since the quantity $\langle a^2\rangle$ is finite and the spectral power density $W(\nu)$ is positive, for high frequencies, $W(\nu)$ must decrease to zero. Starting from the auto correlation function ρ, its Fourier transform shall be denoted w,

$$w(\nu) = \int_{-\infty}^\infty \rho(\tau)\ \exp(-2\pi\,\mathrm{i}\,\nu\,\tau)\,\mathrm{d}\tau. \tag{J.16}$$

Also,

$$\rho(\tau) = \int_{-\infty}^\infty w(\nu)\ \exp(2\pi\,\mathrm{i}\,\nu\,\tau)\,\mathrm{d}\nu. \tag{J.17}$$

Using $\tau = 0$ in this equation, we have obtained an equation similar to (J.15). With (J.12) $w(\nu)$ can be identified as a spectral power density. Due to (J.13), w is a real and even function and we find for the noise power spectrum W in (J.15) $W = 2\,w$ (Wiener-Khintchine theorem),

$$W(\nu) = 2\int_{-\infty}^\infty \rho(\tau)\ \exp(-2\pi\,\mathrm{i}\,\nu\,\tau)\,\mathrm{d}\tau = 4\int_0^\infty \rho(\tau)\ \cos(-2\pi\,\nu\,\tau)\,\mathrm{d}\tau. \tag{J.18}$$

The noise power is practically measured in a finite frequency range, often in a narrow band of width B (with varying central frequency). If the frequency dependence of W can be neglected within B around the frequency ν_0, the variance is given by

$$\langle a^2 \rangle(\nu_0, B) = \int_{\nu_0 - B/2}^{\nu_0 + B/2} W(\nu)\, \mathrm{d}\nu \approx W(\nu_0)\, B. \tag{J.19}$$

Typical noise mechanisms and spectra are discussed in the following sections.

J.3.1 Thermal Noise

Finite temperature induces random motion of particles, e.g. as known from the theory of ideal gases and diffusion. In the case of charge carriers such motions lead to fluctuations of current or at a resistor to fluctuation of voltage. This happens also in the case of zero bias (no external fields). Such 'thermal noise' at a resistor was experimentally found by Johnson [1529, 1530] and theoretically derived by Nyquist [1531].

Using the general result from Langevin theory of motion under a fluctuating force, the mobility[11] is given as

$$\mu(\omega) = \frac{e}{kT} \int_0^\infty \langle v(t)\, v(0) \rangle \, \exp(\mathrm{i}\omega t)\, \mathrm{d}t. \tag{J.20}$$

Now we restrict ourselves to times much longer than the relaxation time constant, and subsequently to frequencies much smaller than $1/\tau$. In this case the conductivity $\sigma(\omega) = e\, n\, \mu(\omega)$ does not depend on frequency and can be taken as its low frequency limit σ_0 (cmp. Sect. 8.5). In a conductor (resistor) of length L and cross section A shall be N electrons ($n = N\,(A\,L)$). With the electron velocities $v_i(t)$, the current is

$$I(t) = \frac{e}{L} \sum_i v_i(t). \tag{J.21}$$

Without external field, $\langle v_i(t) \rangle = 0$ and $\langle I(t) \rangle = 0$ and we name this fluctuating current $i(t)$. If all electrons move independently of each other,

$$\langle i(\tau)\, i(0) \rangle = N\, \frac{e^2}{L^2}\, \langle v(\tau)\, v(0) \rangle. \tag{J.22}$$

The power spectrum of $i(t)$ is according to (J.18),

[11] In the Langevin theory the mobility is the ratio of velocity v and the force K, here the mobility is the ratio of v and the field E with $K = -e\,E$.

$$W(\omega) = 2\int_{-\infty}^{\infty} \langle i(\tau)\, i(0)\rangle \, \exp(\mathrm{i}\omega\tau)\, \mathrm{d}\tau = 4N\,\frac{e^2}{L^2}\int_0^{\infty} \langle v(\tau)\, v(0)\rangle \, \exp(\mathrm{i}\omega\tau)\, \mathrm{d}\tau$$
$$= 4\,\frac{N\,e^2}{L^2}\,\frac{\mu\, kT}{e} = 4\,\frac{N\,e^2}{L^2}\,\frac{\sigma_0\, kT\, A\, L}{e^2\, N} = 4\,\sigma_0\,\frac{A}{L}\,kT. \tag{J.23}$$

Then, using the conductance $G = R^{-1} = \sigma_0\, A/L$, we find the frequency independent spectral power

$$W = 4\,kT\,G. \tag{J.24}$$

Therefore the fluctuation of the current induced by the thermal motion is

$$\langle i^2\rangle = 4\,kT\,G\,B, \tag{J.25}$$

and the variance of the fluctuating voltage at a resistor with resistance R in a frequency range B is ($i = u/R$)

$$\langle u^2\rangle = 4\,kT\,R\,B. \tag{J.26}$$

At room temperature ($T_0 = 293\,\mathrm{K}$), the quantity $k\,T_0$ is about 26 meV; in the context here, the unit W s = W/Hz is the appropriate one, and $k\,T_0 = 4.04 \times 10^{-21}$ W/Hz. This represents a fundamental limit to noise in devices. Since the power density is independent of frequency, this noise is 'white' noise. The formulas (J.26) and (J.25) are valid for frequencies $h\,\nu \ll kT$; for larger frequencies the quantum nature of electromagnetic radiation and photon statistics play a role. For practical purposes even cooled devices at $T = 4\,\mathrm{K}$ fulfill the limit condition for frequencies up in the 100 GHz regime. In the cases of heated electron (or hole) gases (cmp. Fig. 10.3), the lattice temperature must be replaced by the temperature of the carrier gas.

For a RC low pass, the power spectrum $W_i = 4\,kT\,G$ at the resistor is converted using $u^2 = |Z|^2\, i^2$ to $W_u = 4\,kT\,R/[1 + (\omega\, R\, C)^2]$.

J.3.2 1/f Noise

For many processes a frequency dependent noise spectral power following a ν^{α}-law is found with α close to -1. Such noise is termed 'pink noise', $1/f$-noise or Flicker noise. The microscopic reasons for such behavior can be manifold and various models have been proposed [1853, 1854]. As an example the noise spectrum of a RuO_2 thick film resistor is depicted in Fig. J.1a; for this system, the fluctuation of tunneling current in metal-insulator-metal units was used to explain the observed frequency (and temperature) dependence of the $1/f$-noise. At high frequencies, the $1/f$ spectral power vanishes and other noise sources such as thermal noise dominate, as depicted in Fig. J.1b for an a-Si thin film transistor. The $1/f$-dependence of the noise spectral power (of a carbon sheet resistor) has been detected for frequencies down to 3×10^{-6} Hz [1855].

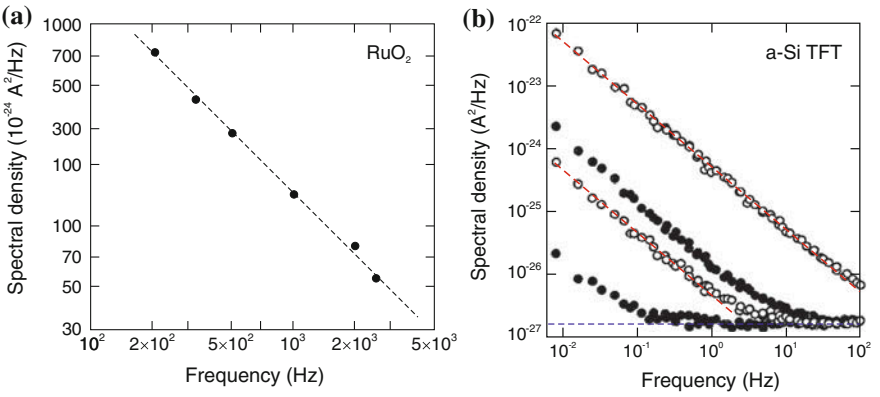


Fig. J.1 (**a**) Noise current density spectrum of a ruthenium oxide resistor (at $T = 300\,\mathrm{K}$ and current of $I = 1\,\mathrm{mA}$), experimental data (*symbols*) and $1/f$-dependency (*dashed line*). Adapted from [1856]. (**b**) Noise current density spectrum of an amorphous silicon thin film transistor, experimental data (*symbols*) for various source-drain voltages, thermal noise (*horizontal blue dashed line*) and $1/f$-dependency (*red dashed lines*). Adapted from [1857]

J.3.3 Shot Noise

A dc current $\langle i \rangle = I_0$ through a resistor is a sequence of electron transfers from one contact to the other. The transit time t_{tr} is given by the length L and the drift velocity v_{D} as $t_{\mathrm{tr}} = L/v_{\mathrm{D}} = L^2/(\mu\, V)$. The event times of these transits are random and thus lead to a noise (ac) component on top of the dc current. This is termed 'shot' noise, after the crackling arrival of shot pellets on a target. For low frequencies ($f \ll t_{\mathrm{tr}}^{-1}$), the noise power is

$$W = 2\, e\, I_0, \tag{J.27}$$

and the current noise thus is given by

$$\langle i^2 \rangle = 2\, e\, I_0\, B, \tag{J.28}$$

This noise term has been first found for vacuum diodes in the saturation regime which also serve as noise normals according to (J.28). It is important for the validity of (J.28) that in each event a full charge e is transferred. The situation in a semiconductor diode is more complicated since various currents contribute; if scattering events occur during transit, also fractional transferred charges can occur. The reverse current of an asymmetric diode if originating from the lowly doped region is due to carriers crossing the depletion layer. If generation in the depletion layer plays no role, the noise is also determined by the shot noise (J.28).

The maximal noise level (J.28) is present in absence of all correlations (Poisson process), both in the injection process as well as in the subsequent transport. Such value has been found, e.g., for intrinsic germanium in [1858] and in the limit of large

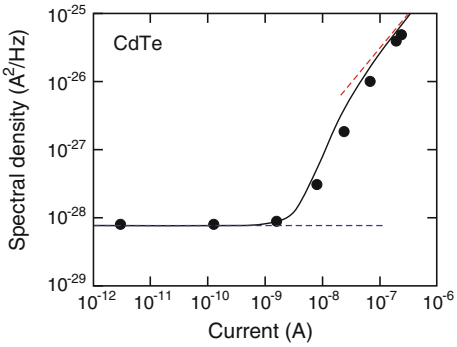


Fig. J.2 Noise of semi-insulating (dark) CdTe detector at $T = 323\,\mathrm{K}$ (at a frequency of about 1–2 kHz when $1/f$-noise plays no role). Experimental data (*symbols*) and detailed theory (*black line*). The *dashed blue line* represents the thermal noise (J.25), the *dashed red line* the shot noise (J.28). Adapted from [1859]

currents for CdTe detectors [1859] (Fig. J.2). In a metallic conductor (or degenerate semiconductor) the noise is reduced to a third of that value due to correlations induced by the Pauli exclusion principle [1860]. The modification in non-degenerate semiconductors on length scales intermediate between the elastic and inelastic mean free paths is discussed in [1861]. The case of shot noise in semiconductors in the presence of transport of electrons and holes has been treated in [1858–1862].

J.3.4 Generation-Recombination Noise

It is a semiconductor specific property that the carrier density is subject to fluctuations due to generation and recombination.[12] A fluctuation in majority carrier density leads to a change of conductivity which will lead to a change in current if a constant voltage is applied. Typical examples of transitions leading to a fluctuation of the carrier density are between bands and localized levels and in between the conduction and valence bands. Usually, the sample remains neutral. Detailed treatments are given in [1863, 1864].

A simple example is the effect of carrier number fluctuation due to transitions between a conduction band and donor levels. This is manifested in the noise spectrum of a n-Si sample at $T = 78\,\mathrm{K}$ (Fig. J.3a) with the plateau at 10^6–$10^7\,\mathrm{Hz}$ on top of the $1/f$ noise [1865] (The plateau at 10^8–$10^9\,\mathrm{Hz}$ is due to velocity fluctuations). The spectral power of the generation-recombination noise contribution is given by

$$W = I_0^2 \, \frac{\langle \delta n^2 \rangle}{\langle n \rangle^2} \, \frac{4\,\tau_0}{1 + (\omega\,\tau_0)^2}, \tag{J.29}$$

[12] A metal exhibits a constant carrier density.

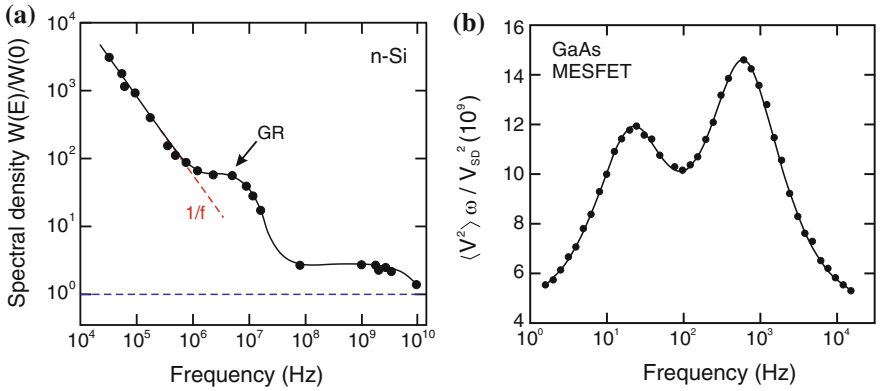


Fig. J.3 (**a**) Current noise spectrum of n-type Si ($T = 78\,\mathrm{K}$, $n = 3 \times 10^{13}\,\mathrm{cm}^{-3}$) for an electric field of $E = 200\,\mathrm{V/cm}$ along the ⟨100⟩ direction, in relative units to the noise spectrum for $E = 0$. The *dashed blue line* indicates the level of thermal noise, The *arrow* labeled 'GR' denotes the contribution of generation-recombination noise. Adapted from [1865]. (**b**) Voltage noise power times frequency of a GaAs MESFET. Experimental data (*symbols*) and fit (*solid line*) including two generation-recombination noise terms of the type (J.29) (times ω) for two different traps. Adapted from [1866]

where τ_0 is the characteristic relaxation time, $\langle n \rangle = \bar{n}$ is the average carrier density (average carrier number per given volume) and $\langle \delta n^2 \rangle = \langle (n - \bar{n})^2 \rangle$ is the fluctuation of the carrier density. In order to better visualize the generation-recombination noise with respect to the $1/f$ noise, the quantity $W \times \omega$ can be plotted (Fig. J.3b) which takes the shape of a peak (for logarithmic frequency axis) [1866].

For a partially compensated semiconductor with $N_\mathrm{D} > N_\mathrm{A}$ it is found (if holes can be neglected) [1864]

$$\frac{\langle \delta n^2 \rangle}{\langle n \rangle} = \left[1 + \frac{\bar{n}\, N_\mathrm{D}}{(\bar{n} + N_\mathrm{A})\,(N_\mathrm{D} - N_\mathrm{A} - \bar{n})} \right]^{-1} \leq 1. \tag{J.30}$$

The fluctuation $\langle \delta n^2 \rangle / \langle n \rangle^2$ is typically smaller than the Poisson value of 1; such sub-Poissonian statistics is typical of a repulsive correlation. For the case $N_\mathrm{D} \gg N_\mathrm{A}$ [1864], (J.30) simplifies to

$$\frac{\langle \delta n^2 \rangle}{\langle n \rangle^2} = \left[1 + \frac{N_\mathrm{D}}{N_\mathrm{D} - \bar{n}} \right]^{-1} = \frac{N_\mathrm{D} - \bar{n}}{2\,N_\mathrm{D} - \bar{n}}. \tag{J.31}$$

In the ambipolar regime, typically close to intrinsic conditions, when only free electrons and holes are important, it is found [1864] ($\mu_\mathrm{n} < 0$)

$$\frac{\langle \delta n^2 \rangle}{\langle n \rangle^2} = \frac{\bar{n}^2\, \bar{p}\, (\mu_\mathrm{p} - \mu_\mathrm{n})^2}{(\bar{n} + \bar{p})\,(\bar{p}\,\mu_\mathrm{p} - \bar{n}\,\mu_\mathrm{n})^2}. \tag{J.32}$$

which simplifies to

$$\frac{\langle \delta n^2 \rangle}{\langle n \rangle^2} = \frac{1}{2}. \tag{J.33}$$

in the intrinsic case ($\bar{n} = \bar{p}$).

References

1. J.R. Chelikowsky, M.L. Cohen, Semiconductors: a pillar of pure and applied physics. J. Appl. Phys. **117**, 112812(8pp) (2015)
2. J.M. Bishop, *How to Win the Nobel Prize: An Unexpected Life in Science* (Harvard University Press, Cambridge, 2003)
3. G. Busch, Early history of the physics and chemistry of semiconductors-from doubts to fact in a hundred years. Eur. J. Phys. **10**, 254–264 (1989)
4. L. Hoddeson, E. Braun, J. Teichmann, S. Weart (eds.), *Out of the Crystal Maze* (Oxford University Press, Oxford, 1992)
5. K.C. Handel, Anfänge der Halbleiterforschung und -entwicklung. Dargestellt an den Biographien von vier deutschen Halbleiterpionieren. Ph.D. thesis, RWTH Aachen (1999)
6. R.S. Feigelson (ed.), *50 Years Progress in Crystal Growth* (Elsevier, Amsterdam, 2004)
7. P.R. Morris, *A History of the World Semiconductor Industry*. IEE History of Technology Series, vol. 12 (Peter Peregrinus Ltd., London, 1990)
8. D. Holbrook, W.M. Cohen, D.A. Hounshell, S. Klepper, The nature, sources, and consequences of firm differences in the early history of the semiconductor industry. Strateg. Manag. J. **21**, 1017–1041 (2000)
9. A. Volta, Del modo di render sensibilissima la più debole Elettricità sia Naturale, sia Artificiale. Philos. Trans. R. Soc. Lond. **72**, 237–280 (1782) (followed by an English translation Of the Method of rendering very sensible the weakest Natural and Artificial Electricity)
10. T.J. Seebeck, Magnetische Polarisation der Metalle und Erze durch Temperaturdifferenz. Abhandl. Deut. Akad. Wiss. Berlin (Physik. Klasse) 265–373 (1822); Ann. Physik **82**, 1–20, 133–160, 253–286 (1826)
11. M. Faraday, Experimental researches in electricity, §10. On conduction power generally. Philos. Trans. R. Soc. Lond. **123**, 507–522 (1833) (Series IV, paragraph 434, p. 519)
12. J. Peltier, Nouvelles Expériences sur la Caloricité des courans électriques. Ann. Chim. **LVI**, 371–387 (1834)
13. W. Smith, The action of light on selenium. J. Soc. Telegr. Eng. **2**, 31–33 (1873); Nature **7**(173), 303 (1873)
14. S. Bidwell, On the sensitiveness of selenium to light, and the development of a similar property in sulphur. Proc. Phys. Soc. **7**, 129–145 (1885)
15. S. Bidwell, The electrical properties of selenium. Proc. Phys. Soc. **13**, 552–579 (1894)
16. W. Schreier, Ferdinand Braun in Leipzig, Zum 150. Geburtstag des Entdeckers des Halbleitereffektes und des Erfinders der Kathodenstrahlröhre. NTM Zeitschrift für Geschichte der Wissenschaften Technik und Medizin **8**, 201–208 (2000)

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

17. F. Braun, Über die Stromleitung durch Schwefelmetalle. J.C. Poggendorf (ed.), Ann. Phys. Chem. **153**, 556–563 (1874)
18. W.G. Adams, R.E. Day, The action of light on selenium. Proc. R. Soc. Lond. A **25**, 113–117 (1876)
19. W. Siemens, On the influence of light upon the conductivity of crystalline selenium. Philos. Mag. **50**(332), 416 (1875)
20. ChE Fritts, On a new form of selenium cell, and some electrical discoveries made by its use. Am. J. Sci. **26**, 465–472 (1883)
21. E.H. Hall, On a new action of the magnet on electric currents. Am. J. Math. **2**, 287–292 (1879)
22. E.H. Hall, On a new action of the magnet on electric currents. Philos. Mag. **9**, 225–230 (1879)
23. H.A. Rowland, Preliminary notes on Mr. Hall's recent discovery. Philos. Mag. **9**, 432–434 (1880)
24. J.Z. Buchwald, The Hall effect and Maxwellian electrodynamics in the 1880's. Part I: the discovery of a new electric field. Centaurus **23**, 51–99 (1979)
25. J.Z. Buchwald, The Hall effect and Maxwellian electrodynamics in the 1880's. Part II: the unification of theory, 1881–1893. Centaurus **23**, 118–162 (1979)
26. J.C. Bose, Detector for electrical disturbances. US Patent 755,860, filed 1901, awarded 1904
27. G.W. Pickard, Means for receiving intelligence communicated by electric waves. US Patent 836,531, filed 1906, awarded 1906
28. G.W. Pickard, Oscillation-receiver. US Patent 888,191, filed 1907, awarded 1908
29. T. Jenkins, A brief history of . . . semiconductors. Phys. Educ. **40**, 430–439 (2005)
30. H.J. Round, A note on carborundum. Electron World **19**, 309 (1907)
31. K. Bädeker, Über die elektrische Leitfähigkeit und die thermoelektrische Kraft einiger Schwermetallverbindungen. Ann. Physik **327**, 749–766 (1907) (Habilitationsschrift, Universität Leipzig)
32. W. Kaiser, Karl Bädekers Beitrag zur Halbleiterforschung. Centaurus **22**, 187–200 (1978)
33. M. Grundmann, Karl Bädeker (1877–1914) and the discovery of transparent conductive materials. Phys. Status Solidi A **212**, 1409–1426 (2015)
34. K. Bädeker, Über eine eigentümliche Form elektrischen Leitvermögens bei festen Körpern. Ann. Physik **334**, 566–584 (1909)
35. K. Steinberg, Über den Halleffekt bei jodhaltigem Kupferjodür. Ann. Physik **340**, 1009–1033 (1911)
36. G.W. Pickard, The discovery of the oscillating crystal. Radio News **6**(7) (Jan 1925 issue), 1166, 1270 (1925)
37. J.A. Ratcliffe, William Henry Eccles. 1875–1966. Biogr. Mem. Fellows R. Soc. **17**, 195–214 (1971)
38. W.H. Eccles, On an oscillation detector actuated solely by resistance-temperature variations. Proc. Phys. Soc. Lond. **22**, 360–368 (1910)
39. J. Weiss, Experimentelle Beiträge zur Elektronentheorie aus dem Gebiet der Thermoelektrizität. Ph.D. Thesis, Albert-Ludwigs Universität Freiburg i.Br. (1910)
40. J. Königsberger, J. Weiss, Über die thermoelektrischen Effekte (Thermokräfte, Thomsonwärme) und die Wärmeleitung in einigen Elementen und Verbindungen und über die experimentelle Prüfung der Elektronentheorien. Ann. Physik **35**, 1–46 (1911)
41. F. Walther, P. Knipping, M. von Laue, Interferenzerscheinungen bei Röntgenstrahlen. Bayerische Akademie der Wissenschaften, Mathematisch-Physikalische Klasse, Sitzungsberichte (1912), pp. 303–322; Ann. Physik **346**, 971–988 (1913)
42. M. von Laue, Eine quantitative Prüfung der Theorie für die Interferenz-Erscheinungen bei Röntgenstrahlen. Bayerische Akademie der Wissenschaften, Mathematisch-Physikalische Klasse, Sitzungsberichte (1912), pp. 363–373; Ann. Physik **346**, 989–1002 (1913)
43. G. Dörfel, *Julius Edgar Lilienfeld und William David Coolidge-Ihre Röntgenröhren und ihre Konflikte* (Max-Planck-Institut für Wissenschaftsgeschichte, Berlin, 2006)

44. Th.L. Thomas, The Twenty Lost Years of Solid-State Physics. Analog (Science Fact → Science Fiction) **LXXV**(1) (March 1965 issue), pp. 8–13, 81 (1965) (Erroneously this article places Universität Leipzig into Poland)
45. C. Kleint, Julius Edgar Lilienfeld: life and profession. Prog. Surf. Sci. **57**, 253–327 (1998)
46. J.E. Lilienfeld, Method and apparatus controlling electric currents. US patent 1,745,175, filed 1926, awarded 1930
47. B.E. Crawford, The invention of the transistor. MSc thesis, The University of Vermont, Burlington (1991)
48. J.E. Lilienfeld, Device for controlling electric current. US patent 1,900,018, filed 1928, awarded 1933
49. J.E. Lilienfeld, Amplifier for electric currents. US patent 1,877,140, filed 1928, awarded 1932
50. W. Gosling, The pre-history of the transistor. Radio Electron. Eng. **43**, 10 (1973)
51. A. Schleede, H. Buggisch, Untersuchungen am Bleiglanz- und Pyritdetektor. Z. Anorg. Allg. Chemie **161**, 85–107 (1927)
52. E. Körner, Über die Darstellungsmethoden und die Lumineszenzfähigkeit reinsten Zinksulfids und Zinkoxydes. Ph.D. thesis, Universität Greifswald (1930)
53. A. Schleede, Über den Strukturellen Bau der Leuchtzentren in den Zink- und Cadmium-Sulfid-Leuchtstoffen. Chem. Ber. **90**, 1162–1175 (1957)
54. F. Bloch, Über die Quantenmechanik der Elektronen in Kristallgittern. Z. Phys. **52**, 555–560 (1928)
55. E.E. Loebner, Subhistories of the light emitting diode. IEEE Trans. Electron Devices **ED-23**, 675–699 (1976)
56. M.A. Nivikov, Oleg Vladimirovich Losev: pioneer of semiconductor electronics. Fiz. Tverd. Tela 46, 5–9 (2004) (Soviet Phys. Solid State **46**, 1–4 (2004))
57. N. Zheludev, The life and times of the LED-a 100-year history. Nat. Photonics **1**, 189–192 (2007)
58. O.V. Lossev, Luminous carborundum detector and detection effect and oscillations with crystals. Philos. Mag. Ser. 7 **6**(39), 1024–1044 (1928)
59. R. Peierls, Zur Theorie der galvanomagnetischen Effekte. Zeitschr. f. Physik **53**, 255–266 (1929)
60. R.E. Peierls, Zur Theorie des Hall-Effekts. Phys. Zeitschrift **30**, 273–274 (1929)
61. M.J.O. Strutt, Zur Wellenmechanik des Atomgitters. Ann. Physik **391**, 319–324 (1928)
62. R. Peierls, Zur Theorie der elektrischen und thermischen Leitfähigkeit von Metallen. Ann. Physik **396**, 121–148 (1930)
63. W. Heisenberg, Zum Paulischen Ausschließungsprinzip. Ann. Physik **402**, 888–904 (1931)
64. R. de L. Kronig, W.G. Penney, Quantum mechanics of electrons in crystal lattices. Proc. R. Soc. Lond. A **130**, 499–513 (1931)
65. A.H. Wilson, *Semi-Conductors and Metals* (Cambridge University Press, Cambridge, 1939)
66. R. Cahn, Silicon: a child of revolution. Interface **14**(1), 15–16 (2005)
67. A.H. Wilson, The theory of electronic semiconductors. Proc. R. Soc. Lond. A **133**, 458–491 (1931)
68. A.H. Wilson, The theory of electronic semiconductors II. Proc. R. Soc. Lond. A **134**, 277–287 (1931)
69. Chemie der Elementarprozesse, C. Wagner, W. Schottky, Theorie der geordneten Mischphasen. Z. Phys. Chem. Abteilung B. Aufbau der Materie **11**, 163–210 (1930)
70. H.H. von Baumbach, C. Wagner, Die elektrische Leitfähigkeit von Zinkoxyd und Cadmiumoxyd. Z. Phys. Chem. B **22**, 199–211 (1933)
71. C. Zener, A theory of electrical breakdown of solid dielectrics. Proc. R. Soc. Lond. A **145**, 523–529 (1934)
72. J. Frenkel, On the absorption of light and the trapping of electrons and positive holes in crystalline dielectrics. Phys. Z. Sowjetunion **9**, 158–186 (1936)
73. B.I. Davydov, On the rectification of current at the boundary between two semiconductors. Dokl. Acad. Nauk SSSR (C.R. Acad. Sci. USSR) **20**, 279–282 (1938)

74. B.I. Davydov, On the theory of solid rectifiers. Dokl. Acad. Nauk SSSR (C.R. Acad. Sci. USSR) **20**, 279–282 (1938)
75. W. Schottky, Halbleitertheorie der Sperrschicht. Naturwissenschaften **26**, 843 (1938)
76. N.F. Mott, Note on the contact between a metal and an insulator or semiconductor. Proc. Camb. Philos. Soc. **34**, 568–572 (1938)
77. N.F. Mott, The theory of crystal rectifiers. Proc. R. Soc. Lond. A **171**, 27–38 (1939)
78. R. Hilsch, R.W. Pohl, Steuerung von Elektronenströmen mit einem Dreielektrodenkristall und ein Modell einer Sperrschicht. Z. Phys. **111**, 399–408 (1938)
79. R.S. Ohl, Light-sensitive electric device. US patent 2,402,662, filed 1941, awarded 1946
80. J.H. Scaff, The role of metallurgy in the technology of electronic materials. Metall. Trans. **1**, 561–573 (1970)
81. M. Riordan, L. Hoddeson, The origins of the pn junction. IEEE Spectr. **34**(6), 46–51 (1997)
82. R.S. Ohl, Alternating current rectifier. US patent 2,402,661, filed 1941, awarded 1946
83. J. Scaff, R.S. Ohl, The development of silicon crystal rectifiers for microwave radar receivers. Bell Syst. Tech. J. **26**, 1–30 (1947)
84. K. Clusius, E. Holz, H. Welker, Elektrische Gleichrichteranordnung mit Germanium als Halbleiter und Verfahren zur Herstellung von Germanium für eine solche Gleichrichteranordnung. German patent DBP 966 387, 21g, 11/02 (1942)
85. H. Welker, Halbleiteranordnung zur kapazitiven Steuerung von Strömen in einem Halbleiterkristall. German patent DBP 980 084, 21g, 11/02 (1945)
86. J. Bardeen, W.H. Brattain, Three-electrode circuit element utilizing semiconductor materials. US patent 2,524,035, filed 1948, awarded 1950
87. M. Riordan, L. Hoddeson, C. Herring, The invention of the transistor. Rev. Mod. Phys. **71**, S336–S345 (1999)
88. J. Bardeen, W.H. Brittain, Physical principles involved in transistor action. Phys. Rev. **75**, 1208–1225 (1949)
89. J. Bardeen, W.H. Brittain, The transistor, a semiconductor triode. Phys. Rev. **74**, 230–231 (1948)
90. W.F. Brinkman, D.E. Haggan, W.W. Troutman, A history of the invention of the transistor and where it will lead us. IEEE J. Solid-State Circuit **SC-32**, 1858–1865 (1997)
91. W.B. Shockley, Circuit element utilizing semiconductor material. US patent 2,569,347, filed 1948, awarded 1951
92. N.A. Goryunova, Grey tin. Ph.D. thesis (in Russian), Leningrad State University (1951)
93. N.A. Goryunova, *The Chemistry of Diamond-Like Semiconductors* (Chapman and Hall, London, 1965) (translation of her book that was published 1963 in Leningrad)
94. H. Welker, Verfahren zur Herstellung eines Halbleiterkristalls mit Zonen verschiedenen Leitungstyps bei $A_{III} - B_V$-Verbindungen. German patent DBP 976 709, 21g, 11/02 (1952)
95. H. Welker, Verfahren zur Herstellung eines Halbleiterkristalls aus einer $A_{III} - B_V$-Verbindung mit Zonen verschiedenen Leitungstyps. German patent DBP 976 791, 12c, 2 (1952)
96. H. Welker, Semiconductor devices and methods of their manufacture. US patent 2,798,989, filed 1952, awarded 1957
97. H. Welker, Über neue halbleitende Verbindungen. Z. Naturf. **7a**, 744–749 (1952)
98. W. Shockley, A unipolar "field-effect" transistor. Proc. IRE **40**, 1365–1376 (1952)
99. G.C. Dacey, I.M. Ross, Unipolar "field-effect" transistor. Proc. IRE **41**, 970–979 (1953)
100. D.M. Chapin, C.S. Fuller, G.L. Pearson, A new silicon p-n junction photocell for converting solar radiation into electrical power. J. Appl. Phys. **25**, 676–677 (1954)
101. J.S. Kilby, Miniaturized electronic circuits. US patent 3,138,743, filed 1959, awarded 1964
102. R.N. Noyce, Semiconductor device-and-lead structure. US patent 2,981,877, filed 1959, awarded 1961
103. A.N. Saxena, *Invention of Integrated Circuits-Untold Important Facts* (World Scientific, Singapore, 2009)
104. G. Mortenson, D.O. Relin, *Three Cups of Tea: One Man's Mission to Fight Terrorism and Build Nations . . . One School at a Time* (Viking Penguin, New York, 2006)

105. J.A. Hoerni, Method of manufacturing semiconductor devices. US patent 3,025,589, filed 1959, awarded 1962
106. J.A. Hoerni, Semiconductor device. US patent 3,064,167, filed 1959, awarded 1962
107. J.A. Hoerni, Planar silicon diodes and transistors, in *International Electron Devices Meeting*, Washington October 1960 (1960), p. 50
108. J.A. Hoerni, Planar silicon diodes and transistors. IRE Trans. Electron Devices **8**, 178 (1961)
109. M. Riordan, The silicon dioxide solution: how physicist Jean Hoerni built the bridge from the transistor to the integrated circuit. IEEE Spectr. **44**(12), 50–56 (2007)
110. Datasheet 2N1613, Silicon npn transistor. Fairchild Semiconductor
111. D. Kahng, M.M. Atalla, *Silicon-silicon dioxide field induced surface device, in IRE Solid-State Device Research Conference* (Carnegie Institute of Technology, Pittsburgh, 1960)
112. D. Kahng, Electric field controlled semiconductor device. US patent 3,102,230, filed 1960, awarded 1963
113. R.N. Hall, G.E. Fenner, J.D. Kingsley, T.J. Soltys, R.O. Carlson, Coherent light emission from GaAs junctions. Phys. Rev. Lett. **9**, 366–368 (1962)
114. R.N. Hall, Stimulated emission semiconductor devices. US patent 3,245,002, filed 1962, awarded 1966
115. M.I. Nathan, W.P. Dumke, G. Burns, F.H. Dill Jr, G. Lasher, Stimulated emission of radiation from GaAs p-n junctions. Appl. Phys. Lett. **1**, 62–64 (1962)
116. T.M. Quist, R.H. Rediker, R.J. Keyes, W.E. Krag, B. Lax, A.L. McWhorter, H.J. Zeigler, Semiconductor maser in GaAs. Appl. Phys. Lett. **1**, 91–92 (1962)
117. N. Holonyak, S.F. Bevacqua, Coherent (visible) light emission from $Ga(As_{1-x}P_x)$ junctions. Appl. Phys. Lett. **1**, 82–83 (1962)
118. Remembering the laser diode, Nat. Photonics **6**, 795 (2012)
119. J.J. Coleman, The development of the semiconductor laser diode after the first demonstration in 1962. Semicond. Sci. Technol. **27**, 090207(10pp) (2012)
120. Zh.I. Alferov, R.F. Kasarinov, Semiconductor laser with electric pumping. Inventor's Certificate No. 181737 (in Russian), Application No. 950840 (1963)
121. H. Kroemer, A proposed class of heterojunction injection lasers. Proc. IEEE **51**, 1782–1783 (1963)
122. J.B. Gunn, Microwave oscillations of current in III-V semiconductors. Solid State Commun. **1**, 88–91 (1963)
123. C.A. Mead, Schottky barrier gate field effect transistor. Proc. IEEE **54**, 307–308 (1966)
124. Zh.I. Alferov, V.I. Korol'kov, V.I. Maslov, A.V. Lishina, M.K. Trukan, A.A. Yakovenko, Investigation of current-voltage characteristics of diffused p-n junctions in $GaAs_{0.65}P_{0.35}$. Fizika i Tekn. Poluprovodn. **1**, 260–264 (1967) (Sov. Phys. Semicond. **1**, 206–209 (1967))
125. Zh.I. Alferov, D.Z. Garbuzov, V.S. Grigor'ev, Yu.V. Zhilyaev, L.V. Kradinova, V.I. Korol'kov, E.P. Morozov, O.A. Ninua, E.L. Portnoy, V.D. Prochukhan, M.K. Trukan, Injection luminescence of epitaxial heterojunctions in the GaP-GaAs system. Fiz. Tverd. Tela **9**, 279–282 (1967) (Sov. Phys. Solid State **9**, 208–210 (1967))
126. W.W. Hooper, W.I. Lehrer, An epitaxial GaAs field effect transistor. Proc. IEEE **55**, 1237–1238 (1967)
127. Zh.I. Alferov, V.M. Andreev, E.L. Portnoi, M.K. Trukan, AlAs-GaAs heterojunction injection lasers with a low room-temperature threshold. Fizika i. Tekn. Poluprovodn. **3**, 1328–1332 (1969) (Sov. Phys. Semicond. **3**, 1107–1110 (1970))
128. I. Hayashi, Heterostructure lasers. IEEE Trans. Electron Devices **ED-31**, 1630–1642 (1984)
129. R.A. Logan, H.G. White, W. Wiegmann, Efficient green electroluminescence in nitrogen-doped GaP p-n junctions. Appl. Phys. Lett. **13**, 139–141 (1968)
130. R.M. Potter, J.M. Blank, A. Addamiano, Silicon carbide light-emitting diodes. J. Appl. Phys. **40**, 2253–2257 (1969)
131. R.S. Pengelly, J.A. Turner, Monolithic broadband GaAs F.E.T. amplifiers. Electron. Lett. **12**, 251–252 (1976)
132. S. Nakamura, Crystal growth method for gallium nitride-based compound semiconductor. US patent 5,290,393, filed: 1992, awarded: 1994

133. S. Nakamura, M. Senoh, N. Iwasa, S. Nagahama, High-brightness InGaN blue, green and yellow light-emitting diodes with quantum-well structures. Jpn. J. Appl. Phys. **34**, L797–L799 (1995)
134. A.J. Moses, *Optical Materials Properties, Handbook of Electronic Materials*, vol. 1 (Plenum, New York, 1971)
135. M. Neuberger, *III-V Semiconducting Compounds, Handbook of Electronic Materials*, vol. 2 (Plenum, New York, 1971)
136. J.T. Milek, *Silicon Nitride for Microelectronic Applications. Part 1 Preparation and Properties. Handbook of Electronic Materials*, vol. 3 (Plenum, New York, 1971)
137. M. Neuberger, D.L. Grigsby, W.H. Veazie Jr, *Niobium Alloys and Compounds, Handbook of Electronic Materials*, vol. 4 (Plenum, New York, 1972)
138. M. Neuberger, *Group IV Semiconducting Materials, Handbook of Electronic Materials*, vol. 5 (Plenum, New York, 1972)
139. J.T. Milek, *Silicon Nitride for Microelectronic Applications. Part 2 Applications and Devices, Handbook of Electronic Materials*, vol. 6 (Plenum, New York, 1972)
140. O. Madelung, *Semiconductors: Data Handbook*, 3rd edn. (Springer, Berlin, 2004)
141. W. Martienssen, Semiconductors, in *Springer Handbook of Condensed Matter and Materials Data*, ed. by W. Martienssen, H. Warlimont (Springer, Berlin, 2005), pp. 575–694
142. S. Adachi, *Handbook on Physical Properties of Semiconductors (Group IV Elemental Semiconductors)*, vol. 1 (Kluwer, New York, 2004)
143. S. Adachi, *Handbook on Physical Properties of Semiconductors (III-V Compound Semiconductors)*, vol. 2 (Kluwer, New York, 2004)
144. S. Adachi, *Handbook on Physical Properties of Semiconductors (II-VI Compound Semiconductors)*, vol. 3 (Kluwer, New York, 2004)
145. S. Adachi, *Optical Constants of Crystalline and Amorphous Semiconductors (Numerical Data and Graphical Information* (Springer, New York, 1999)
146. O. Madelung, M. Schulz, H. Weiss (eds.), *Semiconductors, Landolt-Börnstein, Numerical Data and Functional Relationships in Science and Technology*. New Series, vol. 17 (Springer, Berlin, 1982)
147. M. Dayah, http://www.dayah.com
148. B. Bryson, *A Short History of Nearly Everything* (Doubleday, London, 2003)
149. J.C. Phillips, *Bonds and Bands in Semiconductors* (Academic Press, New York, 1973)
150. J.C. Phillips, G. Lucovsky, *Bonds and Bands in Semiconductors*, 2nd edn. (Momentum Press, New York, 2009)
151. P. Evers, *Die wundersame Welt der Atomis: 10 Jahre in den Physikalischen Blättern* (Wiley-VCH, Weinheim, 2002)
152. M.T. Yin, M.L. Cohen, Microscopic theory of the phase transformation and lattice dynamics of Si. Phys. Rev. Lett. **45**, 1004–1007 (1980)
153. J. Bardeen, W. Shockley, Deformation potentials and mobilities in non-polar crystals. Phys. Rev. **80**, 72–80 (1950)
154. G.E. Kimball, The electronic structure of diamond. J. Chem. Phys. **3**, 560–564 (1935)
155. N. Bernstein, M.J. Mehl, D.A. Papaconstantopoulos, N.I. Papanicolaou, M.Z. Bazant, E. Kaxiras, Energetic, vibrational, and electronic properties of silicon using a nonorthogonal tight-binding model. Phys. Rev. B **62**, 4477–4487 (2000)
156. R.M. Roberts, *Serendipidty, Accidental Discoveries in Science* (Wiley, New York, 1989), pp. 75–81
157. H.J.F. Jansen, A.J. Freeman, Structural properties and electron density of NaCl. Phys. Rev. B **33**, 8629–8631 (1986)
158. T. Soma, T. Umenai, Madelung energies of tetragonal and orthorhombic systems. Phys. Status Solidi B **78**, 229–240 (1976)
159. J.A. Van Vechten, Quantum dielectric theory of electronegativity in covalent systems. I. Electronic dielectric constant. Phys. Rev. **182**, 891–905 (1969)
160. J.C. Phillips, J.A. Van Vechten, Dielectric classification of crystal structures, ionization potentials, and band structures. Phys. Rev. Lett. **22**, 705–708 (1969)

161. C.R.A. Catlow, A.M. Stoneham, Ionicity in solids. J. Phys. C: Solid State Phys. **16**, 4321–4338 (1983)
162. H.B. Callen, Electric breakdown in ionic solids. Phys. Rev. **76**, 1394–1402 (1949)
163. J.C. Phillips, J.A. Van Vechten, Charge redistribution and piezoelectric constants. Phys. Rev. Lett. **23**, 1115–1117 (1969)
164. S. Takeuchi, K. Suzuki, Phys. Status Solidi A **171**, 99–103 (1999)
165. J.-B. Romé de l'Isle, *Cristallographie, ou Description Des Formes Propre a Tous les Corps du Regne Minéral*, vol. 1 (1783), p. 91
166. M.S. Dresselhaus, G. Dresselhaus, A. Jorio, *Group Theory, Application to the Physics of Condensed Matter* (Springer, Berlin, 2008)
167. D. Shechtman, I. Blech, D. Gratias, J.W. Cahn, Metallic phase with long-range orientational order and no translational symmetry. Phys. Rev. Lett. **53**, 1951–1953 (1984)
168. I.R. Fisher, K.O. Cheon, A.F. Panchula, P.C. Canfield, M. Chernikov, H.R. Ott, K. Dennis, Magnetic and transport properties of single-grain R-Mg-Zn icosahedral quasicrystals [$R =$ Y, $(Y_{1-x}Gd_x)$, $(Y_{1-x}Tb_x)$, Tb, Dy, Ho, and Er]. Phys. Rev. B **59**, 308–321 (1999)
169. K. Kimura, A. Hori, M. Takeda, H. Yamashita, H. Ino, Possibility of semiconducting quasicrystal in boron-rich solids. J. Non-Cryst. Solids **153&154**, 398–402 (1993)
170. K. Kirihara, K. Kimura, Covalency, semiconductor-like and thermoelectric properties of Al-based quasicrystals: icosahedral cluster solids. Sci. Technol. Adv. Mater. **1**, 227–236 (2000)
171. G. Burns, A.M. Glazer, *Space Groups for Solid State Scientists*, 3rd edn. (Academic Press, San Diego, 2013)
172. R.E. Proano, D.G. Ast, Effects of the presence/absence of HCl during gate oxidation on the electrical and structural properties of polycrystalline silicon thin film transistors. J. Appl. Phys. **66**, 2189–2199 (1989)
173. A.H. Clark, Electrical and optical properties of amorphous germanium. Phys. Rev. **154**, 750–757 (1967)
174. R.C. Chittick, J.H. Alexander, H.F. Sterling, The preparation and properties of amorphous silicon. J. Electrochem. Soc. **116**, 77–81 (1969)
175. E.P. Denton, H. Rawson, J.E. Stanworth, Seed production. Testing and distribution in European countries. Nature **173**, 1030–1032 (1954)
176. H. Hosono, Ionic amorphous oxide semiconductors: material design, carrier transport, and device application. J. Non-Cryst. Solids **352**, 851–858 (2006)
177. B. Kramer, Electronic structure and optical properties of amorphous semiconductors. Adv. Solid State Phys. (Festkörperprobleme) **12**, 133–182 (1972)
178. M.H. Brodsky (ed.), *Amorphous Semiconductors, in Topics in Applied Physics*, vol. 36 (Springer, Berlin, 1979)
179. R. Hull (ed.), *Properties of Crystalline Silicon* (IET, London, 1999)
180. A.J. Haq, P.R. Munroe, Phase transformations in (111) Si after spherical indentation. J. Mater. Res. **24**, 1967–1975 (2009)
181. G.A. Busch, R. Kern, Semiconducting properties of gray tin. Solid State Phys. **11**, 1–40 (1960)
182. J.R. Chelikowski, Chemical trends in the structural stability of binary crystals. Phys. Rev. B **34**, 5295–5304 (1986)
183. J.M. Besson, J.P. Itié, A. Polian, G. Weill, J.L. Mansot, J. Gonzalez, High-pressure phase transition and phase diagram of gallium arsenide. Phys. Rev. B **44**, 4214–4234 (1991)
184. P. Lawaetz, Stability of the Wurtzite structure. Phys. Rev. B **5**, 4039–4045 (1972)
185. R.F. Davis, M.S. Shur, *GaN-based Materials And Devices: Growth, Fabrication, Characterization & Performance* (World Scientific, Singapore, 2004)
186. C. Jagadish, S.J. Pearton (eds.), *Zinc Oxide Bulk, Thin Films and Nanostructures: Processing, Properties and Applications* (Elsevier, Amsterdam, 2006)
187. M. Shur, S. Rumyantsev, M. Levinshtein (eds.), *SiC Materials and Devices*, vol. 1 (World Scientific, Singapore, 2006); M. Shur, S. Rumyantsev, M. Levinshtein (eds.), *SiC Materials And Devices*, vol. 2 (World Scientific, Singapore, 2007)

188. S. Siebentritt, U. Rau (eds.), *Wide-Gap Chalcopyrites* (Springer, Berlin, 2006)
189. J.L. Shay, J.H. Wernick, *Ternary Chalcopyrites* (Pergamon, Oxford, 1975)
190. A. Zunger, Order-disorder transformation in ternary tetrahedral semiconductors. Appl. Phys. Lett. **50**, 164–166 (1987)
191. J.E. Jaffe, A. Zunger, Theory of the band-gap anomaly in ABC_2 chalcopyrite semiconductors. Phys. Rev. B **29**, 1882–1906 (1984)
192. H.F. Liu, A.S.W. Wong, G.X. Hu, H. Gong, Observation of interfacial reactions and recrystallization of extrinsic phases in epitaxial grown ZnO/GaAs heterostructures. J. Cryst. Growth **310**, 4305–4308 (2008)
193. K.-W. Park, Y.-H. Yun, S.-C. Choi, Photoluminescence characteristics of $ZnGa_2O_4$ thin films prepared by chemical solution method. Solid State Ionics **177**, 1875–1878 (2006)
194. T. Maitra, R. Valentí, Ferromagnetism in the Fe-substituted spinel semiconductor $ZnGa_2O_4$. J. Phys.: Condens. Matter **17**, 7417–7431 (2005)
195. K.E. Sickafus, J.M. Wills, N.W. Grimes, Structure of Spinel. J. Am. Ceram. Soc. **82**, 3279–3292 (1999)
196. H.St.C. O'Neill, A. Navrotsky, Simple spinels: crystallographic parameters, cation radii, lattice energies, and cation distribution. Am. Mineral. **68**, 181–194 (1983)
197. J.E. Jaffe, R.A. Bachorz, M. Gutowski, Low-temperature polymorphs of ZrO_2 and HfO_2: a density-functional theory study. Phys. Rev. B **72**, 144107(9pp) (2005)
198. P. Duwez, F.H. Brown Jr, F. Odell, The Zirconia-Yttria system. J. Electrochem. Soc. **98**, 356–362 (1951)
199. R. Ruh, K.S. Mazdiyasni, P.G. Valentine, H.O. Bielstein, Phase relations in the system $ZrO_2 - Y_2O_3$ at Low Y_2O_3 contents. J. Am. Ceram. Soc. **67**, C-190-C-192 (1984)
200. X. Nie, S.-H. Wei, S.B. Zhang, Bipolar doping and band-gap anomalies in delafossite transparent conductive oxides. Phys. Rev. Lett. **88**, 066405(4pp) (2002)
201. C.Y. Yang, Z.R. Huang, W.H. Yang, Y.Q. Zhou, K.K. Fung, Crystallographic study of lanthanum aluminate by convergent-beam electron diffraction. Acta Cryst. A **47**, 703–706 (1991)
202. A. Guivarc'h, R. Guérin, J. Caulet, A. Poudoulec, J. Fontanille, Metallurgical study of Ni/GaAs contacts. II. Interfacial reactions of Ni thin films on (111) and (001) GaAs. J. Appl. Phys. **66**, 2129–2136 (1989)
203. S. Yoshioka, H. Hayashi, A. Kuwabara, F. Oba, K. Matsunaga, I. Tanaka, Structures and energetics of Ga_2O_3 polymorphs. J. Phys.: Condens. Matter **19**, 346211(11pp) (2007)
204. E.S. Dana, W.E. Ford, *A Textbook of Mineralogy with an Extended Treatise on Crystallography and Physical Mineralogy*, 3rd edn. (Wiley, New York, 1922)
205. W. Borchardt-Ott, *Kristallographie: Eine Einführung für Naturwissenschaftler*, 8th edn. (Springer, Berlin, 2013)
206. W. Borchardt-Ott, R.O. Gould, *Crystallography*, 3rd edn. (Springer, Berlin, 2012)
207. Th. Hahn (ed.), *International Tables for Crystallography, Volume A: Space-Group Symmetry*, 5th edn. (Springer, Dordrecht, 2002)
208. Naval Research Laboratory, http://cst-www.nrl.navy.mil/lattice
209. Th. Höche, private communication (2008)
210. Y. Lifshitz, X.F. Duan, N.G. Shang, Q. Li, L. Wan, I. Bello, S.T. Lee, Epitaxial diamond polytypes on silicon. Nature **412**, 404 (2001)
211. M.J. Kozielski, Polytype single crystals of $Zn_{1-x}Cd_xS$ and $ZnS_{1-x}Se_x$ solid solutions grown from the melt under high argon pressure by Bridgman's method. J. Cryst. Growth **30**, 86–92 (1975)
212. P.F. Fewster, *X-Ray Scattering from Semiconductors* (Imperial College Press, London, 2003)
213. W.H. Miller, *Treatise on Crystallography* (Deighton, Cambridge, 1839)
214. J.F. Zheng, J.D. Walker, M.B. Salmeron, E.R. Weber, Interface segregation and clustering in strained-layer InGaAs/GaAs heterostructures studied by cross-section scanning tunneling microscopy. Phys. Rev. Lett. **72**, 2414–2417 (1994)
215. K.-J. Chao, Ch-K Shih, D.W. Gotthold, B.G. Streeman, Determination of 2D pair correlations and pair interaction energies of in atoms in molecular beam epitaxially grown InGaAs alloys. Phys. Rev. Lett. **79**, 4822–4825 (1997)

216. C. Bundesmann, M. Schubert, D. Spemann, A. Rahm, H. Hochmuth, M. Lorenz, M. Grundmann, Infrared dielectric function and phonon modes of Mg-rich cubic $Mg_xZn_{1-x}O$ ($x \geq 0.67$) thin films on sapphire (0001). Appl. Phys. Lett. **85**, 905–907 (2004)
217. M.M. Kreitman, D.L. Barnett, Probability tables for clusters of foreign atoms in simple lattices assuming next-nearest-neighbor interactions. J. Chem. Phys. **43**, 364–371 (1965)
218. O. Okada, Magnetic susceptibilities of $Mg_{1-p}Mn_pTe_2$. J. Phys. Soc. Jpn. **48**, 391–398 (1980)
219. S.-H. Wei, L.G. Ferreira, A. Zunger, First-principles calculation of temperature-composition phase diagrams of semiconductor alloys. Phys. Rev. B **41**, 8240–8269 (1990)
220. I.-H. Ho, G.B. Stringfellow, Solid phase immiscibility in GaInN. Appl. Phys. Lett. **69**, 2701–2703 (1996)
221. I.V. Maznichenko, A. Ernst, M. Bouhassoune, J. Henk, M. Däne, M. Lüders, P. Bruno, W. Hergert, I. Mertig, Z. Szotek, W.M. Temmerman, Structural phase transitions and fundamental band gaps of $Mg_xZn_{1-x}O$ alloys from first principles. Phys. Rev. B **80**, 144101(11pp) (2009)
222. J.C. Mikkelsen Jr, J.B. Boyce, Atomic-scale structure of random solid solutions: extended X-ray absorption fine-structure study of $Ga_{1-x}In_xAs$. Phys. Rev. Lett. **49**, 1412–1415 (1982)
223. J.C. Mikkelsen Jr, J.B. Boyce, Extended x-ray absorption fine-structure study of $Ga_{1-x}In_xAs$ random solid solutions. Phys. Rev. B **28**, 7130–7140 (1983)
224. O. Ambacher, M. Eickhoff, A. Link, M. Hermann, M. Stutzmann, F. Bernardini, V. Fiorentini, Y. Smorchkova, J. Speck, U. Mishra, W. Schaff, V. Tilak, L.F. Eastman, Electronics and sensors based on pyroelectric AlGaN/GaN heterostructures, part A: polarization and pyroelectronics. Phys. Status Solidi C 1878–1907 (2003)
225. A. Zunger, S. Mahajan, *Atomic Ordering and Phase Separation in III-V Alloys*. Handbook on Semiconductors, vol. 3 (Elsevier, Amsterdam, 1994), pp. 1399–1513
226. H.S. Lee, J.Y. Lee, T.W. Kim, D.U. Lee, D.C. Choo, H.L. Park, Coexistence behavior of the $CuPt_B$-type and the CuAu-I-type ordered structures in highly strained $Cd_xZn_{1-x}Te$/GaAs heterostructures. Appl. Phys. Lett. **79**, 1637–1639 (2001)
227. C. Domke, M. Ph Ebert, K. Urban, Heinrich, Direct determination of the interaction between vacancies on InP(110) surfaces. Phys. Rev. B **54**, 10288–10291 (1996)
228. X. Ph Ebert, M. Chen, M. Heinrich, K. Simon, M.G. Lagally, Urban. Phys. Rev. Lett. **76**, 2089–2092 (1996)
229. R.M. Feenstra, J.M. Woodall, G.D. Pettit, Observation of bulk defects by scanning tunneling microscopy and spectroscopy: arsenic antisite defects in GaAs. Phys. Rev. Lett. **71**, 1176–1179 (1993)
230. R.B. Capaz, K. Cho, J.D. Joannopoulos, Signatures of bulk and surface arsenic antisite defects in GaAs(110). Phys. Rev. Lett. **75**, 1811–1814 (1995)
231. R. Jones, A. Carvalho, J.P. Goss, P.R. Briddon, The self-interstitial in silicon and germanium. Mater. Sci. Eng. B **159**, 112–116 (2008)
232. S.J. Clark, Complex structure in tetrahedral semiconductors. Ph.D. thesis, University of Edinburgh (1994)
233. M. Lannoo, J. Bourgoin, *Point Defects in Semiconductors I* (Springer, Berlin, 1981)
234. H. Bracht, N.A. Stolwijk, H. Mehrer, Properties of intrinsic point defects in silicon determined by zinc diffusion experiments under nonequilibrium conditions. Phys. Rev. B **52**, 16542–16560 (1995)
235. V. Ranki, K. Saarinen, Formation of thermal vacancies in highly As and P doped Si. Phys. Rev. Lett. **93**, 255502(4pp) (2004)
236. J. Gebauer, M. Lausmann, F. Redmann, R. Krause-Rehberg, H.S. Leipner, E.R. Weber, Ph. Ebert, Determination of the Gibbs free energy of formation of Ga vacancies in GaAs by positron annihilation. Phys. Rev. B **67**, 235207(8pp) (2003)
237. F. Morehead, N. Stolwijk, W. Meyberg, U. Gösele, Self-interstitial and vacancy contributions to silicon self-diffusion determined from the diffusion of gold in silicon. Appl. Phys. Lett. **42**, 690–692 (1983)
238. S. Dannefaer, P. Mascher, D. Kerr, Monovacancy formation enthalpy in silicon. Phys. Rev. Lett. **56**, 2195–2198 (1986)

239. P.M. Fahey, P.B. Griffin, J.D. Plummer, Point defects and dopant diffusion in silicon. Rev. Mod. Phys. **61**, 289–384 (1989)
240. P. Pichler, *Intrinsic Point Defects, Impurities, and Their Diffusion in Silicon* (Springer, Berlin, 2004)
241. G.D. Watkins, The vacancy in silicon: identical diffusion properties at cryogenic and elevated temperatures. J. Appl. Phys. **103**, 106106(2pp) (2008)
242. Y. Shimizu, M. Uematsu, K.M. Itoh, Experimental evidence of the vacancy-mediated silicon self-diffusion in single-crystalline silicon. Phys. Rev. Lett. **98**, 095901(4pp) (2007)
243. C.S. Nichols, C.G. van der Walle, S.T. Pantelides, Mechanisms of dopant impurity diffusion in silicon. Phys. Rev. B **40**, 5484–5496 (1989)
244. A. Ural, P.B. Griffin, J.D. Plummer, Fractional contributions of microscopic diffusion mechanisms for common dopants and self-diffusion in silicon. J. Appl. Phys. **85**, 6440–6446 (1999)
245. J.-W. Jeong, A. Oshiyama, Atomic and electronic structures of a Boron impurity and its diffusion pathways in crystalline Si. Phys. Rev. B **64**, 235204(9pp) (2001)
246. W. Windl, Concentration dependence of self-interstitial and boron diffusion in silicon. Appl. Phys. Lett. **92**, 202104(3pp) (2008)
247. K.-S. Yoon, C.-O. Wang, J.-H. Yoo, T. Won, First-principles investigation of indium diffusion in a silicon substrate. J. Korean Phys. Soc. **48**, 535–539 (2006)
248. X.-Y. Liu, W. Windl, K.M. Beardmore, M.P. Masquelier, First-principles study of phosphorus diffusion in silicon: interstitial- and vacancy-mediated diffusion mechanisms. Appl. Phys. Lett. **82**, 1839–1841 (2003)
249. L. Rayleigh, On the distillation of binary mixtures. Philos. Mag. **4**, 521–537 (1902)
250. C.H. Li, Normal freezing of ideal ternary liquid solutions of the pseudobinary type. J. Phys. D: Appl. Phys. **7**, 2003–2008 (1974)
251. W. Zulehner, The growth of highly pure silicon crystals. Metrologia **31**, 255–261 (1994)
252. I. Yonenaga, T. Ayuzawa, Segregation coefficients of various dopants in Si_xGe_{1-x} ($0.93 < x < 0.96$) single crystals. J. Cryst. Growth **297**, 14–19 (2006)
253. W.C. O'Mara, R.B. Herring, L.P. Hunt (eds.), *Handbook of Semiconductor Silicon Technology* (Noyes, Berkshire, 1990)
254. J.J. Dowd, R.L. Rouse, Distribution coefficient of indium in germanium on crystallization. Proc. Phys. Soc. B **66**, 60–61 (1953)
255. W.W. Tyler, R. Newman, H.H. Woodbury, Properties of germanium doped with nickel. Phys. Rev. **98**, 461–465 (1955)
256. W.C. Dunlap Jr, Gold as acceptor in germanium. Phys. Rev. **97**, 614–629 (1955)
257. W.G. Pfann, Techniques of zone melting and crystal growing. Solid State Phys. **4**, 423–521 (1957)
258. M. Nastasi, J.W. Mayer, *Ion Implantation and Synthesis of Materials* (Springer, Berlin, 2006)
259. A. Benninghoven, F.G. Rudenauer, H.W. Werner (eds.), *Secondary Ion Mass Spectrometry* (Wiley, New York, 1987)
260. S. Uppal, A.F.W. Willoughby, J.M. Bonar, A.G.R. Evans, N.E.B. Cowern, R. Morris, M.G. Dowsett, Diffusion of ion-implanted boron in germanium. J. Appl. Phys. **90**, 4293–4295 (2001)
261. R.F. Lever, K.W. Brannon, A low energy limit to boron channeling in silicon. J. Appl. Phys. **69**, 6369–6372 (1991)
262. http://www.srim.org
263. J.F. Ziegler, J.P. Biersack, M.D. Ziegler, *SRIM-The Stopping and Range of Ions in Matter* (Lulu Press, Morrisville, 2006)
264. K. Kimura, Y. Oota, K. Nakajima, T.H. Büyüklimanli, High-resolution depth profiling of ultrashallow boron implants in silicon using high-resolution RBS. Current Appl. Phys. **3**, 9–11 (2003)
265. R. Wittmann, Miniaturization problems in CMOS technology: investigation of doping profiles and reliability. Ph.D. thesis, TU Wien (2007)

266. J. Yokota, Lattice distorsion around charged impurities in semiconductors. J. Phys. Soc. Jpn. **19**, 1487 (1964)
267. J.A. Vergés, D. Glötzel, M. Cardona, O.K. Andersen, Absolute hydrostatic deformation potentials of tetrahedral semiconductors. Phys. Status Solidi B **113**, 519–534 (1982)
268. M. Cardona, N.E. Christensen, Acoustic deformation potentials and heterostructure band offsets in semiconductors. Phys. Rev. B **35**, 6182–6194 (1987)
269. D.A.W. Soarez, C.A. Pimentel, Precision interplanar spacing measurements of boron-doped silicon. J. Appl. Cryst. **16**, 486–492 (1983)
270. J.L. Hastings, S.K. Estreicher, P.A. Fedders, Vacancy aggregates in silicon. Phys. Rev. B **56**, 10215–10220 (1997)
271. M. Albrecht, R. Schewski, K. Irmscher, Z. Galazka, T. Markurt, M. Naumann, T. Schulz, R. Uecker, R. Fornari, S. Meuret, M. Kociak, Coloration and oxygen vacancies in wide band gap oxide semiconductors: absorption at metallic nanoparticles induced by vacancy clustering–a case study on indium oxide. J. Appl. Phys. **115**, 053504(12pp) (2014)
272. S. Lee, G.S. Hwang, Theoretical determination of stable fourfold coordinated vacancy clusters in silicon. Phys. Rev. B **78**, 125310(6pp) (2008)
273. H.R. Vydyanath, J.S. Lorenzo, F.A. Kröger, Defect pairing diffusion, and solubility studies in selenium-doped silicon. J. Appl. Phys. **49**, 5928–5937 (1978)
274. P. Deák, A. Gali, A. Sólynom, A. Buruzs, Th Fraunheim, Electronic structure of boron-interstitial clusters in silicon. J. Phys.: Condens. Matter **17**, S2141–S2153 (2005). (and references therein)
275. B.P. Uberuaga, G. Henkelman, H. Jónsson, S.T. Dunham, W. Windl, R. Stumpf, Theoretical studies of self-diffusion and dopant clustering in semiconductors. Phys. Status Solidi B **233**, 24–30 (2002)
276. M.J. Caturla, M.D. Johnson, T.D. de la Rubia, The fraction of substitutional boron in silicon during ion implantation and thermal annealing. Appl. Phys. Lett. **72**, 2736–2738 (1998)
277. F.A. Trumbore, Solid solubilities of impurity elements in germanium and silicon. Bell Syst. Tech. J. **39**, 205–233 (1960)
278. V.E. Borisenko, S.G. Yudin, Steady-state solubility of substitutional impurities in silicon. Phys. Status Solidi A **101**, 123–127 (1987)
279. H.-J. Queisser, Deep impurities. Adv. Solid State Phys. (Festkörperprobleme) **11**, 45–64 (1971)
280. J. Adey, R. Jones, P.R. Briddon, Enhanced dopant solubility in strained silicon. J. Phys.: Condens. Matter **16**, 9117–9126 (2004)
281. S. Fischler, Correlation between maximum solid solubility and distribution coefficient for impurities in Ge and Si. J. Appl. Phys. **33**, 1615 (1962)
282. F.S. Shishiyanu, V.G. Gheorghiu, S.K. Palazov, Diffusion, solubility, and electrical activity of Co and Fe in InP. Phys. Status Solidi A **40**, 29–35 (1977)
283. M. Luysberg, R. Göbel, H. Janning, FeP precipitates in hydride-vapor phase epitaxially grown InP:Fe. J. Vac. Sci. Technol. B **12**, 2305–2309 (1994)
284. N.A. Smith, I.R. Harris, B. Cockayne, W.R. MacEwan, The identification of precipitate phases in Fe-doped InP single crystals. J. Cryst. Growth **68**, 517–522 (1984)
285. J.P. Hirth, A brief history of dislocation theory. Metall. Mater. Trans. A **16**, 2085–2090 (1985)
286. S. Kret, P. Dłużewski, P. Dłużewski, J.-Y. Laval, On the measurement of dislocation core distributions in a GaAs/ZnTe/CdTe heterostructure by high-resolution transmission electron microscopy. Philos. Mag. **83**, 231–244 (2003)
287. A.R. Smith, V. Ramachandran, R.M. Feenstra, D.W. Greve, M.-S. Shin, M. Skowronski, J. Neugebauer, J.E. Northrup, Wurtzite GaN surface structures studied by scanning tunneling microscopy and reflection high energy electron diffraction. J. Vac. Sci. Technol. A **16**, 1641–1645 (1998)
288. S.N.G. Chu, W.T. Tsang, T.H. Chiu, A.T. Macrander, Lattice-mismatch-generated dislocation structures and their confinement using superlattices in heteroepitaxial GaAs/InP and InP/GaAs grown by chemical beam epitaxy. J. Appl. Phys. **66**, 520–530 (1989)

289. M. Horn-von Hoegen, F.K. LeGoues, M. Copel, M.C. Reuter, R.M. Tromp, Defect self-annihilation in surfactant-mediated epitaxial growth. Phys. Rev. Lett. **67**, 1130–1133 (1991)
290. R.B. Heimann, *Auflösung von Kristallen*. Applied Mineralogy, vol. 8 (Springer, Wien, 1975)
291. K. Sato, M. Shikida, T. Yamashiro, K. Asaumi, Y. Iriye, M. Yamamoto, Anisotropic etching rates of single-crystal silicon for TMAH water solution as a function of crystallographic orientation. Sens. Actuators **73**, 131–137 (1999)
292. J.G. Grabmaier, C.B. Watson, Dislocation etch pits in single crystal GaAs. Phys. Status Solidi **32**, K13–K15 (1969)
293. H. Richter, M. Schulz, Versetzungsnachweis auf {100}-Flächen von GaAs-Einkristallen. Kristall und Technik **9**, 1041–1050 (1974)
294. K. Ishida, H. Kawano, Different etch pit shapes revealed by Molten KOH etching on the (001) GaAs surface and their dependence on the Burgers vectors. Phys. Status Solidi A **98**, 175–181 (1986)
295. J.L. Weyher, J. van de Ven, Selective etching and photoetching of GaAs in CrO_3-HF aqueous solutions. III. Interpretation of defect-related etch figures. J. Cryst. Growth **78**, 191–217 (1986)
296. A.F. Bogenschütz, *Ätzpraxis für Halbleiter* (Carl Hanser, München, 1967)
297. M. Köhler, *Etching in Microsystem Technology* (Wiley-VCH, Weinheim, 1999)
298. A.R. Clawson, Guide to references on III-V semiconductor chemical etching. Mater. Sci. Eng. **31**, 1–438 (2001)
299. J. Frühauf, *Shape and Functional Elements of the Bulk Silicon Microtechnique: A Manual of Wet-Etched Silicon Structures* (Springer, Berlin, 2005)
300. D.M. Manos, D.L. Flamm (eds.), *Plasma Etching: An Introduction* (Academic Press, San Diego, 1989)
301. G.S. May, C.J. Spanos, *Fundamentals of Semiconductor Manufacturing and Process Control* (Wiley, Hoboken, 2006)
302. J.W. Coburn, *Plasma Etching and Reactive Ion Etching: Fundamentals and Applications (American Vacuum Society Monograph Series* (AVS, New York, 1982)
303. V.M Donnelly, A. Konrblit, Plasma etching: yesterday, today, and tomorrow. J. Vac. Sci. Technol. A **31**, 050825(47pp) (2013)
304. V. Smaminathan, A.S. Jordan, Dislocations in III/V compounds. Semicond. Semimet. **38**, 293–341 (1993)
305. T. Kamejima, J. Matsui, Y. Seki, H. Watanabe, Transmission electron microscopy study of microdefects in dislocation-free GaAs and InP crystals. J. Appl. Phys. **50**, 3312–3321 (1979)
306. L. Wang, W. Jie, Y. Yang, G. Xu, L. Fu, Defect characterization and composition distributions of mercury indium telluride single crystals. J. Cryst. Growth **310**, 2810–2814 (2008)
307. NREL, http://www.nrel.gov/measurements/trans.html
308. Z. Liliental-Weber, H. Sohn, J. Washburn, Structural defects in epitaxial III/V layers. Semicond. Semimet. **38**, 397–447 (1993)
309. the role of mass diffusion in vapor-liquid-solid growth, Y. Hao, G. Meng, Ch. ZhL Wang, L. Zhang Ye, Periodically twinned nanowires and polytypic nanobelts of ZnS. Nano Lett. **6**, 1650–1655 (2006)
310. S. Mader, A.E. Blakeslee, Extended dislocations in $GaAs_{0.7}P_{0.3}$. Appl. Phys. Lett. **25**, 365–367 (1974)
311. M. Sato, K. Sumino, K. Hiraga, Impurity effect in stacking fault energy of silicon crystals studied by high resolution electron microscopy. Phys. Status Solidi A **68**, 567–577 (1981)
312. A. Gomez, D.J.H. Cockayne, P.B. Hirsch, V. Vitek, Dissociation of near-screw dislocations in germanium and silicon. Philos. Mag. **31**, 105–113 (1975)
313. D. Gerthsen, C.B. Carter, Stacking-fault energies of GaAs. Phys. Status Solidi A **136**, 29–43 (1993)
314. P. Pirouz, D.J.H. Cockayne, N. Shimada, P. Hirsch, A.R. Lang, Dissociation of dislocations in diamond. Proc. R. Soc. Lond. A **386**, 241–249 (1983)
315. H. Gottschalk, G. Patzer, H. Alexander, Stacking fault energy and ionicity of cubic III-V compounds. Phys. Status Solidi A **45**, 207–217 (1978)

316. S. Takeuchi, K. Suzuki, K. Maeda, Stacking-fault energy of II-VI compounds. Philos. Mag. A **50**, 171–178 (1984)
317. S. Amelinckx, W. Dekeyser, The structure and properties of grain boundaries. Solid State Phys. **8**, 325–499 (1959)
318. C.R.M. Grovenor, Grain boundaries in semiconductors. J. Phys. C: Solid State Phys. **18**, 4079–4119 (1985)
319. C. Fontaine, D.A. Smith, On the atomic structure of the $\Sigma = 3\{112\}$ twin in silicon. Appl. Phys. Lett. **40**, 153–154 (1982)
320. H. Sawada, H. Ichinose, M. Kohyama, Imaging of a single atomic column in silicon grain boundary. J. Electron Microsc. **51**, 353–357 (2002)
321. F.L. Vogel, W.G. Pfann, H.E. Corey, E.E. Thomas, Observations of dislocations in lineage boundaries in germanium. Phys. Rev. **90**, 489–490 (1953)
322. H. Föll, http://www.tf.uni-kiel.de/matwis/amat/def_en/kap_7/backbone/r7_2_2.html
323. H. Föll, D. Ast, TEM observations on grain boundaries in sintered silicon. Philos. Mag. A **40**, 589–610 (1979)
324. M. Grundmann, A. Krost, D. Bimberg, LP-MOVPE growth of antiphase domain free InP on (001) Si using low temperature processing. J. Cryst. Growth **107**, 494–495 (1991)
325. W. Mader, Universität Bonn, private communication (2006)
326. F. Wolf, W. Mader, Microdomain structure in Fe_2O_3-doped ZnO and the formation of spinel. Optik **110**(Suppl. 8), 66 (1999)
327. F. Wolf, B.H. Freitag, W. Mader, Inversion domain boundaries in ZnO with additions of Fe_2O_3 studied by high-resolution ADF imaging. Micron **38**, 549–552 (2007)
328. G.P. Srivastava, *The Physics of Phonons* (Taylor & Francis, New York, 1990)
329. J.L.T. Waugh, G. Dolling, Crystal dynamics of gallium arsenide. Phys. Rev. **132**, 2410–2412 (1963)
330. P.H. Borcherds, K. Kunc, G.F. Alfrey, R.L. Hall, The lattice dynamics of gallium phosphide. J. Phys. C: Solid State Phys. **12**, 4699–4706 (1979)
331. J.L. Yarnell, J.L. Warren, R.G. Wenzel, P.J. Dean, *Lattice dynamics of gallium phosphide, in Neutron Inelastic Scattering*, vol. 1 (IAEA, Vienna, 1968), pp. 301–313
332. S. Bağcı, S. Duman, H.M. Tütüncü, G.P. Srivastava, G. Uğur, Ab initio calculation of the structural and dynamical properties of the zinc-blende BN and its (110) surface. Diam. Relat. Mater. **15**, 1161–1165 (2006)
333. T. Tomoyose, A. Fukuchi, M. Kobayashi, Influence of quadrupolar deformability on phonon dispersion in silver and copper halides. Solid State Ionics **167**, 83–90 (2004)
334. E. Ruiz, S. Alvarez, P. Alemany, Electronic structure and properties of AlN. Phys. Rev. B **49**, 7115–7123 (1994)
335. A. Göbel, T. Ruf, A. Fischer, K. Eberl, M. Cardona, J.P. Silveira, F. Briones, Optical phonons in isotope superlattices of GaAs, GaP, and GaSb studied by Raman scattering. Phys. Rev. B **59**, 12612–12621 (1999)
336. M. Schubert, Universität Leipzig, private communication (2006)
337. A.S. Barker, A.J. Sievers, Optical studies of the vibrational properties of disordered solids, Rev. Mod. Phys. **47**(Suppl. No. 2), S1–S179 (1975)
338. R.C. Newman, Local vibrational mode spectroscopy of defects in III/V compounds. Semicond. Semimet. **38**, 117–187 (1993)
339. M. Stavola, Vibrational spectroscopy of light element impurities in semiconductors. Semicond. Semimet. **51B**, 153–224 (1999)
340. M.D. McCluskey, Local vibrational modes of impurities in semiconductors. J. Appl. Phys. **87**, 3593–3617 (2000)
341. L. Hoffmann, J.C. Bach, B. Bech, Nielsen, P. Leary, R. Jones, S. Öberg, Substitutional carbon in germanium. Phys. Rev. B **55**, 11167–11173 (1997)
342. F. Thompson, R.C. Newman, Localized vibrational modes of light impurities in gallium phosphide. J. Phys. C: Solid State Phys. **4**, 3249–3257 (1971)
343. M. Vandevyver, D.N. Talwar, Grenn's function theory of impurity vibrations due to defect complexes in elemental and compound semiconductors. Phys. Rev. B **21**, 3405–3431 (1980)

344. I.F. Chang, S.S. Mitra, Application of a modified random-element-isodisplacement model to long-wavelength optic phonons of mixed crystals. Phys. Rev. **172**, 924–933 (1968)
345. L.I. Deych, A. Yamilov, A.A. Lisyansky, Concept of local polaritons and optical properties of mixed polar crystals. Phys. Rev. B **62**, 6301–6316 (2000)
346. H. von Wenckstern, R. Schmidt-Grund, C. Bundesmann, A. Müller, C.P. Dietrich, M. Stölzel, M. Lange, M. Grundmann, The (Mg, Zn)O Alloy, in *Handbook of Zinc Oxide and Related Materials*, vol. 1, ed. by Z.C. Feng (Taylor and Francis/CRC Press, Florida, USA, 2012), pp. 257–320
347. M. Born, I. Huang, *Dynamical Theory of Crystal Lattices* (Clarendon Press, Oxford, 1960)
348. H.M. Kayaga, T. Soma, Temperature dependence of the linear thermal expansion coefficient for Si and Ge. Phys. Status Solidi B **129**, K5–K8 (1985)
349. A.S. Saada, *Elasticity, Theory and Applications* (Pergamon, New York, 1974)
350. P.N. Keating, Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure. Phys. Rev. **145**, 637–645 (1966)
351. A.-B. Chen, A. Sher, W.T. Yost, Elastic constants and related properties of semiconductor compounds and their alloys. Semicond. Semimet. **37**, 1–77 (1992)
352. R.S. Lakes, Foam structures with a negative Poisson's ratio. Science **235**, 1038–1040 (1987)
353. K.E. Evans, M.A. Nkansah, I.J. Hutchinson, S.C. Rogers, Molecular network design. Nature **353**, 124 (1991)
354. R.H. Baughman, Avoiding the shrink. Nature **425**, 667 (2003)
355. R. Gatt, J.N. Grima, Negative compressibility. Phys. Status Solidi RRL **2**, 236–238 (2008)
356. R.M. Martin, Elastic properties of ZnS structure semiconductors. Phys. Rev. B **1**, 4005–4011 (1970)
357. R.M. Martin, Relation between elastic tensors of wurtzite and zinc-blende structure material. Phys. Rev. B **6**, 4546–4553 (1972)
358. K. Kim, W.R.L. Lambrecht, B. Segall, Elastic constants and related properties of tetrahedrally bonded BN, AlN, GaN, and InN. Phys. Rev. B **53**, 16310–16326 (1996)
359. A. Polian, M. Grimsditch, I. Grzegory, Elastic constants of gallium nitride. J. Appl. Phys. **79**, 3343–3344 (1996)
360. L.E. McNeil, M. Grimsditch, R.H. French, Vibrational spectroscopy of aluminum nitride. J. Am. Ceram. Soc. **76**, 1132–1136 (1993)
361. C.F. Cline, H.L. Dunegan, G.W. Henderson, Elastic constants of hexagonal BeO, ZnS, and CdSe. J. Appl. Phys. **38**, 1944–1948 (1967)
362. G. Carlotti, D. Fioretto, G. Socino, E. Verona, Brillouin scattering determination of the whole set of elastic constants of a single transparent film of hexagonal symmetry. J. Phys.: Condens. Matter **7**, 9147–9153 (1995)
363. ChG Van de Walle, Band lineups and deformation potentials in the model-solid theory. Phys. Rev. B **39**, 1871–1883 (1989)
364. A.E. Romanov, T.J. Baker, S. Nakamura, J.S. Speck, Strain-induced polarization in wurtzite III-nitride semipolar layers. J. Appl. Phys. **100**, 023522(10pp) (2006)
365. M. Grundmann, J. Zúñiga-Pérez, Pseudomorphic ZnO-based heterostructures: from polar through all semipolar to nonpolar orientations. Phys. Status Solidi B (2015). doi:10.1002/pssb.201552535
366. J.D. Eshelby, The determmination of the elastic field of an ellipsoidal inclusion and related problems. Proc. R. Soc. Lond. A **241**, 376–396 (1957)
367. M. Grundmann, O. Stier, D. Bimberg, InAs/GaAs pyramidal quantum dots: strain distribution, optical phonons, and electronic structure. Phys. Rev. B **52**, 11969–11981 (1995)
368. L.B. Freund, J.A. Floro, E. Chason, Extensions of the stoney formula for substrate curvature to configurations with thin substrates or large deformations. Appl. Phys. Lett. **74**, 1987–1989 (1999)
369. M. Grundmann, Nanoscroll formation from strained layer heterostructures. Appl. Phys. Lett. **83**, 2444–2446 (2003)
370. G.G. Stoney, The tension of metallic films deposited by electrolysis. Proc. R. Soc. Lond. A **82**, 172–175 (1909)

371. R. Beresford, J. Yin, K. Tetz, E. Chason, Real-time measurements of stress relaxation in InGaAs/GaAs. J. Vac. Sci. Technol. B **18**, 1431–1434 (2000)
372. A. Krost, A. Dadgar, F. Schulze, R. Clos, K. Haberland, T. Zettler, Heteroepitaxy of GaN on silicon. In situ measurements. Mater. Sci. Forum **483–485**, 1051–1056 (2005)
373. V.Y. Prinz, V.A. Seleznev, A.K. Gutakovsky, in *Proceedings of 24th International Conference on Physics of Semiconductors*, Jerusalem, Israel, 1998 (World Scientific, Singapore, 1998), Th3-D5
374. X. Li, Strain induced semiconductor nanotubes: from formation process to device applications. J. Phys. C: Solid State Phys. **41**, 193001(12pp) (2008)
375. B. Schmidt, Minimal energy configurations of strained multi-layers. Calc. Var. **30**, 477–497 (2007)
376. O.G. Schmidt, N. Schmarje, C. Deneke, C. Müller, N.-Y. Jin-Phillipp, Three-dimensional nano-objects evolving from a two-dimensional layer technology. Adv. Mater. **13**, 756–759 (2001)
377. L. Zhang, E. Ruh, D. Grützmacher, L. Dong, D.J. Bell, B.J. Nelson, C. Schönenberger, Anomalous coiling of SiGe/Si and SiGe/Si/Cr helical nanobelts. Nano Lett. **6**, 1311–1317 (2006)
378. J. Zang, F. Liu, Modified Timoshenko formula for bending of ultrathin strained bilayer films. Appl. Phys. Lett. **92**, 021905(3pp) (2008)
379. S. Mendach, University of Hamburg, private communication (2006)
380. J.C. Bean, L.C. Feldman, A.T. Fiory, S. Nakahara, I.K. Robinson, Ge_xSi_{1-x}/Si strained-layer superlattice grown by molecular beam epitaxy. J. Vac. Sci. Technol. A **2**, 436–440 (1984)
381. R. Hull, J.C. Bean, Nucleation of misfit dislocations in strained-layer epitaxy in the Ge_xSi_{1-x}/Si system. J. Vac. Sci. Technol. A **7**, 2580–2585 (1989)
382. J.W. Matthews, A.E. Blakeslee, Defects in epitaxial multilayers. I. Misfit dislocations. J. Cryst. Growth **27**, 118–125 (1974)
383. Appl. Phys. Lett. Calculation of critical layer thickness versus lattice mismatch for Ge_xSi_{1-x}/Si strained-layer heterostructures. **47**, 322–324 (1985)
384. F.C. Frank, J. van der Merwe, One-dimensional dislocations. I. Static theory. Proc. R. Soc. Lond. A **198**, 205–216 (1949)
385. F.C. Frank, J. van der Merwe, One-dimensional dislocations. II. Misfitting monolayers and oriented overgrowth. Proc. R. Soc. Lond. A **198**, 216–225 (1949)
386. J.H. van der Merwe, Crystal interfaces. Part II. Finite overgrowths. J. Appl. Phys. **34**, 123–127 (1962)
387. J.R. Willis, S.C. Jain, R. Bullough, The energy of an array of dislocations: implications for strain relaxation in semiconductor heterostructures. Philos. Mag. A **62**, 115–129 (1990)
388. B.W. Dodson, J.Y. Tsao, Relaxation of strained-layer semiconductor structures via plastic flow. Appl. Phys. Lett. **51**, 1325–1327 (1987)
389. E. Kasper, H.J. Herzog, H. Kibbel, A one-dimensional SiGe superlattice grown by UHV epitaxy. Appl. Phys. **8**, 199–205 (1975)
390. K.L. Kavanagh, M.A. Capano, L.W. Hobbs, J.C. Barbour, P.M.J. Marée, W. Schaff, J.W. Mayer, D. Pettit, J.M. Woodall, J.A. Stroscio, R.M. Feenstra, Asymmetries in dislocation densities, surface morphology, and strain of GalnAs/GaAs single heterolayers. J. Appl. Phys. **64**, 4843–4852 (1988)
391. M. Grundmann, U. Lienert, D. Bimberg, A. Fischer-Colbrie, J.N. Miller, Anisotropic and inhomogeneous strain relaxation in pseudomorphic $In_{0.23}Ga_{0.77}As$/GaAs quantum wells. Appl. Phys. Lett. **55**, 1765–1767 (1989)
392. P. Quadbeck, K. Ph Ebert, J. Urban, R. Krause-Rehberg, Gebauer, Effect of dopant atoms on the roughness of III-V semiconductor cleavage surfaces. Appl. Phys. Lett. **76**, 300–302 (2000)
393. J.E. Ayers, Compliant substrates for heteroepitaxial semiconductor devices: theory, experiment, and current directions. J. Electron. Mater. **37**, 1511–1523 (2008)
394. E. Weber, UC Berkeley, private communication

395. X.F. Brun, S.N. Melkote, Analysis of stresses and breakage of crystalline silicon wafers during handling and transport. Sol. Energy Mater. Sol. Cells **93**, 1238–1247 (2009)
396. H.J. Möller, C. Funke, M. Rinio, S. Scholz, Multicrystalline silicon for solar cells. Thin Solid Films **487**, 179–187 (2005)
397. P.S. Dominguez, J.M. Fernandez, Introduction of thinner monocrystalline silicon wafers in an industrial cell-manufacturing facility, in Proceedings of the 20th EU PVSEC, Barcelona, Spain (2005), pp. 903–905
398. P.-Y. Chen, M.-H. Tsai, W.-K. Yeh, M.-H. Jing, Y. Chang, Investigation of the relationship between whole-wafer strength and control of its edge engineering. Jpn. J. Appl. Phys. **48**, 126503(6pp) (2009)
399. Ch. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963)
400. G. Dresselhaus, Spin-orbit coupling effects in zinc blende structures. Phys. Rev. **100**, 580–586 (1955)
401. J.-W. Luo, G. Bester, A. Zunger, Full-zone spin splitting for electrons and holes in bulk GaAs and GaSb. Phys. Rev. Lett. **102**, 056405(4pp) (2009)
402. E.I. Rashba, Properties of semiconductors with an extremum loop. I. Cyclotron and combinatorial resonance in a magnetic field perpendicular to the plane of the loop. Fiz. Tverd. Tela **2**, 1224–1238 (1960) (Sov. Phys. Solid State **2**, 1109–1122 (1960))
403. Yu.A. Bychkov, É.I. Rashba, Properties of a 2D electron gas with lifted spectral degeneracy. Pis'ma Zh. Èksp. Teor. Fiz. **39**, 66 (1984) (JETP Lett. **39**, 78–81 (1984))
404. S.G. Ganichev, L.E. Golub, Interplay of Rashba/Dresselhaus spin splittings probed by photogalvanic spectroscopy-a review. Phys. Status Solidi B **251**, 1801–1823 (2014)
405. G.F. Koster, J.O. Dimmock, R.G. Wheeler, H. Statz, *Properties of the Thirty-Two Point Groups* (MIT, Cambridge, 1963)
406. D.G. Bell, Group theory and crystal lattices. Rev. Mod. Phys. **26**, 311–320 (1954)
407. R.H. Parmentier, Symmetry properties of the energy bands of the zinc blende structure. Phys. Rev. **100**, 573–579 (1955)
408. L.P. Bouckaert, R. Smoluchowski, E. Wigner, Theory of Brillouin zones and symmetry properties of wave functions in crystals. Phys. Rev. **50**, 58–67 (1936)
409. J.R. Chelikowsky, M.L. Cohen, Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. Phys. Rev. B **14**, 556–582 (1976)
410. R. Dalven, Electronic structure of PbS. PbSe, and PbTe. Solid State Phys. **28**, 179–224 (1973)
411. A. Schleife, F. Fuchs, J. Furthmüller, F. Bechstedt, First-principles study of ground- and excited-state properties of MgO, ZnO, and CdO polymorphs. Phys. Rev. B **73**, 245212(14pp) (2006)
412. H. Dixit, R. Saniz, S. Cottenier, D. Lamoen, B. Partoens, Electronic structure of transparent oxides with the Tran-Blaha modified Becke-Johnson potential. J. Phys.: Condens. Matter **24**, 205503(9pp) (2012)
413. J.E. Jaffe, A. Zunger, Electronic structure of the ternary chalcopyrite semiconductors $CuAlS_2$, $CuGaS_2$, $CuInS_2$, $CuAlSe_2$, $CuGaSe_2$, and $CuInSe_2$. Phys. Rev. B **28**, 5822–5847 (1983)
414. S. Limpijumnong, S.N. Rashkeev, W.R.L. Lambrecht, Electronic structure and optical properties of $ZnGeN_2$. MRS Internet J. Nitride Semicond. Res. **4S1**, G6.11(6pp) (1999)
415. W. Rehwald, Band structure of spinel-type semiconductors. Phys. Rev. **155**, 861–868 (1967)
416. H. Dixit, N. Tandon, S. Cottenier, R. Saniz, D. Lamoen, B. Partoens, V. Van Speybroeck, M. Waroquier, Electronic structure and band gap of zinc spinel oxides beyond LDA: $ZnAl_2O_4$, $ZnGa_2O_4$ and $ZnIn_2O_4$. New J. Phys. **13**, 063002(11pp) (2011)
417. R. Ahuja, O. Eriksson, B. Johansson, Electronic and optical properties of $BaTiO_3$ and $SrTiO_3$. J. Appl. Phys. **90**, 1854–1859 (2001)
418. B. Seraphin, Über ein ein-dimensionales Modell halbleitender Verbindungen vom Typus $A^{III}B^{V}$. Z. Naturf. **9a**, 450–456 (1954)

419. J. Wu, W. Walukiewicz, Band gaps of InN and group III nitride alloys. Superlatt. Microstruct. **34**, 63–75 (2003)
420. J.E. Bernard, A. Zunger, Electronic structure of ZnS, ZnSe, ZnTe, and their pseudobinary alloys. Phys. Rev. B **36**, 3199–3228 (1987)
421. E.W. Williams, V. Rehn, Electroreflectance studies of InAs, GaAs and (Ga, In)As alloys. Phys. Rev. **172**, 798–810 (1968)
422. K.-R. Schulze, H. Neumann, K. Unger, Band structure of $Ga_{1-x}In_xAs$. Phys. Status Solidi B **75**, 493–500 (1976)
423. J. Wu, W. Walukiewicz, K.M. Yu, J.W. Ager III, S.X. Li, E.E. Haller, H. Lu, W.J. Schaff, Universal bandgap bowing in group-III nitride alloys. Solid State Commun. **127**, 411–414 (2003)
424. R. Braunstein, A.R. Moore, F. Herman, Intrinsic optical absorption in germanium-silicon alloys. Phys. Rev. **109**, 695–710 (1958)
425. D.J. Wolford, W.Y. Hsu, J.D. Dow, B.G. Streetman, Nitrogen trap in the semiconductor alloys $GaAs_{1-x}P_x$ and $Al_xGa_{1-x}As$. J. Lumin. **18**(19), 863–867 (1979)
426. R. Schmidt, B. Rheinländer, M. Schubert, D. Spemann, T. Butz, J. Lenzner, E.M. Kaidashev, M. Lorenz, M. Grundmann, Dielectric functions (1 to 5 eV) of wurtzite $Mg_xZn_{1-x}O$ ($x \leq 0.29$) thin films. Appl. Phys. Lett. **82**, 2260–2262 (2003)
427. R. Schmidt-Grund, A. Carstens, B. Rheinländer, D. Spemann, H. Hochmut, G. Zimmermann, M. Lorenz, M. Grundmann, C.M. Herzinger, M. Schubert, Refractive indices and band-gap properties of rocksalt $Mg_xZn_{1-x}O$ ($0.68 \leq x \leq 1$). J. Appl. Phys. **99**, 123701(7pp) (2006)
428. S. Larach, R.E. Shrader, C.F. Stocker, Anomalous variation of band gap with composition in zinc sulfo- and seleno-tellurides. Phys. Rev. **108**, 587–589 (1957)
429. S. Merita, T. Krämer, B. Mogwitz, B. Franz, A. Polity, B.K. Meyer, Oxygen in sputter-deposited ZnTe thin films. Phys. Status Solidi C **3**, 960–963 (2006)
430. B. Kramer, A Pseudopotential approach for the Green's function of electrons in amorphous solids. Phys. Status Solidi **41**, 649–658 (1970)
431. B. Kramer, Electronic structure and optical properties of amorphous germanium and silicon. Phys. Status Solidi B **47**, 501–510 (1971)
432. J.A. van Vechten, *A Simple Man's View of the Thermochemistry of Semiconductors*. Handbook on Semiconductors, vol. 3 (North Holland, Amsterdam, 1980), pp. 1–111
433. J. Hartung, L.Å. Hansson, J. Weber, in *Proceedings of the 20th International Conference on the Physics of Semiconductors, Thessaloniki, Greece*, ed. by E.M. Anastassakis, J.D. Joannopoulos (World Scientific, Singapore, 1990), p. 1875
434. Y.W. Tsang, M.L. Cohen, Calculation of the temperature dependence of the energy gaps of PbTe and SnTe. Phys. Rev. B **3**, 1254–1261 (1971)
435. I. Tanaka, K. Sugimoto, D. Kim, H. Nishimura, M. Nakayama, Control of temperature dependence of exciton energies in CuI-CuBr alloy thin films grown by vacuum deposition. Int. J. Mod. Phys. B **15**, 3977–3980 (2001)
436. J. Serrano, Ch. Schweitzer, C.T. Lin, K. Reinmann, M. Cardona, D. Fröhlich, Electron-phonon renormalization of the absorption edge of the cuprous halides. Phys. Rev. B **65**, 125110(7pp) (2002)
437. J. Bhosale, A.K. Ramdas, A. Burger, A. Muñoz, A.H. Romero, M. Cardona, R. Lauck, R.K. Kremer, Temperature dependence of band gaps in semiconductors: electron-phonon interaction. Phys. Rev. B **86**, 195208(10pp) (2012)
438. Y. Varshni, Temperature dependence of the energy gap in semiconductors. Physica **34**, 149–154 (1967)
439. K.P. O'Donnell, X. Chen, Temperature dependence of semiconductor band gaps. Appl. Phys. Lett. **58**, 2924–2926 (1991)
440. R. Pässler, Dispersion-related description of temperature dependencies of band gaps in semiconductors. Phys. Rev. B **66**, 085201(18pp) (2002)
441. N. Garro, A. Cantarero, M. Cardona, A. Göbel, T. Ruf, K. Eberl, Dependence of the lattice parameters and the energy gap of zinc-blende-type semiconductors on isotopic masses. Phys. Rev. B **54**, 4732–4740 (1996)

442. E.O. Kane, Band structure of indium antimonide. J. Phys. Chem. Solids **1**, 249–261 (1957)
443. C. Hermann, C. Weisbuch, $\overrightarrow{k} \cdot \overrightarrow{p}$ perturbation theory in III-V compounds and alloys: a reexamination. Phys. Rev. B **15**, 823–833 (1977)
444. Y.-N. Xu, W.Y. Ching, Electronic, optical, and structural properties of some wurtzite crystals. Phys. Rev. B **48**, 4335–4351 (1993)
445. M. Oshikiri, F. Aryasetiawan, Y. Imanaka, G. Kido, Quasiparticle effective-mass theory in semiconductors. Phys. Rev. B **66**, 125204(4pp) (2002)
446. F. Nava, C. Canali, C. Jacoboni, L. Reggiani, S.F. Kozlov, Electron effective masses and lattice scattering in natural diamond. Solid State Commun. **33**, 475–477 (1980)
447. G. Dresselhaus, A.F. Kip, C. Kittel, Cyclotron resonance of electrons and holes in silicon and germanium crystals. Phys. Rev. **98**, 368–384 (1955)
448. S. Shokhovets, G. Gobsch, O. Ambacher, Conduction band parameters of ZnO. Superlatt. Microstruct. **39**, 299–305 (2006)
449. W.S. Baer, R.N. Dexter, Electron cyclotron resonance in CdS. Phys. Rev. **135**, A1388–A1393 (1964)
450. W. Shockley, Cyclotron resonance, magnetoresistance, and Brillouin zones in semiconductors. Phys. Rev. **90**, 491 (1953)
451. P. Pfeffer, W. Zawadzki, Conduction electrons in GaAs: five-level k·p theory and polaron effects. Phys. Rev. B **41**, 1561–1576 (1990)
452. M. Cardona, N.E. Christensen, G. Fasol, Relativistic band structure and spin-orbit splitting of znic-blende-type semiconductors. Phys. Rev. B **38**, 1806–1827 (1988)
453. D. Bimberg, private communication, original authorship unknown
454. G. Ottaviani, L. Reggiani, C. Canali, F. Nava, A. Alberigi-Quaranta, Hole drift velocity in silicon. Phys. Rev. B **12**, 3318–3329 (1975)
455. M. Cardona, F.H. Pollak, Energy-band structure of germanium and silicon: the k·p method. Phys. Rev. **142**, 530–543 (1966)
456. M.I. Eremets, Semiconducting diamond. Semicond. Sci. Technol. **6**, 439–444 (1991)
457. J.J. Hopfield, Fine structure in the optical absorption edge of anisotropic crystals. J. Phys. Chem. Solids **15**, 97–107 (1960)
458. J.L. Shay, B. Tell, L.M. Schiavone, H.M. Kasper, F. Thiel, Energy bands of $AgInS_2$ in the chalcopyrite and orthorhombic structures. Phys. Rev. B **9**, 1719–1723 (1974)
459. W. Feng, D. Xiao, J. Ding, Y. Yuo, Three-dimensional topological insulators in I-III-VI_2 and II-IV-V_2 chalcopyrite semiconductors. Phys. Rev. Lett. **106**, 016402(4pp) (2011)
460. G. Bastard, Dielectric anomalies in extremely non-parabolic zero-gap semiconductors. J. Phys. C: Solid State Phys. **14**, 839–845 (1981)
461. G.L. Bir, G.E. Pikus, *Symmetry and Strain-induced Effects in Semiconductors* (Wiley, New York, 1974)
462. G.E. Pikus, G.L. Bir, Effect of deformation on the hole energy spectrum of germanium and silicon. Fiz. Tverd. Tela **1**, 1642–1658 (1956) (Sov. Phys. Solid State **1**, 1502–1517 (1959))
463. N.E. Christensen, Electronic structure of GaAs under strain. Phys. Rev. B **30**, 5753–5765 (1984)
464. Y. Zhang, Motion of electrons in semiconductors under inhomogeneous strain with application to laterally confined quantum wells. Phys. Rev. B **49**, 14352–14366 (1994)
465. A.R. Goñi, K. Strössner, K. Syassen, M. Cardona, Pressure dependence of direct and indirect optical absorption in GaAs. Phys. Rev. B **36**, 1581–1587 (1987)
466. W. Shan, W. Walukiewicz, J.W. Ager III, E.E. Haller, J.F. Geisz, D.J. Friedman, J.M. Olson, S.R. Kurtz, Band anticrossing in GaInNAs alloys. Phys. Rev. Lett. **82**, 1221–1224 (1999)
467. S.L. Chuang, C.S. Chang, k·p method for strained wurtzite semiconductors. Phys. Rev. B **54**, 2491–2504 (1996)
468. M. Kumagai, S.L. Chuang, H. Ando, Analytical solutions of the block-diagonalized Hamiltonian for strained wurtzite semiconductors. Phys. Rev. B **57**, 15303–15314 (1998)
469. C. Herring, E. Vogt, Transport and deformation-potential theory for many-valley semiconductors with anisotropic scattering. Phys. Rev. **101**, 944–961 (1956)

470. M.V. Fischetti, S.E. Laux, Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys. J. Appl. Phys. **80**, 2234–2252 (1996)
471. D. Aspnes, M. Cardona, Strain dependence of effective masses in tetrahedral semiconductors. Phys. Rev. B **17**, 726–740 (1978)
472. P.R.C. Kent, A. Zunger, Theory of electronic structure evolution in GaAsN and GaPN alloys. Phys. Rev. B **64**, 115208(23pp) (2001)
473. J. Wu, W. Shan, W. Walukiewicz, Band anticrossing in highly mismatched III-V semiconductor alloys. Semicond. Sci. Technol. **17**, 860–869 (2002)
474. J. Chelikowsky, D.J. Chadi, M.L. Cohen, Calculated valence-band densities of states and photoemission spectra of diamond and zinc-blende semiconductors. Phys. Rev. B **8**, 2786–2794 (1973)
475. N.F. Mott, Electrons in disordered structures. Adv. Phys. **16**, 49–144 (1967)
476. M.H. Cohen, H. Fritzsche, S.R. Ovshinsky, Simple band model for amorphous semiconductor alloys. Phys. Rev. Lett. **22**, 1065–1068 (1969)
477. E.A. Davis, N.F. Mott, Conduction in non-crystalline systems. V. Conductivity, optical absorption and photoconductivity in amorphous semiconductors. Philos. Mag. **22**, 903–922 (1970)
478. J.M. Marshall, A.E. Owen, Drift mobility studies in vitreous arsenic triselenide. Philos. Mag. **24**, 1281–1305 (1971)
479. D.A. Drabold, U. Stephan, J. Dong, S.M. Nakhmanson, The structure of electronic states in amorphous silicon. J. Mol. Graph. Model. **17**(285–291), 330–332 (1999)
480. W. Körner, C. Elsässer, Density-functional theory study of stability and subgap states of crystalline and amorphous Zn-Sn-O. Thin Solid Films **555**, 81–86 (2014)
481. S. Sallis, K.T. Butler, N.F. Quackenbush, D.S. Williams, M. Junda, D.A. Fischer, J.C. Woicik, N.J. Podraza, B.E. White, Jr, A. Walsh, L.F.J. Piper, Origin of deep subgap states in amorphous indium gallium zinc oxide: chemically disordered coordination of oxygen. Appl. Phys. Lett. **104**, 232108(4pp) (2014)
482. W. Pauli 1931 in a letter from New York to R. Peierls, in: Pauli, Briefwechsel II, 1985, Nr. 287, p. 94; cf. G. Busch, Condens. Matter News 2, 15 (1993). Lateron Pauli complains to Peierls "... daß Sie immer noch nicht von der Physik des festen Körpers losgekommen sind." (in: Pauli an Peierls, 22. Mai 1933, in: Pauli, Briefwechsel II, 1985, Nr. 310, p. 163)
483. W. Shockley, *Electrons and Holes in Semiconductors* (D. van Nostrand, New York, 1950)
484. X. Aymerich-Humet, F. Serra-Mestres, J. Millán, An analytical approximation for the Fermi-Dirac integral $F_{3/2}(\eta)$. Solid State Electron. **24**, 981–982 (1981)
485. D. Bednarczyk, J. Bernarczyk, The approximation of the Fermi-Dirac integral $F_{1/2}(\eta)$. Phys. Lett. **64A**, 409–401 (1978)
486. K. Unger, Reversible formulae to approximate Fermi integrals. Phys. Status Solidi B **149**, K141–K144 (1988)
487. J. McDougall, E.C. Stoner, The computation of Fermi-Dirac functions. Philos. Trans. R. Soc. Lond. **237**, 67–104 (1938)
488. A.J. MacLeod, Algorithm 779: Fermi-Dirac functions of order −1/2, 1/2, 3/2, 5/2. ACM Trans. Math. Softw. **24**, 1–12 (1998)
489. A.B. Sproul, M.A. Green, J. Zhao, Improved value for the silicon intrinsic carrier concentration at 300 K. Appl. Phys. Lett. **57**, 255–257 (1990)
490. A.B. Sproul, M.A. Green, Intrinsic carrier concentration and minority-carrier mobility of silicon from 77 to 300 K. J. Appl. Phys. **73**, 1214–1225 (1993)
491. M.A. Green, Intrinsic concentration, effective densities of states, and effective mass in silicon. J. Appl. Phys. **67**, 2944–2954 (1990)
492. K. Misiakos, D. Tsamakis, Accurate measurements of the silicon intrinsic carrier density from 78 to 340 K. J. Appl. Phys. **74**, 3293–3297 (1993)
493. B. Gudden, Über die Elektrizitätsleitung in Halbleitern. Sitzungsber. Phys.-Med. Soz. Erlangen **62**, 289–302 (1930)
494. M. Grundmann, F.-L. Schein, M. Lorenz, T. Böntgen, J. Lenzner, H. von Wenckstern, Cuprous iodide-a p-type transparent semiconductor: history and novel applications. Phys. Status Solidi A **210**, 1671–1703 (2013)

495. A. Zunger, Practical doping principles. Appl. Phys. Lett. **83**, 57–59 (2003)
496. M. Ramamoorthy, S.T. Pantelides, Complex dynamical phenomena in heavily arsenic doped silicon. Phys. Rev. Lett. **76**, 4753–4756 (1996)
497. S. Lany, J. Osorio-Guillén, A. Zunger, Origins of the doping asymmetry in oxides: hole doping in NiO versus electron doping in ZnO. Phys. Rev. B **75**, 241203(4pp) (2007)
498. W. Kohn, Shallow impurity states in semiconductors-the early years. Phys. B **146**, 1–5 (1987)
499. A.K. Ramdas, Spectroscopy of shallow centers in semiconductors: progress since 1960. Phys. B **146**, 6–18 (1987)
500. S.M. Sze, K.K. Ng, *Physics of Semiconductor Devices*, 3rd edn. (Wiley, New York, 2007)
501. E.F. Schubert, *Doping in III-V Semiconductors* (Cambridge University Press, Cambridge, 1993)
502. U. Heim, Evidence for donor-acceptor recombination in InP by time-resolved photoluminescence spectroscopy. Solid State Commun. **7**, 445–447 (1969)
503. W. Kohn, J.M. Luttinger, Theory of donor states in silicon. Phys. Rev. **98**, 915–922 (1955)
504. G. Feher, D.K. Wilson, E.A. Gere, Electron spin resonance experiments on shallow donors in germanium. Phys. Rev. Lett. **3**, 25–28 (1959)
505. R. Kalish, The search for donors in diamond. Diamond Relat. Mater. **10**, 1749–1755 (2001)
506. V.A. Karasyuk, D.G.S. Beckett, M.K. Nissen, A. Villemarie, T.W. Steiner, M.L.W. Thewalt, Fourier-transform magnetoluminescence spectroscopy of donor-bound excitons in GaAs. Phys. Rev. B **49**, 16381–16397 (1994)
507. U. Kaufmann, J. Schneider, Point defects in GaP, GaAs, and InP. Adv. Electron. Electron Phys. **58**, 81–141 (1982)
508. W. Götz, N.M. Johnson, C. Chen, H. Liu, C. Kuo, W. Imler, Activation energies of Si donors in GaN. Appl. Phys. Lett. **68**, 3144–3146 (1996)
509. A.J. Ptak, L.J. Holbert, L. Ting, C.H. Swartz, M. Moldovan, N.C. Giles, T.H. Myersa, P. Van Lierde, C. Tian, R.A. Hockett, S. Mitha, A.E. Wickenden, D.D. Koleske, R.L. Henry, Controlled oxygen doping of GaN using plasma assisted molecular-beam epitaxy. Appl. Phys. Lett. **79**, 2740–2742 (2001)
510. S. Teitler, R.F. Wallis, Note on semiconductor statistics. J. Phys. Chem. Solids **16**, 71–75 (1960)
511. B. Šantić, On the determination of the statistical characteristics of the magnesium acceptor in GaN. Superlatt. Microstruct. **36**, 445–453 (2004)
512. R.M. Dickstein, S.L. Titcomb, R.L. Anderson, Carrier concentration model for n-type silicon at low temperatures. J. Appl. Phys. **66**, 2437–2441 (1989)
513. J.S. Blakemore, Radiative capture by impurities in semiconductors. Phys. Rev. **163**, 809–815 (1967)
514. H.B. Bebb, Comments on radiative capture by impurities in semiconductors. Phys. Rev. B **5**, 4201–4203 (1972)
515. P.P. Debye, E.M. Conwell, Electrical properties of N-type germanium. Phys. Rev. **93**, 693–706 (1954)
516. G.W. Brown, H. Grube, M.E. Hawley, Observation of buried phosphorus dopants near clean Si(100)-(2×1) surfaces with scanning tunneling microscopy. Phys. Rev. B **70**, 121301(4pp) (2004)
517. A.T. Collins, A.W.S. Williams, The nature of the acceptor centre in semiconducting diamond. J. Phys. C: Solid State Phys. **4**, 1789–1800 (1971)
518. K. Thonke, The boron acceptor in diamond. Semicond. Sci. Technol. **18**, S20–S26 (2003)
519. H. Alves: Defects, doping and compensation in wide bandgap semiconductors. Ph.D. thesis, Universität Giessen (2003)
520. S. Fischer, C. Wetzel, E.E. Haller, B.K. Meyer, On p-type doping in GaN-acceptor binding energies. Appl. Phys. Lett. **67**, 1298–1300 (1995)
521. W. Kohn, Shallow impurity states in silicon and germanium. Solid State Phys. **5**, 257–320 (1957)
522. K. Ya, Shtivel'man, R.G. Useinov, Degeneracy multiplicity of the states of two-charge acceptors in semiconductors. Soviet Phys. J. **17**, 1439–1440 (1974)

523. D.C. Look, *Electrical Characterization of GaAs Materials and Devices* (Wiley, New York, 1989)
524. E.M. Conwell, Hall effect and density of states in germanium. Phys. Rev. **99**, 1195–1198 (1955)
525. E.E. Haller, W.L. Hansen, F.S. Goulding, Physics of ultra-pure germanium. Adv. Phys. **30**, 93–138 (1981)
526. M.B. Johnson, O. Albrektsen, R.M. Feenstra, H.W.M. Salemink, Direct imaging of dopants in GaAs with cross-sectional scanning tunneling microscopy. Appl. Phys. Lett. **63**, 2923–2925 (1993); Appl. Phys. Lett. **64**, 1454 (1994) (erratum)
527. A.M. Yakunin, A.Yu. Silov, P.M. Koenraad, J.H. Wolter, W. Van Roy, J. De Boeck, J.-M. Tang, M.E. Flatté, Spatial structure of an individual Mn acceptor in GaAs. Phys. Rev. Lett. **92**, 216806(4pp) (2004)
528. J.-M. Tang, M. Flatté, Multiband tight-binding model of local magnetism in $Ga_{1-x}Mn_xAs$. Phys. Rev. Lett. **92**, 047201(4pp) (2004)
529. J.S. Blakemore, *Semiconductor Statistics* (Pergamon Press, Oxford, 1962)
530. N.B. Hannay (ed.), *Semiconductors* (Reinhold Publishing Corporation, New York, 1959)
531. G. Leibiger, $A^{III}B^{V}$-Mischkristallbildung mit Stickstoff und Bor. Ph.D. thesis, Universität Leipzig (2003)
532. E. Ziegler, W. Siegel, Determination of two donor (majority) levels in semiconductors from hall effect measurements. Cryst. Res. Technol. **17**, 1015–1024 (1982)
533. H.-J. Hoffmann, Defect-level analysis of semiconductors by a new differential evaluation of n(1/T)-characteristics. Appl. Phys. **19**, 307–312 (1979)
534. M. Tao, A kinetic model for metalorganic chemical vapor deposition from trimethylgallium and arsine. J. Appl. Phys. **87**, 3554–3562 (2000)
535. G. Weyer, J.W. Petersen, S. Damgaard, H.L. Nielsen, J. Heinemeier, Site-selective doping of compound semiconductors by ion implantation of radioactive nuclei. Phys. Rev. Lett. **44**, 155–157 (1980)
536. E. Antoncik, B.L. Gu, On the Mössbauer isomer shift studies of the electronic structure of Sn implanted $A^{III}B^{V}$ compounds. Hyperfine Interact. **14**, 257–269 (1983)
537. R. Noufi, R. Axton, C. Herrington, S.K. Deb, Electronic properties versus composition of thin films of $CuInSe_2$. Appl. Phys. Lett. **45**, 668–670 (1994)
538. T.R. Paudel, A. Zakutayev, S. Lany, M. d'Avezac, A. Zunger, Doping rules and doping prototypes in A_2BO_4 spinel oxides. Adv. Funct. Mater. **21**, 4493–4501 (2011)
539. J.D. Perkins, T.R. Paudel, A. Zakutayev, P. Ndione, P.A. Parilla, S. Lany, D.S. Ginley, Y. Shi, J.S. Bettinger, M.F. Toney, Inverse design approach to hole doping in ternary oxides: enhancing p-type conductivity in cobalt oxide spinels. Phys. Rev. B **84**, 205207(8pp) (2011)
540. N.F. Mott, *Metal-Insulator Transitions* (Taylor & Francis, London, 1990)
541. V.I. Fistul, *Highly Doped Semiconductors* (Springer, New York, 1969)
542. B.L. Crowder, W.N. Hammer, Shallow acceptor states in ZnTe and CdTe. Phys. Rev. **150**, 541–545 (1966)
543. G.L. Pearson, J. Bardeen, Electrical properties of pure silicon and silicon alloys containing boron and phosphorous. Phys. Rev. **75**, 865–883 (1949)
544. G.E. Stillman, L.W. Cook, T.J. Roth, T.S. Low, B.J. Skromme, High-purity material, in *GaInAsP Alloy Semiconductors*, ed. by T.P. Pearsall (Wiley, New York, 1982), pp. 121–166
545. T.F. Lee, T.C. McGill, Variation of impurity-to-band activation energies with impurity density. J. Appl. Phys. **46**, 373–380 (1975)
546. M.L. Young, S.J. Bass, The electrical properties of undoped and oxygen- doped GaP grown by the liquid encapsulation technique. J. Phys. D: Appl. Phys. **4**, 995–1005 (1971)
547. H.C. Casey Jr, F. Ermanis, K.B. Wolfstirn, Variation of electrical properties with Zn concentration in GaP. J. Appl. Phys. **40**, 2945–2958 (1969)
548. K. Pakula, M. Wojdak, M. Palczewska, B. Suchanek, J.M. Baranowski, Luminescence and ESR spectra of GaN: Si below and above Mott transition. MRS Internet J. Nitride Semicond. Res. **3**, e34(4pp) (1998)

549. G.R. James, A.W.R. Leitch, F. Omnès, M.C. Wagener, M. Leroux, Correlation of transport and optical properties of Si-doped $Al_{0.23}Ga_{0.77}N$. J. Appl. Phys. **96**, 1047–1052 (2004) (The printed paper title reads '$Al_{0.23}Ga_{0.77}N$')
550. M. Brandt, H. von Wenckstern, C. Meinecke, T. Butz, H. Hochmuth, M. Lorenz, M. Grundmann, Dopant activation in homoepitaxial MgZnO: P thin films. J. Vac. Sci. Technol. B **27**, 1604–1608 (2009)
551. H. Kato, M. Sano, K. Miyamoto, T. Yao, Growth and characterization of Ga-doped ZnO layers on a-plane sapphire substrates grown by molecular beam epitaxy. J. Cryst. Growth **237–239**, 538 (2002)
552. H. Kato, A. Ogawa, H. Kotani, M. Sano, T. Yao, Effects of polarity on MBE growth of undoped, Ga- and N-doped ZnO films, in *MRS Fall Meeting 2006, Boston, Symposium K ('Zinc Oxide and Related Materials'), K5.6* (2006)
553. Z.-L. Lu, W.-Q. Zou, M.-X. Xu, F.-M. Zhang, Y.-W. Du, Structural and electrical properties of single crystalline Ga-doped ZnO thin films grown by molecular beam epitaxy. Chin. Phys. Lett. **26**, 116102(4pp) (2009)
554. M.C. Wu, Y.K. Su, K.Y. Cheng, C.Y. Chang, Electrical and optical properties of heavily doped Mg- and Te-GaAs grown by liquid-phase epitaxy. Solid State Electron. **31**, 251–256 (1988)
555. M. Ogawa, T. Baba, Heavily Si-doped GaAs and AlAs/n-GaAs superlattice grown by molecular epitaxy. Jpn. J. Appl. Phys. **24**, L572–L574 (1985)
556. T. Yamada, E. Tokumitsu, K. Saito, T. Akatsuka, M. Miyauchi, M. Konagai, K. Takahashi, Heavily carbon doped p-Type GaAs and GaAlAs grown by metalorganic molecular beam epitaxy. J. Cryst. Growth **95**, 145–149 (1989)
557. J.L. Liévin, F. Alexandre, C. Dubon-Chevallier, Molecular beam epitaxy of $Ga_{0.99}Be_{0.01}As$ for very high speed heterojunction bipolar transistors, in *Properties of Impurity States in Superlattice Semiconductors. NATO ASI Series B: Physics*, vol. 183, ed. by C.Y. Fong, I.P. Batra, S. Ciraci (Plenum, New York, 1988), pp. 19–28
558. D.J. Thouless, Electrons in disordered systems and the theory of localization. Phys. Rep. **13**, 93–142 (1974)
559. T.F. Rosenbaum, K. Andres, G.A. Thomas, R.N. Bhatt, Sharp metal-insulator transition in a random solid. Phys. Rev. Lett. **45**, 1723–1726 (1980)
560. P.W. Anderson, Absence of diffusion in certain random lattices. Phys. Rev. **109**, 1492–1505 (1958)
561. J. Bourgoin, M. Lannoo, *Point Defects in Semiconductors II* (Springer, Berlin, 1983)
562. S.T. Pantelides (ed.), *Deep Centers in Semiconductors* (Gordon & Breach, New York, 1986)
563. J.-W. Chen, A.G. Milnes, Energy levels in silicon. Ann. Rev. Mater. Sci. **10**, 157–228 (1980)
564. K. Lischka, Deep level defects in narrow gap semiconductors. Phys. Status Solidi B **133**, 17–46 (1986)
565. J.-M. Spaeth, H. Overhof, *Point Defects in Semiconductors and Insulators, Determination of Atomic and Electronic Structure from Paramagnetic Hyperfine Interactions* (Springer, Berlin, 2003)
566. P.T. Landsberg, Degeneracy factors of traps from solubility data for semiconductors. J. Phys. D: Appl. Phys. **10**, 2467–2471 (1977)
567. H.I. Ralph, The degeneracy factor of the gold acceptor level in silicon. J. Appl. Phys. **49**, 672–675 (1978)
568. R. Kassing, L. Cohausz, P. van Staa, W. Mackert, H.J. Hoffman, Determination of the entropy-factor of the gold donor level in silicon by resistivity and DLTS measurements. Appl. Phys. A **34**, 41–47 (1984)
569. H.H. Woodbury, G.W. Ludwig, Spin resonance of transition metals in silicon. Phys. Rev. **117**, 102–108 (1960)
570. S. Greulich-Weber, J.R. Niklas, E.R. Weber, J.M. Spaeth, Electron nuclear double resonance of interstitial iron in silicon. Phys. Rev. B **30**, 6292–6299 (1984)
571. H. Feichtinger, J. Waltl, A. Gschwandtner, Localization of the Fe^0-Level in silicon. Solid State Commun. **27**, 867–871 (1978)

572. Y.H. Lee, R.L. Kleinhenz, J.W. Corbett, EPR of a thermally induced defect in silicon. Appl. Phys. Lett. **31**, 142–144 (1977)
573. H.G. Grimmeiss, E. Janzén, H. Ennen, O. Schirmer, J. Schneider, R. Wörner, C. Holm, E. Sirtl, P. Wagner, Tellurium donors in silicon. Phys. Rev. B **24**, 4571–4586 (1981)
574. H.G. Grimmeiss, L. Montelius, K. Larsson, Chalcogens in germanium. Phys. Rev. B **37**, 6916–6928 (1988)
575. R.K. Franks, J.B. Robertson, Magnesium as a donor impurity in silicon. Solid State Commun. **5**, 479–481 (1967)
576. U. Kaufmann, J. Schneider, R. Wörner, T.A. Kennedy, N.D. Wilsey, The deep double donor P_{Ga} in GaP. J. Phys. C: Solid State Phys. **14**, L951–L955 (1981)
577. R.J. Wagner, J.J. Krebs, G.H. Strauss, A.M. White, Submillimeter EPR evidence for the As antisite defect in GaAs. Solid State Commun. **36**, 15–17 (1980)
578. C.H. Champness, The statistics of divalent impurity centres in a semiconductor. Proc. Phys. Soc. B **69**, 1335–1339 (1956)
579. R. Schaub, G. Pensl, M. Schulz, C. Holm, Donor states in tellurium-doped silicon. Appl. Phys. A **34**, 215–222 (1984)
580. R.O. Carlson, Double-acceptor behavior of zinc in silicon. Phys. Rev. **108**, 1390–1393 (1957)
581. W.W. Tyler, H.H. Woodbury, Scattering of carriers from doubly charged impurity sites in germanium. Phys. Rev. **102**, 647–655 (1956)
582. M.D. Sturge, The Jahn-Teller effect in solids. Solid State Phys. **20**, 91–211 (1968)
583. G.D. Watkins, Negative-U properties for defects in solids. Adv. Solid State Phys. (Festkörperprobleme) **24**, 163–189 (1984)
584. G.D. Watkins, J.R. Troxell, Negative-U properties for point defects in silicon. Phys. Rev. Lett. **44**, 593–596 (1980)
585. P.W. Anderson, Model for the electronic structure of amorphous semiconductors. Phys. Rev. Lett. **34**, 953–955 (1975)
586. R.D. Harris, J.L. Newton, G.D. Watkins, Negative-U defect: interstitial boron in silicon. Phys. Rev. B **36**, 1094–1104 (1987)
587. G.A. Baraff, E.O. Kane, M. Schlüter, Theory of the silicon vacancy: an Anderson negative-U system. Phys. Rev. B **21**, 5662–5686 (1980)
588. M. Sprenger, S.H. Muller, E.G. Sieverts, C.A.J. Ammerlaan, Vacancy in silicon: hyperfine interactions from electron-nuclear double resonance measurements. Phys. Rev. B **35**, 1566–1581 (1987)
589. A.F. Wright, Density-functional-theory calculations for the silicon vacancy. Phys. Rev. B **74**, 165116(8pp) (2006)
590. A. Fazzio, A. Janotti, A.J.R. da Silva, Microscopic picture of the single vacancy in germanium. Phys. Rev. B **61**, R2401–R2404 (2000)
591. P.M. Mooney, N.S. Caswell, S.L. Wright, The capture barrier of the DX center in Si-doped $Al_xGa_{1-x}As$. J. Appl. Phys. **62**, 4786–4797 (1987)
592. P.M. Mooney, Deep donor levels (DX centers) in III-V semiconductors. J. Appl. Phys. **67**, R1–R26 (1990)
593. J.C. Bourgoin (ed.), *Physics of DX Centers in GaAs Alloys* (Sci-Tech, Lake Isabella, 1990)
594. D.V. Lang, R.A. Logan, Large-lattice-relaxation model for persistent photoconductivity in compound semiconductors. Phys. Rev. Lett. **39**, 635–639 (1977)
595. J. Mäkinen, T. Laine, K. Saarinen, P. Hautojärvi, C. Corbel, V.M. Araksinen, J. Nagle, Microscopic structure of the DX center in Si-doped $Al_xGa_{1-x}As$: observation of a vacancy by positron-annihilation spectroscopy. Phys. Rev. B **52**, 4870–4883 (1995)
596. J. Dabrowski, M. Scheffler, Isolated arsenic-antisite defect in GaAs and the properties of EL2. Phys. Rev. B **40**, 10391–10401 (1989)
597. G.M. Martin, Optical assessment of the main electron trap in bulk semi-insulating GaAs. Appl. Phys. Lett. **39**, 747–748 (1981)
598. J.S. Blakemore, Semiconducting and other major properties of gallium arsenide. J. Appl. Phys. **53**, R123–R181 (1982)

599. A. Rohatgi, R.H. Hopkins, J.R. Davis, R.B. Campbell, H.C. Mollenkopf, J.R. McCormick, The impact of molybdenum on silicon and silicon solar cell performance. Solid State Electron. **23**, 1185–1190 (1980)
600. D.C. Look, The electrical characterization of semi-insulating GaAs: a correlation with mass-spectrographic analysis. J. Appl. Phys. **48**, 5141–5148 (1977)
601. O. Mizuno, H. Watanabe, Semi-insulating properties of Fe-doped InP. Electron. Lett. **11**, 118–119 (1975)
602. Y. Toudic, B. Lambert, R. Coquille, G. Grandpierre, M. Gauneau, Chromium-doped p-type semi-insulating InP. Semicond. Sci. Technol. **3**, 464–468 (1988)
603. R.P. Tapster, M.S. Skolnick, R.G. Humphreys, P.J. Dean, B. Cockayne, W.T. MacEwan, Optical and capacitance spectroscopy of InP:Fe. J. Phys. C: Solid State Phys. **14**, 5069–5079 (1981)
604. A. Juhl, A. Hoffmann, D. Bimberg, H.J. Schulz, Bound-exciton-related fine structure in charge transfer spectra of InP: Fe detected by calorimetric absorption spectroscopy. Appl. Phys. Lett. **50**, 1292–1294 (1987)
605. A.M. Hennel, Transition metals in III/V compounds. Semicond. Semimet. **38**, 189–234 (1993)
606. H.J. Schulz, Optical properties of 3d transition metals in II-VI compounds. J. Cryst. Growth **59**, 65–80 (1982)
607. J. Cheng, S.R. Forrest, B. Tell, D. Wilt, B. Schwartz, P.D. Wright, Semi-insulating properties of Fe-implanted InP. I. Current-limiting properties of n^+-semi-insulating-n^+ structures. J. Appl. Phys. **58**, 1780–1786 (1985)
608. D.G. Knight, B. Watt, R. Bruce, D.A. Clark, Characterization of Fe doped InP grown by low pressure MOCVD, in *Proceedings of the 6th Conference on Semi-Insulating III-V Materials*, ed. by A. Milnes, C.J. Miner (Piscataway, IEEE, 1990), pp. 83–88
609. A. Dadgar, O. Stenzel, A. Näser, M. Zafar, Iqbal, D. Bimberg, H. Schumann, Ruthenium: a superior compensator of InP. Appl. Phys. Lett. **73**, 3878–3880 (1998)
610. D. Söderström, G. Fornuto, A. Buccieri, Studies on InP: Fe growth in a close-spaced showerhead MOVPE reactor, in *Proceedings of the 10th European Workshop on MOVPE*, Lecce (Italy), PS.IV.01(4pp) (2003)
611. R.A. Faulkner, Toward a theory of isoelectronic impurities in semiconductors. Phys. Rev. **175**, 991–1009 (1968)
612. P.J. Dean, Recombination processes associated with "Deep states" in gallium phosphide. J. Lumin. **1–2**, 398–419 (1970)
613. R. Schwabe, W. Seifert, F. Bugge, R. Bindemann, V.F. Agekyan, S.V. Pogarev, Photoluminescence of nitrogen-doped VPE GaAs. Solid State Commun. **55**, 167–173 (1985)
614. X. Liu, M.-E. Pistol, L. Samuelson, Excitons bound to nitrogen pairs in GaAs. Phys. Rev. B **42**, 7504–7512 (1990)
615. B. Gil, J.P. Albert, J. Camassel, H. Mathieu, C. Benoit à la Guillaume, Model calculation of nitrogen properties in III-V compounds. Phys. Rev. B **33**, 2701–2712 (1986)
616. D.G. Thomas, J.J. Hopfield, Isoelectronic traps due to nitrogen in gallium phosphide. Phys. Rev. **150**, 680–689 (1966)
617. W. Mönch, *Semiconductor Surfaces and Interfaces* (Springer, Berlin, 2001)
618. Y. Rosenwaks, R. Shikler, Th. Glatzel, S. Sadewasser, Kelvin probe force microscopy of semiconductor surface defects. Phys. Rev. B **70**, 085320(7pp) (2004)
619. E. Mollwo, Die Wirkung von Wasserstoff auf die Leitfähigkeit und Lumineszenz von Zinkoxydkristallen. Z. Phys. **138**, 478–488 (1954)
620. J.I. Pankove, N.M. Johnson (eds.), Hydrogen in semiconductors. Semicond. Semimet. **34** (1991)
621. S.J. Pearton, J.W. Corbett, M. Stavola, *Hydrogen in Crystalline Semiconductors* (Springer, Berlin, 1992)
622. T. Sakurai, H.D. Hagstrom, Hydrogen chemisorption on the silicon (110) 5×1 surface. J. Vac. Sci. Technol. **13**, 807–809 (1976)

623. J.I. Pankove, Photoluminescence recovery in rehydrogenated amorphous silicon. Appl. Phys. Lett. **32**, 812–813 (1978)
624. M.H. Brodsky, M. Cardona, J.J. Cuomo, Infrared and Raman spectra of the silicon-hydrogen bonds in amorphous silicon prepared by glow discharge and sputtering. Phys. Rev. B **16**, 3556–3571 (1977)
625. P.J.H. Denteneer, C.G. Van de Walle, S.T. Pantelides, Microscopic structure of hydrogen-boron complex in crystalline silicon. Phys. Rev. B **39**, 10809–10824 (1989)
626. C.P. Herrero, M. Stutzmann, Microscopic structure of boron-hydrogen complexes in crystalline silicon. Phys. Rev. B **38**, 12668–12671 (1988)
627. M. Stavola, K. Bergmann, S.J. Pearton, J. Lopata, Hydrogen motion in defect complexes: reorientation kinetics of the B-H complex in silicon. Phys. Rev. Lett. **61**, 2786–2789 (1988)
628. V.P. Markevich, A.R. Peaker, J. Coutinho, R. Jones, V.J.B. Torres, S. Öberg, P.R. Briddon, L.I. Murin, L. Dobaczewski, N.V. Abrosimov, Structure and properties of vacancy-oxygen complexes in $Si_{1-x}Ge_x$ alloys. Phys. Rev. B **69**, 125218(11pp) (2004)
629. V.P. Markevich, L.I. Murin, M. Suezawa, J.L. Lindström, J. Coutinho, R. Jones, P.R. Briddon, S. Öberg, Observation and theory of the V-O-H_2 complex in silicon. Phys. Rev. B **61**, 12964–12969 (2000)
630. A.S. Yapsir, P. Deák, R.K. Singh, L.C. Snyder, J.W. Corbett, T.-M. Lu, Hydrogen passivation of a substitutional sulfur defect in silicon. Phys. Rev. B **38**, 9936–9940 (1988)
631. K. Seeger, *Semiconductor Physics, An Introduction*, 9th edn. (Springer, Berlin, 2004)
632. M. Lundstrom, *Fundamentals of Carrier Transport*, 2nd edn. (Cambridge University Press, Cambridge, 2000)
633. H.-D. Liu, Y.-P. Zhao, G. Ramanath, S.P. Murarka, G.-C. Wang, Thickness dependent electrical resistivity of ultrathin (<40 nm) Cu films. Thin Solid Films **384**, 151–156 (2001)
634. C. Jacoboni, L. Reggiani, Bulk hot-electron properties of cubic semiconductors. Adv. Phys. **28**, 493–553 (1979)
635. E.M. Conwell, V. Weisskopf, Theory of impurity scattering in semiconductors. Phys. Rev. **77**, 388–390 (1950)
636. H. Brooks, Scattering by ionized impurities in semiconductors. Phys. Rev. **83**, 879 (1951)
637. B.K. Ridley, Reconciliation of the Conwell-Weisskopf and Brooks-Herring formulae for charged-impurity scattering in semiconductors: third-body interference. J. Phys. C: Solid State Phys. **10**, 1589–1593 (1977)
638. D.C. Chattopadhyay, H.J. Queisser, Electron scattering by ionized impurities in semiconductors. Rev. Mod. Phys. **53**, 745–768 (1981)
639. B.K. Ridley, *Quantum Processes in Semiconductors*, 2nd edn. (Clarendon, Oxford, 1988)
640. D.B.M. Klaasen, A unified model for device simulation-I. Model equations and concentration dependence. Solid State Electron. **35**, 953–959 (1992)
641. D.B.M. Klaasen, A unified model for device simulation-II. Temperature dependence of carrier mobility and lifetime. Solid State Electron. **35**, 961–967 (1992)
642. J.H. You, H.T. Johnson, Effect of dislocations on electrical and optical properties of GaAs and GaN. Solid State Phys. **61**, 143–261 (2009)
643. G.L. Pearson, W.T. Read, F.J. Morin, Dislocations in plastically deformed germanium. Phys. Rev. **93**, 666–667 (1954)
644. W.T. Read, Theory of dislocations in germanium. Philos. Mag. **45**, 775–796 (1954)
645. W.T. Read, Scattering of electrons by charged dislocations in semiconductors. Philos. Mag. **46**, 111–131 (1954)
646. B. Pödör, Electron mobility in plastically deformed germanium. Phys. Status Solidi **16**, K167–K170 (1966)
647. H.W. Choi, J. Zhang, S.J. Chua, Dislocation scattering in n-GaN. Mater. Sci. Semicond. Process. **4**, 567–570 (2001)
648. J.Y.W. Seto, The electrical properties of polycrystalline silicon films. J. Appl. Phys. **46**, 5247–5254 (1975)
649. C.H. Seager, T.G. Castner, Zero-bias resistance of grain boundaries in neutron-transmutation-doped polycrystalline silicon. J. Appl. Phys. **49**, 3879–3889 (1978)

650. M.R. Murti, K.V. Reddy, Grain boundary effects on the carrier mobility of polysilicon. Phys. Status Solidi A **119**, 237–240 (1990)
651. J.W. Orton, M.J. Powell, The Hall effect in polycrystalline and powdered semiconductors. Rep. Prog. Phys. **43**, 1263–1307 (1980)
652. N. Gupta, B.P. Tyagi, An analytical model of the influence of grain size on the mobility and transfer characteristics of polysilicon thin-film transistors (TFTs). Phys. Scr. **71**, 225–228 (2005)
653. R.L. Petritz, Theory of photoconductivity in semiconductor films. Phys. Rev. **104**, 1508–1516 (1956)
654. C.M. Wolfe, G.E. Stillman, W.T. Lindley, Electron mobility in high-purity GaAs. J. Appl. Phys. **41**, 3088–3091 (1970)
655. H. v. Wenckstern, S. Weinhold, G. Biehne, R. Pickenhain, H. Schmidt, H. Hochmuth, M. Grundmann, Donor levels in ZnO. Adv. Solid State Phys. (Festkörperprobleme) **45**, 263–274 (2006)
656. H.K. Sy, D.K. Desai, C.K. Ong, Electron screening and mobility in heavily doped silicon. Phys. Status Solidi B **130**, 787–792 (1985)
657. G. Masetti, M. Severi, S. Solmi, Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon. IEEE Trans. Electron Devices **ED-30**, 764–769 (1983)
658. C.S. Smith, Piezoresistance effect in germanium and silicon. Phys. Rev. **94**, 42–49 (1954)
659. R.W. Keyes, The effect of elastic deformation on the electrical conductivity of semiconductors. Solid State Phys. **11**, 149–221 (1960)
660. Y. Ohmura, Piezoresistance effect in p-type Si. Phys. Rev. B **42**, 9178–9181 (1990)
661. Y. Kanda, A graphical representation of the piezoresistance coefficients in silicon. IEEE Trans. Electron Devices' **ED-29**, 64–70 (1982)
662. M. Zerbst, Piezowiderstandseffekt in Galliumarsenid. Z. Naturf. **17a**, 649–651 (1962)
663. A. Sagar, Piezoresistance in n-type GaAs. Phys. Rev. **112**, 1533 (1958)
664. G. Bauer, Determination of electron temperatures and of hot electron distribution functions in semiconductors, in *Solid-State Physics*. Springer Tracts in Modern Physics, vol. 74 (Springer, Berlin, 1974), pp. 1–106
665. T. Kunikiyo, M. Takenaka, Y. Kamakura, M. Yamaji, H. Mizuno, M. Morifuji, K. Taniguchi, C. Hamaguchi, A Monte Carlo simulation of anisotropic electron transport in silicon including full band structure and anisotropic impact-ionization model. J. Appl. Phys. **75**, 297–312 (1994)
666. E.M. Conwell, *High Field Transport in Semiconductors* (Academic Press, New York, 1967)
667. C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, A. Alberigi-Quaranta, Electron drift velocity in silicon. Phys. Rev. B **12**, 2265–2284 (1975)
668. C. Jacoboni, C. Canali, G. Ottaviani, A. Alberigi-Quaranta, A review of some charge transport properties of silicon. Solid State Electron. **20**, 77–89 (1977)
669. C. Jacoboni, F. Nava, C. Canali, G. Ottaviani, Electron drift velocity and diffusivity in germanium. Phys. Rev. B **24**, 1014–1026 (1981)
670. T. González Sánchez, J.E. Velázquez Pérez, P.M. Gutiérrez Conde, D. Pardo Collantes, Electron transport in InP under high electric field conditions. Semicond. Sci. Technol. **7**, 31–36 (1992)
671. V. Balynas, A. Krotkus, A. Stalnionis, A.T. Gorelionok, N.M. Shmidt, J.A. Tellefsen, Time-resolved, hot-electron conductivity measurement using an electro-optic sampling technique. Appl. Phys. A **51**, 357–360 (1990)
672. J.D. Albrecht, P.P. Ruden, S. Limpijumnong, W.R. Lambrecht, K.F. Brennan, High field electron transport properties of bulk ZnO. J. Appl. Phys. **86**, 6864–6867 (1999)
673. B.K. Ridley, T.B. Watkins, The possibility of negative resistance effects in semiconductors. Proc. Phys. Soc. **78**, 293–304 (1961)
674. P.N. Butcher, The Gunn effect. Rep. Prog. Phys. **30**, 97–148 (1967)
675. T.K. Ishii (ed.), *Handbook of Microwave Technology* (Academic Press, San Diego, 1995)

676. J. Požela, A. Reklaitis, Electron transport properties in GaAs at high electric fields. Solid State Electron. **23**, 927–933 (1980)
677. B. Kramer, A. Mircea, Determination of saturated electron velocity in GaAs. Appl. Phys. Lett. **26**, 623–625 (1975)
678. S. Sridharan, P.D. Yoder, Anisotropic transient and stationary electron velocity in bulk wurtzite GaN. IEEE Electron Device Lett. **29**, 1190–1192 (2008)
679. C.L. Anderson, C.R. Crowell, Threshold energies for electron-hole pair production by impact ionization in semiconductors. Phys. Rev. B **5**, 2267–2272 (1972)
680. T. Pearsall, Threshold energies for impact ionization by electrons and holes in InP. Appl. Phys. Lett. **35**, 168–170 (1979)
681. Y. Kamakura, H. Mizuno, M. Yamaji, M. Morifuji, K. Taniguchi, C. Hamaguchi, T. Kunikiyo, M. Takenaka, Impact ionization model for full band Monte Carlo simulation. J. Appl. Phys. **75**, 3500–3506 (1994)
682. T. Kunikiyo, M. Takenaka, M. Morifuji, K. Taniguchi, C. Hamaguchi, A model of impact ionization due to the primary hole in silicon for a full band Monte Carlo simulation. J. Appl. Phys. **79**, 7718–7725 (1996)
683. A. Kuligk, N. Fitzer, R. Redmer, Ab initio impact ionization rate in GaAs, GaN, and ZnS. Phys. Rev. B **71**, 085201(6pp) (2005)
684. J.T. Devreese, Polarons. Encycl. Appl. Phys. **14**, 383–409 (1996)
685. D. Emin, *Polarons* (Cambrige University Press, Cambridge, 2013)
686. J. Appel, Polarons. Solid State Phys. **21**, 193–391 (1968)
687. R.P. Feynman, Slow electrons in a polar crystal. Phys. Rev. **97**, 660–665 (1955)
688. G.D. Mahan, *Many-Particle Physics* (Springer, New York, 2000)
689. H. Fröhlich, H. Pelzer, S. Zienau, Properties of slow electrons in polar materials. Philos. Mag. **41**, 221–242 (1950)
690. T.D. Schultz, Slow electrons in polar crystals: self-energy, mass, and mobility. Phys. Rev. **116**, 526–543 (1959)
691. S. Adachi, *Properties of Group-IV, III-V and II-VI Semiconductors* (Wiley, Chichester, 2005)
692. O.F. Schirmer, O^- bound small polarons in oxide materials. J. Phys.: Condensed. Matter **18**, R667–R704 (2006)
693. M. Albrecht, J. Varley, T. Remmele, Z. Gałaska, R. Uecker, C. Van de Walle, R. Fornari, In-situ observation of small polarons in gallium oxide by aberration corrected high resolution transmission electron microscopy, in *The 15th European Microscopy Congress*, Manchester, UK, PS2.3(2pp) (2012)
694. B.J. Morgan, D.O. Scanlon, G.W. Watson, Small polarons in Nb- and Ta-doped rutile and anatase TiO_2. J. Mater. Chem. **19**, 5175–5178 (2009)
695. S.J.F. Byrnes, Basic theory and phenomenology of polarons (2008), http://sjbyrnes.com/FinalPaper--Polarons.pdf
696. N.F. Mott, E.A. Davis, *Electronic Properties in Non-Crystalline Materials* (Clarendon Press, Oxford, 1971)
697. A.L. Efros, B.I. Shklovskii, *Electronic Properties of Doped Semiconductors* (Springer, Berlin, 1984)
698. M. Pollak, B.I. Shklovskii (eds.), *Hopping Transport in Solids* (Elsevier/North-Holland, Amsterdam, 1990)
699. J.J. Hauser, Electrical properties and anisotropy in amorphous Si and $Si_{0.5}Ge_{0.5}$ alloy. Phys. Rev. B **8**, 3817–3823 (1973)
700. N.F. Mott, Conduction in glasses containing transition metal ions. J. Non-Cryst. Solids **1**, 1–17 (1968)
701. A.L. Efros, B.I. Shklovskii, Coulomb gap and low temperature conductivity of disordered system. J. Phys. C: Solid State Phys. **8**, L49–L51 (1975)
702. A. Yildiz, N. Serin, T. Serin, M. Kasap, Crossover from nearest-neighbor hopping conduction to Efros-Shklovskii variable-range hopping conduction in hydrogenated amorphous silicon films. Jpn. J. Appl. Phys. **48**, 111203(5pp) (2009)
703. N. Mott, The mobility edge since 1967. J. Phys. C: Solid State Phys. **20**, 3075–3102 (1987)

704. B.C.H. Steele, A. Heinzel, Materials for fuel-cell technologies. Nature **414**, 345–352 (2001)
705. N. Sata, K. Eberman, K. Eberl, J. Maier, Mesoscopic fast ion conduction in nanometre-scale planar heterostructures. Nature **408**, 946–949 (2000)
706. J. Garcia-Barriocanal, A. Rivera-Calzada, M. Varela, Z. Sefrioui, E. Iborra, C. Leon, S.J. Pennycook, J. Santamaria, Colossal ionic conductivity at interfaces of epitaxial ZrO_2:Y_2O_3/$SrTiO_3$ heterostructures. Science **321**, 676–680 (2008)
707. J.X.M. Zheng-Johansson, R.L. McGreevy, A molecular dynamics study of ionic conduction in CuI. II. Local ionic motion and conduction mechanisms. Solid State Ionics **83**, 35–48 (1996)
708. M. Yashima, Q. Xu, A. Yoshiasa, S. Wada, Crystal structure, electron density and diffusion path of the fast-ion conductor copper iodide CuI. J. Mater. Chem. **16**, 4393–4396 (2006)
709. T. Jow, J.B. Wagner, On the electrical properties of cuprous iodide. J. Electrochem. Soc. **125**, 613–620 (1978)
710. P.T. Landsberg, On the diffusion theory of rectification. Proc. R. Soc. Lond. A **213**, 226–237 (1952)
711. N.G. Nilsson, An accurate approximation of the generalized einstein relation for degenerate semiconductors. Phys. Status Solidi A **19**, K75–K78 (1973)
712. E. Spenke, *Elektronische Halbleiter, Eine Einführung in die Physik der Gleichrichter und Transistoren* (Springer, Berlin, 1955)
713. F.A. Lindholm, R.W. Ayers, Generalized Einstein relation for degenerate semiconductors. Proc. IEEE **56**, 371–372 (1968)
714. H.S. Carslaw, J.C. Jaeger, *Conduction of Heat in Solids* (Clarendon Press, Oxford, 1959)
715. T.H. Geballe, G.W. Hull, Isotopic and other types of thermal resistance in germanium. Phys. Rev. **110**, 773–775 (1958)
716. W.S. Capinski, H.J. Maris, E. Bauser, I. Silier, M. Asen-Palmer, T. Ruf, M. Cardona, E. Gmelin, Thermal conductivity of isotopically enriched Si. Appl. Phys. Lett. **71**, 2109–2111 (1997)
717. T. Ruf, R.W. Henn, M. Asen-Palmer, E. Gmelin, M. Cardona, H.J. Pohl, G.G. Devyatykh, P.G. Sennikov, Thermal conductivity of isotopically enriched silicon. Solid State Commun. **115**, 243–247 (2000)
718. G.O. Mahan, Good thermoelectrics. Solid State Phys. **51**, 81–157 (1997)
719. H. Böttner, G. Chen, R. Venkatasubramanian, Aspects of thin-film superlattice thermoelectric materials, devices, and applications. MRS Bull. **31**, 211–217 (2006)
720. V.A. Johnson, K. Lark-Horowitz, Theory of thermoelectric power in semiconductors with applications to germanium. Phys. Rev. **92**, 226–232 (1953)
721. T.H. Geballe, G.W. Hull, Seebeck effect in silicon. Phys. Rev. **98**, 940–947 (1955)
722. H. Fritzsche, A general expression for the thermoelectric power. Solid State Commun. **9**, 1813–1815 (1971)
723. Z. Wang, S. Wang, S. Obukhov, N. Vast, J. Sjakste, V. Tjuterev, N. Mingo, Thermoelectric transport properties of silicon: toward an ab initio approach. Phys. Rev. B **83**, 205208(5pp) (2011)
724. G.D. Mahan, L. Lindsay, D.A. Broido, The Seebeck coefficient and phonon drag in silicon. J. Appl. Phys. **116**, 245102(7pp) (2014)
725. M. Bouguer, Essai d'optique sur la gradation de la lumiere (Claude Jombert, Paris, 1729)
726. A. Beer, Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten. Ann. Phys. Chem. **162**, 78–88 (1852)
727. N. ChM Wolfe, G.E. Holonyak Jr, *Stillman, Physical Properties of Semiconductors* (Prentice Hall, Englewood Cliffs, 1989)
728. B. Jogai, Absorption coefficient of wurtzite GaN calculated from an empirical tight binding model. Solid State Commun. **116**, 153–157 (2000)
729. M. Cardona, *Modulation Spectroscopy. Solid State Physics Supplement*, vol. 11 (Academic, New York, 1969)
730. M.D. Sturge, Optical absorption of gallium arsenide between 0.6 and 2.75 eV. Phys. Rev. **127**, 768–773 (1962); Phys. Rev. **129**, 2835 (1963) (erratum)

731. R.G. Ulbrich, Band edge spectra of highly excited gallium arsenide. Adv. Solid State Phys. (Festkörperprobleme) **25**, 299–307 (1985)
732. T. Matsuyama, H. Horinaka, K. Wada, T. Kondo, M. Hangyo, T. Nakanishi, S. Okumi, K. Togowa, Spin-dependent luminescence of highly polarized electrons generated by two-photon absorption in semiconductors. Jpn. J. Appl. Phys. **40**, L555–L557 (2001)
733. P.J. Dean, D.G. Thomas, Intrinsic absorption-edge spectrum of gallium phosphide. Phys. Rev. **150**, 690–703 (1966)
734. L.H. Hall, J. Bardeen, F.J. Blatt, Infrared absorption spectrum of germanium. Phys. Rev. **95**, 559–560 (1954)
735. J. Geist, A. Migdall, H.P. Baltes, Analytic representation of the silicon absorption coefficient in the indirect transition region. Appl. Opt. **27**, 3777–3779 (1988)
736. G.G. Macfarlane, V. Roberts, Infrared absorption of germanium near the lattice edge. Phys. Rev. **97**, 1714–1716 (1955)
737. G.G. Macfarlane, T.P. McLean, J.E. Quarrington, V. Roberts, Fine structure in the absorption-edge spectrum of Si. Phys. Rev. **111**, 1245–1254 (1958)
738. B.N. Brockhouse, Lattice vibrations in silicon and germanium. Phys. Rev. Lett. **2**, 256–258 (1959)
739. J.I. Pankove, P. Aigrain, Optical absorption of arsenic-doped degenerate germanium. Phys. Rev. **126**, 956–962 (1962)
740. G.A. Cox, G.G. Roberts, R.H. Tredgold, The optical absorption edge of barium titanate. British J. Appl. Phys. **17**, 743–745 (1966)
741. F. Urbach, The long-wavelength edge of photographic sensitivity and of the electronic absorption of solids. Phys. Rev. **92**, 1324 (1953)
742. S.R. Johnson, T. Tiedje, Temperature dependence of the Urbach edge in GaAs. J. Appl. Phys. **78**, 5609–5613 (1995)
743. M. Beaudoin, A.J.G. DeVries, S.R. Johnson, H. Laman, T. Tiedje, Optical absorption edge of semi-insulating GaAs and InP at high temperatures. Appl. Phys. Lett. **70**, 3540–3542 (1997)
744. T.S. Moss, T.D.F. Hawking, Infrared absorption in gallium arsenide. Infrared Phys. **1**, 111–115 (1961)
745. J. Stuke, Review of optical and electrical properties of amorphous semiconductors. J. Non-Cryst. Solids **4**, 1–26 (1970)
746. A. Baldereschi, N.O. Lipari, Energy levels of direct excitons in semiconductors with degenerate bands. Phys. Rev. B **3**, 439–451 (1971)
747. N.O. Lipari, M. Altarelli, Theory of indirect excitons in semiconductors. Phys. Rev. B **15**, 4883–4897 (1977)
748. N.O. Lipari, Exciton energy levels in wurtzite-type crystals. Phys. Rev. B **4**, 4535–4538 (1971)
749. D. Birkedal, J. Singh, V.G. Lyssenko, J. Erland, J.M. Hvam, Binding of quasi-two-dimensional biexcitons. Phys. Rev. Lett. **76**, 672–675 (1996)
750. E.F. Gross, Excitons and their motion in crystal lattices. Usp. Fiz. Nauk **76**, 433–466 (1962); (Sov. Phys. Usp. **5**, 195–218 (1962))
751. T. Kazimierczuk, D. Fröhlich, S. Scheel, H. Stolz, M. Bayer, Giant Rydberg excitons in the copper oxide Cu_2O. Nature **514**, 343–347 (2014)
752. Ch. Uihlein, D. Fröhlich, R. Kenklies, Investigation of exciton fine structure in Cu_2O. Phys. Rev. B **23**, 2731–2740 (1981)
753. R.J. Elliott, Intensity of optical absorption by excitons. Phys. Rev. **108**, 1384–1389 (1957)
754. A. Shikanai, T. Azuhata, T. Sota, S. Chichibu, A. Kuramata, K. Horino, S. Nakamura, Biaxial strain dependence of exciton resonance energies in wurtzite GaN. J. Appl. Phys. **81**, 417–424 (1997)
755. S. Rudin, T.L. Reinecke, B. Segall, Temperature-dependent exciton linewidths in semiconductors. Phys. Rev. B **42**, 11218–11231 (1990)
756. A.J. Fischer, W. Shan, J.J. Song, Y.C. Chang, R. Horning, B. Goldenberg, Temperature-dependent absorption measurements of excitons in GaN epilayers. Appl. Phys. Lett. **71**, 1981–1983 (1997)

757. A.J. Fischer, D.S. Kim, J. Hays, W. Shan, J.J. Song, D.B. Eason, J. Ren, J.F. Schetzina, H. Luo, J.K. Furdyna, Z.Q. Zhu, T. Yao, W. Schäfer, Femtosecond coherent spectroscopy of bulk ZnSe and ZnCdSe/ZnSe quantum wells. Phys. Rev. Lett. **73**, 2368–2371 (1994)
758. A.K. Viswanath, J.I. Lee, D. Kim, C.R. Lee, J.Y. Leem, Exciton-phonon interactions, exciton binding energy, and their importance in the realization of room-temperature semiconductor lasers based on GaN. Phys. Rev. B **58**, 16333–16339 (1998)
759. D.-S. Kim, J. Shah, J.E. Cunnigham, T.C. Damen, W. Schäfer, M. Hartmann, S. Schmitt-Rink, Giant exciton resonance in time-resolved four-wave mixing in quantum wells. Phys. Rev. Lett. **68**, 1006–1009 (1992)
760. R. Hauschild, H. Priller, M. Decker, J. Brückner, H. Kalt, C. Klingshirn, Temperature dependent band gap and homogeneous line broadening of the exciton emission in ZnO. Phys. Status Solidi C **3**, 976–979 (2006)
761. J.J. Hopfield, D.G. Thomas, Polariton absorption lines. Phys. Rev. Lett. **15**, 22–25 (1965)
762. J.J. Hopfield, Resonant scattering of polaritons as composite particles. Phys. Rev. **182**, 945–952 (1969)
763. A.A. Maradudin, D.L. Mills, Effect of spatial dispersion on the properties of a semi-infinite dielectric. Phys. Rev. B **7**, 2787–2810 (1973)
764. Y. Toyozawa, On the dynamical behavior of an exciton. Prog. Theor. Phys. Suppl. **12**, 111–140 (1959)
765. U. Heim, P. Wiesner, Direct evidence for a bottleneck of exciton-polariton relaxation in CdD. Phys. Rev. Lett. **30**, 1205–1207 (1973)
766. B. Gil, Oscillator strengths of A, B, and C excitons in ZnO films. Phys. Rev. B **64**, 201310(3pp) (2001)
767. T. Soma, H.-M. Kagaya, The metallic and ionic contributions to lattice vibrations of III-V covalent crystals. Phys. Status Solidi B **118**, 245–254 (1983)
768. A. Göldner, Nichtstrahlende Relaxationsprozesse bandkantennaher Zustände in II-VI- und III-V-Halbleiterstrukturen. Ph.D. thesis, TU Berlin (2000) (Wissenschaft und Technik Verlag, Berlin, 2000)
769. I. Broser, M. Rosenzweig, Determination of excitonic parameters of the A polariton in CdS from magnetoreflectance spectroscopy. Phys. Rev. B **22**, 2000–2007 (1980)
770. M. Rosenzweig, Exzitonische Polaritonen in CdS–Optische Eigenschaften räumlich dispersiver Medien. Ph.D. thesis, Technische Universität Berlin (1982)
771. G. Blattner, G. Kurtze, G. Schmieder, C. Klingshirn, Infuence of magnetic fields up to 20 T on excitons and polaritons in CdS and ZnO. Phys. Rev. B **25**, 7413–7427 (1982)
772. B. Gil, S. Clur, O. Briot, The exciton-polariton effect on the photoluminescence of GaN on sapphire. Solid State Commun. **104**, 267–270 (1997)
773. A.E. Yunovich, Radiative recombination and optical properties of GaP, in Radiative Recombination in Semiconductors, ed. by Ya.E. Pokrovskii (Moscow, 1972) (in Russian); A.E. Yunovich, Strahlende Rekombination und optische Eigenschaften von GaP. Fortschritte der Physik **23**, 317–398 (1975)
774. V. Vashishta, R.K. Kalia, Universal behavior of exchange-correlation energy in electron-hole liquid. Phys. Rev. B **25**, 6492–6495 (1982)
775. R. Zimmermann, Nonlinear optics and the Mott transition in semiconductors. Phys. Status Solidi B **146**, 371–384 (1988)
776. H.-E. Swoboda, M. Sence, F.A. Majumder, M. Rinker, J.-Y. Bigot, J.B. Grun, C. Klingshirn, Properties of electron-hole plasma in II-VI compounds as a function of temperature. Phys. Rev. B **39**, 11019–11027 (1989)
777. J.P. Löwenau, S. Schmitt-Rink, H. Haug, Many-body theory of optical bistability in semiconductors. Phys. Rev. Lett. **49**, 1511–1514 (1982)
778. L.V. Keldysh, in *Proceedings of the of the 9th International Conference on the Physics of Semiconductors, Moscow* (Nauka, Leningrad, 1968), pp. 1303–1312
779. W.F. Brinkman, T.M. Rice, P.W. Anderson, S.T. Chui, Metallic state of the electron-hole liquid. particularly in germanium. Phys. Rev. Lett. **28**, 961–964 (1972)

780. G.A. Thomas, T.M. Rice, J.C. Hensel, Liquid-gas phase diagram of an electron-hole fluid. Phys. Rev. Lett. **33**, 219–222 (1974)
781. T.L. Reinecke, S.C. Ying, Droplet model of electron-hole liquid condensation in semiconductors. Phys. Rev. Lett. **35**, 311–315 (1975)
782. R.S. Markiewicz, J.P. Wolfe, C.D. Jeffries, Strain-confined electron-hole liquid in germanium. Phys. Rev. B **15**, 1988–2005 (1977)
783. L.V. Butov, C.W. Lai, A.L. Ivanov, A.C. Gossard, D.S. Chemla, Towards Bose-Einstein condensation of excitons in potential traps. Nature **417**, 47–52 (2002)
784. L.V. Butov, A.C. Gossard, D.S. Chemla, Macroscopically ordered state in an exciton system. Nature **418**, 751–754 (2002)
785. K.E. O'Hara, L.Ó. Súilleabháin, J.P. Wolfe, Strong nonradiative recombination of excitons in Cu_2O and its impact on Bose-Einstein statistics. Phys. Rev. B **60**, 10565–10568 (1999)
786. M. Skolnick, A.I. Tartakovskii, R. Butté, R.M. Stevenson, J.J. Baumberg, D.M. Whittaker, High occupancy effects and condensation phenomena in semiconductor microcavities and bulk semiconductors, in *Nano-Optoelectronics, Concepts, Physics and Devices*, ed. by M. Grundmann (Springer, Berlin, 2002), pp. 273–296
787. H. Mahr, Two-photon absorption spectroscopy, in *Quantum Electronics: A Treatise*, ed. by H. Rabin, C.L. Tang, vol. I, Part A, (Academic, New York, 1975), pp. 285–361
788. H.J. Fossum, D.B. Chang, Two-photon excitation rate in indium antimonide. Phys. Rev. B **8**, 2842–2849 (1973)
789. J.P. van der Ziel, Two-photon absorption spectra of GaAs with $2\hbar\omega_1$ near the direct band gap. Phys. Rev. B **16**, 2775–2780 (1977)
790. C.J. Summers, R. Dingle, D.E. Hill, Far-infrared donor absorption and photoconductivity in epitaxial n-type GaAs. Phys. Rev. B **1**, 1603–1606 (1970)
791. T.M. ShM Kogan, Lifshits, Photoelectric spectroscopy-a new method of analysis of impurities in semiconductors. Phys. Status Solidi A **39**, 11–39 (1977)
792. L.T. Ho, A.K. Ramdas, Excitation spectra and piezospectroscopic effects of magnesium donors in Silicon. Phys. Rev. B **5**, 462–474 (1972)
793. R.A. Cooke, R.A. Hoult, R.F. Kirkman, R.A. Stradling, The characterisation of the donors in GaAs epitaxial films by far-infrared photoconductive techniques. J. Phys. D: Appl. Phys. **11**, 945–953 (1978)
794. B.L. Cardozo, E.E. Haller, L.A. Reichertz, J.W. Beemann, Far-infrared absorption in GaAs: Te liquid phase epitaxial films. Appl. Phys. Lett. **83**, 3990–3992 (2003)
795. M. Kleverman, K. Bergmann, H.G. Grimmeiss, Photothermal investigations of magnesium-related donors in silicon. Semicond. Sci. Technol. **1**, 49–52 (1986)
796. U. Fano, On the absorption spectrum of noble gases at the arc spectrum limit. Nuovo Cimento **12**, 156–161 (1935)
797. G. Breit, E. Wigner, Capture of slow neutrons. Phys. Rev. **49**, 519–531 (1936)
798. G. Lucovsky, On the photoionization of deep impurity centers in semiconductors. Solid State Commun. **3**, 299–302 (1965)
799. P. Drude, Zur Ionentheorie der Metalle. Z. Phys. **1**, 161–165 (1900)
800. C.R. Pidgeon, *Free Carrier Optical Properties of Semiconductors*. Handbook on Semiconductors, vol. 2 (North Holland, Amsterdam, 1980), pp. 223–328
801. H.Y. Fan, Effects of free carriers on the optical properties. Semicond. Semimet. **3**, 405–419 (1967)
802. R. von Baltz, W. Escher, Quantum theory of free carrier absorption. Phys. Status Solidi B **51**, 499–507 (1972)
803. P. Kleinert, M. Giehler, Theory of free-carrier infrared absorption in GaAs. Phys. Status Solidi B **136**, 763–777 (1986)
804. Z.G. Hu, M.B.M. Rinzan, S.G. Matsik, A.G.U. Perera, G. Von Winckel, A. Stintz, S. Krishna, Optical characterizations of heavily doped p-type $Al_xGa_{1-x}As$ and GaAs epitaxial films at terahertz frequencies. J. Appl. Phys. **97**, 093529(7pp) (2005)
805. H.R. Chandrasekhar, A.K. Ramdas, Nonparabolicity of the conduction band and the coupled plasmon-phonon modes in n-GaAs. Phys. Rev. B **21**, 1511–1515 (1980)

806. E. Burstein, Anomalous optical absorption limit in InSb. Phys. Rev. **93**, 632–633 (1954)
807. T.S. Moss, The interpretation of the properties of indium antimonide. Proc. Phys. Soc. B **76**, 775–782 (1954)
808. T.J. Coutts, D.L. Young, X. Li, Characterization of transparent conducting oxides. MRS Bull. **25**, 58–65 (2000)
809. G.N. Childs, S. Brand, R.A. Abram, Intervalence band absorption in semiconductor laser materials. Semicond. Sci. Technol. **1**, 116–120 (1986)
810. J. Taylor, V. Tolstikhin, Intervalence band absorption in InP and related materials for optoelectronic device modeling. J. Appl. Phys. **87**, 1054–1059 (2000)
811. R. Braunstein, E.O. Kane, The valence band structure of the III-V compounds. J. Phys. Chem. Solids **23**, 1423–1431 (1962)
812. A. Chandola, R. Pino, P.S. Dutta, Below bandgap optical absorption in tellurium-doped GaSb. Semicond. Sci. Technol. **20**, 886–893 (2005)
813. W.P. Dumke, M.R. Lorenz, G.D. Pettit, Intra- and interband free-carrier absorption and the fundamental absorption edge in n-type InP. Phys. Rev. B **1**, 4668–4673 (1970)
814. H. Peelaers, E. Kioupakis, C.G. Van de Walle, Fundamental limits on optical transparency of transparent conducting oxides: free-carrier absorption in SnO_2. Appl. Phys. Lett. **100**, 011914(3pp) (2012)
815. A. Schleife, J.B. Varley, F. Fuchs, C. Rödl, F. Bechstedt, P. Rinke, A. Janotti, C.G. Van de Walle, Tin dioxide from first principles: quasiparticle electronic states and optical properties. Phys. Rev. B **83**, 035116(9pp) (2011)
816. R. Zallen, *Crystal Structures*. Handbook on Semiconductors, vol. 1 (North Holland, Amsterdam, 1980), pp. 1–27
817. W.G. Spitzer, Multiphonon lattice absorption. Semicond. Semimet. **3**, 17–69 (1967)
818. E.S. Koteles, W.R. Datars, Two-phonon absorption in InSb, InAs, and GaAs. Can. J. Phys. **54**, 1676–1682 (1976)
819. A. Mooradian, G.B. Wright, Observation of the interaction of plasmons with longitudinal optical phonons in GaAs. Phys. Rev. Lett. **16**, 999–1001 (1966)
820. P.T. Landsberg, *Recombination in Semiconductors* (Cambridge University Press, Cambridge, 1991)
821. E. Göbel, Recombination without k-selection rules in dense electron-hole plasmas in high-purity GaAs lasers. Appl. Phys. Lett. **24**, 492–494 (1974)
822. X. Sun, J. Liu, L.C. Kimerling, J. Michel, Direct gap photoluminescence of n-type tensile-strained Ge-on-Si. Appl. Phys. Lett. **95**, 011911(3pp) (2009)
823. J. Shah, R.C.C. Leite, Radiative recombination from photoexcited hot carriers in GaAs. Phys. Rev. Lett. **22**, 1304–1307 (1969)
824. W. Michaelis, M. Pilkuhn, Radiative recombination in silicon p-n junctions. Phys. Status Solidi **36**, 311–319 (1969)
825. W. Shockley, W.T. Read Jr, Statistics of the recombination of holes and electrons. Phys. Rev. **87**, 835–842 (1952)
826. R.N. Hall, Electron-hole recombination in germanium. Phys. Rev. **87**, 387 (1952)
827. V.K. Malyutenko, Negative luminescence in semiconductors: a retrospective view. Phys. E **20**, 553–557 (2004)
828. J.F. Muth, J.H. Lee, I.K. Shmagin, R.M. Kolbas, H.C. Casey Jr, B.P. Keller, U.K. Mishra, S.P. DenBaars, Absorption coefficient, energy gap, exciton binding energy, and recombination lifetime of GaN obtained from transmission measurements. Appl. Phys. Lett. **71**, 2572–2774 (1997)
829. W. Gerlach, H. Schlangenotto, H. Maeder, On the radiative recombination rate in silicon. Phys. Status Solidi A **13**, 277–283 (1972)
830. A. Galeckas, J. Linnros, V. Grivickas, U. Lindefelt, C. Hallin, Auger recombination in 4H-SiC: unusual temperature behavior. Appl. Phys. Lett. **71**, 3269–3271 (1997)
831. V. Palankovski, Simulation of heterojunction bipolar transistors. Ph.D. thesis, Technische Universität Wien (2002)

832. R.K. Akrenkiel, Minority-carrier lifetime in III-V semiocnductors. Semicond. Semimet. **39**, 39–150 (1993)
833. P.J. Dean, J.R. Haynes, W.F. Flood, New radiative recombination processes involving neutral donors and acceptors in silicon and germanium. Phys. Rev. **161**, 711–729 (1967)
834. G. Davies, The optical properties of luminescence centres in silicon. Phys. Rep. **176**, 83–188 (1989)
835. G.D. Gilliland, Photoluminescence spectroscopy of crystalline semiconductors. Mater. Sci. Eng. R **18**, 99–400 (1997)
836. P.J. Dean, D.C. Herbert, Bound excitons in semiconductors, in *Excitons*, ed. by K. Cho (Springer, Berlin, 1979), pp. 55–182
837. P.J. Dean, Lithium donors and the binding of excitons at neutral donors and acceptors in gallium phosphide, in *Luminescence of Crystals, Molecules and Solutions*, ed. by F. Williams (Plenum, New York, 1973), pp. 538–552
838. S. Permogorov, A. Reznitsky, A. Naumov, H. Stolz, W. von der Osten, Localisation of excitons at small Te clusters in diluted $ZnSe_{1-x}Te_x$ solid solutions. J. Phys.: Condens. Matter **1**, 5125–5137 (1989)
839. T. Skettrup, M. Suffczynski, W. Gorzkowski, Properties of excitons bound to ionized donors. Phys. Rev. B **4**, 512–517 (1971)
840. R.G. Ulbrich, Low density photoexcitation phenomena in semiconductors: aspects of theory and experiment. Solid State Electron. **21**, 51–59 (1978)
841. D.E. Hill, Exciton recombination radiation of GaAs:Zn. Phys. Rev. B **1**, 1863–1864 (1970)
842. D.C. Reynolds, D.C. Look, B. Jogai, V.M. Phanse, R.P. Vaudo, Identification of an ionized-donor-bound-exciton transition in GaN. Solid State Commun. **103**, 533–535 (1997)
843. N. Nepal, M.L. Nakarmi, K.B. Nam, J.Y. Lin, H.X. Jiang, Acceptor-bound exciton transition in Mg-doped AlN epilayer. Appl. Phys. Lett. **85**, 2271–2273 (2004)
844. D.G. Thomas, J.J. Hopfield, Optical properties of bound exciton complexes in cadmium sulfide. Phys. Rev. **128**, 2135–2148 (1962)
845. J.L. Merz, H. Kukimoto, K. Nassau, J.W. Shiever, Optical properties of substitutional donors in ZnSe. Phys. Rev. B **6**, 545–556 (1972)
846. P.J. Dean, D.C. Herbert, C.J. Werkhoven, B.J. Fitzpatrick, R.N. Bhargava, Donor bound-exciton excited states in zinc selenide. Phys. Rev. B **23**, 4888–4901 (1981)
847. B.K. Meyer, J. Sann, S. Lautenschläger, M.R. Wagner, A. Hoffmann, Ionized and neutral donor-bound excitons in ZnO. Phys. Rev. B **76**, 184120(4pp) (2007)
848. J.R. Haynes, Experimental proof of the existence of a new electronic complex in silicon. Phys. Rev. Lett. **4**, 361–363 (1960)
849. B.K. Meyer, H. Alves, D.M. Hofmann, W. Kriegseis, D. Forster, F. Bertram, J. Christen, A. Hoffmann, M. Straßburg, M. Dworzak, U. Haboeck, A.V. Rodina, Bound exciton and donor-acceptor pair recombinations in ZnO. Phys. Status Solidi B **241**, 231–260 (2004)
850. S. Müller, D. Stichtenoth, M. Uhrmacher, H. Hofsäss, C. Ronning, J. Röder, Unambiguous identification of the PL-I_9 line in zinc oxide. Appl. Phys. Lett. **90**, 012107(3pp) (2007)
851. P.J. Dean, M. Skolnick, Donor discrimination and bound exciton spectra in InP. J. Appl. Phys. **54**, 346–359 (1983)
852. F.A.J.M. Driessen, H.G.M. Lochs, S.M. Olsthoorn, L.J. Giling, An analysis of the two electron satellite spectrum of GaAs in high magnetic fields. J. Appl. Phys. **69**, 906–912 (1991)
853. D. Karaiskaj, M.L.W. Thewalt, T. Ruf, M. Cardona, H.-J. Pohl, G.G. Deviatych, P.G. Senniko, H. Riemann, Photoluminescence of isotopically purified silicon: how sharp are bound exciton transitions? Phys. Rev. Lett. **86**, 6010–6013 (2001)
854. D. Karaiskaj, M.L.W Thewalt, T. Ruf, M. Cardona, M. Konuma, "Intrinsic" acceptor ground state splitting in silicon: an isotopic effect. Phys. Rev. Lett. **89**, 016401(4pp) (2002)
855. V.A. Karasyuk, M.L.W. Thewalt, S. An, E.C. Lightowlers, A.S. Kaminskii, Fourier-transform photoluminescence spectroscopy of excitons bound to group-III acceptors in silicon: zeeman effect. Phys. Rev. B **54**, 10543–10558 (1996)

856. M.L.W. Thewalt, A. Yang, M. Steger, D. Karaiskaj, M. Cardona, H. Riemann, N.V. Abrosimov, A.V. Gusev, A.D. Bulanov, I.D. Kovalev, A.K. Kaliteevskii, O.N. Godisov, P. Becker, H.J. Pohl, E.E. Haller, J.W. Ager III, K.M. Itoh, Direct observation of the donor nuclear spin in a near-gap bound exciton transition: ^{31}P in highly enriched ^{28}Si. J. Appl. Phys. **101**, 081724(5pp) (2007)
857. J.D. Cuthbert, D.G. Thomas, Fluorescent decay times of excitons bound to isoelectronic traps in GaP and ZnTe. Phys. Rev. **154**, 763–771 (1967)
858. J.P. Laurenti, P. Roentgen, K. Wolter, K. Seibert, H. Kurz, J. Camassel, Indium-doped GaAs: a very dilute alloy system. Phys. Rev. B **37**, 4155–4163 (1988)
859. D. Bimberg, M. Sondergeld, E. Grobe, Thermal dissociation of excitons bound to neutral acceptors in high-purity GaAs. Phys. Rev. B **4**, 3451–3455 (1971)
860. M. Grundmann, C.P. Dietrich, Lineshape theory of photoluminescence from semiconductor alloys. J. Appl. Phys. **106**, 123521(10pp) (2009)
861. D.G. Chtchekine, Z.C. Feng, S.J. Chua, G.D. Gilliland, Temperature-varied photoluminescence and magnetospectroscopy study of near-band-edge emissions in GaN. Phys. Rev. B **63**, 125211(7pp) (2001)
862. V.D. Kulakovskiĭ G.E. Pikus, V.B. Timofeev, Multiexciton complexes in semiconductors. Usp. Fiz. Nauk **135**, 237–284 (1981) (Sov. Phys. Usp. **24**, 815–840 (1981))
863. A.S. Kaminskiĭ Ya.E. Pokrovskiĭ Recombination radiation of the condensed phase on nonequlibrium carriers in silicon. Pis'ma Zh. Eksp. Teor. Fiz. **11**, 381 (1970) (JETP Lett. **11**, 255–257 (1970))
864. M.L.W. Thewalt, Details of the structure of bound excitons and bound multiexciton complexes in Si. Can. J. Phys. **55**, 1463–1480 (1977)
865. M.L.W. Thewalt, J.A. Rostworowski, G. Kirczenow, Piezospectroscopic studies of phosphorus-, boron-, and lithium-doped silicon. Can. J. Phys. **57**, 1898–1923 (1979)
866. O. Goede, L. John, D. Hennig, Compositional disorder-induced broadening for free excitons in II-VI semiconducting mixed crystals. Phys. Status Solidi B **89**, K183–K186 (1978)
867. E.F. Schubert, E.O. Göbel, Y. Horikoshi, K. Ploog, H.J. Queisser, Alloy broadening in photoluminescence spectra of $Al_xGa_{1-x}As$. Phys. Rev. B **30**, 813–820 (1984)
868. S. Heitsch, G. Zimmermann, D. Fritsch, C. Sturm, R. Schmidt-Grund, C. Schulz, H. Hochmuth, D. Spemann, G. Benndorf, B. Rheinländer, Th. Nobis, M. Lorenz, M. Grundmann, Luminescence and surface properties of $Mg_xZn_{1-x}O$ thin films grown by pulsed laser deposition. J. Appl. Phys. **101**, 083521(6pp) (2007)
869. R. Zimmermann, Theory of the exciton linewidth in II-VI semiconductor mixed crystals. J. Cryst. Growth **101**, 346–349 (1990)
870. J.M. Langer, R. Buczko, A.M. Stoneham, Alloy broadening of the near-gap luminescence and the natural band offset in semiconductor alloys. Semicond. Sci. Technol. **7**, 547–551 (1992)
871. A. Müller, M. Stölzel, G. Benndorf, M. Lorenz, M. Grundmann, Origin of the near-band-edge luminescence in $Mg_xZn_{1-x}O$ alloys. J. Appl. Phys. **107**, 013704(6pp) (2010)
872. M.C. Wagener, G.R. James, A.W.R. Leitch, F. Omnès, On the nature of Si-doping in AlGaN alloys. Phys. Status Solidi C **1**, 2322–2327 (2004)
873. B. Segall, G.D. Mahan, Phonon-assisted recombination of free excitons in compound semiconductors. Phys. Rev. **171**, 935–948 (1968)
874. J. Conradi, R.R. Haering, Oscillatory exciton emission in CdS. Phys. Rev. Lett. **20**, 1344–1346 (1968)
875. Y.S. Park, J.R. Schneider, Oscillations in exciton emission in the excitation spectra of ZnSe and CdS. Phys. Rev. Lett. **21**, 798–800 (1968)
876. D. Kovalev, B. Averboukh, D. Volm, B.K. Meyer, H. Amano, I. Akasaki, Free exciton emission in GaN. Phys. Rev. B **54**, 2518–2522 (1996)
877. S. Permogorov, Optical emission due to exciton scattering by LO phonons in semiconductors, in *Excitons*, ed. by E.I. Rashba, M.D. Sturge (North-Holland, 1982)
878. M. Wojdak, A. Wysmołek, K. Pakuła, J.M. Baranowski, Emission due to exciton scattering by LO-phonons in gallium nitride. Phys. Status Solidi B **216**, 95–99 (1999)

879. R. Dingle, Luminescent transitions associated with divalent copper impurities and the green emission from semiconducting Zinc oxide. Phys. Rev. Lett. **23**, 579–581 (1969)
880. Th. Agne, Identifikation und Untersuchung von Defekten in ZnO Einkristallen. Ph.D. thesis, Universität des Saarlandes, Saarbrücken (2004)
881. K. Huang, A. Rhys, Theory of light absorption and non-radiative transitions in F-centres. Proc. R. Soc. Lond. A **204**, 406–423 (1950)
882. J.J. Hopfield, A theory of edge-emission phenomena in CdS, ZnS and ZnO. J. Phys. Chem. Solids **10**, 110–119 (1959)
883. M. Lax, The Franck-Condon principle and its application to crystals. J. Chem. Phys. **20**, 1752–1760 (1952)
884. J.L. Merz, Isoelectronic oxygen trap in ZeTe. Phys. Rev. **176**, 961–968 (1968)
885. C. Klingshirn, The luminescence of ZnO under high one- and two-quantum excitation. Phys. Status Solidi B **71**, 547–556 (1975)
886. P.J. Dean, C.H. Henry, C.J. Frosch, Infrared donor-acceptor pair spectra involving the deep oxygen donor in gallium phosphide. Phys. Rev. **168**, 812–816 (1968)
887. A. Juhl, Calorimetrische Absorptionsspektroskopie (CAS) - Eine neue Methode zur Charakterisierung der optischen Eigenschaften von Halbleitersystemen. Ph.D. thesis, Technische Universität Berlin (1987)
888. W.H. Koschel, U. Kaufmann, S.G. Bishop, Optical an ESR analysis of the Fe acceptor in InP. Solid State Commun. **21**, 1069–1072 (1977)
889. W. Lochmann, A. Haug, Phonon-assisted Auger recombination in Si with direct calculation of the overlap integrals. Solid State Commun. **35**, 553–556 (1980)
890. M. Takeshima, Theory of phonon-assisted Auger recombination in semiconductors. Phys. Rev. B **23**, 6625–6637 (1981)
891. M. Takeshima, Phonon-assisted Auger recombination in a quasi-two-dimensional structure semiconductor. Phys. Rev. B **30**, 3302–3308 (1984)
892. P.C. Findlay, C.R. Pidgeon, H. Pellemans, R. Kotitschke, B.N. Murdin, T. Ashley, A.D. Johnson, A.M. White, C.T. Elliott, Auger recombination dynamics of $In_xGa_{1-x}Sb$. Semicond. Sci. Technol. **14**, 1026–1030 (1999)
893. D. Vignaud, J.F. Lampin, E. Lefebvre, M. Zaknoune, F. Mollot, Electron lifetime of heavily Be-doped $In_{0.53}Ga_{0.47}As$ as a function of growth temperature and doping density. Appl. Phys. Lett. **80**, 4151–4153 (2002)
894. D.B. Laks, G.F. Neumark, S.T. Pantelides, Accurate interband-Auger-recombination rates in silicon. Phys. Rev. B **42**, 5176–5185 (1990)
895. P. Blood, J. Orton, *The Electrical Characterization of Semiconductors: Majority Carriers and Electron States* (Academic Press, San Diego, 1992)
896. D. Macdonald, A. Cuevas, Validity of simplified Shockley-Read-Hall statistics for modeling carrier lifetimes in crystalline silicon. Phys. Rev. B **67**, 075203(7pp) (2003)
897. D.H. Auston, Picosecond photoconductivity: high-speed measurements of devices and materials. Semicond. Semimet. **28**, 85–134 (1990)
898. A.A. Istratov, E.R. Weber, Electrical properties and recombination activity of copper, nickel and cobalt in silicon. Appl. Phys. A **66**, 123–136 (1998)
899. A.A. Istratov, H. Hieslmair, E.R. Weber, Iron and its complexes in silicon. Appl. Phys. A **69**, 13–44 (1999)
900. A.A. Istratov, H. Hieslmair, E.R. Weber, Iron contamination in silicon technology. Appl. Phys. A **70**, 489–534 (2000)
901. J.S. Blakemore, Lifetime in p-type silicon. Phys. Rev. **110**, 1301–1308 (1958)
902. M.S. Tyagi, R. van Overstraeten, Minority carrier recombination in heavily doped silicon. Solid State Electron. **26**, 577–597 (1983)
903. G. Bemski, Lifetime of electrons in p-type silicon. Phys. Rev. **100**, 523–524 (1955)
904. Q. Dai, Q. Shan, J. Wang, S. Chhajed, J. Cho, E.F. Schubert, M.H. Crawford, D.D. Koleske, M.-H. Kim, Y. Park, Carrier recombination mechanisms and efficiency droop in GaInN/GaN light-emitting diodes. Appl. Phys. Lett. **97**, 133507(3pp) (2010)

905. S. Karpov, ABC-model for interpretation of internal quantum efficiency and its droop in III-nitride LEDs: a review. Opt. Quantum Electron. **47**, 1293–1303 (2015)
906. J. Frenkel, On pre-breakdown phenomena in insulators and electronic semi-conductors. Phys. Rev. **54**, 647–648 (1938)
907. R.D. Harris, J.L. Newton, G.D. Watkins, Negative-U properties for interstitial boron in silicon. Phys. Rev. Lett. **48**, 1271–1274 (1982)
908. S. Bothra, S. Tyagi, S.K. Chandhi, J.M. Borrego, Surface recombination velocity and lifetime in InP. Solid State Electron. **34**, 47–50 (1991)
909. J. Schmidt, A.G. Aberle, Accurate method for the determination of bulk minority-carrier lifetimes of mono- and multicrystalline silicon wafers. J. Appl. Phys. **81**, 6186–6199 (1997)
910. M.J. Kerr, J. Schmidt, A. Cuevas, J.H. Bultman, Surface recombination velocity of phosphorus-diffused silicon solar cell emitters passivated with plasma enhanced chemical vapor deposited silicon nitride and thermal silicon oxide. J. Appl. Phys. **89**, 3821–3826 (2001)
911. O. Hahneiser, M. Kunst, Theoretical and experimental study of charge carrier kinetics in crystalline silicon. J. Appl. Phys. **85**, 7741–7754 (1999)
912. D.E. Aspnes, Recombination at semiconductor surfaces and interfaces. Surf. Sci. **132**, 406–421 (1983)
913. C. Donolato, Theory of beam induced current characterization of grain boundaries in polycrystalline solar cells. J. Appl. Phys. **54**, 1314–1322 (1983)
914. T. Kieliba, S. Riepe, W. Warta, Effect of dislocations on minority carrier diffusion length in practical silicon solar cells. J. Appl. Phys. **100**, 063706(12pp) (2006)
915. J. Palm, Local investigation of recombination at grain boundaries in silicon by grain boundary-electron beam induced current. J. Appl. Phys. **74**, 1169–1178 (1993)
916. T.F. Ciszek, T.H. Wang, R.W. Burrows, X. Wu, J. Alleman, Y.S. Tsuo, T. Bekkedahl, *Grain boundary and dislocation effects on the PV performance of high-purity silicon, in 23th IEEE Photovoltaic Specialists Conference Record* (IEEE, New York, 1993), pp. 101–105
917. R. Corkish, T. Puzzer, A.B. Sproul, K.L. Luke, Quantitative interpretation of electron-beam-induced current grain boundary contrast profiles with application to silicon. J. Appl. Phys. **84**, 5473–5481 (1998)
918. A.D. Kurtz, S.A. Kulin, B.L. Averbach, Effect of dislocations on the minority carrier lifetime in semiconductors. Phys. Rev. **101**, 1285–1291 (1956)
919. E.A. Fitzgerald, D.G. Ast, P.D. Kirchner, G.D. Pettit, J.M. Woodall, Structure and recombination in InGaAs/GaAs heterostructures. J. Appl. Phys. **63**, 693–703 (1988)
920. influence on lifetime and diffusion of excess excitons, M. Grundmann, J. Christen, D. Bimberg, A. Fischer-Colbrie, R. Hull, Misfit dislocations in pseudomorphic $In_{0.23}Ga_{0.77}As$/GaAs quantum wells. J. Appl. Phys. **66**, 2214–2216 (1989)
921. M. Ni, M.K.H. Leung, D.Y.C. Leung, K. Sumathy, A review and recent developments in photocatalytic water-splitting using TiO_2 for hydrogen production. Renew. Sustain. Energy Rev. **11**, 401–425 (2007)
922. S. Cho, J.-W. Jang, K.-H. Lee, J.S. Lee, Research update: strategies for efficient photoelectrochemical water splitting using metal oxide photoanodes. APL Mater. **2**, 010703(14pp) (2014)
923. J. Watson, The tin oxide gas sensor and its applications. Sens. Act. **5**, 29–42 (1984)
924. G. Eranna, B.C. Joshi, D.P. Runthala, R.P. Gupta, Oxide materials for development of integrated gas sensors-a comprehensive review. Crit. Rev. Solid State Mater. Sci. **29**, 111–188 (2004)
925. X. Liu, S. Cheng, H. Liu, S. Hu, D. Zhang, H. Ning, A survey on gas sensing technology. Sensors **12**, 9635–9665 (2012)
926. H. Ibach, *Physics of Surfaces and Interfaces* (Springer, Berlin, 2006)
927. A. Zangwill, *Physics at Surfaces* (Cambridge University Press, Cambridge, 2012)
928. R.H. Kingston, *Semiconductor Surface Physics* (Literary Licensing, Whitefish, 2012)
929. P.J. Morandi, *Symmetry Groups: The Classification of Wallpaper Patterns, From Group Cohomology to Escher's Tessellations* (New Mexico State University 2007)

930. A.A. Stekolnikov, F. Bechstedt, Shape of free and constrained group-IV crystallites: influence of surface energies. Phys. Rev. B **72**, 125326(9pp) (2005)
931. J.M. Bermond, J.J. Métois, X. Egéa, F. Floret, The equilibrium shape of silicon. Surf. Sci. **330**, 48–60 (1995)
932. C.B. Duke, Semiconductor surface reconstruction: the structural chemistry of two-dimensional durface compounds. Chem. Rev. **96**, 1237–1259 (1996)
933. J. Fritsch, U. Schröder, Density functional calculation of semiconductor surface phonons. Phys. Rep. **309**, 209–331 (1999)
934. K. Hermann, Fritz-Haber-Institut, Berlin
935. W.G. Schmidt, F. Bechstedt, J. Bernholc, GaAs(001) surface reconstructions: geometries, chemical bonding and optical properties. Appl. Surf. Sci. **190**, 264–268 (2002)
936. K. Takayanagi, Y. Tanishiro, S. Takahashi, M. Takahashi, Surface analysis of Si(111)-7×7. Surf. Sci. **164**, 367–392 (1985)
937. R. Wiesendanger, G. Tarrach, L. Scandella, H.-J. Güntherodt, scanning tunneling microscopy on laser- and thermal-annealed Si(111): transitions from 7×7 reconstructed to disordered surface structures. Ultramicroscopy **32**, 291–295 (1990)
938. F.J. Giessibl, S. Hembacher, H. Bielefeldt, J. Mannhart, Subatomic features on the silicon (111)-(7×7) surface observed by atomic force microscopy. Science **289**, 422–425 (2000)
939. B.S. Swartzentruber, N. Kitamura, M.G. Lagally, M.B. Webb, Behavior of steps on Si(001) as a function of vicinality. Phys. Rev. B **47**, 13432–13441 (1993)
940. O. Dulub, L.A. Boatner, U. Diebold, STM study of the geometric and electronic structure of ZnO(0001)-Zn, (000$\bar{1}$)-O, (10$\bar{1}$0), and (11$\bar{2}$0) surfaces. Surf. Sci. **519**, 201–217 (2002)
941. J.J. Métois, A. Saúl, P. Müller, Measuring the surface stress polar dependence. Nat. Mater. **4**, 238–242 (2005)
942. M. Syväjärvi, R. Yakimova, E. Janzén, Step-bunching in SiC epitaxy: anisotropy and influence of growth temperature. J. Cryst. Growth **236**, 297–304 (2002)
943. A.A. Baski, S.C. Erwin, L.J. Whitman, The structure of silicon surfaces from (001) to (111). Surf. Sci. **392**, 69–85 (1997)
944. R.H. Ritchie, Plasma losses by fast electrons in thin films. Phys. Rev. **106**, 874–881 (1957)
945. J.M. Pitarke, V.M. Silkin, E.V. Chulkov, P.M. Echenique, Theory of surface plasmons and surface-plasmon polaritons. Rep. Prog. Phys. **70**, 1–87 (2007)
946. J. Zhang, L. Zhang, W. Xu, Surface plasmon polaritons: physics and applications. J. Phys. D: Appl. Phys. **45**, 113001(19pp) (2012)
947. https://www.wikipedia.de
948. S. Kalusniak, S. Sadofev, F. Henneberger, ZnO as a tunable metal: new types of surface plasmon polaritons. Phys. Rev. Lett. **112**, 137401 (2014)
949. J.E. Northrup, M.S. Hybertsen, S.G. Louie, Many-body calculation of the surface-state energies for Si(111)2×1. Phys. Rev. Lett. **66**, 500–503 (1991)
950. I. Ivanov, A. Mazur, J. Pollmann, The ideal (111), (110), and (100) surfaces of Si, Ge and GaAs: a comparison of their electronic structure. Surf. Sci. **92**, 365–384 (1980)
951. R.I.G. Uhrberg, G.V. Hansson, Electronic structure of silicon surfaces: clean and with ordered overlayers. Crit. Rev. Solid State Mater. Sci. **17**, 133–186 (1991)
952. P. Mårtensson, W.-X. Ni, G.V. Hansson, J.M. Nicholls, B. Reihl, Surface electronic structure of Si(111)7×7-Ge and Si(111)5×5-Ge studied with photoemission and inverse photoemission. Phys. Rev. B **36**, 5974–5981 (1987)
953. S.R. Varier, P.S. Mandal, N. Sahadev, D.N. Biswas, K. Maiti, Study of the surface electronic structure of Si(111) surface using spin and angle resolved photoemission spectroscopy. AIP Conf. Proc. **1512**, 818–819 (2013)
954. S.-Q. Shen, *Topological Insulators, Dirac Equation in Condensed Matter* (Springer, Heidelberg, 2012)
955. Z. Zhu, Y. Cheng, U. Schwingenschlögl, Band inversion mechanism in topological insulators: a guideline for materials design. Phys. Rev. B **85**, 235401(5pp) (2012)
956. L. Fu, C.L. Kane, E.J. Mele, Topological insulators in three dimensions. Phys. Rev. Lett. **98**, 106803(4pp) (2007)

957. H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Zh Fang, S-Ch. Zhang, Topological insulators in Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3 with a single Dirac cone on the surface. Nat. Phys. **5**, 438–442 (2009)
958. M.Z. Hasan, C.L. Kane, Topological insulators. Rev. Mod. Phys. **82**, 3045–3067 (2010)
959. Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y.S. Hor, R.J. Cava, M.Z. Hasan, Observation of a large-gap topological-insulator class with a single Dirac cone on the surface. Nat. Phys. **5**, 398–402 (2009)
960. Y. Xia, Photoemission studies of a new topological insulator class: experimental discovery of the Bi_2X_3 topological insulator class. Ph.D. thesis, Princeton University (2010)
961. H. Kroemer, *Quasi-Electric Fields and Band-Offsets: Teaching Electrons New Tricks*. Nobel Lecture, vol. 8 (Dec 2000)
962. J.E. Ayers, *Heteroepitaxy of Semiconductors-Theory, Growth, Characterization* (CRC Press, Boca Raton, 2007)
963. U. Pohl, *Epitaxy of Semiconductors* (Springer, Berlin, 2013)
964. G. Dhanaraj, K. Byrappa, V. Prasad, M. Dudley (eds.), *Springer Handbook of Crystal Growth* (Springer, Heidelberg, 2010)
965. J.Y. Tsao, *Materials Fundamentals of Molecular Beam Epitaxy* (Academic Press, San Diego, 1993)
966. M.L. Hitchman, K.F. Jensen (eds.), *Chemical Vapor Deposition* (Academic Press, San Diego, 1993)
967. P. George, *Chemical Vapor Deposition: Simulation and Optimization* (VDM Verlag Dr. Mueller, Saarbrücken, 2008)
968. G. Stringfellow, *Organometallic Vapor-Phase Epitaxy: Theory and Practice* (Academic Press, San Diego, 1999)
969. R. Eason (ed.), *Pulsed Laser Deposition of Thin Films: Applications-Led Growth of Functional Materials* (Wiley, Hoboken, 2006)
970. P. Capper, M. Mauk (eds.), *Liquid Phase Epitaxy of Electronic, Optical and Optoelectronic Materials* (Wiley, Chichester, 2007)
971. B.A. Joyce, J.H. Neave, P.J. Dobson, P.K. Larsen, Analysis of reflection high-energy electron-diffraction data from reconstructed semiconductor surfaces. Phys. Rev. B **29**, 814–819 (1984)
972. J.-T. Zettler, W. Richter, K. Ploska, M. Zorn, J. Rumberg, C. Meyne, M. Pristovsek, Real time diagnostics of semiconductor surface modifications by reflectance anisotropy spectroscopy, in *Semiconductor Characterization-Present Status and Future Needs*, ed. by W. Bullis, D. Seiler, A. Diebold (AIP Press, Woodbury, 1996), pp. 537–543
973. J.-T. Zettler, Characterization of epitaxial semiconductor growth by reflectance anisotropy spectroscopy and ellipsometry. Prog. Cryst. Growth Charact. Mater. **35**, 27–98 (1997)
974. W. Zulehner, Status and future of silicon crystal growth. Mater. Sci. Eng. B **4**, 1–10 (1989)
975. W. Zulehner, Historical overview of silicon crystal pulling development. Mater. Sci. Eng. B **73**, 7–15 (2000)
976. B. Depuydt, A. Theuwis, I. Romandie, Germanium: from the first application of Czochralski crystal growth to large diameter dislocation-free wafers. Mater. Sci. Semicond. Process. **9**, 437–443 (2006)
977. Virgina Semiconductor, http://www.virginiasemi.com
978. G.K. Teal, J.B. Little, Growth of germanium single crystals. Phys. Rev. **78**, 647 (1950)
979. W.R. Runyan, Growth of large diameter silicon and germanium crystals by the teal-little method. Rev. Sci. Instrum. **30**, 535–540 (1959)
980. W.G. Pfann, *Zone Melting* (Wiley, New York, 1966)
981. H.J. Scheel, T. Fukuda (eds.), *Crystal Growth Technology* (Wiley, New York, 2004)
982. Siltronic AG
983. K.A. Reinhardt, W. Kern, *Handbook of Silicon Wafer Cleaning Technology* (William Andrew, New York, 2008)
984. J. Ruzyllo, Semiconductor cleaning technology: forty years in the making. Interface **19**, 44–46 (2010)

985. W. Kern, Purifying Si and SiO_2 surfaces with hydrogen peroxide. Semicond. Int. **7**, 94–99 (1984)
986. A. Ishizaka, Y. Shiraki, Low temperature surface cleaning of silicon and its application to silicon MBE. J. Electrochem. Soc. **133**, 666–671 (1986)
987. J.C. Kim, J.-Y. Ji, J.S. Kline, J.R. Tucker, T.-C. Shen, Preparation of atomically clean and flat Si(100) surfaces by low-energy ion sputtering and low-temperature annealing. Appl. Surf. Sci. **220**, 293–297 (2003)
988. H. Hartnagel, B.L. Weiss, A contribution to etch polishing of GaAs. J. Mater. Sci. **8**, 1061–1063 (1973)
989. Z.F. Tomashik, N.V. Kusyak, V.N. Tomashik, Chemical etching of InAs, InSb, and GaAs in H_2O_2-HBr solutions. Inorg. Mater. **38**, 434–437 (2002)
990. H. von Wenckstern, H. Schmidt, C. Hanisch, M. Brandt, C. Czekalla, G. Benndorf, G. Biehne, A. Rahm, H. Hochmuth, M. Lorenz, M. Grundmann, Homoepitaxy of ZnO by pulsed-laser deposition. Phys. Status Solidi RRL **1**, 129–131 (2007)
991. C. Weisbuch, Fundamental properties of III-V semiconductor two-dimensional quantized structures: the basis for optical and electronic device applications. Semicond. Semimet. **24**, 1–133 (1987)
992. C.H.J. van den Brekel, Growth rate anisotropy and morphology of autoepitaxial silicon films from $SiCl_4$. J. Cryst. Growth **23**, 259–266 (1974)
993. J.-I. Nishizawa, T. Terasaki, M. Shimbo, Layer growth in silicon epitaxy. J. Cryst. Growth **13**(14), 297–301 (1972)
994. J.G.E. Gardiniers, C.H. Klein, Douwel, L.J. Giling, Reduced pressure silicon CVD on hemispherical substrates. J. Cryst. Growth **108**, 319–334 (1991)
995. S. Nishino, Y. Nishio, Y. Masuda, Y. Chen, C. Jacob, Morphological stability of 6H-SiC epitaxial layer on hemispherical substrates prepared by chemical vapor deposition. Mater. Sci. Forum **338–342**, 197–200 (2000)
996. L. Hollan, C. Schiller, Étude de l'Anisotropie de la croissance épitaxiale de GaAs en phase vapeur. J. Cryst. Growth **13**(14), 319–324 (1972)
997. K. Morizane, Antiphase domain structures in GaP and GaAs epitaxial layers grown on Si and Ge. J. Cryst. Growth **38**, 249–254 (1977)
998. M. Volmer, A. Weber, Keimbildung in übersättigten Gebilden. Z. Phys. Chem. **119**, 277–301 (1926)
999. I.N. Stranski, L. Krastanow, Zur Theorie der orientierten Ausscheidung von Ionenkristallen aufeinander. Sitzungsberichte d. Akad. d. Wissenschaften in Wien, Abt. IIb **146**, 797–810 (1938)
1000. E. Bauer, Phänomenologische Theorie der Kristallabscheidung an Oberflächen. I. Z. Kristallogr. **110**, 372–394 (1958)
1001. T. Shitara, D.D. Vvedensky, M.R. Wilby, J. Zhang, J.H. Neave, B.A. Joyce, Step-density variations and reflection high-energy electron-diffraction intensity oscillations during epitaxial growth on vicinal GaAs(001). Phys. Rev. B **46**, 6815–6824 (1992)
1002. A. Pimpinelli, J. Villains, *Physics of Crystal Growth* (Cambridge University Press, Cambridge, 1999)
1003. E. Ohshima, H. Ogino, I. Niikura, K. Maeda, M. Sato, M. Ito, T. Fukuda, Growth of the 2-in-size bulk ZnO single crystals by the hydrothermal method. J. Cryst. Growth **260**, 166–170 (2004)
1004. F. Scholz, U. Kaiser, Universität Ulm
1005. S. Heitsch, C. Bundesmann, G. Wagner, G. Zimmermann, A. Rahm, H. Hochmuth, G. Benndorf, H. Schmidt, M. Schubert, M. Lorenz, M. Grundmann, Low temperature photoluminescence and infrared dielectric functions of pulsed laser deposited ZnO thin films on silicon. Thin Solid Films **496**, 234–239 (2006)
1006. M. Grundmann, T. Böntgen, M. Lorenz, Occurrence of rotation domains in heteroepitaxy. Phys. Rev. Lett. **105**, 146102(4pp) (2010)
1007. M. Grundmann, Formation of epitaxial domains: unified theory and survey of experimental results. Phys. Status Solidi B **248**, 805–824 (2011)

1008. Y. Zhang, C. McAleese, H. Xiu, C.J. Humphreys, R.R. Lieten, B. Degroote, G. Borghs, Misoriented domains in (0001)-GaN/(111)-Ge grown by molecular beam epitaxy. Appl. Phys. Lett. **91**, 092125(2pp) (2007)
1009. H. Kawanami, A. Hatayama, Y. Hayashi, Antiphase boundary of GaAs films grown on Si(001) substrates by molecular beam epitaxy. J. Electron. Mater. **17**, 341–349 (1988)
1010. J. Ohta, H. Fujioka, M. Oshima, K. Fujiwara, A. Ishii, Experimental and theoretical investigation on the structural properties of GaN grown on sapphire. Appl. Phys. Lett. **83**, 3075–3077 (2003)
1011. P.J. Schuck, M.D. Mason, R.D. Grober, O. Ambacher, A.P. Lima, C. Miskys, R. Dimitrov, M. Stutzmann, Spatially resolved photoluminescence of inversion domain boundaries in GaN-based lateral polarity heterostructures. Appl. Phys. Lett. **79**, 952–954 (2001)
1012. B.J. Rodriguez, A. Gruveman, A.I. Kingon, R.J. Nemanich, O. Ambacher, Piezoresponse force microscopy for polarity imaging of GaN. Appl. Phys. Lett. **80**, 4166–4168 (2002)
1013. A. Strittmatter, S. Rodt, L. Reißmann, D. Bimberg, H. Schröder, E. Obermeier, T. Riemann, J. Christen, A. Krost, Maskless epitaxial lateral overgrowth of GaN layers on structured Si(111) substrates. Appl. Phys. Lett. **78**, 727–729 (2001)
1014. H. Lahrèche, P. Vennéguès, B. Beaumont, P. Gibart, Growth of high-quality GaN by low-pressure metal-organic vapour phase epitaxy (LP-MOVPE) from 3D islands and lateral overgrowth. J. Cryst. Growth **205**, 245–252 (1999)
1015. T. Riemann, T. Hempel, J. Christen, P. Veit, R. Clos, A. Dadgar, A. Krost, U. Haboeck, A. Hoffmann, Optical and structural microanalysis of GaN grown on SiN submonolayers. J. Appl. Phys. **99**, 123518(8pp) (2006)
1016. H.G. Chen, T.S. Ko, S.C. Ling, T.C. Lu, H.C. Kuo, S.C. Wang, Y.H. Wu, L. Chang, Dislocation reduction in GaN grown on stripe patterned r-plane sapphire substrates. Appl. Phys. Lett. **91**, 021914(3pp) (2007)
1017. N. Okada, K. Tadatomo, Characterization and growth mechanism of nonpolar and semipolar GaN layers grown on patterned sapphire substrates. Semicond. Sci. Technol. **27**, 024003(9pp) (2012)
1018. F. Tendille, P. De Mierry, P. Vennéguès, S. Chenot, M. Teisseire, Defect reduction method in (11–22) semipolar GaN grown on patterned sapphire substrate by MOCVD: toward heteroepitaxial semipolar GaN free of basal stacking faults. J. Cryst. Growth **404**, 177–183 (2014)
1019. Y.H. Lo, New approach to grow pseudomorphic structures over the critical thickness. Appl. Phys. Lett. **59**, 2311–2313 (2005)
1020. C. Lynch, E. Chason, R. Beresford, L.B. Freund, K. Tetz, K.W. Schwarz, Limits of strain relaxation in InGaAs/GaAs probed in real time by in situ wafer curvature measurement. J. Appl. Phys. **98**, 073532(7pp) (2005)
1021. A. Dadgar, J. Bläsing, A. Diez, A. Alam, M. Heuken, A. Krost, Metalorganic chemical vapor phase epitaxy of crack-free GaN on Si(111) exceeding 1 μm thickness. Jpn. J. Appl. Phys. **39**, L1183–L1185 (2000)
1022. J. Bläsing, A. Reiher, A. Dadgar, A. Dietz, A. Krost, The origin of stress reduction by low-temperature AlN interlayers. Appl. Phys. Lett. **81**, 2722–2724 (2002)
1023. K. Sakamoto, T. Sakamoto, S. Nagao, G. Hashiguchi, K. Kuniyoshi, Y. Bando, Reflection high-energy electron diffraction intensity oscillations during Ga_xSi_{1-x} MBE growth on Si(001) substrates. Jpn. J. Appl. Phys. **26**, 666–670 (1987)
1024. P.M.J. Marée, K. Nakagawa, F.M. Mulders, J.F. van der Veen, Thin epitaxial Ge-Si(111) films: study and control of morphology. Surf. Sci. **191**, 305–328 (1987)
1025. M. Copel, M.C. Reuter, E. Kaxiras, R.M. Tromp, Surfactants in epitaxial growth. Phys. Rev. Lett. **63**, 632–635 (1989)
1026. D. Kandel, E. Kaxiras, The surfactant effect in semiconductor thin-film growth. Solid State Phys. **54**, 219–262 (2000)
1027. M. Ilg, D. Eißler, C. Lange, K. Ploog, Surfactant-mediated molecular beam epitaxy of high-quality (111)B-GaAs. Appl. Phys. A **56**, 397–399 (1993)

1028. X. Yang, M.J. Jurkovic, J.B. Heroux, W.I. Wang, Low threshold InGaAsN/GaAs single quantum well lasers grown by molecular beam epitaxy using Sb surfactant. Electron. Lett. **35**, 1082–1083 (1999)
1029. T. Kageyama, T. Miyamoto, M. Ohta, T. Matsuura, Y. Matsui, T. Furuhata, F. Koyama, Sb surfactant effect on GaInAs/GaAs highly strained quantum well lasers emitting at 1200 nm range grown by molecular beam epitaxy. J. Appl. Phys. **96**, 44–48 (2004)
1030. H.J. Osten, J. Klatt, G. Lippert, B. Dietrich, E. Bugiel, Surfactant-controlled solid phase epitaxy of germanium and silicon. Phys. Rev. Lett. **69**, 450–453 (1992)
1031. H. Kroemer, Heterostructure devices: a device physicist looks at interfaces. Surf. Sci. **132**, 543–576 (1983)
1032. ChG van der Walle, J. Neugebauer, Universal alignment of hydrogen levels in semiconductors, insulators and solutions. Nature **423**, 626–628 (2003)
1033. K. Xu, H. Sio, O.A. Kirillov, L. Dong, M. Xu, P.D. Ye, D. Gundlach, N.V. Nguyen, Band offset determination of atomic-layer-deposited Al_2O_3 and HfO_2 on InP by internal photoemission and spectroscopic ellipsometry. J. Appl. Phys. **113**, 024504(5pp) (2013)
1034. W. Wei, Z. Qin, S. Fan, Z. Li, K. Shi, Q. Zhu, G. Zhang, Valence band offset of β-Ga_2O_3/wurtzite GaN heterostructure measured by X-ray photoelectron spectroscopy. Nanoscale Res. Lett. **7**, 562(5pp) (2013)
1035. X. Zhang, Q. Zhang, F. Lu, Energy band alignment of an In_2O_3:Mo/Si heterostructure. Semicond. Sci. Technol. **22**, 900–904 (2007)
1036. A. Walsh, J.L.F. Da Silva, S.-H. Wei, C. Körber, A. Klein, L.F.J. Piper, A. DeMasi, K.E. Smith, G. Panaccione, P. Rorelli, D.J. Payne, A. Bourlange, R.G. Egdell, Nature of the band gap of In_2O_3 revealed by first-principles calculations and X-ray spectroscopy. Phys. Rev. Lett. **100**, 167402(4pp) (2008)
1037. K. Muraki, S. Fukatsu, Y. Shiraki, R. Ito, Surface segregation of In atoms during molecular beam epitaxy and its influence on the energy levels in InGaAs/GaAs quantum wells. Appl. Phys. Lett. **61**, 557–559 (1992)
1038. D. Hernández-Maldonado, M. Herrera, P. Alonso-González, Y. González, L. González, J. Gazquez, M. Varela, S.J. Pennycook, M. de la Paz, Guerrero-Lebrero, J. Pizarro, P.L. Galindo, S.I. Molina, Compositional analysis with atomic column spatial resolution by 5th-order aberration-corrected scanning transmission electron microscopy. Microsc. Microanal. **17**, 578–581 (2011)
1039. V. Gottschalch, G. Wagner, University of Leipzig, private communication (2006)
1040. P. Offermans, Study of III-V semiconductor nanostructures by cross-sectional scanning tunneling microscopy. Ph.D. thesis, Technische Universiteit Eindhoven (2005)
1041. O. von Roos, Position-dependent effective masses in semiconductor theory. Phys. Rev. B **27**, 7547–7552 (1983)
1042. D.J. BenDaniel, C.B. Duke, Space-charge effects on electron tunneling. Phys. Rev. **152**, 683–692 (1966)
1043. P. Harrison, *Quantum Wells, Wires and Dots: Theoretical and Computational Physics of Semiconductor Nanostructures*, 3rd edn. (Wiley, Chichester, 2009)
1044. M. Ehrhardt, Th. Koprucki (eds.), *Multi-Band Effective Mass Approximations: Advanced Mathematical Models and Numerical Techniques*. Lecture Notes in Computational Science and Engineering, vol. 94 (Springer, Heidelberg, 2014)
1045. Y.C. Chang, J.N. Schulman, Interband optical transitions in $GaAs/Ga_{1-x}Al_xAs$ and InAs-GaSb superlattices. Phys. Rev. B **31**, 2069–2079 (1985)
1046. R.C. Miller, D.A. Kleinman, A.C. Gossard, Energy-gap discontinuities and effective masses for $GaAs/Al_xGa_{1-x}As$ quantum wells. Phys. Rev. B **29**, 7085–7087 (1984)
1047. S.L. Chuang, *Physics of Photonic Devices*, 2nd edn. (Wiley, New York, 2009)
1048. R.C. Miller, A.C. Gossard, W.T. Tsang, O. Munteanu, Extrinsic luminescence from GaAs quantum wells. Phys. Rev. B **25**, 3871–3877 (1982)
1049. W.T. Masselink, Y-Ch. Chang, H. Morkoç, Binding energy of acceptors in $GaAs/Al_xGa_{1-x}As$ as quantum wells. Phys. Rev. B **28**, 7373–7376 (1983)

1050. Effect on exciton binding energy, D.B. Tran Thoai, R. Zimmermann, M. Grundmann, D. Bimberg, Image charges in semiconductor quantum wells. Phys. Rev. B **42**, 5906–5909 (1990)
1051. M.J.L.S. Haines, N. Ahmed, S.J.A. Adams, K. Mitchell, I.R. Agool, C.R. Pidgeon, B.C. Cavenett, E.P. O'Reilly, A. Ghiti, M.T. Emeny, Exciton-binding-energy maximum in $Ga_{1-x}In_xAs/GaAs$ quantum wells. Phys. Rev. B **43**, 11944–11949 (1991)
1052. K.J. Moore, G. Duggan, K. Woodbridge, C. Roberts, Observations and calculations of the exciton binding energy in (In, Ga)As/GaAs strained-quantum-well heterostructures. Phys. Rev. B **41**, 1090–1094 (1990)
1053. L. Esaki, A perspective in superlattice development, in *Recent Topics, in Semiconductor Physics*, ed. by H. Kamimura, Y. Toyozawa (World Scientific, Singapore, 1983), pp. 1–19
1054. R. Dingle, A.C. Gossard, W. Wiegmann, Direct observation of superlattice formation in a semiconductor heterostructure. Phys. Rev. Lett. **34**, 1327–1330 (1975)
1055. F. Stern, S. Das, Sarma, Electron energy in GaAs-$Ga_{1-x}Al_x$ heterojunctions. Phys. Rev. B **30**, 840–848 (1984)
1056. A.L. Efros, F.G. Pikus, G.G. Samsonidze, Maximum low-temperature mobility of two-dimensional electrons in heterojunctions with a thick spacer layer. Phys. Rev. B **41**, 8295–8301 (1990)
1057. L. Pfeiffer, K.W. West, The role of MBE in recent quantum Hall effect physics discoveries. Physica E **20**, 57–64 (2003)
1058. J. Christen, D. Bimberg, Line shapes of intersubband and excitonic recombination in quantum wells: influence of final-state interaction, statistical broadening, and momentum conservation. Phys. Rev. B **42**, 7213–7219 (1990)
1059. J. Feldmann, G. Peter, E.O. Göbel, P. Dawson, K. Moore, C. Foxon, R.J. Elliott, Linewidth dependence of radiative exciton lifetimes in quantum wells. Phys. Rev. Lett. **59**, 2337–2340 (1987); Phys. Rev. Lett. **60**, 243 (1988) (erratum)
1060. M.S. Skolnick, J.M. Rorison, K.J. Nash, D.J. Mowbray, P.R. Tapster, S.J. Bass, A.D. Pitt, Observation of a many-body edge singularity in quantum-well luminescence spectra. Phys. Rev. Lett. **58**, 2130–2133 (1987)
1061. D. Gammon, S. Rudin, T.L. Reinecke, D.S. Katzer, C.S. Kyono, Phonon broadening of excitons in $GaAs/Al_xGa_{1-x}As$ quantum wells. Phys. Rev. B **51**, 16785–16789 (1995)
1062. L. Béaur, T. Bretagnon, B. Gil, A. Kavokin, T. Guillet, C. Brimont, D. Tainoff, M. Teisseire, J.-M. Chauveau, Exciton radiative properties in nonpolar homoepitaxial ZnO/(Zn, Mg)O quantum wells. Phys. Rev. B **84**, 165312(8pp) (2011)
1063. E. Runge, R. Zimmermann, Optical properties of localized excitons in nanostructures: theoretical aspects. Adv. Solid State Phys. (Festkörperprobleme) **38**, 251–263 (1999)
1064. E. Runge, Solid State Phys. **57**, 149–305 (2002)
1065. S.T. Davey, E.G. Scott, B. Wakefield, G.J. Davies, A photoluminescence study of $Ga_{1-x}In_xAs/Al_{1-y}In_yAs$ quantum wells grown by MBE. Semicond. Sci. Technol. **3**, 365–371 (1988)
1066. M. Grassi Alessi, F. Fragano, A. Patané, M. Capizzi, E. Runge, R. Zimmermann, Competition between radiative decay and energy relaxation of carriers in disordered $In_xGa_{1-x}As/GaAs$ quantum wells. Phys. Rev. B **61**, 10985–10993 (2000)
1067. Q. Li, S.J. Xu, W.C. Cheng, M.H. Xie, S.Y. Tong, C.M. Che, H. Yang, Thermal redistribution of localized excitons and its effect on the luminescence band in InGaN ternary alloys. Appl. Phys. Lett. **79**, 1810–1812 (2001)
1068. J. Hegarty, L. Goldner, M.D. Sturge, Localized and delocalized two-dimensional excitons in GaAs-AlGaAs multiple-quantum-well structures. Phys. Rev. B **30**, 7346–7348 (1984)
1069. N.F. Mott, E.A. Davies, *Electronic Processes in Noncrystalline Materials*, 2nd edn. (Oxford University Press, New York, 1979)
1070. T. Takeuchi, S. Sota, M. Katsuragawa, M. Komori, H. Takeuchi, H. Amano, I. Akasaki, Quantum-confined Stark effect due to piezoelectric fields in GaInN strained quantum wells. Jpn. J. Appl. Phys. **36**, Part 2, L382–L385 (1997)

1071. S.F. Chichibu, A.C. Abare, M.S. Minsky, S. Keller, S.B. Fleischer, J.E. Bowers, E. Hu, U.K. Mishra, L.A. Coldren, S.P. DenBaars, Effective band gap inhomogeneity and piezoelectric field in InGaN/GaN multiquantum well structures. Appl. Phys. Lett. **73**, 2005–2008 (1998)
1072. C. Morhain, T. Bretagnon, P. Lefebvre, X. Tang, P. Valvin, T. Guillet, B. Gil, T. Taliercio, M. Teisseire-Doninelli, B. Vinter, C. Deparis, Internal electric field in wurtzite $ZnO/Zn_{0.78}Mg_{0.22}O$ quantum wells. Phys. Rev. B **72**, 241305(R)(4pp) (2005)
1073. J.A. Davis, L.V. Dao, X. Wen, C. Ticknor, P. Hannaford, V.A. Coleman, H.H. Tan, C. Jagadish, K. Koike, S. Sasa, M. Inoue, M. Yano, Suppression of the internal electric field effects in $ZnO/Zn_{0.7}Mg_{0.3}O$ quantum wells by ion-implantation induced intermixing. Nanotechnology **19**, 055205(4pp) (2008)
1074. E. Berkowicz, D. Gershoni, G. Bahir, E. Lakin, D. Shilo, E. Zolotoyabko, A.C. Abare, S.P. Denbaars, L.A. Coldren, Measured and calculated radiative lifetime and optical absorption of $In_xGa_{1-x}N/GaN$ quantum structures. Phys. Rev. B **61**, 10994–11008 (2000)
1075. N. Akopian, G. Bahir, D. Gershoni, M.D. Craven, J.S. Speck, S.P. DenBaars, Optical evidence for lack of polarization in $(11\bar{2}0)$-oriented GaN/(AlGa)N quantum structures. Appl. Phys. Lett. **86**, 202104(3pp) (2005)
1076. J. Spitzer, T. Ruf, M. Cardona, W. Dondl, R. Schorer, G. Abstreiter, E.E. Haller, Raman scattering by optical phonons in isotopic $^{70}(Ge)_n{}^{74}(Ge)_n$ superlattices. Phys. Rev. Lett. **72**, 1565–1568 (1994)
1077. A. Reznicek, R. Scholz, S. Senz, U. Gösele, Comparative TEM study of bonded silicon/silicon interfaces fabricated by hydrophilic, hydrophobic and UHV wafer bonding. Mater. Chem. Phys. **81**, 277–280 (2003)
1078. P. Kopperschmidt, S. Senz, G. Kästner, D. Hesse, U.M. Gösele, Materials integration of gallium arsenide and silicon by wafer bonding. Appl. Phys. Lett. **72**, 3181–3183 (1998)
1079. Q.-Y. Tong, U. Gösele, *Semiconductor Wafer Bonding* (Wiley, New York, 1998)
1080. M. Alexe, U. Gösele (eds.), *Wafer Bonding* (Springer, Berlin, 2004)
1081. A. Jaeger: Exzitonen und Franz-Keldysh-Effekt im quaternären Halbleiter InGaAsP/InP. Ph.D. thesis, Phillips-Universität Marburg (1997)
1082. D.A.B. Miller, D.S. Chemla, T.C. Damen, A.C. Gossard, W. Wiegmann, T.H. Wood, C.A. Burrus, Electric field dependence of optical absorption near the band gap of quantum-well structures. Phys. Rev. B **32**, 1043–1060 (1985)
1083. J.M. Luttinger, Quantum theory of cyclotron resonance in semiconductors: general theory. Phys. Rev. **102**, 1030–1041 (1956)
1084. N. Miura, *Physics of Semiconductors in High Magnetic Fields* (Oxford University Press, Oxford, 2008)
1085. M. Oestreich, S. Hallstein, A.P. Heberle, K. Eberl, E. Bauser, W.W. Rühle, Temperature and density dependence of the electron Landé g factor in semiconductors. Phys. Rev. B **53**, 7911–7916 (1996)
1086. M.J. Snelling, E. Blackwood, C.J. McDonagh, R.T. Harley, C.T.B. Foxon, Exciton, heavy-hole, and electron g factors in type-I $GaAs/Al_xGa_{1-x}As$ quantum wells. Phys. Rev. B **45**, 3922–3925 (1992)
1087. C.L. Chien, C.R. Westgate (eds.), *The Hall Effect and Its Application* (Plenum Press, New York, 1980)
1088. H.-J. Lippmann, F. Kuhrt, Der Geometrieeinfluß auf den Hall-Effekt bei rechteckigen Halbleiterplatten. Z. Naturf. **13a**, 474–483 (1958)
1089. L.J. van der Pauw, A method of measuring specific resistivity and Hall effect of discs of arbitrary shape. Philips Res. Rep. **13**, 1–9 (1958)
1090. L.J. van der Pauw, A method of measuring the resistivity and Hall coefficient on lamellae of arbitrary shape. Philips Tech. Rev. **20**, 220–224 (1958)
1091. D.S. Perloff, Four-point sheet resistance correction factors for thin rectangular samples. Solid State Electron. **20**, 681–687 (1977)
1092. O. Breitenstein, W. Warta, M. Langenkamp, *Lock-in Thermography, Basics and Use for Evaluating Electronic Devices and Materials*, 2nd edn. (Springer, Heidelberg, 2010)

1093. O. Madelung, H. Weiss, Die elektrischen Eigenschaften von Indiumantimonid II. Z. Naturf. **9a**, 527–534 (1954)
1094. H.G. Svavarsson, J.T. Gudmundsson, H.P. Gislason, Impurity band in lithium-diffused and annealed GaAs: conductivity and Hall effect measurements. Phys. Rev. B **67**, 205213(6pp) (2003)
1095. B. Arnaudov, T. Paskova, S. Evtimova, E. Valcheva, M. Heuken, B. Monemar, Multilayer model for Hall effect data analysis of semiconductor structures with step-changed conductivity. Phys. Rev. B **67**, 045314(10pp) (2003)
1096. D.C. Look, J.R. Sizelove, Dislocation scattering in GaN. Phys. Rev. Lett. **82**, 1237–1240 (1999)
1097. H. von Wenckstern, M. Brandt, H. Schmidt, G. Biehne, R. Pickenhain, H. Hochmuth, M. Lorenz, M. Grundmann, Donor like defects in ZnO substrate materials and ZnO thin films. Appl. Phys. A **88**, 135–139 (2007)
1098. D.C. Look, Two-layer Hall-effect model with arbitrary surface-donor profiles: application to ZnO. J. Appl. Phys. **104**, 063718(7pp) (2008)
1099. J. Antoszweski, D.J. Seymour, L. Faraone, J.R. Meyer, C.A. Hoffmann, Magneto-transport characterization using quantitative mobility-spectrum analysis. J. Electron. Mater. **24**, 1255–1262 (1995)
1100. I. Vurgaftman, J.R. Meyer, C.A. Hoffman, D. Redfern, J. Antoszewski, L. Faraone, J.R. Lindemuth, Improved quantitative mobility spectrum analysis for Hall characterization. J. Appl. Phys. **84**, 4966–4973 (1998)
1101. J. Rothman, J. Meilhan, G. Perrais, J.-P. Belle, O. Gravrand, Maximum entropy mobility spectrum analysis of HgCdTe heterostructures. J. Electron. Mater. **35**, 1174–1184 (2006)
1102. J. Antoszewski, L. Faraone, Quantitative mobility spectrum analysis (QMSA) in multi-layer semiconductor structures. Opto-Electron. Rev. **12**, 347–352 (2004)
1103. the Hall mobility of the small polaron, L. Friedman, T. Holstein, Studies of polaron motion. Part III. Ann. Phys. **21**, 494–549 (1963)
1104. Y.M. Gal'perin, E.P. German, V.G. Karpov, Hall effect under hopping conduction conditions. Zh. Eksp. Teor. Fiz. **99**, 343–356 (1991) (Sov. Phys. JETP **72**, 193–200 (1991))
1105. P.G. Le Comber, D.I. Jones, W.E. Spear, Hall effect and impurity conduction in substitutionally doped amorphous silicon. Philos. Mag. **35**, 1173–1187 (1977)
1106. I. Crupi, S. Mirabella, D. D'Angelo, S. Gibilisco, A. Grasso, S. Di Marco, F. Simone, A. Terrasi, Anomalous and normal hall effect in hydrogenated amorphous Si prepared by plasma enhanced chemical vapor deposition. J. Appl. Phys. **107**, 043503(6pp) (2010)
1107. M. Grünewald, P. Thomas, D. Würtz, The sign anomaly of the Hall effect in amorphous tetrahedrally bonded semiconductors: a chemical-bond orbital approach. J. Phys. C: Solid State Phys. **14**, 4083–4093 (1981)
1108. M. Schubert, *Infrared Ellipsometry on Semiconductor Layer Structures: Phonons, Plasmons and Polaritons* (Springer, Heidelberg, 2004)
1109. M. Schubert, T. Hofmann, C.M. Herzinger, Generalized far-infrared magneto-optic ellipsometry for semiconductor layer structures: determination of free-carrier effective-mass, mobility, and concentration parameters in n-type GaAs. J. Opt. Soc. Am. A **20**, 347–356 (2003)
1110. B. Lax, J.G. Mavroides, H.J. Zeiger, R.J. Keys, Cylcotron resonance in indium antimonide at high magnetic fields. Phys. Rev. **122**, 31–35 (1961)
1111. W. Zawadski, R. Lassnig, Specific heat and magneto-thermal oscillations of two-dimensional electron gas in a magnetic field. Solid State Commun. **50**, 537–539 (1984)
1112. H.L. Störmer, R. Dingle, A.C. Gossard, W. Wiegmann, R.A. Logan, Electronic properties of modulation-doped GaAs-$Al_xGa_{1-x}As$ superlattices. Inst. Phys. Conf. Ser. **43**, 557–560 (1979)
1113. A.B. Fowler, F.F. Fang, W.E. Howard, P.J. Stiles, Magneto-oscillatory conductance in silicon surfaces. Phys. Rev. Lett. **16**, 901–903 (1966)
1114. M.A. Paalanen, D.C. Tsui, A.C. Gossard, Quantized Hall effect at low temperatures. Phys. Rev. B **25**, 5566–5569 (1982)

1115. Physikalisch-Technische Bundesanstalt (PTB), Braunschweig
1116. K. v. Klitzing, G. Dorda, M. Pepper, New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. Phys. Rev. Lett. **45**, 494–497 (1980)
1117. G. Landwehr, 25 Years quantum Hall effect: how it all came about. Phys. E **20**, 1–13 (2003)
1118. E.O. Göbel, Semiconductor applications in metrology. Adv. Solid State Phys. (Festkörperprobleme) **39**, 1–12 (1999)
1119. H. Bachmair, E.O. Göbel, G. Hein, J. Melcher, B. Schumacher, J. Schurr, L. Schweitzer, P. Warnecke, The von Klitzing resistance standard. Phys. E **20**, 14–23 (2003)
1120. J.K. Jain, *Composite Fermions* (Cambridge University Press, Cambridge, 2007)
1121. H. Aoki, T. Ando, Effect of localization of the Hall conductivity in the two-dimensional system in strong magnetic fields. Solid State Commun. **38**, 1079–1082 (1981)
1122. E. Ahlswede, P. Weitz, J. Weis, K. v. Klitzing, K. Eberl, Hall potential profiles in the quantum Hall regime measured by a scanning force microscope. Phys. B **298**, 562–566 (2001)
1123. R.E. Prange, Quantized Hall resistance and the meaurement of the fine-structure constant. Phys. Rev. B **23**, 4802–4805 (1981)
1124. M. Büttiker, Absence of backscattering in the quantum Hall effect in multiprobe conductors. Phys. Rev. B **38**, 9375–9389 (1988)
1125. K. Lier, R.R. Gerhardts, Self-consistent calculation of edge channels in laterally confined two-dimensional electron systems. Phys. Rev. B **50**, 7757–7767 (1994)
1126. R.B. Laughlin, Quantized Hall conductivity in two dimensions. Phys. Rev. B **23**, 5632–5633 (1981)
1127. J.P. Eisenstein, H.L. Störmer, The fractional quantum Hall effect. Science **248**, 1510–1516 (1990)
1128. J.K. Jain, Composite fermion approach for the fractional quantum Hall effect. Phys. Rev. Lett. **63**, 199–202 (1989)
1129. J.K. Jain, Theory of the fractional quantum Hall effect. Phys. Rev. B **41**, 7653–7665 (1990)
1130. H.L. Störmer, D. Tsui, A.C. Gossard, The fractional quantum Hall effect. Rev. Mod. Phys. **71**, S298–S305 (1999)
1131. D. Weiss, M.L. Roukes, A. Menschig, P. Frambow, K.v. Klitzing, G. Weimann, Electron pinball and commensurate orbits in a periodic array of scatterers. Phys. Rev. Lett. **66**, 2790–2793 (1991)
1132. R.P. Feynman, There's plenty of room at the bottom, in *After-Dinner Speech on December 29th 1959 at the Annual Meeting of the American Physical Society at the California Institute of Technology*
1133. D. Bimberg, M. Grundmann, N.N. Ledentsov, *Quantum Dot Heterostructures* (Wiley, Chichester, 1999)
1134. M. Grundmann (ed.), *Nano-Optoelectronics, Concepts, Physics and Devices* (Springer, Heidelberg, 2002)
1135. E. Kapon, M. Walther, J. Christen, M. Grundmann, C. Caneau, D.M. Hwang, E. Colas, R. Bhat, G.H. Song, D. Bimberg, Quantum wire heterostructures for optoelectronic applications. Superlatt. Microstruct. **12**, 491–499 (1992)
1136. M. Grundmann, J. Christen, M. Joschko, O. Stier, D. Bimberg, E. Kapon, Recombination kinetics and intersubband relaxation in semiconductor quantum wires. Semicond. Sci. Technol. **9**, 1939–1945 (1994)
1137. L. Pfeiffer, K.W. West, H.L. Störmer, J.P. Eisenstein, K.W. Baldwin, D. Gershoni, J. Spector, Formation of a high quality two-dimensional electron gas on cleaved GaAs. Appl. Phys. Lett. **56**, 1697–1699 (1990)
1138. M. Grundmann, D. Bimberg, Formation of quantum dots in twofold cleaved edge overgrowth. Phys. Rev. B **55**, 4054–4056 (1997)
1139. W. Wegscheider, G. Schedelbeck, G. Abstreiter, M. Rother, M. Bichler, Atomically precise GaAsyAlGaAs quantum dots fabricated by twofold cleaved edge overgrowth. Phys. Rev. Lett. **79**, 1917–1920 (1997)
1140. A.P. Levitt (ed.), *Whisker Technology* (Wiley, New York, 1970)

1141. Zh.L. Wang (ed.), Nanowires and Nanobelts-Materials, Properties and Devices, vol. I (Metal and Semiconductor Nanowires) and, vol. II (Nanowires and Nanobelts of Functional Materials) (Kluwer Academic, Boston, 2004)
1142. M.T. Björk, B.J. Ohlsen, T. Sass, A.I. Perrson, C. Thelander, M.H. Magnusson, K. Deppert, L.R. Wallenberg, L. Samuelson, One-dimensional heterostructures in semiconductor nanowhiskers. Appl. Phys. Lett. **80**, 1058–1060 (2002)
1143. M.H. Huang, S. Mao, H. Feick, H. Yan, Y. Wu, H. Kind, E. Weber, R. Russo, P. Yang, Room-temperature ultraviolet nanowire nanolasers. Science **292**, 1897–1899 (2001)
1144. X. Duan, Y. Huang, R. Agarwal, C. Lieber, Single-nanowire electrically driven lasers. Nature **421**, 241–245 (2003)
1145. J. ZhL Wang, Song, Piezoelectric nanogenerators based on zinc oxide nanowire arrays. Science **312**, 242–246 (2006)
1146. M. Lorenz, J. Lenzner, E.M. Kaidashev, H. Hochmuth, M. Grundmann, Cathodoluminescence of selected single ZnO nanowires on sapphire. Ann. Physik **13**, 39–42 (2004)
1147. B.Q. Cao, J. Zúñiga-Pérez, N. Boukos, C. Czekalla, H. Hilmer, J. Lenzner, A. Travlos, M. Lorenz, M. Grundmann, Homogeneous core/shell ZnO/ZnMgO quantum well heterostructures on vertical ZnO nanowires. Nanotechnology **20**, 305701 (2009)
1148. M.V. Barton, The circular cylinder with a band of uniform pressure on a finite length of the surface. J. Appl. Mech. **8**, A97–A104 (1941)
1149. E. Ertekin, P.A. Greaney, D.C. Chrzan, T.D. Sands, Equilibrium limits of coherency in strained nanowire heterostructures. J. Appl. Phys. **97**, 114325(10pp) (2005)
1150. F. Glas, Critical dimensions for the plastic relaxation of strained axial heterostructures in free-standing nanowires. Phys. Rev. B **74**, 121302(4pp) (2006)
1151. X.Y. Kong, ZhL Wang, Polar-surface dominated ZnO nanobelts and the electrostatic energy induced nanohelixes, nanosprings, and nanospirals. Appl. Phys. Lett. **84**, 975–977 (2004)
1152. ZhL Wang, Novel zinc oxide nanostructures discovery by electron microscopy. J. Phys.: Conf. Ser. **26**, 1–6 (2006)
1153. M. Grundmann, E. Kapon, J. Christen, D. Bimberg, Electronic and optical properties of quasi one-dimensional carriers in quantum wires. J. Nonlinear Opt. Phys. Mater. **4**, 99–140 (1995)
1154. M. Grundmann, O. Stier, D. Bimberg, Electronic states in strained cleaved-edge-overgrowth quantum wires and quantum dots. Phys. Rev. B **58**, 10557–10561 (1998)
1155. H.J. Xiang, J. Yang, J.G. Hou, Q. Zhu, Piezoelectricity in ZnO nanowires: a first-principles study. Appl. Phys. Lett. **89**, 223111(3pp) (2006)
1156. O. Stier, M. Grundmann, D. Bimberg, Electronic and optical properties of strained quantum dots modeled by 8-band k·p theory. Phys. Rev. B **59**, 5688–5701 (1999)
1157. R. Santoprete, B. Koiller, R.B. Capaz, P. Kratzer, Q.K.K. Liu, M. Scheffler, Tight-binding study of the influence of the strain on the electronic properties of InAs/GaAs quantum dots. Phys. Rev. B **68**, 235311(9pp) (2003)
1158. Th. Maltezopoulos, A. Bolz, C. Meyer, C. Heyn, W. Hansen, M. Morgenstern, R. Wiesendanger, Wave-function mapping of InAs quantum dots by scanning tunneling spectroscopy. Phys. Rev. Lett. **91**, 196804(4pp) (2003)
1159. O. Stier, Theory of the optical properties of InGaAs/GaAs quantum dots, in *Nano-Optoelectronics. Concepts, Physics, Devices*, ed. by M. Grundmanned (Berlin, Grundmann (Springer, 2002), pp. 167–202
1160. L.P. Kouwenhoven, N.C. van der Vaart, A.T. Johnson, W. Kool, C.J.P.M. Harmans, J.G. Williamson, A.A.M. Staring, C.T. Foxon, Conductance oscillations and transport spectroscopy of a quantum dot. Z. Phys. B **85**, 357–66 (1991)
1161. S. Tarucha, D.G. Austing, T. Honda, R.J. van der Hage, L.P. Kouwenhoven, Shell filling and spin effects in a few electron quantum dot. Phys. Rev. Lett. **77**, 3613–3616 (1996)
1162. N. Horiguchi, T. Futatsugi, Y. Nakata, N. Yokoyama, Electron transport properties through InAs self-assembled quantum dots in modulation doped structures. Appl. Phys. Lett. **70**, 2294–2296 (1997)

1163. A. Forchel, R. Steffen, M. Michel, A. Pecher, T.L. Reinecke, in *Proceedings of the 23rd International Conference on the Physics of Semiconductors, Berlin (Germany)*, ed. by M. Scheffler, R. Zimmermann (World Scientific, Singapore, 1996), pp. 1285–1292
1164. K.C. Rajkumar, K. Kaviani, J. Chen, P. Chen, A. Madhukar, D. Rich, In-situ growth of three-dimensionally confined structures on patterned GaAs(111)B substrates. Proc. Mater. Res. Soc. **263**, 163 (1992)
1165. W. Weller, private communication (2006)
1166. M. Grundmann, Pseudomorphic InAs/GaAs quantum dots on low index planes. Adv. Solid State Phys. (Festkörperprobleme) **35**, 123–154 (1996)
1167. V.A. Shchukin, N.N. Ledentsov, D. Bimberg, *Epitaxy of Nanostructures* (Springer, Heidelberg, 2004)
1168. V.A. Shchukin, N.N. Ledentsov, P.S. Kop'ev, D. Bimberg, Spontaneous ordering of arrays of coherent strained Islands. Phys. Rev. Lett. **75**, 2968–2971 (1995)
1169. N. Moll, M. Scheffler, E. Pehlke, Influence of surface stress on the equilibrium shape of strained quantum dots. Phys. Rev. B **58**, 4566–4571 (1998)
1170. D. Leonard, M. Krishnamurthy, C.M. Reaves, S.P. Denbaars, P.M. Petroff, Direct formation of quantum-sized dots from uniform coherent islands of InGaAs on GaAs surfaces. Appl. Phys. Lett. **63**, 3203–3205 (1993)
1171. D.M. Bruls, P.M. Koenraad, H.W.M. Salemink, J.H. Wolter, M. Hopkinson, M.S. Skolnick, Stacked low-growth-rate InAs quantum dots studied at the atomic level by cross-sectional scanning tunneling microscopy. Appl. Phys. Lett. **82**, 3758–3760 (2003)
1172. Q. Xie, A. Madhukar, P. Chen, N. Kobayashi, Vertically self-organized InAs quantum box Islands on GaAs(100). Phys. Rev. Lett. **75**, 2542–2545 (1995)
1173. S. Facsko, T. Dekorsy, C. Koerdt, C. Trappe, H. Kurz, A. Vogt, H.L. Hartnagel, Formation of ordered nanoscale semiconductor dots by ion sputtering. Science **285**, 1551–1553 (1999)
1174. F. Frost, A. Schindler, F. Bigl, Roughness evolution of ion sputtered rotating InP surfaces: pattern formation and scaling laws. Phys. Rev. Lett. **85**, 4116–4419 (2000)
1175. R. Gago, L. Vásquez, R. Cuerno, M. Varela, C. Ballesteros, J.M. Albella, Production of ordered silicon nanocrystals by low-energy ion sputtering. Appl. Phys. Lett. **78**, 3316–3318 (2001)
1176. R. Gago, L. Vásquez, O. Plantevin, T.H. Metzger, J. Munoz-García, R. Cuerno, M. Castro, Order enhancement and coarsening of self-organized silicon nanodot patterns induced by ion-beam sputtering. Appl. Phys. Lett. **89**, 233101(3pp) (2006)
1177. B. Ziberi, F. Frost, B. Rauschenbach, Th. Höche, Highly ordered self-organized dot patterns on Si surfaces by low-energy ion-beam erosion. Appl. Phys. Lett. **87**, 033113(3pp) (2005)
1178. W.L. Chan, E. Chason, Making waves: kinetic processes controlling surface evolution during low energy ion sputtering. J. Appl. Phys. **101**, 121301(46pp) (2007)
1179. B. Ziberi, F. Frost, Th. Höche, B. Rauschenbach, Ripple pattern formation on silicon surfaces by low-energy ion-beam erosion: experiment and theory. Phys. Rev. B **72**, 235310(7pp) (2005)
1180. F. Findeis, A. Zrenner, G. Böhm, G. Abstreiter, Optical spectroscopy on a single InGaAs/GaAs quantum dot in the few-exciton limit. Solid State Commun. **114**, 227–230 (2000)
1181. O. Stier, A. Schliwa, R. Heitz, M. Grundmann, D. Bimberg, Stability of biexcitons in pyramidal InAs/GaAs quantum dots. Phys. Status Solidi B **224**, 115–118 (2001)
1182. D.J.P. Ellis, R.M. Stevenson, R.J. Young, A.J. Shields, P. Atkinson, D.A. Ritchie, Control of fine-structure splitting of individual InAs quantum dots by rapid thermal annealing. Appl. Phys. Lett. **90**, 011907(3pp) (2007)
1183. B. Urbaszek, R.J. Warburton, K. Karrai, B.D. Gerardot, P.M. Petroff, J.M. Garcia, Fine structure of highly charged excitons in semiconductor quantum dots. Phys. Rev. Lett. **90**, 247403(4pp) (2003)
1184. R.J. Warburton, C. Schäflein, D. Haft, F. Bickel, A. Lorke, K. Karrai, J.M. Garcia, W. Schoenfeld, P.M. Petroff, Optical emission from a charge-tunable quantum ring. Nature **405**, 926–929 (2000)

1185. L. Besombes, Y. Léger, L. Maingault, D. Ferrand, H. Mariette, Probing the spin state of a single magnetic ion in an individual quantum dot. Phys. Rev. Lett. **93**, 207403(4pp) (2004)
1186. M. Fiebig, Revival of the magnetoelectric effect. J. Phys. D: Appl. Phys. **38**, R123–R152 (2005)
1187. W. Eerenstein, N.D. Mathur, J.F. Scott, Multiferroic and magnetoelectric materials. Nature **442**, 759–765 (2006)
1188. Nils Ashcroft, Cornell University, private communication (2006)
1189. V.M. Fridkin, *Ferroelectric Semiconductors (translated from Russian)* (Plenum, New York, 1980)
1190. Y. Xu, *Ferroelectric Materials and Their Applications* (North Holland, Amsterdam, 1991)
1191. K. Rabe, Ch.H. Ahn, J.-M. Triscone (eds.), Physics of Ferroelectrics, A Modern Perspective. Topics in Applied Physics, vol. 105 (Springer, Berlin, 2007)
1192. R. Comes, M. Lambert, A. Guinier, The chain structure of $BaTiO_3$ and $KNbO_3$. Solid State Commun. **6**, 715–719 (1968)
1193. R.E. Cohen, Origin of ferroelectricity in perovskite oxides. Nature **358**, 136–138 (1992)
1194. G. Shirane, S. Hoshino, On the phase transition in lead titanate. J. Phys. Soc. Jpn. **6**, 265–270 (1951)
1195. W.J. Merz, The electric and optical behavior of $BaTiO_3$ single-domain crystals. Phys. Rev. **76**, 1221–1225 (1949)
1196. D.K. Agrawal, C.H. Perry, Long-wavelength optical phonons and phase transitions in SbSI. Phys. Rev. B **4**, 1893–1902 (1971)
1197. G. Rupprecht, R.O. Bell, Dielectric constant in paraelectric perovskites. Phys. Rev. **135**, A748–A752 (1964)
1198. M.E. Lines, Statistical theory for displacement Ferroelectrics. III. Comparison with experiment for lithium tantalate. Phys. Rev. **177**, 819–829 (1969)
1199. W.J. Merz, Double hysteresis loop of $BaTiO_3$ at the Curie point. Phys. Rev. **91**, 513–517 (1953)
1200. P.W. Forsbergh Jr, Domain structures and phase transitions in barium titanate. Phys. Rev. **76**, 1187–1201 (1949)
1201. Ch. Gähwiller, Einfluß des elektrischen Feldes auf die fundamentale Absorptionskante von Bariumtitanat. Phys. Kondens. Materie **6**, 269–289 (1967)
1202. G. Arlt, P. Quadflieg, Piezoelectricity in III-V compounds with a phenomenological analysis of the piezoelectric effect. Phys. Status Solidi **25**, 323–330 (1968)
1203. M. Grundmann, O. Stier, D. Bimberg, Symmetry breaking in pseudomorphic V-groove quantum wires. Phys. Rev. B **50**, 14187–14192 (1994)
1204. D.L. Smith, C. Mailhiot, Piezoelectric effects in strained-layer superlattices. J. Appl. Phys. **63**, 2717–2719 (1988)
1205. T. Hanada, Basic properties of ZnO, GaN, and related materials, in *Oxide and Nitride Semiconductors: Processing, Properties and Applications*, ed. by T. Yao, S.-K. Hong (Springer, Heidelberg, 2009)
1206. M. Catti, Y. Noel, R. Dovesi, Full piezoelectric tensors of wurtzite and zinc blende ZnO and ZnS by first-principles calculations. J. Phys. Chem. Solids **64**, 2183–2190 (2003)
1207. M.S. Shur, B. Gelmont, A. Khan, Electron mobility in two-dimensional electron gas in AlGaN/GaN heterostructures and in bulk GaN. J. Electron. Mater. **25**, 777–785 (1996)
1208. F. Bernardini, V. Fiorentini, D. Vanderbilt, Spontaneous polarization and piezoelectric constants of III-V nitrides. Phys. Rev. B **56**, R10024–R10027 (1997)
1209. O. Ambacher, private communication (2005)
1210. P. Waltereit, O. Brandt, A. Trampert, H.T. Grahn, J. Menniger, M. Ramsteiner, M. Reiche, K.H. Ploog, Nitride semiconductors free of electrostatic fields for efficient white light-emitting diodes. Nature **406**, 865–868 (2000)
1211. C.Q. Chen, V. Adivarahan, J.W. Yang, M. Shatalov, E. Kuokstis, M.A. Khan, Ultraviolet light emitting diodes using non-polar a-plane GaN-AlGaN multiple quantum wells. Jpn. J. Appl. Phys. **42**, L1039–L1040 (2003) (Part 2)

1212. R. Sharma, P.M. Pattison, H. Masui, R.M. Farrel, T.J. Baker, B.A. Haskell, F. Wu, S.P. DenBaars, J.S. Speck, S. Nakamura, Demonstration of a semipolar ($10\bar{1}\bar{3}$) InGaN/GaN green light emitting diode. Appl. Phys. Lett. **87**, 231110(3pp) (2005)
1213. S.-H. Park, Crystal orientation effects on electronic properties of wurtzite InGaN/GaN quantum wells. J. Appl. Phys. **91**, 9904–9908 (2002)
1214. H. Sato, R.B. Chung, H. Hirasawa, N. Fellows, H. Masui, F. Wu, M. Saito, K. Fujito, J.S. Speck, S.P. DenBaars, S. Nakamura, Optical properties of yellow light-emitting diodes grown on semipolar ($11\bar{2}2$) bulk GaN substrates. Appl. Phys. Lett. **92**, 221110(3pp) (2008)
1215. M. Grundmann, Theory of semiconductor solid and hollow nano- and microwires with hexagonal cross-section under torsion. Phys. Status Solidi B **252**, 773–785 (2015)
1216. A.D. Andreev, E.P. O'Reilly, Theory of the electronic structure of GaN/AlN hexagonal quantum dots. Phys. Rev. B **62**, 15851–15870 (2000)
1217. J.K. Furdyna, Diluted magnetic semiconductors. J. Appl. Phys. **64**, R29–R64 (1988)
1218. J.K. Furdyna, J. Kossut (eds.), Diluted magnetic semiconductors. Semicond. Semimet. **25** (1988)
1219. D. Awshalom, D. Loss, N. Samarth (eds.), *Semiconductor Spintronics and Quantum Computation, Nanoscience and Technology* (Springer, Berlin, 2002)
1220. S.J. Pearton, C.R. Abernathy, M.E. Overberg, G.T. Thaler, D.P. Norton, N. Theodoropoulou, A.F. Hebard, Y.D. Park, F. Ren, J. Kim, L.A. Boatner, Wide band gap ferromagnetic semiconductors and oxides. J. Appl. Phys. **93**, 1–13 (2003)
1221. T. Dietl, H. Ohno, Dilute ferromagnetic semiconductors: physics and spintronic structures. Rev. Mod. Phys. **86**, 187–251 (2014)
1222. I. Tsubokawa, On the magnetic properties of a $CrBr_3$ single crystal. J. Phys. Soc. Jpn. **15**, 1664–1668 (1960)
1223. B.T. Matthias, R.M. Bozorth, J.H. van Vleck, Ferromagnetic interaction in EuO. Phys. Rev. Lett. **7**, 160–161 (1961)
1224. G. Güntherodt, Optical properties and electronic structure of europium chalcogenides. Phys. Condens. Matter **18**, 37–78 (1974)
1225. P.G. Steeneken, New light on EuO thin films: preparation, transport, magnetism and spectroscopy of a ferromagnetic semiconductor. Ph.D. thesis, Rijksuniversiteit Groningen (2002)
1226. P. Wachter, Europium chalcogenides: EuO, EuS, EuSe, and EuTe, in *Handbook on the Physics and Chemistry of Rare Earths*, ed. by K.A. Gschneider Jr, L. Eyring (North-Holland, Amsterdam, 1979), pp. 507–574
1227. T. Fukumura, H. Toyosaki, Y. Yamada, Magnetic oxide semiconductors. Semicond. Sci. Technol. **20**, S103–S111 (2005)
1228. W. Giriat, J.K. Furdyna, Crystal structure, composition, and materials preparation of diluted magnetic semiconductors in Ref. [1217], pp. 1–34
1229. C. Rigaux, Magnetooptics in narrow gap diluted magnetic semiconductors in Ref. [1217], pp. 229–274
1230. C. Domb, N.W. Dalton, Crystal statistics with long-range forces. I. The equivalent neighbour model. Proc. Phys. Soc. **89**, 859–871 (1966)
1231. C. Zener, Interaction between the d shells in the transition metals. Phys. Rev. **81**, 440–444 (1951)
1232. H. Akai, Ferromagnetism and its stability in the diluted magnetic semiconductor (In, Mn)As. Phys. Rev. Lett. **81**, 3002–3005 (1998)
1233. T. Dietl, H. Ohno, F. Matsukura, J. Cibert, D. Ferrand, Zener model description of ferromagnetism in zinc-blende magnetic semiconductors. Science **287**, 1019–1022 (2000)
1234. H. Munekata, H. Ohno, S. von Molnár, A. Segmüller, L.L. Chang, L. Esaki, Diluted magnetic III-V semiconductors. Phys. Rev. Lett. **63**, 1849–1852 (1989)
1235. H. Ohno, M. Munekata, T. Penney, S. von Molnár, L.L. Chang, Magnetotransport properties of p-type (In, Mn)As diluted magnetic III-V semiconductors. Phys. Rev. Lett. **68**, 2664–2667 (1992)
1236. H. Ohno, A. Shen, F. Matsukura, A. Oiwa, A. Endo, S. Katsumoto, Y. Iye, (Ga, Mn)As: a new diluted magnetic semiconductor based on GaAs. Appl. Phys. Lett. **69**, 363–365 (1996)

1237. Y.-J. Zhao, P. Mahadevan, A. Zunger, Comparison of predicted ferromagnetic tendencies of Mn substituting the Ga site in III-V's and in I-III-VI_2 chalcopyrite semiconductors. Appl. Phys. Lett. **84**, 3753–3755 (2004)
1238. H. Ohno, D. Chiba, F. Matsukura, T. Omiya, E. Abe, T. Dietl, Y. Ohno, K. Ohtani, Electric-field control of ferromagnetism. Nature **408**, 944–946 (2000)
1239. S.T.B. Goennenwein, Th.A. Wassner, H. Huebl, M.S. Brandt, J.B. Philipp, M. Opel, R. Gross, A. Koeder, W. Schoch, A. Waag, Hydrogen control of ferromagnetism in a dilute magnetic semiconductor. Phys. Rev. Lett. **92**, 227202(4pp) (2004)
1240. P. Sharma, A. Gupta, K.V. Rao, F.J. Owens, R. Sharma, R. Ahuja, J.M. Osorio, Guillen, B. Johansson, G.A. Gehring, Ferromagnetism above room temperature in bulk and transparent thin films of Mn-doped ZnO. Nat. Mater. **2**, 673–677 (2003)
1241. M. Diaconu, H. Schmidt, H. Hochmuth, M. Lorenz, G. Benndorf, D. Spemann, A. Setzer, P. Esquinazi, A. Pöppl, H. von Wenckstern, K.-W. Nielsen, R. Gross, H. Schmid, W. Mader, G. Wagner, M. Grundmann, Room-temperature ferromagnetic Mn-alloyed ZnO films obtained by pulsed laser deposition. J. Magn. Magn. Mater. **307**, 212–221 (2006)
1242. E.H. Hall, On the "Rotational coefficient" in nickel and cobalt. Philos. Mag. Ser. 5 12(74), 157–172 (1881)
1243. N. Nagaosa, J. Sinova, S. Onoda, A.H. MacDonald, N.P. Ong, Anomalous Hall effect. Rev. Mod. Phys. **82**, 1539–1592 (2010)
1244. M.I. Miah, Pure anomalous Hall effect in nonmagnetic zinc-blende semiconductors. Opt. Quantum Electron. **40**, 1033–1042 (2008)
1245. E.I. Rashba, Electron spin operation by electric fields: spin dynamics and spin injection. Phys. E **20**, 189–195 (2004)
1246. M.I. D'yakonov, V.I. Perel', Spin relaxation of conduction electrons in noncentrosymetric semiconductors. Fiz. Tverd. Tela **13**, 3581–3585 (1971) (Sov. Phys. Solid State **13**, 3023–3026 (1972))
1247. S. Datta, B. Das, Electronic analog of the electroptic modulator. Appl. Phys. Lett. **56**, 665–667 (1990)
1248. A.T. Hanbicki, O.M.J. van 't Erve, R. Magno, G. Kioseoglou, C.H. Li, B.T. Jonker, Analysis of the transport process providing spin injection through an Fe/AlGaAs Schottky barrier. Appl. Phys. Lett. **82**, 4092–4094 (2003)
1249. B.T. Jonker, S.C. Erwin, A. Petrou, A.G. Petukhov, Electrical spin injection and transport in semiconductor spintronic device. MRS Bull. **28**, 740–748 (2003)
1250. C.L. Braun, Organic Semiconductors. Handbook on Semiconductors, vol. 3 (North Holland, Amsterdam, 1980), pp. 857–873
1251. W. Brütting (ed.), *Physics of Organic Semiconductors* (Wiley-VCH, Weinheim, 2005)
1252. ThU Kampen, *Low Molecular Weight Organic Semiconductors* (Wiley-VCH, Weinheim, 2009)
1253. A. Curioni, W. Andreoni, R. Treusch, F.J. Himpsel, E. Haskal, P. Seidler, C. Heske, S. Kakar, T. van Buuren, L.J. Terminello, Atom-resolved electronic spectra for Alq_3 from theory and experiment. Appl. Phys. Lett. **72**, 1575–1577 (1998)
1254. D. Braun, A. Heeger, Visible light emission from semiconducting polymer diodes. Appl. Phys. Lett. **58**, 1982–1984 (1991)
1255. A. McL, J.M. Mathieson, V.C.Sinclair Robertson, The crystal and molecular structure of anthracene. I. X-ray measurements. Acta Cryst. **3**, 245–250 (1950)
1256. J.M. Robertson, V.C. Sinclair, J. Trotter, The crystal and molecular structure of tetracene. Acta Cryst. **14**, 697–704 (1961)
1257. N. Karl, High purity organic molecular crystals, in *Crystals: Growth, Properties and Applications*, vol. 4, ed. by H.C. Freyhardt (Springer, Berlin, 1980), pp. 1–100
1258. N. Karl, Growth and electrical properties of high purity organic molecular crystals. J. Cryst. Growth **99**, 1009–1016 (1990)
1259. Ch. Kloc, P.G. Simpkins, T. Siegrist, R.A. Laudise, Physical vapor growth of centimeter-sized crystals of α-hexathiophene. J. Cryst. Growth **182**, 416–427 (1997)

1260. R.A. Laudise, Ch. Kloc, P.G. Simpkins, T. Siegrist, Physical vapor growth of organic semiconductors. J. Cryst. Growth **187**, 449–454 (1998)
1261. M. Mas-Torrent, M. Durkut, P. Hadley, X. Ribas, C. Rovira, High mobility of dithiophene-tetrathiafulvalene single-crystal organic field effect transistors. J. Am. Chem. Soc. **126**, 984–985 (2004)
1262. M. Campione, R. Ruggerone, S. Tavazzi, M. Moret, Growth and characterisation of centimetre-sized single crystals of molecular organic materials. J. Mater. Chem. **15**, 2437–2443 (2005)
1263. M. Oehzelt, A. Aichholzer, R. Resel, G. Heimel, E. Venuti, R.G. Della Valle, Crystal structure of oligoacenes under high pressure. Phys. Rev. B **74**, 104103(7pp) (2006)
1264. B. Part, A. Kahn, N. Koch, W. Gao, Electronic structure and electrical properties of interfaces between metals and π-conjugated molecular films. J. Polym. Sci. Polym. Phys. **41**, 2529–2548 (2003)
1265. M. Pfeiffer, A. Beyer, T. Fritz, K. Leo, Controlled doping of phthalocyanine layers by cosublimation with acceptor molecules: a systematic seebeck and conductivity study. Appl. Phys. Lett. **73**, 3202–3204 (1998)
1266. B.A. Gregg, S.-G. Chen, R.A. Cormier, Coulomb forces and doping in organic semiconductors. Chem. Mater. **16**, 4586–4599 (2004)
1267. G.F. Neumark, Concentration and temperature dependence of impurity-to-band activation energies. Phys. Rev. B **5**, 408–414 (1972)
1268. B.A. Gregg, S.-G. Chen, H.M. Branz, On the superlinear increase in conductivity with dopant concentration in excitonic semiconductors. Appl. Phys. Lett. **84**, 1707–1709 (2004)
1269. A.J. Heeger, S. Kivelson, J.R. Schrieffer, W.-P. Su, Solitons in conducting polymers. Rev. Mod. Phys. **60**, 781–850 (1988)
1270. P.W.M. Blom, M.C.J.M. Vissenberg, Charge transport in poly(p-phenylene vinylene) light-emitting diodes. Mater. Sci. Eng. **27**, 53–94 (2000)
1271. L.B. Schein, Temperature independent drift mobility along the molecular direction of As_2Se_3. Phys. Rev. B **15**, 1024–1034 (1977)
1272. L.B. Schein, C.B. Duke, A.R. McGhie, Observation of the band-hopping transition for electrons in naphthalene. Phys. Rev. Lett. **40**, 197–200 (1978)
1273. L. Li, G. Meller, H. Kosina, Temperature and field-dependence of hopping conduction in organic semiconductors. Microelectron. J. **38**, 47–51 (2007)
1274. Y. Kawasumi, I. Akai, T. Karasawa, Photoluminescence and dynamics of excitons in Alq_3 single crystals. Int. J. Mod. Phys. B **15**, 3825–3828 (2001)
1275. M.A. Baldo, S.R. Forrest, Transient analysis of organic electrophosphorescence: I. Transient analysis of triplet energy transfer. Phys. Rev. B **62**, 10958–10966 (2000)
1276. W. Humbs, H. Zhang, M. Glasbeek, Femtosecond fluorescence upconversion spectroscopy of vapor-deposited tris(8-hydroxyquinoline) aluminum films. Chem. Phys. **254**, 319–327 (2000)
1277. Z. Vardeny, E. Ehrenfreund, J. Shinar, F. Wudl, Photoexcitation spectroscopy of polythiophene. Phys. Rev. B **35**, 2498–2500 (1987)
1278. J.-W. van der Horst, P.A. Bobbert, M.A.J. Michels, Electronic and optical excitations in crystalline conjugated polymers. Phys. Rev. B **66**, 035206(7pp) (2002)
1279. A. Baldo, D.F. O'Brien, M.E. Thompson, S.R. Forrest, Excitonic singlet-triplet ratio in a semiconducting organic thin film. Phys. Rev. B **60**, 14422–14428 (1999)
1280. A. Baldo, D.F. O'Brien, Y. You, A. Shoustikov, S. Sibley, M.E. Thompson, S.R. Forrest, Highly efficient phosphorescent emission from organic electroluminescent devices. Nature **395**, 151–154 (1998)
1281. C. Adachi, M.A. Baldo, M.E. Thompson, S.R. Forrest, Nearly 100% internal phosphorescence efficiency in an organic light emitting device. J. Appl. Phys. **90**, 5048–5051 (2001)
1282. M.A. Baldo, S. Lamansky, P.E. Burrows, M.E. Thompson, S.R. Forrest, Very high-efficiency green organic light-emitting devices based on electrophosphorescence. Appl. Phys. Lett. **75**, 4–6 (1999)

1283. M.A. Baldo, M.E. Thompson, S.R. Forrest, High-efficiency fluorescent organic light-emitting devices using a phosphorescent sensitizer. Nature **403**, 750–753 (2000)
1284. T. Förster, Transfer mechanisms of electron excitation. Discuss. Faraday Soc. **27**, 7–17 (1959)
1285. K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, Y. Zhang, S.V. Dubonos, I.V. Grigorieva, A.A. Firsov, Electric field effect in atomically thin carbon films. Science **306**, 666–669 (2004)
1286. K.S. Novoselov, D. Jiang, F. Schedin, T.J. Booth, V.V. Khotkevich, S.V. Morozov, A.K. Geim, Two-dimensional atomic crystals. PNAS **102**, 10451–10453 (2005)
1287. J.S. Bunch, A.M. van der Zande, S.S. Verbridge, I.W. Frank, D.M. Tanenbaum, J.M. Parpia, H.G. Craighead, P.L. McEuen, Electromechanical resonators from graphene sheets. Science **315**, 490–493 (2007)
1288. University of Manchester, http://www.condmat.physics.manchester.ac.uk
1289. L. Wirtz, A. Rubio, The phonon dispersion of graphite revisited. Solid State Commun. **131**, 141–152 (2004)
1290. ChL Kane, Erasing electron mass. Nature **438**, 168–170 (2005)
1291. K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, M.I. Katsnelson, I.V. Grigorieva, S.V. Dubonos, A.A. Firsov, Two-dimensional gas of massless Dirac fermions in graphene. Nature **438**, 197–200 (2005)
1292. P.R. Wallace, The band theory of graphite. Phys. Rev. **71**, 622–634 (1947)
1293. R. Saito, G. Dresselhaus, M.S. Dresselhaus, *Physical Properties of Carbon Nanotubes* (Imperial, London, 1998)
1294. S. Reich, J. Maultzsch, C. Thomsen, P. Ordejón, Tight-binding description of graphene. Phys. Rev. B **66**, 035412(5pp) (2002)
1295. A. Bostwick, T. Ohta, K. Th Seyller, E.Rotenberg Horn, Quasiparticle dynamics in graphene. Nat. Phys. **3**, 36–40 (2006)
1296. B. Partoens, F.M. Peeters, From graphene to graphite: electronic structure around the K point. Phys. Rev. B **74**, 075404(11pp) (2006)
1297. T. Ohta, A. Bostwick, K. Th Seyller, E. Rotenberg, Horn, Controlling the electronic structure of bilayer graphene. Science **313**, 951–954 (2006)
1298. S.V. Vonsovsky, M.I. Katsnelson, *Quantum Solid-State Physics* (Springer, New York, 1989)
1299. M.S. Sercheli, Y. Kopelevich, R. Ricardo da Silva, J.H.S. Torres, C. Rettori, Evidence for internal field in graphite: a conduction electron spin-resonance study. Solid State Commun. **121**, 579–583 (2002)
1300. K.I. Bolotin, K.J. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim, H.L. Störmer, Ultrahigh electron mobility in suspended graphene. Solid State Commun. **146**, 351–355 (2008)
1301. K.S. Novoselov, Z. Jiang, Y. Zhang, S.V. Morozov, H.L. Störmer, U. Zeitler, J.C. Maan, G.S. Boebinger, P. Kim, A.K. Geim, Room-temperature quantum Hall effect in graphene. Science **315**, 1379 (2007)
1302. O. Klein, Die Reflexion yon Elektronen an einem Potentialsprung nach der relativistischen Dynamik von Dirac. Z. Phys. **53**, 157–165 (1929)
1303. A. Calogeracos, N. Dombey, History and physics of the Klein paradox. Contemp. Phys. **40**, 313–321 (1999)
1304. M.I. Katsnelson, K.S. Novoselov, A.K. Geim, Chiral tunnelling and the Klein paradox in graphene. Nat. Phys. **2**, 620–625 (2006)
1305. C. Tusche, H.L. Meyerheim, J. Kirschner, Observation of depolarized ZnO(0001) monolayers: formation of unreconstructed planar sheets. Phys. Rev. Lett. **99**, 026102(4pp) (2007)
1306. Y. Miyamoto, M.L. Cohen, S.G. Louie, Ab initio calculation of phonon spectra for graphite, BN, and BC_2N sheets. Phys. Rev. B **52**, 14971–14975 (1995)
1307. S. Iijima, Helical microtubules of graphitic carbon. Nature **354**, 56–58 (1991)
1308. S. Iijima, T. Ichihasi, Single-shell carbon nanotubes of 1-nm diameter. Nature **363**, 603–605 (1993)

1309. S. Reich, C. Thomsen, J. Maultzsch, *Carbon Nanotubes: Basic Concepts and Physical Properties* (Wiley-VCH, Berlin, 2004)
1310. J. Kürti, V. Zólyomi, M. Kertesz, G. Sun, The geometry and the radial breathing mode of carbon nanotubes: beyond the ideal behaviour. New J. Phys. **5**, 125(21pp) (2003)
1311. R. Saito, M. Fujita, G. Dresselhaus, M.S. Dresselhaus, Electronic structure of chiral graphene tubules. Appl. Phys. Lett. **60**, 2204–2206 (1992)
1312. M.-F. Yu, B.S. Files, S. Arepalli, R.S. Ruoff, Tensile loading of ropes of single wall carbon nanotubes and their mechanical properties. Phys. Rev. Lett. **84**, 5552–5555 (2000)
1313. S. Ogata, Y. Shibutani, Ideal tensile strength and band gap of single-walled carbon nanotubes. Phys. Rev. B **68**, 165409(4pp) (2003)
1314. R. Saito, M. Fujita, G. Dresselhaus, M.S. Dresselhaus, Electronic structure of carbon fibers based on C_{60}. MRS Symp. Proc. **247**, 333 (1992)
1315. N. Hamada, S.-I. Sawada, A. Oshiyama, New one-dimensional conductors: graphitic microtubule. Phys. Rev. Lett. **68**, 1579–1581 (1992)
1316. X. Blase, L.X. Benedict, E.L. Shirley, S.G. Louie, Hybridization effects and metallicity in small radius carbon nanotubes. Phys. Rev. Lett. **72**, 1878–1881 (1994)
1317. J.W. Mintmire, C.T. White, Universal density of states for carbon nanotubes. Phys. Rev. Lett. **81**, 2506–2509 (1989)
1318. A. Grüneis, R. Saito, G.G. Samsonidze, T. Kimura, M.A. Pimenta, A. Jorio, A.G. Souza Filho, G. Dresselhaus, M.S. Dresselhaus, Inhomogeneous optical absorption around the K point in graphite and carbon nanotubes. Phys. Rev. B **67**, 165402(7pp) (2003)
1319. S.M. Bachilo, M.S. Strano, C. Kittrell, R.H. Hauge, R.E. Smalley, R.B. Weisman, Structure-assigned optical spectra of single-walled carbon nanotubes. Science **298**, 2361–2366 (2002)
1320. H. Telg, J. Maultzsch, S. Reich, F. Hennrich, C. Thomsen, Chirality distribution and transition energies of carbon nanotubes. Phys. Rev. Lett. **93**, 177401(4pp) (2004)
1321. C. Thomsen, S. Reich, Raman Scattering, in *Carbon Nanotubes. Light Scattering in Solids IX*, ed. by M. Cardona, R. Merlin Topics in Applied Physics, vol. 108, (Springer, Berlin, 2007), pp. 115–234
1322. N.G. Chopra, R.J. Luyken, K. Cherrey, V.H. Crespi, M.L. Cohen, S.G. Louie, A. Zettl, Boron nitride nanotubes. Science **269**, 966–967 (1995)
1323. R.S. Lee, J. Gavillet, M.L. de la Chapelle, A. Loiseau, J.-L. Cochon, D. Pigache, J. Thibault, F. Willaime, Catalyst-free synthesis of boron nitride single-wall nanotubes with a preferred zig-zag configuration. Phys. Rev. B **64**, 121405(R)(4pp) (2001)
1324. A. Rubio, J.L. Corkill, M.L. Cohen, Theory of graphitic boron nitride nanotubes. Phys. Rev. B **49**, 5081–5084 (1994)
1325. L. Wirtz, V. Olevano, A.G. Marinopoulos, L. Reining, A. Rubio, Optical absorption in small BN and C nanotubes. AIP Conference Proceedings **685**, 406–410 (2003)
1326. R.R. Willey, *Practical Design and Production of Optical Thin Films* (CRC Press, London, 2002)
1327. J.M. Bendickson, J.P. Dowling, M. Scalora, Analytic expressions for the electromagnetic mode density in finite, one-dimensional, photonic band-gap structures. Phys. Rev. E **53**, 4107–4121 (1996)
1328. J. Gottmann, A. Husmann, T. Klotzbücher, E.W. Kreutz, Optical properties of alumina and zirconia thin films grown by pulsed laser deposition. Surf. Coat. Technol. **100–101**, 415–419 (1998)
1329. J. Sellmann, Ch. Sturm, R. Schmidt-Grund, Ch. Czekalla, J. Lenzner, H. Hochmuth, B. Rheinländer, M. Lorenz, M. Grundmann, Structural and optical properties of ZrO_2 and Al_2O_3 thin films and Bragg reflectors grown by pulsed laser deposition. Phys. Status Solidi C **5**, 1240–1243 (2008)
1330. B. Rauschenbach, Leibniz-Institut für Oberflächenmodifizierung, Leipzig, private communication (2006)
1331. N. Kaiser, Fraunhofer-Institute for Applied Optics, Jena, Germany
1332. C.M. Soukoulis (ed.), *Photonic Crystals and Light Localization in the 21st Century*. NATO Science Series C: Mathematical and Physical Sciences, vol. 563 (Kluwer Academic Publishers, Dordrecht, 1996)

1333. J.D. Joannopoulos, R.D. Meade, J.N. Winn, *Photonic Crystals* (Princeton University Press, Princeton, 1995)
1334. K. Busch, S. Lölkes, R.B. Wehrspohn, H. Föll (eds.), *Photonic Crystals* (Wiley-VCH, Berlin, 1994)
1335. E.-X. Ping, Transmission of electromagnetic waves in planar, cylindrical, and spherical dielectric layer systems and their applications. J. Appl. Phys. **76**, 7188–7194 (1994)
1336. K.M. Ho, C.T. Chan, C.M. Soukoulis, Existence of a photonic gap in periodic dielectric structures. Phys. Rev. Lett. **65**, 3152–3155 (1990)
1337. O. Toader, S. John, Proposed square spiral microfabrication architecture for large three-dimensional photonic band gap crystals. Science **292**, 1133–1135 (2001)
1338. S.R. Kennedy, M.J. Brett, O. Toader, S. John, Fabrication of tetragonal square spiral photonic crystals. Nano Lett. **2**, 59–62 (2002)
1339. K. Busch, S. John, Photonic band gap formation in certain self-organizing systems. Phys. Rev. E **58**, 3896–3908 (1998)
1340. E. Yablonovitch, T.J. Gmitter, K.M. Leung, Photonic band structure: the face-centered-cubic case employing nonspherical atoms. Phys. Rev. Lett. **67**, 2295–2298 (1991)
1341. K.M. Ho, C.T. Chan, C.M. Soukoulis, R. Biswas, M. Sigalas, Photonic band gaps in three dimensions: new layer-by-layer periodic structures. Solid State Commun. **89**, 413–416 (1994)
1342. A. Chutinan, S. Noda, Spiral three-dimensional photonic-band-gap structure. Phys. Rev. B **57**, R2006–R2008 (1998)
1343. S. Fan, P.R. Villeneuve, R.D. Meade, J.D. Joannopoulos, Design of three-dimensional photonic crystals at submicron lengthscales. Appl. Phys. Lett. **65**, 1466–1468 (1994)
1344. H.S. Sözüer, J.W. Haus, Photonic bands: simple-cubic lattice. J. Opt. Soc. Am. B **10**, 296–302 (1993)
1345. H.S. Sözüer, J.W. Haus, R. Inguva, Photonic bands: convergence problems with the plane-wave method. Phys. Rev. B **45**, 13962–13972 (1992)
1346. S. John, K. Busch, Photonic bandgap formation and tunability in certain self-organizing systems. IEEE J. Lightwave Technol. **17**, 1931–1943 (1999)
1347. F. Garcia-Santamaria, C. López, F. Meseguer, F. López-Tejeira, J. Sánchez-Dehesa, H.T. Miyazaki, Opal-like photonic crystal with diamond lattice. Appl. Phys. Lett. **79**, 2309–2311 (2001)
1348. A. Blanco, E. Chomski, S. Grabtchak, M. Ibisate, S. John, S.W. Leonard, C. Lopez, F. Meseguer, H. Miguez, J.P. Mondia, G.A. Ozin, O. Toader, H.M. van Driel, Large-scale synthesis of a silicon photonic crystal with a complete three-dimensional bandgap near 1.5 micrometres. Nature **405**, 437–440 (2000)
1349. A.F. Koenderink, A. Lagendijk, W.L. Vos, Optical extinction due to intrinsic structural variations of photonic crystals. Phys. Rev. B **72**, 153102(4pp) (2005)
1350. M. Lončar, D. Nedeljković, T. Doll, J. Vučović, A. Scherer, T.P. Pearsall, Waveguiding in planar photonic crystals. Appl. Phys. Lett. **77**, 1937–1939 (2000)
1351. Y. Zhu, Q. Wu, S. Morin, T.W. Mossberg, Observation of a two-photon gain feature in the strong-probe absorption spectrum of driven two-level atoms. Phys. Rev. Lett. **65**, 1200–1203 (1990)
1352. C. Weisbuch, M. Nishioka, A. Ishikawa, Y. Arakawa, Observation of the coupled exciton-photon mode splitting in a semiconductor quantum microcavity. Phys. Rev. Lett. **69**, 3314–3317 (1992)
1353. V. Savona, Z. Hradil, A. Quattropani, P. Schwendimann, Quantum theory of quantum-well polaritons in semiconductor microcavities. Phys. Rev. B **49**, 8774–8779 (1994)
1354. G. Khitrova, H.M. Gibbs, F. Jahnke, M. Kira, S.W. Koch, Nonlinear optics of normal-mode-coupling semiconductor microcavities. Rev. Mod. Phys. **71**, 1591–1639 (1999)
1355. M.S. Skolnick, T.A. Fisher, D.M. Whittaker, Strong coupling phenomena in quantum microcavity structures. Semicond. Sci. Technol. **13**, 645–669 (1998)
1356. P.J. Klar, G. Rowland, P.J.S. Thomas, A. Onischenko, T.E. Sale, T.J.C. Hosea, R. Grey, Photomodulated reflectance study of $In_xGa_{1-x}As/GaAs/AlAs$ microcavity vertical-cavity surface emitting laser structures in the weak-coupling regime: the cavity/ground-state-exciton resonance. Phys. Rev. B **59**, 2894–2901 (1999)

1357. J.J. Baumberg, P.G. Savvidis, R.M. Stevenson, A.I. Tartakovskii, M.S. Skolnick, D.M. Whittaker, J.S. Roberts, Parametric oscillation in a vertical microcavity: a polariton condensate or micro-optical parametric oscillation. Phys. Rev. B **62**, R16247–R16250 (2000)
1358. A. Kavokin, G. Malpuech, B. Gil, Semiconductor microcavities: towards polariton lasers. MRS Internet J. Nitride Semicond. Res. **8**, 3(25pp) (2003)
1359. L.S. Dang, D. Heger, R. André, F. Bœuf, R. Romestain, Stimulation of polariton photoluminescence in semiconductor microcavity. Phys. Rev. Lett. **81**, 3920–3923 (1998)
1360. A. Pawlis, D.J. As, D. Schikora, J. Schörmann, K. Lischka, Photonic devices based on wide gap semiconductors for room temperature polariton emission. Phys. Status Solidi C **1**, S202–S209 (2004)
1361. T. Tawara, H. Gotoh, T. Akasaka, N. Kobayashi, T. Saitoh, Cavity polaritons in InGaN microcavities at room temperature. Phys. Rev. Lett. **92**, 256402(4pp) (2004)
1362. N. Antoine-Vincent, F. Natali, D. Byrne, P. Disseix, A. Vasson, J. Leymarie, F. Semond, J. Massies, Potentialities of GaN-based microcavities in strong coupling regime at room temperature. Superlattice Microstruct. **36**, 599–606 (2004)
1363. E. Feltin, G. Christmann, R. Butté, J.-F. Carlin, M. Mosca, N. Grandjean, Room temperature polariton luminescence from a GaN/AlGaN quantum well microcavity. Appl. Phys. Lett. **89**, 071107(3pp) (2006)
1364. F. Semond, I.R. Sellers, F. Natali, D. Byrne, M. Leroux, J. Massies, N. Ollier, J. Leymarie, P. Disseix, A. Vasson, Strong light-matter coupling at room temperature in simple geometry GaN microcavities grown on silicon. Appl. Phys. Lett. **87**, 021102(3pp) (2005)
1365. R. Schmidt-Grund, B. Rheinländer, C. Czekalla, G. Benndorf, H. Hochmuth, M. Lorenz, M. Grundmann, Exciton-polariton formation at room temperature in a planar ZnO resonator structure. Appl. Phys. A **93**, 331–337 (2008)
1366. P.G. Savvidis, J.J. Baumberg, R.M. Stevenson, M.S. Skolnick, D.M. Whittaker, J.S. Roberts, Angle-resonant stimulated polariton amplifier. Phys. Rev. Lett. **84**, 1547–1550 (2000)
1367. Y. Yamamoto, T. Tassone, H. Cao, *Semiconductor Cavity Quantum Electrodynamics* (Springer, Berlin, 2000)
1368. D. Sanvitto, V. Timofeev, *Exciton Polaritons in Microcavities* (Springer, Berlin, 2012)
1369. S.L. McCall, A.F.J. Levi, R.E. Slusher, S.J. Pearton, R.A. Logan, Whispering-gallery mode microdisk lasers. Appl. Phys. Lett. **60**, 289–291 (1992)
1370. R.P. Wang, M.-M. Dumitrescu, Theory of optical modes in semiconductor microdisk lasers. J. Appl. Phys. **81**, 3391–3397 (1997)
1371. A.F.J. Levi, R.E. Slusher, S.L. McCall, J.L. Glass, S.J. Pearton, R.A. Logan, Directional light coupling from microdisk lasers. Appl. Phys. Lett. **62**, 561–563 (1993)
1372. S.-K. Kim, S.-H. Kim, G.-H. Kim, H.-G. Park, D.-J. Shin, Y.-H. Lee, Highly directional emission from few-micron-size elliptical microdisks. Appl. Phys. Lett. **84**, 861–863 (2004)
1373. E. Peter, P. Senellart, D. Martrou, A. Lemaître, J. Hours, J.M. Gérard, J. Bloch, Exciton-photon strong-coupling regime for a single quantum dot embedded in a microcavity. Phys. Rev. Lett. **95**, 067401(4pp) (2005)
1374. E.M. Purcell, Spontaneous emission probabilities at radio frequencies. Phys. Rev. **69**, 681 (1946)
1375. J.M. Gérard, B. Sermage, B. Gayral, B. Legrand, E. Costard, V. Thierry-Mieg, Enhanced spontaneous emission by quantum boxes in a monolithic optical microcavity. Phys. Rev. Lett. **81**, 1110–1113 (1998)
1376. A. Kiraz, P. Michler, C. Becher, B. Gayral, A. Imamoğlu, L. Zhang, E. Hu, W.V. Schoenfeld, P.M. Petroff, Cavity-quantum electrodynamics using a single InAs quantum dot in a microdisk structure. Appl. Phys. Lett. **78**, 3932–3934 (2001)
1377. J.U. Nöckel, A.D. Stone, Ray and wave chaos in asymmetric resonant optical cavities. Nature **385**, 45–47 (1997)
1378. C. Gmachl, F. Capasso, E.E. Narimanov, J.U. Nöckel, A.D. Stone, J. Faist, D.L. Sivco, A.Y. Cho, High-power directional emission from microlasers with chaotic resonators. Science **280**, 1556–1564 (1998)

1379. G.D. Chern, H.E. Tureci, A. Douglas, Stone, R.K. Chang, M. Kneissl, N.M. Johnson, Unidirectional lasing from InGaN multiple-quantum-well spiral-shaped micropillars. Appl. Phys. Lett. **83**, 1710–1712 (2003)
1380. S.-Y. Lee, S. Rim, J.-W. Ryu, T.-Y. Kwon, M. Choi, C.-M. Kim, Quasiscarred resonances in a spiral-shaped microcavity. Phys. Rev. Lett. **93**, 164102(4pp) (2004)
1381. M. Kneissl, M. Teepe, N. Miyashita, N.M. Johnson, G.D. Chern, R.K. Chang, Current-injection spiral-shaped microcavity disk laser diodes with unidirectional emission. Appl. Phys. Lett. **84**, 2485–2498 (2004)
1382. T. Nobis, E.M. Kaidashev, A. Rahm, M. Lorenz, M. Grundmann, Whispering gallery modes in nanosized dielectric resonators with hexagonal cross section. Phys. Rev. Lett. **93**, 103903(4pp) (2004)
1383. F. Leiter, H. Zhou, F. Henecker, A. Hofstaetter, D.M. Hofmann, B.K. Meyer, Magnetic resonance experiments on the green emission in undoped ZnO crystals. Phys. B **308–310**, 908–911 (2001)
1384. J. Wiersig, Hexagonal dielectric resonators and microcrystal lasers. Phys. Rev. A **67**, 023807(12pp) (2003)
1385. C. Czekalla, C. Sturm, R. Schmidt-Grund, B. Cao, M. Lorenz, M. Grundmann, Whispering gallery mode lasing in zinc oxide microwires. Appl. Phys. Lett. **92**, 241102(3pp) (2008)
1386. C.P. Dietrich, M. Lange, T. Böntgen, M. Grundmann, The corner effect in hexagonal whispering gallery microresonators. Appl. Phys. Lett. **101**, 141116(5pp) (2012)
1387. F. Streintz, Ueber die elektrische Leitfähigkeit vom gepressten Pulvern. Ann. Physik **314**, 854–885 (1902)
1388. H.L. Hartnagel, A.L. Dawar, A.K. Jain, C. Jagadish, *Semiconducting Transparent Thin Films* (Institute of Physics Publishing, Philadelphia, 1995)
1389. MRS Bull. **25**(8), issue on Transparent Conducting Oxides
1390. T. Minami, Transparent conducting oxide semiconductors for transparent electrodes. Semicond. Sci. Technol. **20**, S35–S44 (2005)
1391. R.G. Gordon, Criteria for choosing transparent conductors. MRS Bull. **25**, 52–57 (2000)
1392. H. Kawazoe, M. Yasukawa, H. Hyodo, M. Kurita, H. Yanagi, H. Hosono, P-type electrical conduction in transparent thin films of $CuAlO_2$. Nature **389**, 939–942 (1997)
1393. M. Dekkers, G. Rijnders, D.H.A. Blank, $ZnIr_2O_4$, a p-type transparent oxide semiconductor in the class of spinel zinc-d^6-transition metal oxide. Appl. Phys. Lett. **90**, 021903(3pp) (2007)
1394. F.-L. Schein, H. von Wenckstern, H. Frenzel, M. Grundmann, ZnO-based n-channel junction field-effect transistor with room-temperature-fabricated amorphous p-type $ZnCo_2O_4$ gate. IEEE Electron Device Lett. **33**, 676–678 (2012)
1395. M. Marezio, Refinement of the crystal structure of In_2O_3 at two wavelengths. Acta Crystallogr. **20**, 723–728 (1966)
1396. S.H. Brewer, S. Franzen, Calculation of the electronic and optical properties of indium tin oxide by density functional theory. Chem. Phys. **300**, 285–293 (2004)
1397. O.N. Mryasov, A.J. Freeman, Electronic band structure of indium tin oxide and criteria for transparent conducting behavior. Phys. Rev. B **64**, 233111(3pp) (2001)
1398. K. Nomura, H. Ohta, A. Takagi, T. Kamiya, M. Hirano, H. Hosono, Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors. Nature **432**, 488–492 (2004)
1399. H. Agura, A. Suzuki, T. Matsushita, T. Aoki, M. Okuda, Low resistivity transparent conducting Al-doped ZnO films prepared by pulsed laser deposition. Thin Solid Films **445**, 263–267 (2003)
1400. K. Ellmer, Resistivity of polycrystalline zinc oxide films: current status and physical limit. J. Phys. D: Appl. Phys. **34**, 3097–3108 (2001)
1401. T. Minami, S. Suzuki, T. Miyata, Electrical conduction mechanism of highly transparent and conductive ZnO thin films. Proc. Mater. Res. Soc. **666**, F1.3.1 (2001)
1402. D.S. Ginley, C. Bright, Transparent conducting oxides. MRS Bull. **8**, 15–18 (2000)
1403. J.R. Bellingham, W.A. Phillips, C.J. Adkins, Intrinsic performance limits in transparent conducting oxides. J. Mater. Sci. Lett. **11**, 263–265 (1992)

1404. J.A. Fleming, *The Thermionic Valve and Its Developments in Radiotelegraphy and Telephony* (The Wireless Press, London, 1919)
1405. G.S. Meikle, The hot cathode argon gas filled rectifier. Gen. Electr. Rev. XIX(4), 297–304 (1916); see also. J. Frankl. Inst. **181**, 704–705 (1916)
1406. H.K. Henisch, Rectifying semiconductor contacts. J. Electrochem. Soc. **103**, 637–643 (1956)
1407. H.K. Henisch, *Rectifying Semiconductor Contacts* (Oxford University Press (Clarendon), London, 1957)
1408. J.R. Macdonald, Accurate solution of an idealized one-carrier metal-semiconductor junction problem. Solid State Electron. **5**, 11–37 (1962)
1409. F.A. Padovani, The voltage-current characteristic of metal-semiconductor contacts. Semicond. Semimet. **7A**, 75–146 (1971)
1410. D.J. Coe, E.H. Rhoderick, P.H. Gerzon, A.W. Tinsley, Silicide formation in Rh-Si Schottky barrier diodes. Inst. Phys. Conf. Ser. **22**, 74–83 (1974)
1411. E.H. Rhoderick, R.H. Williams, *Metal-Semiconductor Contacts*, 2nd edn. (Oxford University Press, Oxford, 1988)
1412. R.T. Tung, Recent advances in Schottky barrier concepts. Mater. Sci. Eng. R **35**, 1–138 (2001)
1413. bibitemmetalwf G.N. Derry, M.E. Kern, E.H. Worth, Recommended values of clean metal surface work functions, J. Vac. Sci. Technol., A **33**, 060801(9pp) (2015)
1414. K.W. Boer, *Introduction to Space Charge Effects in Semiconductors* (Springer, Heidelberg, 2010)
1415. A.M. Cowley, S.M. Sze, Surface states and barrier height of metal-semiconductor systems. J. Appl. Phys. **36**, 3212–3220 (1965)
1416. A. Dimoulas, P. Tsipas, A. Sotiropoulos, E.K. Evangelou, Fermi-level pinning and charge neutrality level in germanium, Appl. Phys. Lett. **89**, 252110(3pp) (2006)
1417. C.A. Mead, Physics of interfaces, in *Ohmic Contacts to Semiconductors*, ed. by B. Schwartz (Electrochemical Society, New York, 1969), pp. 3–16
1418. S. Kurtin, T.C. McGill, C.A. Mead, Fundamental transition in the electronic nature of solids. Phys. Rev. Lett. **22**, 1433–1436 (1969)
1419. C.A. Mead, Surface barriers on ZnSe and ZnO. Phys. Lett. **18**, 218 (1965)
1420. J. Bardeen, Surface states and rectification at a metal-semiconductor contact. Phys. Rev. **71**, 717–727 (1947)
1421. W. Mönch, 125 Years of metal-semiconductor contacts: where do we stand? Adv. Solid State Phys. (Festkörperprobleme) **39**, 13–24 (1999)
1422. M. Schmidt, R. Pickenhain, M. Grundmann, Exact solutions for the capacitance of space charge regions at semiconductor interfaces. Solid State Electron. **51**, 1002–1004 (2007)
1423. W. Schottky, Über den Einfluß von Strukturwirkungen, besonders der Thomsonschen Bildkraft, auf die Elektronenemission der Metalle. Phys. Zeitschrift **15**, 872–878 (1914)
1424. E.D. Courant, Image force and tunnel effect in crystal rectifers. Phys. Rev. **69**, 684 (1946)
1425. W.E. Meyerhof, Contact potential difference in silicon crystal rectifiers. Phys. Rev. **71**, 727–735 (1947)
1426. S.M. Sze, C.R. Crowell, D. Kahng, Photoelectric determination of the image force dielectric constant for hot electrons in Schottky barriers. J. Appl. Phys. **35**, 2534–2536 (1964)
1427. A.M. Goodman, Metal-semiconductor barrier height measurement by the differential capacitance method-one carrier system. J. Appl. Phys. **34**, 329–338 (1963)
1428. C. van Opdorp, Evaluation of doping profiles from capacitance measurements. Solid State Electron. **11**, 397–406 (1968)
1429. S. Dueñas, M. Jaraiz, J. Vicente, E. Rubio, L. Bailón, J. Barbolla, Optical admittance spectroscopy: a new method for deep level characterization. J. Appl. Phys. **61**, 2541–2545 (1987)
1430. D.L. Losee, Admittance spectroscopy of impurity levels in Schottky barriers. J. Appl. Phys. **46**, 2204–2214 (1975)
1431. M.P. Verkhovodov, H.P. Peka, D.A. Pulemyotov, Capacitance behaviour of junctions with frozen dopant levels. Semicond. Sci. Technol. **8**, 1842–1847 (1993)

1432. F.D. Auret, J.M. Nel, M. Hayes, L. Wu, W. Wesch, E. Wendler, Electrical characterization of growth-induced defects in bulk-grown ZnO. Superlattices Microstruct. **39**, 17–23 (2006)
1433. M. Schmidt, M. Ellguth, C. Czekalla, H. von Wenckstern, R. Pickenhain, M. Grundmann, G. Brauer, W. Skorupa, M. Helm, Q. Gu, ChCh. Ling, Defects in zinc-implanted ZnO thin films. J. Vac. Sci. Technol. B **27**, 1597–1600 (2009)
1434. C. Wagner, Zur Theorie der Gleichrichterwirkung. Phys. Zeitschrift **32**, 641–645 (1931)
1435. W. Schottky, E. Spenke, Zur quantitativen Durchführung der Raumladungs- und Randschichttheorie der Kristallgleichrichter. Wiss. Veröff. Siemens Werke **18**, 225–291 (1939)
1436. H.A. Bethe, Theory of the boundary layer of crystal rectifiers. MIT Radiation Laboratory Report 43–12 (26pp) (1942)
1437. P.T. Landsberg, The theory of direct-current characteristics of rectifiers. Proc. R. Soc. Lond. A **206**, 463–477 (1951)
1438. P.T. Landsberg, Contributions to the theory of heterogeneous barrier layer rectifiers. Proc. R. Soc. Lond. A **206**, 477–488 (1951)
1439. H.C. Torrey, C.A. Whitmer, Crystal Rectifiers. MIT Radiation Laboratory Series, vol. 15 (McGraw-Hill Book Company, New York, 1948)
1440. M.W. Allen, S.M. Durbin, X. Weng, J.M. Redwing, K. Sarpatwari, S.E. Mohney, H. von Wenckstern, M. Grundmann, Temperature-dependent properties of nearly ideal ZnO Schottky diodes. IEEE Trans. Electron Devices **ED-56**, 2160–2164 (2009)
1441. L.F. Wagner, R.W. Young, A. Sugerman, A note on the correlation between the Schottky diode barrier height and the ideality factor as determined from I-V measurements. IEEE Electron Device Lett. **4**, 320–322 (1983)
1442. H.J. Yearian, D.C. Characteristics of Ge and Si crystal rectifiers. Phys. Rev. **69**, 682 (1946)
1443. R.G. Sachs, Theory of crystal rectifiers. Phys. Rev. **69**, 682 (1946)
1444. V.A. Johnson, R.N. Smith, H.J. Yearian, Semi-quantitative explanation of D.C. characteristics of crystal rectifiers. Phys. Rev. **69**, 682–683 (1946)
1445. H.J. Yearian, D.C. characteristics of silicon and germanium point contact crystal rectifiers. Part I. Experimental. J. Appl. Phys. **21**, 214–221 (1950)
1446. V.A. Johnson, R.N. Smith, H.J. Yearian, D.C. characteristics of silicon and dermanium point contact crystal rectifiers. Part II. The multicontact theory. J. Appl. Phys. **21**, 283–289 (1950)
1447. J.H. Werner, H.H. Güttler, Barrier inhomogeneities at Schottky contacts. J. Appl. Phys. **69**, 1522–1533 (1991)
1448. V.B. Bondarenko, S.G. YuA Kudinov, V.V.Korablev Ershov, Natural nonuniformities in the height of a Schottky barrier. Semiconductor **32**, 495–496 (1998)
1449. D. Defives, O. Noblanc, C. Dua, C. Brylinski, M. Barthula, V. Aubry- Fortuna, F. Meyer, Barrier inhomogeneities and electrical characteristics of Ti/4H-SiC schottky rectifiers. IEEE Trans. Electron Devices **ED-46**, 449–455 (1999)
1450. S. Müller, H. von Wenckstern, O. Breitenstein, J. Lenzner, M. Grundmann, Microscopic identification of hot spots in multibarrier Schottky contacts on pulsed laser deposition grown zinc oxide thin films. IEEE Trans. Electron Devices **ED-59**, 536–541 (2012)
1451. H. von Wenckstern, G. Biehne, R. Abdel Rahman, H. Hochmuth, M. Lorenz, M. Grundmann, Mean barrier height of Pd Schottky contacts on ZnO thin films. Appl. Phys. Lett. **88**, 092102(3pp) (2006)
1452. J.H. Werner, Schottky barrier and pn-junction I/V plots-small signal evaluation. Appl. Phys. A **47**, 291–200 (1988)
1453. F.A. Padovani, G.G. Sumner, Experimental study of gold-gallium arsenide Schottky barriers. J. Appl. Phys. **36**, 3744–3747 (1965)
1454. F.A. Padovani, R. Stratton, Field and thermionic-field emission in Schottky barriers. Solid State Electron. **9**, 695–707 (1966)
1455. W. Mönch, Barrier heights of real Schottky contacts explained by metal-induced gap states and lateral inhomogeneities. J. Vac. Sci. Technol. B **17**, 1867–1876 (1999)
1456. C.R. Crowell, S.M. Sze, Current transport in metal-semiconductor barriers. Solid State Electron. **9**, 1035–1048 (1966)

1457. C.R. Crowell, M. Beguwala, Recombination velocity effects on current diffusion and imref in Schottky barriers. Solid State Electron. **14**, 1149–1157 (1971)
1458. A. Lajn, H. von Wenckstern, M. Grundmann, G. Wagner, P. Barquinha, E. Fortunato, R. Martins, Comparative study of transparent rectifying contacts on semiconducting oxide single crystals and amorphous thin films. J. Appl. Phys. **113**, 044511(13pp) (2013)
1459. J.M. Andrews, M.P. Lepselter, Reverse current-voltage characteristics of metal-silicide Schottky diodes. Solid State Electron. **13**, 1011–1023 (1970)
1460. R. Holm, Electrical Contacts. Theory and Applications, 4th edn. (Springer, Berlin, 1981)
1461. A.Y.C. Yu, Electron tunneling and contact resistance of metal-silicon contact barriers. Solid State Electron. **13**, 239–247 (1970)
1462. C.Y. Chang, Y.K. Fang, S.M. Sze, Specific contact resistance of metal-semiconductor barriers. Solid State Electron. **14**, 541–550 (1971)
1463. T. Sanada, O. Wada, Ohmic contacts to p-GaAs with Au/Zn/Au structure. Jpn. J. Appl. Phys. **19**, L491–L494 (1980)
1464. N. Braslau, J.B. Gunn, J.L. Staples, Metal-semiconductor contacts for GaAs bulk effect devices. Solid State Electron. **10**, 381–383 (1967)
1465. C.L. Chen, L.J. Mahoney, M.C. Finn, R.C. Brooks, A. Chu, J.G. Mavroides, Low resistance Pd/Ge/Au and Ge/Pd/Au ohmic contacts to n-type GaAs. Appl. Phys. Lett. **48**, 535–537 (1986)
1466. V.L. Rideout, A review of the theory and technology for ohmic contacts to group III-V compound semiconductors. Solid State Electron. **18**, 541–550 (1975)
1467. B.C. Sharma, Ohmic contacts to III-V compound semiconductors. Semicond. Semimetals **15**, 1–38 (1981)
1468. G.Y. Robinson, Metallurgical and electrical properties of alloyed Ni/Au-Ge films on n-type GaAs. Solid State Electron. **18**, 331–342 (1975)
1469. C.M. Heller, I.H. Campbell, D.L. Smith, N.N. Barashkov, J.P. Ferraris, Chemical potential pinning due to equilibrium electron transfer at metal/C_{60}-doped polymer interfaces. J. Appl. Phys. **81**, 3227–3231 (1997)
1470. I.H. Campbell, T.W. Hagler, D.L. Smith, J.P. Ferraris, Direct measurement of conjugated polymer electronic excitation energies using metal/polymer/metal structures. Phys. Rev. Lett. **76**, 1900–1903 (1996)
1471. I.H. Campbell, D.L. Smith, Physics of organic electronic devices. Solid State Phys. **55**, 1–117 (2001)
1472. Frieder Baumann, Lucent Technologies Bell Labs
1473. C.G.B. Garrett, W.H. Brattain, Physical theory of semiconductor surfaces. Phys. Rev. **99**, 376–387 (1955)
1474. A.S. Grove, B.E. Deal, E.H. Snow, C.T. Shah, Investigation of thermally oxidised silicon surfaces using metal-oxide-semiconductor structures. Solid State Electron. **8**, 145–163 (1965)
1475. B.E. Deal, E.H. Snow, C.A. Mead, Barrier energies in metal-silicon dioxide-silicon structures. J. Phys. Chem. Solids **27**, 1873–1879 (1966)
1476. W.M. Werner, The work function difference of the MOS-system with aluminum field plates and polycrystalline silicon field plates. Solid State Electron. **17**, 769–775 (1974)
1477. J.F. Guilleomoles, I. Lubomirsky, I. Riess, D. Cahen, Thermodynamic stability of p/n junctions. J. Phys. Chem. **99**, 14486–14493 (1995)
1478. W.F. O'Hearn, Y.F. Chang, An analysis on the frequency dependence of the capacitance of abrupt P-N junction semiconductor devices. Solid State Electron. **13**, 473–483 (1970)
1479. Y. Yoshizumi, S. Hashimoto, T. Tanabe, M. Kiyama, High-breakdown-voltage pn-junction diodes on GaN substrates. J. Cryst. Growth **298**, 875–878 (2007)
1480. C.-T. Sah, R.N. Noyce, W. Shockley, Carrier generation and recombination in P-N junctions and P-N junction characteristics. Proc. IRE **45**, 1228–1243 (1957)
1481. M. Grundmann, The bias dependence of the non-radiative recombination current in p-n diodes. Solid State Electron. **49**, 1446–1448 (2005)
1482. H.K. Gummel, Hole-electron product of pn junctions. Solid State Electron. **10**, 209–212 (1967)

1483. J.L. Moll, *Physics of Semiconductors* (McGraw-Hill, New York, 1964)
1484. M.J.O. Strutt, Properties of semiconductor diodes, in Semiconductor Devices, vol. 1 (Academic Press, New York, 1966), English translation of W. Guggenbühl, M.J.O. Strutt, W. Wunderlin, Halbleiterbauelemente, I. Halbleiter und Halbleiterdioden (Springer, Basel, 1962)
1485. S. Mahadevan, S.M. Hardas, G. Suryan, Electrical breakdown in semiconductors. Phys. Status Solidi A **8**, 335–374 (1971)
1486. D. Lausch, K. Petter, H. von Wenckstern, M. Grundmann, Correlation of pre-breakdown sites and bulk defects in multicrystalline silicon solar cells. Phys. Status Solidi RRL **3**, 70–72 (2009)
1487. S.M. Sze, G. Gibbons, Avalanche breakdown voltages of abrupt and linearly graded p-n junctions in Ge, Si. GaAs and GaP. Appl. Phys. Lett. **8**, 111–113 (1966)
1488. W. Fulop, Calculation of avalanche breakdown voltages of silicon p-n junctions. Solid State Electron. **10**, 39–43 (1967)
1489. B.J. Baliga, S.K. Ghandhi, Analytical solutions for the breakdown voltage of abrupt cylindrical and spherical junctions. Solid State Electron. **19**, 739–744 (1976)
1490. M.H. Lee, S.M. Sze, Orientation dependence of breakdown voltage in GaAs. Solid State Electron. **23**, 1007–1009 (1980)
1491. A. Goetzberger, B. McDonald, R.H. Haitz, R.M. Scarlet, Avalanche effects in silicon p-n junctions. II. Structurally perfect junctions. J. Appl. Phys. **34**, 1591–1600 (1963)
1492. B.J. Baliga, *Modern Power Devices* (Wiley, New York, 1987)
1493. B.J. Baliga, *Fundamentals of Power Semiconductor Devices* (Springer, Berlin, 2008)
1494. S.M. Sze, G. Gibbons, Effect of junction curvature on breakdown voltage in semiconductors. Solid State Electron. **9**, 831–845 (1966)
1495. A.G. Milnes, D.L. Feucht, *Heterojunctions and Metal Semiconductor Junctions* (Academic Press, New York, 1972)
1496. H.J.A. Bluyssen, L.J. van Ruyven, F. Williams, Effects of quantum confinement and compositional grading on the band structure of heterojunctions. Solid State Electron. **22**, 573–579 (1979)
1497. F.-L. Schein, H. von Wenckstern, M. Grundmann, Transparent p-CuI/n-ZnO heterojunction diodes. Appl. Phys. Lett. **102**, 092109(4pp) (2013)
1498. M. Grundmann, R. Karsthof, H. von Wenkstern, M. Grundmann, Interface recombination current in type II heterostructure bipolar diodes. ACS Appl. Mater. Interface **6**, 14785–14789 (2014)
1499. J. Yang, J. Shen, Effects of discrete trap levels on organic light emitting diodes. J. Appl. Phys. **85**, 2699–2705 (1999)
1500. J. Staudigel, M. Stößel, F. Steuber, J. Simmerer, A quantitative numerical model of multilayer vapor-deposited organic light emitting diodes. J. Appl. Phys. **86**, 3895–3910 (1999)
1501. J. Blochwitz, Organic light-emitting diodes with doped charge transport layers. Ph.D. thesis, Technische Universität Dresden (2001)
1502. K. Harada, A.G. Werner, M. Pfeiffer, C.J. Bloom, C.M. Elliott, K. Leo, Organic homojunction diodes with a high built-in potential: interpretation of the current-voltage characteristics by a generalized Einstein relation. Phys. Rev. Lett. **94**, 036601(4pp) (2005)
1503. S.A. Van Slyke, C.H. Chen, C.W. Tang, Organic electroluminescent devices with improved stability. Appl. Phys. Lett. **69**, 2160–2162 (1996)
1504. T. Oyamada, H. Yoshizaki, H. Sasabe, C. Adachi, Efficient electron injection characteristics of triazine derivatives for transparent OLEDs (TOLEDs). Chem. Lett. **33**, 1034–1035 (2004)
1505. B.L. Sharma, *Metal-Semiconductor Schottky Barrier Junctions and Their Applications* (Plenum Press, New York, 1984)
1506. Virginia Diodes Inc. http://www.virginiadiodes.com
1507. A. Dearn, L. Devlin, Plextek Ltd., UK
1508. Tuning varactors, Application Note, MicroMetrics Inc., Londonderry, NH, USA
1509. BY329 Product Specifications, Philips Semiconductors (1998)
1510. Step recovery diodes, Application Note, MicroMetrics Inc., Londonderry, NH, USA

1511. M.R.T. Tan, S.Y. Yang, D.E. Mars, J.L. Moll, *A 12 psec GaAs Double Heterostructure Step Recovery Diode, HPL-91-187 (11 pp)* (Hewlett-Packard Co., Palo Alto, 1991)
1512. E.O. Kane, Zener tunneling in semiconductors. J. Phys. Chem. Solids **12**, 181–188 (1959)
1513. L. Esaki, Y. Miyahara, A new device using the tunneling process in narrow p-n junctions. Solid State Electron. **1**, 13–21 (1960)
1514. N. Holonyak Jr, I.A. Lesk, R.N. Hall, J.J. Tiemann, H. Ehrenreich, Direct observation of phonons during tunneling in narrow junction diodes. Phys. Rev. Lett. **3**, 167–168 (1959)
1515. A.G. Chynoweth, R.A. Logan, D.E. Thomas, Phonon-assisted tunneling in silicon and germanium Esaki junctions. Phys. Rev. **125**, 877–881 (1962)
1516. M. Bao, K.L. Wang, Accurately measuring current-voltage characteristics of tunnel diodes. IEEE Trans. Electron Devices **ED-53**, 2564–2568 (2006)
1517. Datasheet Tunneling Diode 1N4396. American Semiconductors, http://www.americanmicrosemi.com
1518. Datasheet Tunneling Diode 1N4397. American Semiconductors, http://www.americanmicrosemi.com
1519. P.J. Bulmann, G.S. Hobson, B.S. Taylor, *Transferred Electron Devices* (Academic, New York, 1972)
1520. A. Khalid, N.J. Pilgrim, G.M. Dunn, M.C. Holland, C.R. Stanley, I.G. Thayne, D.R.S. Cumming, A planar Gunn diode operating above 100 GHz. IEEE Electron Device Lett. **28**, 849–851 (2007)
1521. V.N. Sokolov, K.W. Kim, V.A. Kochelap, D.L. Woolard, Terahertz generation in submicron GaN diodes within the limited space-charge accumulation regime. J. Appl. Phys. **98**, 064507(7pp) (2005)
1522. A. Rydberg, Second and third-harmonic millimeter-wave InP-TED oscillators at frequencies above 170 GHz. IEEE Electron Device Lett. **11**, 439–441 (1990)
1523. H. Eisele, Dual Gunn device oscillator with 10 mW at 280 GHz. Electron. Lett. **43**, 636–637 (2007)
1524. S. Yngvesson, Transferred electron devices (TED)–Gunn devices, in Ref. [1524], pp. 23–58
1525. R. Van Zyl, W. Perold, R. Botha, The Gunn-diode: fundamentals and fabrication, in *Proceedings of the 1998 South African Symposium on Communications and Signal Processing* (1998), pp. 407–412
1526. S. Yngvesson, Microwave Semiconductor Devices. Springer International Series in Engineering and Computer Science, vol. 134 (Springer, New York, 1991)
1527. A.L. Linsebigler, G. Lu, J.T. Yates, Photocatalysis on TiO_2 surfaces: principles, mechanisms, and selected results. Chem. Rev. **95**, 735–758 (1995)
1528. D. Duonghong, E. Borgarello, M. Grätzel, Dynamics of light-induced water cleavage in colloidal systems. J. Am. Chem. Soc. **103**, 4685–4690 (1981)
1529. K.Y. Jung, YCh. Kang, S.B. Park, Photodegradation of trichloroethylene using nanometre-sized ZnO particles prepared by spray pyrolysis. J. Mater. Sci. Lett. **16**, 1848–1849 (1997)
1530. J.L. Yang, S.J. An, W.I. Park, G-Ch. Yi, W. Choi, Photocatalysis using ZnO thin films and nanoneedles grown by metal-organic chemical vapor deposition. Adv. Mater. **16**, 1661–1664 (2004)
1531. J.B. Johnson, Thermal agitation of electricity in conductors. Phys. Rev. **29**, 367 (1927)
1532. J.B. Johnson, Thermal agitation of electricity in conductors. Phys. Rev. **32**, 97 (1928)
1533. H. Nyquist, Thermal agitation of electric charge in conductors. Phys. Rev. **32**, 110 (1928)
1534. A. van der Ziel, *Fluctuation Phenomena in Semiconductors* (Butterworth Scientific Publications, London, 1959)
1535. R.W. Boyd, *Radiometry and Detection of Optical Radiation* (Wiley, New York, 1983)
1536. S. Nudelman, The detectivity of infrared photodetectors. Appl. Opt. **1**, 627–636 (1962)
1537. R.C. Jones, Method of rating the performance of photoconductive cells. Proc. IRIS (Infrared Radiation Information Symposium) **2**, 9–12 (1957)
1538. R.C. Jones, Proposal of the detectivity D^{**} for detectors limited by radiation noise. J. Opt. Soc. Am. **50**, 1058 (1960)
1539. Canon Inc., http://www.canon.com/technology

1540. B.F. Levine, Quantum-well infrared photodetectors. J. Appl. Phys. **74**, R1–R81 (1993)
1541. H.C. Liu, R. Dudek, A. Shen, E. Dupont, C.Y. Song, Z.R. Wasilewski, M. Buchanan, High absorption (>90%) quantum-well infrared photodetectors. Appl. Phys. Lett. **79**, 4237–4239 (2001)
1542. A. Rogalski, Progress in focal plane array technologies. Prog. Quantum Electron. **36**, 342–473 (2012)
1543. Fraunhofer-Institut für Angewandte Festkörperphysik, Freiburg, http://www.iaf.fraunhofer.de
1544. M.D. Petroff, M.G. Stapelbroek, Blocked impurity band detectors. US Patent 4,568,960, filed 1980, awarded 1986
1545. N.M. Haegel, BIB detector development for the far infrared: from Ge to GaAs. Proc. SPIE **4999**, 182–194 (2003)
1546. L.A. Reichertz, J.W. Beeman, B.L. Cardozo, N.M. Haegel, E.E. Haller, G. Jakob, R. Katterloher, GaAs BIB photodetector development for far-infrared astronomy. Proc. SPIE **5543**, 231–238 (2004)
1547. F. Szmulowicz, F.L. Madarsz, Blocked impurity band detectors-an analytical model: figures of merit. J. Appl. Phys. **62**, 2533–2540 (1987)
1548. H. Melchior, Demodulation and Photodetection Techniques. Laser Handbook, vol. 1, ed. by F.T. Arecchi, E.O. Schulz-Dubois (North-Holland, Amsterdam, 1972), pp. 725–835
1549. G.E. Stillman, C.M. Wolfe, Avalanche photodiodes. Semicond. Semimetals **12**, 291–393 (1977)
1550. Datasheet Position Sensitive Photodiodes, DL-100-7-KER pin (2002), Silicon Sensor GmbH, Berlin (Germany), http://www.silicon-sensor.com
1551. S.M. Sze, D.J. Coleman, A. Loya, Current transport in metal-semiconductor-metal (MSM) structures. Solid State Electron. **14**, 1209–1218 (1971)
1552. D. Kuhl, Herstellung und Charakterisierung von MSM Detektoren. Ph.D. thesis, Technische Universität Berlin (1992)
1553. S.Y. Chou, M.Y. Liu, Nanoscale tera-hertz metal-semiconductor-metal photodetectors. IEEE J. Quantum Electron. QE- **28**, 2358–2368 (1992)
1554. R.J. McIntyre, Multiplication noise in uniform avalanche diodes. IEEE Trans. Electron Devices **ED-13**, 164–168 (1966)
1555. R.D. Baertsch, Noise and ionization rate measurements in silicon photodiodes. IEEE Trans. Electron Devices **ED-13**, 987 (1966)
1556. Datasheet APD Silicon Sensor AD500-12 TO, Silicon Sensor International AG (2010)
1557. G.M. Williams, M. Compton, D.A. Ramirez, M.M. Hayat, A.S. Huntington, Multi-gain-stage InGaAs avalanche photodiode with enhanced gain and reduced excess noise. IEEE J. Electron Device Soc. **1**, 54–65 (2013)
1558. K.S. Giboney, M.J.W. Rodwell, J.E. Bowers, Traveling-wave photodetectors. IEEE Photonics Technol. Lett. **4**, 1363–1365 (1992)
1559. J.-W. Shi, K.-G. Gan, Y.-J. Chiu, Y.-H. Chen, C.-K. Sun, Y.-J. Yang, J.E. Bowers, Metal-semiconductor-metal traveling-wave photodetectors. IEEE Photonics Technol. Lett. **16**, 623–635 (2001)
1560. W.S. Boyle, G.E. Smith, Charge coupled semiconductor devices. Bell Syst. Tech. J. **49**, 587–593 (1970)
1561. G.F. Amelio, M.F. Tompsett, G.E. Smith, Experimental verification of the charge coupled device concept. Bell Syst. Tech. J. **49**, 593–600 (1970)
1562. J.D.E. Beynon, D.R. Lamb (eds.), *Charge-coupled Devices and Their Applications* (McGraw-Hill, Maidenhead, 1977)
1563. A.J.P. Theuwissen, *Solid-State Imaging with Charge-Coupled Devices* (Kluwer, New York, 1995)
1564. A. Goetzberger, E.H. Nicollian, Temperature dependence of inversion-layer frequency response in silicon. Bell Syst. Tech. J. **46**, 513–522 (1967)
1565. J.D.E. Beynon, The basic principles of charge-coupled devices. Microelectronics **7**, 7–13 (1975)

1566. Charge-coupled device (CCD) image sensors, MTD/PS-0218, Rev. 2 (2008), Eastman Kodak Company, Rochester, http://www.kodak.com
1567. J.E. Carnes, W.F. Kosonocky, E.G. Ramberg, Free charge transfer in charge-coupled devices. IEEE Trans. Electron Devices **ED-19**, 798–808 (1972)
1568. C.H. Sequin, M.F. Tompsett, *Charge Transfer Devices* (Academic, New York, 1975)
1569. D.J. Burt, Basic Operation of the Charge Coupled Device (University of Edinburgh, Int. Conf. Technol. Appl. CCD, 1974), pp. 1–12
1570. SONY Corporation, http://www.sony.net
1571. B.E. Bayer, Color imaging array. US Patent 3,971,065, filed 1975, awarded 1976
1572. Datasheet CMOS linear image sensor, S9226 (2003), http://www.hamamatsu.com
1573. Foveon Inc., http://www.foveon.com
1574. R.B. Merrill, Color separation in an active pixel cell imaging array using a triple-well structure. US Patent 5,965,875, filed 1998, awarded 1999
1575. Datasheet Avalanche Photodiode Array, AD-LA-16-9-DIL 18 (2002), Silicon Sensor GmbH, Berlin, Germany, http://www.silicon-sensor.com
1576. Datasheet InGaAs Linear Photodiode Array, SU1024LE-1.7 (2003), Sensors Unlimited Inc. http://www.sensorsinc.com
1577. Datasheet Quadrant Photodiode with Position Sensing, QD50-0-SD (2004), OSI Optoelectronics, Hawthorne, CA, USA, http://www.osioptoelectronics.com
1578. J.J. Loferski, The first forty years: a brief history of the modern photovoltaic age. Prog. Photovolt: Res. Appl. **1**, 67–78 (1993)
1579. M.A. Green, K. Emery, Y. Hishikawa, W. Warta, E.D. Dunlop, The up to date table at printing (version 45) can be found. Prog. Photovolt: Res. Appl. **23**, 1–9 (2015); Older tables are e.g. (version 33) M.A. Green, K. Emery, Y. Hishikawa, W. Warta, Prog. Photovolt: Res. Appl. **17**, 85–94 (2009) (version 42) M.A. Green, K. Emery, Y. Hishikawa, W. Warta, E.D. Dunlop, Prog. Photovolt: Res. Appl. **21**, 827–837 (2013)
1580. R. Leemans, W.P. Cramer, *The IIASA Database for Mean Monthly Values of Temperature, Precipitation and Cloudiness on a Global Terrestrial Grid, Research Report RR-91-18* (International Institute of Applied Systems Analyses, Laxenburg, 1991)
1581. W.W. Gärtner, Depletion-layer photoeffects in semiconductors. Phys. Rev. **116**, 84–87 (1959)
1582. X. Liu, J. Sites, Solar-cell collection efficiency and its variation with voltage. J. Appl. Phys. **75**, 577–581 (1994)
1583. H.H. Hovel, Carrier collection, spectral response, and photocurrent. Semicond. Semimetals **11**, 8–47 (1975)
1584. S. Hegedus, D. Desai, C. Thompson, Voltage dependent photocurrent collection in CdTe/CdS solar cells. Prog. Photovolt: Res. Appl. **15**, 587–602 (2007)
1585. C.H. Henry, Limiting efficiencies of ideal single and multiple energy gap terrestrial solar cells. J. Appl. Phys. **51**, 4494–4500 (1980)
1586. W. Shockley, H.-J. Queisser, Detailed balance limit of efficiency of p-n junction solar cells. J. Appl. Phys. **32**, 510–519 (1961)
1587. H.-J. Queisser, Detailed balance limit for solar cell efficiency. Mater. Sci. Eng. B **159**, 322–328 (2008)
1588. U. Th Kirchartz, Rau, Detailed balance and reciprocity in solar cells. Phys. Status Solidi A **205**, 2737–2751 (2008)
1589. J.H. Werner, S. Kolodinski, H.-J. Queisser, Novel optimization principles and efficiency limits for semiconductor solar cells. Phys. Rev. Lett. **72**, 3851–3854 (1994)
1590. T. Markvart, Solar cell as a heat engine: energy-entropy analysis of photovoltaic conversion. Phys. Status Solidi A **205**, 2752–2756 (2008)
1591. M.B. Prince, Silicon solar energy converters. J. Appl. Phys. **26**, 534–540 (1955)
1592. R.T. Ross, Some thermodynamics of photochemical systems. J. Chem. Phys. **46**, 4590–4593 (1967)
1593. O.D. Miller, E. Yablonovitch, S.R. Kurtz, Strong internal and external luminescence as solar cells approach the Shockley-Queisser limit. IEEE J. Photovolt. **2**, 303–311 (2012)

1594. J. Mandelkorn, J.H. Lamneck, *Simplified fabrication of back surface electric field silicon cells and novel characteristics of such cells, in Conference Record of the 9th IEEE Photovoltaic Specialists Conference* (IEEE, New York, 1972), pp. 83–90
1595. A. Hauser, I. Melnyk, P. Fath, S. Narayanan, S. Roberts, T.M. Bruton, A simplified process for isotropic texturing of mc-Si, in *Proceedings of the 3rd World Conference on Photovoltaic Energy Conversion*, vol. 2 (2003), pp. 1447–1450
1596. K. Wambach, S. Schlenker, I. Röver, Deutsche Solar AG, Freiberg
1597. R.A. Arndt, J.F. Allison, J.G. Haynos, A. Meulenberg Jr, *Optical properties of the COMSAT Non-Reflective cell, in 11th IEEE Photovoltaic Specialists Conference Record* (IEEE, New York, 1975), pp. 40–43
1598. NREL, http://www.nrel.gov/gis/solar.html
1599. Australian CRC for Renewable Energy Ltd. (ACRE), http://acre.murdoch.edu.au
1600. A. Woyte, J. Nijs, R. Belmans, Partial shadowing of photovoltaic arrays with different system configurations: literature review and field test results. Sol. Energy **74**, 217–233 (2003)
1601. J.W. Bishop, Computer simulation of the effects of electrical mismatches in photovoltaic cell interconnection circuits. Sol. Cells **25**, 73–89 (1988)
1602. J.H. Werner, K. Taretto, U. Rau, Grain boundary recombination in thin-film silicon solar cells. Solid State Phenom. **80–81**, 299–304 (2001)
1603. M. Imaizumi, T. Ito, M. Yamaguchi, K. Kaneko, Effect of grain size and dislocation density on the performance of thin film polycrystalline silicon solar cells. J. Appl. Phys. **81**, 7635–7640 (1997)
1604. A.K. Ghosh, C. Fishman, T. Feng, Theory of the electrical and photovoltaic properties of polycrystalline silicon. J. Appl. Phys. **51**, 446–454 (1980)
1605. K.R. Taretto, Modeling and characterization of polycrystalline silicon for solar cells and microelectronics. Ph.D. thesis, Universität Stuttgart (2003)
1606. R.R. King, D.C. Law, K.M. Edmondson, C.M. Fetzer, G.S. Kinsey, H. Yoon, R.A. Sherif, N.H. Karam, 40% efficient metamorphic GaInP/GaInAs/Ge multijunction solar cells. Appl. Phys. Lett. **90**, 183516 (2007)
1607. V. Probst, J. Palm, S. Visbeck, T. Niesen, R. Tölle, A. Lerchenberger, M. Wendl, H. Vogt, H. Calwer, W. Stetter, F. Karg, New developments in Cu(In, Ga)(S, Se)$_2$ thin film modules formed by rapid thermal processing of stacked elemental layers. Sol. Energy Mater. Sol. Cells **90**, 3115–3123 (2006)
1608. H. Hoppe, N.S. Sariciftci, Organic solar cells: an overview. J. Mater. Res. **19**, 1924–1944 (2004)
1609. S.E. Shaheen, C.J. Brabec, N.S. Sariciftci, F. Padinger, T. Fromherz, J.C. Hummelen, 2.5% efficient organic plastic solar cells. Appl. Phys. Lett. **78**, 841–843 (2001)
1610. V. Shrotriya, G. Li, Y. Yao, T. Moriarty, K. Emery, Y. Yang, Accurate measurement and characterization of organic solar cells. Adv. Funct. Mater. **16**, 2016–2023 (2006)
1611. Solarion AG, Leipzig (Germany), http://www.solarion.net
1612. T. Trupke, M.A. Green, P. Würfel, Improving solar cell efficiencies by down-conversion of high-energy photons. J. Appl. Phys. **92**, 1668–1674 (2002)
1613. T. Trupke, M.A. Green, P. Würfel, Improving solar cell efficiencies by up-conversion of sub-band-gap light. J. Appl. Phys. **92**, 4117–4122 (2002)
1614. A. Luque, A. Martí, Increasing the efficiency of ideal solar cells by photon induced transitions at intermediate levels. Phys. Rev. Lett. **78**, 5014–5017 (1997)
1615. K.M. Yu, W. Walukiewicz, J. Wu, W. Shan, J.W. Beeman, M.A. Scarpulla, O.D. Dubon, P. Becla, Diluted II-VI oxide semiconductors with multiple band gaps. Phys. Rev. Lett. **91**, 246403(4pp) (2003)
1616. A. Franceschetti, S. Lanya, G. Bester, Quantum-dot intermediate-band solar cells with inverted band alignment. Phys. E **41**, 15–17 (2008)
1617. A. Luque, A. Martí, A.J. Nozik, Solar cells based on quantum dots: multiple exciton generation and intermediate bands. MRS Bull. **32**, 236–241 (2007)
1618. S. Suraprapapich, S. Thainoi, S. Kanjanachuchai, S. Panyakeow, Quantum dot integration in heterostructure solar cells. Sol. Energy Mater. Sol. Cells **90**, 2968–2974 (2006)

1619. F. Dimroth, High-efficiency solar cells from III-V compound semiconductors. Phys. Status Solidi C **3**, 373–379 (2006)
1620. R.R. King, D.E. Joslin, H. Karam, Multijunction photovoltaic cell with thin 1st (top) subcell and thick 2nd subcell of same or similar semiconductor material. US patent 6,316,715, filed 2000, awarded 2001
1621. Gergö Létay: Modellierung von III-V solarzellen. Ph.D. thesis, Universität Konstanz (2003)
1622. Spectrolab, http://www.spectrolab.com
1623. F. Dimroth, S. Kurtz, High-efficiency multijunction solar cells. MRS Bull. **32**, 230–235 (2007)
1624. M.A. Green, A. Ho-Baillie, H.J. Snaith, The emergence of perovskite solar cells. Nat. Photonics **8**, 506–514 (2014)
1625. H.J. Snaith, Perovskites: the emergence of a new era for low-cost. High-efficiency solar cells. Phys. Chem. Lett. **4**, 3623–3630 (2013)
1626. US Department of Energy, http://www.eere.energy.gov/solar
1627. J.H. Werner, Second and third generation photovoltaics-dreams and reality. Adv. Solid State Phys. (Festkörperprobleme) **44**, 51–66 (2004)
1628. S.E. Derenzo, M.J. Weber, E. Bourret-Courchesne, M.K. Klintenberg, The quest for the ideal inorganic scintillator. Nucl. Instrum. Meth. A **505**, 111–117 (2003)
1629. Y.A. Ono, *Electroluminescent Displays* (World Scientific, Singapore, 1995)
1630. G.F. Knoll, *Radiation Detection and Measurement*, 4th edn. (Wiley, New York, 2010)
1631. G. Hoffmann, FH Emden
1632. M. Nikl, V.V. Laguta, A. Vedda, Complex oxide scintillators: material defects and scintillation performance. Phys. Status Solidi B **245**, 1701–1722 (2008)
1633. S.E. Derenzo, W.W. Moses, M.J. Weber, A.C. West, Methods for a systematic, comprehensive search for fast, heavy scintillator materials. Proc. Mater. Res. Soc. **348**, 39 (1994)
1634. E.F. Schubert, Light-emitting Diodes (Cambridge University Press, 2003)
1635. A.A. Bergh, P.J. Dean, Light-emitting diodes. Proc. IEEE **60**, 156–223 (1972)
1636. M.R. Krames, O.B. Shchekin, R. Mueller-Mach, G.O. Mueller, L. Zhou, G. Harbers, M.G. Craford, Status and future of high-power light-emitting diodes for solid-state lighting. IEEE J. Disp. Technol. **3**, 160–175 (2007)
1637. R.D. Dupuis, M.R. Krames, History, development, and applications of high-brightness visible light-emitting diodes. IEEE J. Lightwave Technol. **26**, 1154–1171 (2008)
1638. H.P. Maruska, D.A. Stevenson, J.I. Pankove, Violet luminescence of Mg-doped GaN. Appl. Phys. Lett. **22**, 303–305 (1973)
1639. N. Koide, H. Kato, S. Yamasaki, K. Manabe, M. Hashimoto, H. Amano, K. Hiramatsu, I. Aksaki, Doping of GaN with Si and properties of blue m/i/n/n^+ GaN LED with Si-doped n^+-layer by MOVPE. J. Cryst. Growth **115**, 639–642 (1991)
1640. P. Santhanam, D.J. Gray, Jr, R.J. Ram, Thermoelectrically pumped light-emitting diodes operating above unity efficiency. Phys. Rev. Lett. **108**, 097403(5pp) (2012)
1641. J. Tauc, The share of thermal energy taken from the surroundings in the electro-luminescent energy radiated from a p-n junction. Czech. J. Phys. **7**, 275–276 (1957)
1642. A. David, C.A. Hurni, R.I. Aldaz, M.J. Cich, B. Ellis, K. Huang, F.M. Steranka, M.R. Krames, High light extraction efficiency in bulk-GaN based volumetric violet light-emitting diodes. Appl. Phys. Lett. **105**, 23111(5pp) (2014)
1643. Soora, http://www.soora.com
1644. W.N. Carr, Photometric figures of merit for semiconductor luminescent sources operating in spontaneous mode. Infrared Phys. **6**, 1–19 (1966)
1645. S.V. Galginaitis, Improving the external efficiency of electroluminescent diodes. J. Appl. Phys. **36**, 460–461 (1965)
1646. M.G. Craford, Recent developments in light-emitting-diode technology. IEEE Trans. Electron Devices **ED-24**, 935–943 (1977)
1647. OSRAM Opto Semiconductors GmbH, Regensburg, Germany (2001), http://www.osram-os.com
1648. LumiLeds Lighting, http://www.lumileds.com

1649. V. Haerle, B. Hahn, S. Kaiser, A. Weimar, S. Bader, F. Eberhard, A. Plössl, D. Eisert, High brightness LEDs for general lighting applications Using the new ThinGaN™-Technol. Phys. Status Solidi A **201**, 2736–2739 (2004)
1650. OSRAM Opto Semiconductors GmbH, Regensburg, Germany (2003), http://www.osram-os.com
1651. O.B. Shchekin, J.E. Epler, T.A. Trottier, T. Margalith, D.A. Steigerwald, M.O. Holcomb, P.S. Martin, M.R. Krames, High performance thin-film flip-chip InGaN/GaN light-emitting diodes. Appl. Phys. Lett. **89**, 071109 (2006)
1652. C.A. Hurni, A. David, M.J. Cich, R.I. Aldaz, B. Ellis, K. Huang, A. Tyagi, R.A. DeLille, M.D. Craven, F.M. Steranka, M.R. Krames, Bulk GaN flip-chip violet light-emitting diodes with optimized efficiency for high-power operation. Appl. Phys. Lett. **106**, 031101 (2015)
1653. F. Akyol, S. Krishnamoorthy, Y. Zhang, S. Rajan, Tunneling-based carrier regeneration in cascaded GaN light emitting diodes to overcome efficiency droop. Appl. Phys. Lett. **103**, 081107 (2013)
1654. F. Akyol, S. Krishnamoorthy, Y. Zhang, S. Rajan, GaN-based three-junction cascaded light-emitting diode with low-resistance InGaN tunnel junctions. Appl. Phys. Express **8**, 082103 (2015)
1655. M.G. Craford, Visible light-emitting diodes: past, present, and very bright future. MRS Bull. **25**, 27–31 (2000)
1656. R. Mueller-Mach, G. Mueller, M.R. Krames, H.A. Höppe, F. Stadler, W. Schnick, P. Th Juestel, Schmidt, Highly efficient all-nitride phosphor-converted white light emitting diode. Phys. Status Solidi A **202**, 1727–1732 (2005)
1657. Balancing White Color, Application note STS-KSE3692 (2pp), Nichia Corp. (2004)
1658. J. Christen, T. Riemann, F. Bertram, D. Rudloff, P. Fischer, A. Kaschner, U. Haboeck, A. Hoffmann, C. Thomsen, Optical micro-characterization of group-III-nitrides: correlation of structural, electronic and optical properties. Phys. Status Solidi C, 1795–1815 (2003)
1659. Power Light Source Luxeon© K2, Technical Datasheet DS51 (6/08), Philips Lumileds Lighting Company (2008)
1660. Y.C. Shen, G.O. Mueller, S. Watanabe, N.F. Gardner, A. Munkholm, M.R. Krames, Auger recombination in InGaN measured by photoluminescence. Appl. Phys. Lett. **91**, 141101(3pp) (2007)
1661. C. Weisbuch, M. Piccardo, L. Martinelli, J. Iveland, J. Peretti, J.S. Speck, The efficiency challenge of nitride light-emitting diodes for lighting. Phys. Status Solidi A **212**, 899–913 (2015)
1662. S. Weise, ThTh Zahner, A. Stich, *Lutz, Reliability of the DRAGON® product family* (Application Note, Osram Opto, Semiconductors, February 2009)
1663. A. Lochmann, E. Stock, O. Schulz, F. Hopfer, D. Bimberg, V.A. Haisler, A.I. Toropov, A.K. Bakarov, A.K. Kalagin, Electrically driven single quantum dot polarised single photon emitter. Electron. Lett. **42**, 774–775 (2006)
1664. R.M. Stevenson, R.J. Young, P. Atkinson, K. Cooper, D.A. Ritchie, A.J. Shields, A semiconductor source of triggered entangled photon pairs. Nature **439**, 179–182 (2006)
1665. R.J. Young, R.M. Stevenson, P. Atkinson, D.A. Ritchie, A.J. Shields, Improved fidelity of triggered entangled photons from single quantum dots. New J. Phys. **8**, 29 (2006)
1666. A.J. Shields, Semiconductor quantum light sources. Nat. Photonics **1**, 215 (2007)
1667. A. Lochmann, E. Stock, O. Schulz, F. Hopfer, D. Bimberg, V.A. Haisler, A.I. Toropov, A.K. Bakarov, M. Scholz, S. Büttner, O. Benson, Electrically driven quantum dot single photon source. Phys. Status Solidi C **4**, 547–550 (2007)
1668. S.K. Ray, K.M. Groom, M.D. Beattie, H.Y. Liu, M. Hopkinson, R.A. Hogg, Broad-band superluminescent light-emitting diodes incorporating quantum dots in compositionally modulated quantum wells. IEEE Photonics Technol. Lett. **18**, 58–60 (2006)
1669. E.U. Rafailov, M.A. Cataluna, W. Sibbett, Mode-locked quantum-dot lasers. Nat. Photonics **1**, 395–401 (2007)
1670. C.W. Tang, S.A. Van Slyke, Organic electroluminescent diodes. Appl. Phys. Lett. **51**, 913–915 (1987)

1671. OSRAM Opto Semiconductors GmbH, Regensburg, Germany (2007), http://www.osram-os.com
1672. Pioneer Electronics, http://www.pioneerelectronics.com
1673. Sony Model XEL-1 (2009), http://www.sony.com
1674. H.C. Casey, Jr, M.B Panish, *Heterostructure Lasers* (Academic Press, New York, 1978) (Two volumes, Part A: Fundamental Principles and Part B: Materials and Operating Characteristics)
1675. H. Kressel, J. Butler, *Semiconductor Lasers and LEDs* (Academic Press, New York, 1977)
1676. J. Faist, F. Capasso, D.L. Sivco, C. Sirtori, A.L. Hutchinson, A.Y. Cho, Quantum cascade laser. Science **264**, 553–556 (1994)
1677. Lasers and Photonics Marketplace Seminar, as published by Laser Focus World, http://www.laserfocusworld.com
1678. H. Haug, S.W. Koch, *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, 5th edn. (World Scientific, Singapore, 2009)
1679. W.W. Chow, P.M. Smowton, P. Blood, A. Girndt, F. Jahnke, S.W. Koch, Comparison of experimental and theoretical GaInP quantum well gain spectra. Appl. Phys. Lett. **71**, 157–159 (1997)
1680. N. Kirstaedter, N.N. Ledentsov, M. Grundmann, D. Bimberg, V.M. Ustinov, S.S. Ruvimov, M.V. Maximov, P.S. Kop'ev, Zh.I. Alferov, U. Richter, P. Werner, U. Gösele J. Heydenreich, Low threshold, large T_0 injection laser emission from (InGa)As quantum dots. Electron. Lett. **30**, 1416 (1994)
1681. M. Grundmann, The present status of quantum dot lasers. Phys. E **5**, 167–184 (2000)
1682. M.B. Panish, I. Hayashi, S. Sumski, Double-heterostructure injection lasers with room-temperature thresholds as low as 2300 A/cm^2. Appl. Phys. Lett. **16**, 326–327 (1970)
1683. L.A. D'Asaro, Advances in GaAs junction lasers with stripe geometry. J. Lumin. **7**, 310–337 (1973)
1684. H. Wenzel, F. Bugge, G. Erbert, R. Hülsewede, R. Staske, G. Tränkle, High-power diode lasers with small vertical beam divergence emitting at 808 nm. Electron. Lett. **37**, 1024–1026 (2001)
1685. H. Yonezu, I. Sakuma, K. Kobayashi, T. Kamejima, M. Ueno, Y. Nannichi, A GaAs/Al_xGa_{1-x}As double heterostructure planar stripe laser. Jpn. J. Appl. Phys. **12**, 1585–1592 (1973)
1686. W.G. Scheibenzuber, *GaN-Based Laser Diodes: Towards Longer Wavelengths and Short Pulses* (Springer, Berlin, 2012)
1687. M. Kamp, J. Hofmann, A. Forchel, S. Lourdudoss, Ultrashort InGaAsP/InP lasers with deeply etched Bragg mirrors. Appl. Phys. Lett. **78**, 4074–4075 (2001)
1688. M. Kamp, private communication
1689. M. Fujita, R. Ushigome, T. Baba, Large spontaneous emission factor of 0.1 in a microdisk injection laser. IEEE Photonics Technol. Lett. **13**, 403–405 (2001)
1690. G. Björk, A. Karlsson, Y. Yamamoto, Definition of a laser threshold. Phys. Rev. A **50**, 1675–1680 (1994)
1691. Y. Yamamoto, S. Machida, G. Björk, Microcavity semiconductor laser with enhanced spontaneous emission. Phys. Rev. A **44**, 657–668 (1991)
1692. N.A. Pikhtin, S.O. Slipchenko, Z.N. Sokolova, A.L. Stankevich, D.A. Vinokurov, I.S. Tarasov, ZhI Alferov, 16 W continuous-wave output power from 100 μm-aperture laser with quantum well asymmetric heterostructure. Electron. Lett. **40**, 1413–1414 (2004)
1693. A. Knigge, G. Erbert, J. Jonsson, W. Pittroff, R. Staske, B. Sumpf, M. Weyers, G. Tränkle, Passively cooled 940 nm laser bars with 73% wall-plug efficiency at 70 W and 25 °C. Electron. Lett. **41**, 250–251 (2005)
1694. D. Garbuzov, I. Kudryashov, A. Komissarov, M. Maiorov, W. Roff, J. Connolly, *14xx nm DFB InGaAsP/InP pump lasers with 500 mW CW output power for WDM combining, in Optical Fiber Communication Conference, OSA Technical Digest Series, WD1* (Optical Society of America, Washington, 2003), pp. 486–487

1695. T. Kimura, M. Nakae, J. Yoshida, S. Iizuka, A. Sato, 14XX nm over 1 W Pump Laser Module with Integrated PBC. Optical Fiber Communication Conference, OSA Technical Digest Series (Optical Society of America, Washington, 2003), pp. 485–486
1696. W.T. Tsang, R.A. Logan, J.P. van der Ziel, Low-current-threshold strip-buried-heterostructure lasers with self-aligned current injection stripes. Appl. Phys. Lett. **34**, 644–647 (1979)
1697. D. Garbuzov, M. Maiorov, R. Menna, A. Komissarov, V. Khalfin, I. Kudryashov, A. Lunev, L. DiMarco, J. Connolly, High-power 1300-nm Fabry-Perot and DFB ridge-waveguide lasers. Proc. SPIE **4651**, 92–100 (2002)
1698. M. Grundmann, O. Stier, S. Bognár, C. Ribbat, F. Heinrichsdorff, D. Bimberg, Optical properties of self-organized quantum dots: modeling and experiments. Phys. Status Solidi A **178**, 255–262 (2000)
1699. M. Grundmann, How a quantum-dot laser turns on. Appl. Phys. Lett. **77**, 1428–1430 (2000)
1700. F. Heinrichsdorff, C. Ribbat, M. Grundmann, D. Bimberg, High-power quantum-dot lasers at 1100 nm. Appl. Phys. Lett. **76**, 556–558 (2000)
1701. Ch. Ribbat, R. Sellin, M. Grundmann, D. Bimberg, High power quantum dot lasers at 1160 nm. Phys. Status Solidi B **224**, 819–822 (2001)
1702. D. Garbuzov, I. Kudryashov, A. Tsekoun, A. Komissarov, W. Roff, M. Maiorov, R. Menna, A. Lunev, J. Connolly, 14xx nm DFB InGaAsP/InP pump lasers with 500 mw CW output power for WDM combining, in Optical Fiber Communication Conference 2002, Technical Digest: ThN6 (2002)
1703. High-Power 1550 nm DFB Source Lasers, A1112, Datasheet DS00-281OPTO-1, Agere Systems (2001)
1704. M.-C. Amann, J. Buus, *Tunable Laser Diodes* (Artech House, Boston, 1998)
1705. J.N. Walpole, A.R. Calawa, T.C. Harman, S.H. Groves, Double-heterostructure PbSnTe lasers grown by molecular-beam epitaxy with cw operation up to 114 K. Appl. Phys. Lett. **28**, 552–554 (1976)
1706. S.-L. Lee, I.-F. Jang, C.-T. Pien, C.-Y. Wang, T.-T. Shih, Sampled grating DBR laser arrays with adjustable 0.8/1.6-nm wavelength spacing. IEEE Photonics Technol. Lett. **11**, 955–957 (1999)
1707. B. Mason, G.A. Fish, S.P. DenBaars, L.A. Coldren, Ridge waveguide sampled grating DBR lasers with 22-nm quasi-continuous tuning range. IEEE Photonics Technol. Lett. **10**, 1211–1213 (1998)
1708. K. Petermann, *Laser Diode Modulation and Noise* (Kluwer, Dordrecht, 1988)
1709. C.J. Hwang, J.C. Dyment, Dependence of threshold and electron lifetime on acceptor concentration in GaAs-$Ga_{1-x}Al_xAs$ lasers. J. Appl. Phys. **44**, 3240–3244 (1973)
1710. N.K. Dutta, S.J. Wang, A.B. Piccirilli, R.F. Karlicek Jr, R.L. Brown, M. Washington, U.K. Chakrabarti, A. Gnauck, Wide-bandwidth and high-power InGaAsP distributed feedback lasers. J. Appl. Phys. **66**, 4640–4644 (1989)
1711. F. Grillot, B. Dagens, J.-G. Provost, H. Su, L.F. Lester, Gain compression and above-threshold linewidth enhancement factor in 1.3-μm InAs-GaAs quantum-dot lasers. IEEE J. Quantum Electron. **44**, 946–951 (2008)
1712. J.S. Gustavsson, Å. Haglund, J. Bengtsson, A. Larsson, High-speed digital modulation characteristics of oxide-confined vertical-cavity surface-emitting lasers-numerical simulations consistent with experimental results. IEEE J. Quantum Electron. QE- **38**, 1089–1096 (2002)
1713. C.H. Henry, Theory of the linewidth of semiconductor lasers. IEEE J. Quantum Electron. **QE-18**, 259–264 (1982)
1714. Y. Yamamoto, H.A. Haus, Commutation relations and laser linewidth. Phys. Rev. A **41**, 5164–5170 (1990)
1715. D. Welford, A. Mooradian, Output power and temperature dependence of the linewidth of single-frequency cw (GaAl)As diode lasers. Appl. Phys. Lett. **40**, 865–867 (1982); Appl. Phys. Lett. **41**, 1007 (1982) (erratum)
1716. Quintessence Photonics Corporation, http://www.qpclasers.com
1717. BinOptics Corporation, Ithaca, NY, now MACOM, http://www.binoptics.com

1718. H.E. Li, K. Iga (eds.), *Vertical-Cavity Surface-Emitting Laser Devices* (Springer, Berlin, 2003)
1719. Sandia National Laboratories, http://www.sandia.gov
1720. Ulm Photonics, now Philips, http://www.ulm-photonics.de
1721. K.D. Choquette, Vertical-cavity surface-emitting lasers: light for the information age. MRS Bull. **27**, 507–511 (2002)
1722. C.J. Chang-Hasnain, Tunable VCSELs. IEEE J. Sel. Top. Quantum Electron. **6**, 978–987 (2000)
1723. Bandwidth 9 (company closed 2005)
1724. F. Römer, C. Prott, J. Daleiden, S. Irmer, M. Strassner, A. Tarraf, H. Hillmer, Micromechanically tunable air gap resonators for long wavelength VCSEL's, in IEEE LEOS International Semiconductor Laser Conference, Garmisch/Germany (2002), pp. 13–14
1725. M. Schulze, J.-M. Pelaprat, Efficiency experts. Photonics Spectra **35**(5), 150–151 (2001)
1726. Sapphire^{TM} 488–20 laser, Datasheet MC-027-10-0M0111Rev.A, Coherent Inc. (2004)
1727. F. Capasso, private communication
1728. J. Faist, private communication
1729. R.F. Kazarinov, R.A. Suris, Possibility of the amplification of electromagnetic waves in a semiconductor with a superlattice. Fiz. Tekh. Poluprovodn. **5**, 797–800 (1971); ((Sov. Phys. Semicond. **5**, 707–709 (1971))
1730. F. Capasso, K. Mohammed, A.Y. Cho, Resonant tunneling through double barriers, perpendicular quantum transport phenomena in superlattices, and their device applications. IEEE J. Quantum Electron. **QE-22**, 1853–1869 (1986)
1731. H. Krömer, Proposed negative-mass microwave amplifier. Phys. Rev. **109**, 1856 (1958)
1732. A.A. Andronov, I.V. Zverev, V.A. Kozlov, Yu.N. Nozdrin, S.A. Pavlov, V.N. Shastin, Stimulated emission in the long-wavelength region from hot holes in Ge in crossed electric and magnetic fields. Pis'ma Zh. Eksp. Teor. Fiz. **40**, 69–72 (1984); (JETP Lett. **40**, 804–807 (1984))
1733. Opt. Quantum Electron. **23**(2); E. Gornik, A.A. Andronov (eds.), Special Issue Far-infrared Semiconductor Lasers (Chapman and Hall, London, 1991)
1734. E. Bründermann, Widely tunable far infrared hot hole semiconductor lasers, in *Long-wavelength Infrared Semiconductor Lasers*, ed. by H.K. Choi (Wiley, New York, 2004), pp. 279–350
1735. E. Bründermann, A.M. Linhart, H.P. Röser, O.D. Dubon, W.L. Hansen, E.E. Haller, Miniaturization of p-Ge lasers: progress toward continuous wave operation. Appl. Phys. Lett. **68**, 1359–1361 (1996)
1736. E. Bründermann, D.R. Chamberlin, E.E. Haller, High duty cycle and continuous terahertz emission from germanium. Appl. Phys. Lett. **76**, 2991–2993 (2000)
1737. N.K. Dutta, Q. Wang, *Semiconductor Optical Amplifiers*, 2nd edn. (World Scientific, Singapore, 2013)
1738. A.J. Collar, G.D. Henshall, J. Farré, B. Mikkelsen, Z. Wang, L. Eskildsen, D.S. Olesen, K.E. Stubkjaer, Low residual reflectivity of angled-facet semiconductor laser amplifiers. IEEE Photonics Technol. Lett. **2**, 553–555 (1990)
1739. G. Jost, *University of Ulm* (Annual Report, Department of Optoelectronics, 1998), p. 64
1740. Ferdinand-Braun-Institut für Höchstfrequenztechnik, Berlin, http://www.fbh-berlin.de
1741. M. Sugawara, N. Hatori, M. Ishida, H. Ebe, Y. Arakawa, T. Akiyama, K. Otsubo, T. Yamamoto, Y. Nakata, Recent progress in self-assembled quantum-dot optical devices for optical telecommunication: temperature-insensitive 10 Gb s^{-1} directly modulated lasers and 40 Gb s^{-1} signal-regenerative amplifiers. J. Phys. D: Appl. Phys. **38**, 2126–2134 (2005)
1742. N. Yasuoka, K. Kawaguchi, H. Ebe, T. Akiyama, M. Ekawa, K. Morito, M. Sugawara, Y. Arakawa, Quantum-dot semiconductor optical amplifiers with polarization-independent gains in 1.5-μm wavelength bands. IEEE Photonics Technol. Lett. **20**, 1908–1910 (2008)
1743. P. Borri, S. Schneider, W. Langbein, U. Woggon, A.E. Zhukov, V.M. Ustinov, N.N. Ledentsov, D. ZhI Alferov, D. Ouyang, Bimberg, Ultrafast carrier dynamics and dephasing in InAs quantum-dot amplifiers emitting near 1.3-μm-wavelength at room temperature. Appl. Phys. Lett. **79**, 2633–2635 (2001)

1744. W.B. Shockley, *Transistor Technology Evokes New Physics*. Nobel Lecture (1956); Nobel Lectures, Physics 1942–1962 (Elsevier, Amsterdam, 1964), pp. 344–374
1745. R.R. Nelson, *The link between science and invention: the case of the transistor, in The Rate and Direction of Inventive Activity: Economic and Social Factors* (Princeton University Press, Princeton, 1962), pp. 549–583
1746. ShS Li, *Semiconductor Physical Electronics*, 2nd edn. (Springer, New York, 2006)
1747. P.G.A Jespers, *Measurements for Bipolar Devices in Process and Device Modelling for Integrated Circuit Design*, ed. by F. van de Wiele, W.L. Engl, P.G. Jespers (Noordhoff, Leyden, 1977)
1748. M.J. Morant, *Introduction to Semiconductor Devices* (Addison-Wesley, Reading, 1964)
1749. H.K. Gummel, H.C. Poon, An integral charge control model of bipolar transistors. Bell Syst. Tech. J. **49**, 827–852 (1970)
1750. J.M. Early, Effects of space-charge layer widening in junction transistors. Proc. IRE **40**, 1401–1406 (1952)
1751. C.C. McAndrew, L.W. Nagel, Early effect modeling in SPICE. IEEE J. Solid-State Circuits **SC-31**, 136–138 (1996)
1752. D.C. Herbert, Extended velocity overshoot in GaAs HBT. Semicond. Sci. Technol. **10**, 682–686 (1995)
1753. Solid State Electronics Laboratory, University of Michigan, http://www.eecs.umich.edu
1754. D. Cui, D. Sawdai, D. Pavlidis, S.H. Hsu, P. Chin, T. Block, High power performance using InAlAs/InGaAs single heterojunction bipolar transistors, in *Proceedings of the 12th International Conference on Indium Phosphide and Related Materials* (2000), pp. 473–476
1755. M. Feng, N. Holonyak Jr, R. Chan, Quantum-well-base heterojunction bipolar light-emitting transistor. Appl. Phys. Lett. **84**, 1952–1954 (2004)
1756. R.R. Bockemuehl, Analysis of field effect transistors with arbitrary charge distribution. IEEE Trans. Electron Devices **ED-10**, 31–34 (1963)
1757. R.D. Middlebrook, I. Richer, Limits on the power-law exponent for field-effect transistor transfer characteristics. Solid State Electron. **6**, 542–544 (1963)
1758. K. Lehovec, R. Zuleeg, Voltage-current characteristics of GaAs-JFET's in the hot electron range. Solid State Electron. **13**, 1415–1425 (1970)
1759. H.C. Pao, C.T. Sah, Effects of diffusion current on characteristics of metal-oxide (Insulator)-semiconductor transistors. Solid State Electron. **9**, 927–937 (1966)
1760. K. Yamaguchi, Field-dependent mobility model for two-dimensional numerical analysis of MOSFET's. IEEE Trans. Electron Devices **ED-26**, 1068–1074 (1979)
1761. R.R. Troutman, Subthreshold design considerations for insulated gate field-effect transistors. IEEE J. Solid-State Circuits **SC-9**, 55–60 (1974)
1762. W. Maly, *Atlas of IC Technologies-An Introduction to VLSI Processes* (Benjamin/Cummings Publishing Company, San Francisco, 1987)
1763. D. Udeshi, E. Maldonado, Y. Xu, M. Tao, W.P. Kirk, Thermal stability of ohmic contacts between Ti and Se-passivated n-type Si(001). J. Appl. Phys. **95**, 4219–4222 (2004)
1764. V. Sverdlov, *Strain-Induced Effects in Advanced MOSFETs* (Springer, Berlin, 2011)
1765. G.E. Moore, Cramming more components onto integrated circuits. Electronics 38(8), 114–117 (1965); reprinted in. Proc. IEEE **86**, 82–85 (1998)
1766. M. Fayolle, F. Romagna, Copper CMP evaluation: planarization issues. Microelectron. Eng. **37**(38), 135–141 (1997)
1767. Z. Stavreva, D. Zeidler, M. Plötner, K. Drescher, Characteristics in chemical-mechanical polishing of copper: comparison of polishing pads. Appl. Surf. Sci. **108**, 39–44 (1997)
1768. C.-H. Tung, G.T.T. Sheng, C.-Y. Lu, *ULSI Semiconductor Technology Atlas* (Wiley, Hoboken, 2003)
1769. K. Hinode, Y. Hanaoka, K.-I. Takeda, S. Kondo, Resistivity increase in ultrafine-line copper conductor for ULSIs. Jpn. J. Appl. Phys. **40**, L1097–L1099 (2001)
1770. T. Harada, A. Ueki, K. Tomita, K. Hashimoto, J. Shibata, H. Okamura, K. Yoshikawa, T. Iseki, M. Higashi, S. Maejima, K. Nomura, K. Goto, T. Shono, S. Muranaka, N. Torazawa, S. Hirao, M. Matsumoto, T. Sasaki, S. Matsumoto, S. Ogawa, M. Fujisawa, A. Ishii, M.

Matsuura, T. Ueda, Extremely low Keff (1.9) Cu interconnects with air gap formed using SiOC, in IEEE International Interconnect Technology Conference (IITC) (2007), pp. 141–143
1771. P. Bai, C. Auth, S. Balakrishnan, M. Bost, R. Brain, V. Chikarmane, R. Heussner, M. Hussein, J. Hwang, D. Ingerly, R. James, J. Jeong, C. Kenyon, E. Lee, S.-H. Lee, N. Lindert, M. Liu, Z. Ma, T. Marieb, A. Murthy, R. Nagisetty, S. Natarajan, J. Neirynck, A. Ott, C. Parker, J. Sebastian, R. Shaheed, S. Sivakumar, J. Steigerwald, S. Tyagi, C. Weber, B. Woolery, A. Yeoh, K. Zhang, M. Bohr, A 65 nm logic technology featuring 35 nm gate lengths, enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57 μm^2 SRAM cell. IEDM Technical Digest (IEEE, Piscataway, 2004), pp. 657–660
1772. http://www.chipworks.com
1773. Y. Taur, CMOS design near the limit of scaling. IBM J. Res. Dev. **46**, 213–222 (2002)
1774. E.J. Nowak, Maintaining the benefits of CMOS scaling when scaling bogs down. IBM J. Res. Dev. **46**, 169–180 (2002)
1775. S.-H. Lo, D.A. Buchanan, Y. Taur, W. Wang, Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET's. IEEE Electron. Device Lett. **18**, 209–211 (1997)
1776. G.D. Wilk, R.M. Wallace, J.M. Anthony, Hafnium and zirconium silicates for advanced gate dielectrics. J. Appl. Phys. **87**, 484–492 (2000)
1777. Ch. Auth, M. Buehler, A. Cappellani, C.-H. Choi, G. Ding, W. Han, S. Joshi, B. McIntyre, M. Prince, P. Ranade, J. Sandford, Ch. Thomas, 45 nm high-k+metal gate strain-enhanced transistors. Intel Technol. J. **12**, 77 (2008)
1778. M.A. Quevedo-Lopez, S.A. Krishnan, D. Kirsch, C.H.J. Li, J.H. Sim, C. Huffman, J.J. Peterson, B.H. Lee, G. Pant, B.E. Gnade, M.J. Kim, R.M. Wallace, D. Guo, H. Bu, T.P. Ma, *High performance gate first HfSiON dielectric satisfying 45 nm node requirements, in IEDM Technical Digest* (Piscataway, IEEE, 2005), pp. 425–428
1779. K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C.-H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Heussner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. McIntyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, K. Zawadzki, *A 45 nm logic technology with high-k+metal gate transistors, strained silicon, 9 Cu interconnect layers, 193 nm dry patterning, and 100% Pb-free packaging, in IEDM Technical Digest* (Piscataway, IEEE, 2007), pp. 247–250
1780. K. Mistry, M. Armstrong, C. Auth, S. Cea, T. Coan, T. Ghani, T. Hoffmann, A. Murthy, J. Sandford, R. Shaheed, K. Zawadzki, K. Zhang, S. Thompson, M. Bohr, Delaying forever: uniaxial strained silicon transistors in a 90 nm CMOS technology, in Symposium on VLSI Technology Digest (2004), pp. 50–51
1781. T. Krishnamohan, C. Jungemann, D. Kim, E. Ungersboeck, S. Selberherr, A.-T. Pham, B. Meinerzhagen, P. Wong, Y. Nishi, K.C. Saraswat, High performance, uniaxially-strained, silicon and germanium, double-gate p-MOSFETs. Microelectron. Eng. **84**, 2063–2066 (2007)
1782. T. Guillaume, M. Mouis, Calculations of hole mass in [110]-uniaxially strained silicon for the stress-engineering of p-MOS transistors. Solid State Electron. **50**, 701–708 (2006)
1783. E. Ungersboeck, S. Dhar, G. Karlowatz, H. Kosina, S. Selberherr, Physical modeling of electron mobility enhancement for arbitrarily strained silicon. J. Comput. Electron. **6**, 55–58 (2007)
1784. M. Bohr, The Invention of Uniaxial Strained Silicon Transistors at Intel, Intel (2007)
1785. Synopsys Inc. http://www.synopsys.com/products/tcad/examples/strain2d.pdf
1786. W. Hansch, C. Fink, J. Schulze, I. Eisele, A vertical MOS-gated Esaki tunneling transistor in silicon. Thin Solid Films **369**, 387–389 (2000)
1787. P.-F. Wang, K. Hilsenbeck, M. Th Nirschl, Ch. Oswald, M. Stepper, D. Weis, W. Hansch, Schmitt-Landsiedel, Complementary tunneling transistor for low power application. Solid State Electron. **48**, 2281–2286 (2004)

1788. W.M. Reddick, G.A.J. Amaratunga, Silicon surface tunnel transistor. Appl. Phys. Lett. **67**, 494–496 (1995)
1789. P.-F. Wang, T. Nirschl, D. Schmitt-Landsiedel, W. Hansch, Simulation of the Esaki-tunneling FET. Solid State Electron. **47**, 1187–1192 (2003)
1790. E.-H. Toh, G.H. Wang, G. Samudra, Y.-C. Yeo, Device physics and design of germanium tunneling field-effect transistor with source and drain engineering for low power and high performance applications. J. Appl. Phys. **103**, 104504(5pp) (2008)
1791. L.M. Grupp, J.D. Davis, S. Swanson, The bleak future of NAND flash memory, in *Proceeding of the 10th USENIX Conference on File and Storage Technologies (FAST'12)* (USENIX Association, Berkeley, 2012), p. 2–2(8pp)
1792. G.R. Fox, R. Bailey, W.B. Kraus, F. Chu, S. Sun, T. Davenport, The current status of FeRAM, in *Ferroelectric Random Access Memories. Topics in Applied Physics*, vol. 93, ed. by H. Ishiwara, M. Okuyama, Y. Arimoto (Springer, Berlin, 2004), pp. 139–148
1793. Y.C. Chen, C.T. Rettner, S. Raoux, G.W. Burr, S.H. Chen, R.M. Shelby, M. Salinga, W.P. Risk, T.D. Happ, G.M. McClelland, M. Breitwisch, A. Schrott, J.B. Philipp, M.H. Lee, R. Cheek, T. Nirschl, M. Lamorey, C.F. Chen, E. Joseph, S. Zaidi, B. Yee, H.L. Lung, R. Bergmann, C. Lam, *Ultra-Thin Phase-Change Bridge Memory Device Using GeSb, IEDM Technical Digest* (Piscataway, IEEE, 2006), pp. 777–780
1794. Zh. Sun, J. Zhou, R. Ahuja, Unique melting behavior in phase-change materials for rewritable data storage. Phys. Rev. Lett. **98**, 055505(4pp) (2007)
1795. J. Hegedüs, S.R. Elliott, Microscopic origin of the fast crystallization ability of Ge-Sb-Te phase-change memory materials. Nat. Mater. **7**, 399–405 (2008)
1796. P.K. Naji, M. Durlam, S. Tehrani, J. Calder, M.F. DeHerrera, A 256kb 3.0V 1T1MTJ non-volatile magnetoresistive RAM, in *ISSCC Digest of Technical Papers* (IEEE, Piscataway, 2001), pp. 122–124
1797. W.J. Gallagher, S.S.P. Parkin, Development of the magnetic tunnel junction MRAM at IBM: from first junctions to a 16-Mb MRAM demonstrator chip. IBM J. Res. Dev. **50**, 5–23 (2006); IBM J. Res. Dev. **50**, 23A (2006) (erratum)
1798. M.N. Kozicki, C. Gopalan, M. Balakrishnan, M. Park, M. Mitkova, Non-volatile memory based on solid electrolytes. IEEE Non-Volatile Memory Technology Symposium (IEEE, Piscataway, 2004), pp. 10–17
1799. M.N. Kozicki, P. Mira, M. Mitkova, Nanoscale memory elements based on solid-state electrolytes. IEEE Trans. Nanotechnol. **4**, 331–338 (2005)
1800. K. Szot, W. Speier, G. Bihlmayer, R. Waser, Switching the electrical resistance of individual dislocations in single-crystalline $SrTiO_3$. Nat. Mater. **5**, 312–320 (2006)
1801. A. Beck, J.G. Bednorz, Ch. Gerber, C. Rossel, D. Widmer, Reproducible switching effect in thin oxide films for memory applications. Appl. Phys. Lett. **77**, 139–141 (2000)
1802. S. Karg, G.I. Meijer, D. Widmer, R. Stutz, J.G. Bednorz, Ch. Rettner, *Nanoscale resistive memory device using $SrTiO_3$ films, in 22nd IEEE Non-Volatile Semiconductor Memory Workshop, Monterey* (Piscataway, IEEE, 2007), pp. 68–70
1803. M. Janousch, G.I. Meijer, U. Staub, B. Delley, S.F. Karg, B.P. Andreasson, Role of oxygen vacancies in Cr-doped $SrTiO_3$ for resistance-change memory. Adv. Mater. **19**, 2232–2235 (2007)
1804. M.A. Reed, C. Zhou, C.J. Muller, T.P. Burgin, J.M. Tour, Conductance of a molecular junction. Science **278**, 252–254 (1997)
1805. J.R. Heath, P.J. Kuekes, G.S. Snider, R.S. Williams, A defect-tolerant computer architecture: opportunities for nanotechnology. Science **280**, 1716–1721 (1998)
1806. A. Aviram, M. Ratner (eds.), *Molecular Electronics: Science and Technology*. Annals of the New York Academy of Sciences, vol. 852 (1998)
1807. E.F. Schubert, Delta-doping of semiconductors: electronic, optical and structural properties of materials and devices. Semicond. Semimetals **40**, 1–151 (1994)
1808. M. Schlechtweg, A. Tessmann, A. Leuther, C. Schwörer, M. Lang, U. Nowotny, O. Kappeler, Integrated circuits based on 300 GHz f_T metamorphic HEMT technology for millimeter-wave and mixed-signal applications, in *Proceedings of the European Gallium Arsenide and*

other Compound Semiconductors Application Symposium (GAAS 2003) (Horizon House, London, 2003), pp. 465–468
1809. D. Lubyshev, W.K. Liu, T.R. Stewart, A.B. Cornfeld, X.M. Fang, X. Xu, P. Specht, C. Kisielowski, M. Naidenkova, M.S. Goorsky, C.S. Whelan, W.E. Hoke, P.F. Marsh, J.M. Millunchick, S.P. Svensson, Strain relaxation and dislocation filtering in metamorphic high electron mobility transistor structures grown on GaAs substrates. J. Vac. Sci. Technol. B **19**, 1510–1514 (2001)
1810. S.D. Brotherton, *Introduction to Thin Film Transistors (Physics and Technology of TFTs* (Springer, Heidelberg, 2013)
1811. Y. Kuo (ed.), *Thin Film Transistors: Materials and Processes* (Two volumes, vol. 1: Amorphous silicon thin film transistors and vol 2: Polycrystalline silicon thin film transistors) (Kluwer, Boston, 2003)
1812. G. Horowitz, Organic field-effect transistors. Adv. Mater. **10**, 365–377 (1998)
1813. C.D. Dimitrakopoulos, P.R.L. Malenfant, Organic thin film transistors for large area electronics. Adv. Mater. **14**, 99–117 (2002)
1814. BCh. Shekar, J. Lee, S.-W. Rhee, Organic thin film transistors: materials, processes and devices. Korean J. Chem. Eng. **21**, 267–285 (2004)
1815. Y.M. Sun, Y.Q. Liu, D.B. Zhu, Advances in organic field-effect transistors. J. Mater. Chem. **15**, 53–65 (2005)
1816. C.T. Liu, Revolution of the TFT LCD technology. J. Disp. Technol. **3**, 342–350 (2007)
1817. U. Köster, Crystallization of amorphous silicon films. Phys. Status Solidi A **48**, 313–321 (1978)
1818. D. Toet, P.M. Smith, T.W. Sigmon, T. Takehara, C.C. Tsai, W.R. Harshbarger, M.O. Thompson, Laser crystallization and structural characterization of hydrogenated amorphous silicon thin films. J. Appl. Phys. **85**, 7914–7918 (1999)
1819. R.B. Iverson, R. Reif, Recrystallization of amorphized polycrystalline silicon films on SiO_2: temperature dependence of the crystallization parameters. J. Appl. Phys. **62**, 1675–1681 (1987)
1820. N. Yamauchi, R. Reif, Polycrystalline silicon thin films processed with silicon ion implantation and subsequent solid-phase crystallization: theory, experiments, and thin-film transistor applications. J. Appl. Phys. **75**, 3235–3257 (1994)
1821. G. Liu, S.J. Fonash, Polycrystalline silicon thin film transistors on Corning 7059 glass substrates using short time, low-temperature processing. Appl. Phys. Lett. **62**, 2554–2556 (1993)
1822. M.S. Haque, H.A. Naseem, W.D. Brown, Aluminum-induced crystallization and counter-doping of phosphorous-doped hydrogenated amorphous silicon at low temperatures. J. Appl. Phys. **79**, 7529–7536 (1996)
1823. L.H. Lee, Y.K. Fang, S.H. Fan, Au metal-induced lateral crystallisation (MILC) of hydrogenated amorphous silicon thin film with very low annealing temperature and fast MILC rate. Electron. Lett. **35**, 1108–1109 (1999)
1824. T.K. Kim, T.H. Ihn, B.I. Lee, S.K. Joo, High-performance low-temperature poly-silicon thin film transistors fabricated by new metal-induced lateral crystallization process. Jpn. J. Appl. Phys. **37**, 4244–4247 (1998)
1825. S.-W. Lee, Y.-C. Jeon, S.-K. Joo, Pd induced lateral crystallization of amorphous Si thin films. Appl. Phys. Lett. **66**, 1671–1673 (1995)
1826. K. Sera, F. Okumura, H. Uchida, S. Itoh, S. Kaneko, K. Hotta, High-performance TFTs fabricated by XeCl excimer laser annealing of hydrogenated amorphous-silicon film. IEEE Trans. Electron Devices **ED-36**, 2868–2872 (1989)
1827. H. Gleskova, S. Wagner, Electron mobility in amorphous silicon thin-film transistors under compressive strain. Appl. Phys. Lett. **79**, 3347–3349 (2001)
1828. N. Yamauchi, N. Kakuda, T. Hisaki, Characteristics of high mobility polysilicon thin-film transistors using very thin sputter-deposited SiO_2 films. IEEE Trans. Electron Devices **ED-41**, 1882–1885 (1994)

1829. M. Zaghdoudi, R. Rogel, N. Alzaied, M. Fathallah, T. Mohammed-Brahim, High polysilicon TFT field effect mobility reached thanks to slight phosphorus content in the active layer. Mater. Sci. Eng. C **28**, 1010–1013 (2008)
1830. J.F. Wager, Transparent electronics. Science **300**, 1245–1246 (2003)
1831. K. Nomura, H. Ohta, K. Ueda, T. Kamiya, M. Hirano, H. Hosono, Thin-film transistor fabricated in single-crystalline transparent oxide semiconductor. Science **300**, 1269–1272 (2003)
1832. E.M.C. Fortunato, P.M.C. Barquinha, A. Pimentel, A.M.F. Gonçalves, A.J.S. Marques, L.M.N. Pereira, R.F.P. Martins, Fully transparent ZnO thin-film transistor produced at room temperature. Adv. Mater. **17**, 590–594 (2005)
1833. H. Frenzel, A. Lajn, M. Brandt, H. von Wenckstern, G. Biehne, H. Hochmuth, M. Lorenz, M. Grundmann, ZnO metal-semiconductor field-effect transistors with Ag-Schottky gates. Appl. Phys. Lett. **92**, 192108 (2008)
1834. M. Grundmann, H. Frenzel, A. Lajn, M. Lorenz, F. Schein, H. von Wenckstern, Transparent semiconducting oxides: materials and devices. Phys. Status Solidi A **207**, 1437–1449 (2010)
1835. H. Frenzel, A. Lajn, H. von Wenckstern, M. Grundmann, Ultrathin gate-contacts for metal-semiconductor field-effect transistor devices: an alternative approach in transparent electronics. J. Appl. Phys. **107**, 114515(6pp) (2010)
1836. S. Jit, P.K. Pandey, P.K. Tiwari, Modeling of the subthreshold current and subthreshold swing of fully depleted short-channel Si-SOI-MESFETs. Solid State Electron. **53**, 57–62 (2009)
1837. S.R. Forrest, Ultrathin organic films grown by organic molecular beam deposition and related techniques. Chem. Rev. **97**, 1793–1896 (1997)
1838. G. Horowitz, Organic thin film transistors: from theory to real devices. J. Mater. Res. **19**, 1946–1962 (2004)
1839. Z. Bao, J. Locklin (eds.), *Organic Field-Effect Transistors* (CRC Press, New York, 2007)
1840. Ch. Wöll (ed.), *Physical and Chemical Aspects of Organic Electronics, From Fundamentals to Functioning Devices: Structural and Electronic Properties of OFETs* (Wiley-VCH, Weinheim, 2009)
1841. G. Horowitz, F. Garnier, A. Yassar, R. Hajlaoui, F. Kouki, Field-effect transistor made with a sexithiophene single crystal. Adv. Mater. **8**, 52–54 (1996)
1842. H.E. Katz, C. Kloc, V. Sundar, J. Zaumseil, A.L. Briseno, Z. Bao, Field-effect transistors made from macroscopic single crystals of tetracene and related semiconductors on polymer dielectrics. J. Mater. Res. **19**, 1995–1998 (2004)
1843. H.A. Kramers, Some remarks on the theory of absorption and refraction of X-rays. Nature **117**, 775 (1926)
1844. R. de L. Kronig, On the theory of dispersion of X-rays. J. Opt. Soc. Am. **12**, 547–557 (1926)
1845. P.-O. Löwdin, A note on the quantum-mechanical perturbation theory. J. Chem. Phys. **19**, 1396–1401 (1951)
1846. G.H. Wannier, The structure of electronic excitation levels in insulating crystals. Phys. Rev. **52**, 191–197 (1937)
1847. M. Cutler, N.F. Mott, Observation of anderson localization in an electron gas. Phys. Rev. **181**, 1336–1140 (1969)
1848. C. Herring, Theory of the thermoelectric power of semiconductors. Phys. Rev. **96**, 1163–1187 (1954)
1849. E.D. Kriger, A.F. Kravchenko, B.V. Morozov, V.G. Polovinkin, E.M. Skok, Investigation of the phonon drag effect in n-GaAs. Phys. Status Solidi A **13**, 389–398 (1972)
1850. M.W. Wu, N.J.M Horing, H.L. Cui, Phonon-drag effects on thermoelectric power. Phys. Rev. B **54**, 5438–5443 (1996)
1851. A. van der Ziel, *Noise Sources, Characterization, Measurement* (Prentice-Hall, Englewood Cliffs, 1970)
1852. J. Engberg, T. Larsen, *Noise Theory of Linear and Nonlinear Circuits* (Wiley, Chichester, 1995)
1853. Sh Kogan, *Electronic Noise and Fluctuations in Solids* (Cambridge University Press, Cambridge, 1996)

1854. B. Schiek, H.-J. Siweris, I. Rolfes, *Noise in High-Frequency Circuits and Oscillators* (Wiley, Hoboken, 2006)
1855. R. Müller, *Rauschen*, 2nd edn. (Springer, Berlin, 2013)
1856. S. Yngvesson, Review of noise processes and noise concepts relevant to microwave semiconductor devices, in: Ref. [1524], pp. 207–228
1857. P. Dutta, P.M. Horn, Low-frequency fluctuations in solids: $1/f$ noise. Rev. Mod. Phys. **53**, 497–516 (1981)
1858. M.B. Weissmann, $1/f$ noise and other slow, nonexponential kinetics in condensed matter. Rev. Mod. Phys. **60**, 537–571 (1988)
1859. T.G.M. Kleinpenning, A.H. de Kuijper, Relation between variance and sample duration of $1/f$ noise signals. J. Appl. Phys. **63**, 43–45 (1988)
1860. T.M. Chen, S.F. Su, D. Smith, $1/f$ Noise in Ru-based thick-film resistor. Solid State Electron. **25**, 821–827 (1982)
1861. J.M. Boudry, L.E. Antonuk, Current-noise-power spectra of amorphous silicon thin-film transistors. J. Appl. Phys. **76**, 2529–2534 (1994)
1862. J.E. Hill, K.M. van Vliet, Ambipolar transport of carrier density fluctuations in germanium. Physica **24**, 709–720 (1958)
1863. G. Ferrari, M. Sampietro, G. Bertuccio, G. Gomila, L. Reggiani, On the origin of shot noise in CdTe detectors. Appl. Phys. Lett. **83**, 2450–2452 (2003)
1864. C.W.J. Beenakker, M. Büttiker, Suppression of shot noise in metallic diffusive conductors. Phys. Rev. B **46**, 1889–1892 (1992)
1865. H. Schomerus, E.G. Mishchenko, C.W.J. Beenakker, Shot-noise, in *non-degenerate semiconductors with energy-dependent elastic scattering, in Statistical and Dynamical Aspects of Mesoscopic Systems*, ed. by D. Reguera, J.M. Rubí, G. Platero, L.L. Bonilla, Lecture Notes, in Physics, vol. 547, (Springer, Berlin, 2000), pp. 96–104
1866. A. van der Ziel, Shot noise in semiconductors. J. Appl. Phys. **24**, 222–223 (1953)
1867. R. Müller, Generation-recombination noise, in Noise in Physical Systems, ed. by D. Wolf. Springer Series in Electrophysics, vol. 2 (1978), pp. 13–25
1868. V. Mitin, L. Reggiani, L. Varani, Generation-recombination noise in semiconductors, in *Noise and Fluctuations Control in Electronic Devices*, ed. by A. Balandin (American Scientific Publishers, Valencia, 2002), pp. 1–19 (Chapter 2)
1869. V. Bareikis, J. Liberis, I. Matulioniene, A. Matulionis, P. Sakalas, Experiments on hot electron noise in semiconductor materials for high-speed devices. IEEE Trans. Electron Devices **41**, 2050–2060 (1994)
1870. B.K. Jones, Low-frequency noise spectroscopy. IEEE Trans. Electron Devices **41**, 2188–2197 (1994)

Index

A
Absorption, 291, 295, 346, 350, 655, 675, 680, 699
  atmospheric, 708
  band–band, 299
  bleaching, 247
  bound exciton, 317, 319
  coefficient, *see* coefficient, absorption
  free-carrier, 296, 328, 448
  impurity, 323
  inter-valley, 335
  intervalence-band, 334
  intra-band, 336
  lattice, 337
  negative, 321
  optical, 246
  phonon, 305
  schematic spectrum, 296
  selfabsorption, 365
  spectrum, 302, 430, 438
  two-photon, 302, 311, 322
Absorption edge, 306
  optical, 166
  shift, 321
Acceleration, 178
Acceptor, 211, 218, 220
  binding energy, 220
  charged, 220
  double, 240
  energy, 220
  neutral, 221, 366
Accumulation, 617
Admittance spectroscopy, 595
Air gap, 780
Air mass, 708
Alkali halogenides, 276
Alloy, 72, 110, 252, 505, 612
  broadening, *see* broadening, alloy
  phonon, *see* phonon, alloy
  quaternary, 73
  random, 72, 359
  ternary, 65, 72, 172
Amplification, 350, 755, 773, 786
  current, *see* current, amplification
  light, 749
  region, 695
Amplifier
  optical, 750, 785
  power, 786
Angle
  Brewster, 295
  critical, 294, 735
  Euler, 844
  off-cut, 390
  taper, 785
Anharmonicity, 492
Annealing, 53, 81, 151
  laser, 838
  thermal, 838
Anode, 583
Anticrossing, 315, 420, 562, 563
Antidot, 460
Antiphase domain, *see* domain, antiphase
Approximation
  abrupt, 589, 626, 805
  adiabatic, 39, 805
  Boltzmann, 204, 205, 217, 607
  Born–Oppenheimer, 39, 364
  electric dipole, 297
  envelope function, 417
  gradual channel, 805
  harmonic, 111
  quasi-cubic, 189
  relaxation time, 256, 275, 328, 876

© Springer International Publishing Switzerland 2016

M. Grundmann, *The Physics of Semiconductors*, Graduate Texts in Physics,
DOI 10.1007/978-3-319-23880-7

two-band, 323
virtual crystal, 77, 170
WKB, 376
Atmosphere, 708
Auger process, 368
Autocompensation, 226
Autodoping, 229
Auxetic, 136
Avalanche multiplication, 645, 678, 792
Average, 881

B
Background radiation, 683
Band
alignment, 416
bending, 424, 587, 617
conduction, 28, 162, 376
minimum, 163, 172
diagram, 788
diode, 584, 616
discontinuity, 417, 419
edge, 324
filled, 183, 207
gap, 29, 162, 167, 168, 178, 207, 464, 497, 867
engineering, 416
fundamental, 305
negative, 189
photonic, 545–564
renormalization, 320
temperature dependence, 174
zero, 189, 507
impurity, 230
lineup, 415
mixing, 192, 420, 473, 675
parabolic, 301
remote, 869, 870
splitting, 34
staggered lineup, 416
straddled lineup, 416
tail, 129, 307
tilted, 437
valence, 28, 162, 784
fine structure, 188
Band structure, 153–193, 297, 523
chalcopyrites, 165
delafossites, 166
direct, 163, 172
extrema, 195
indirect, 163, 172, 321
lead salt, 163
multivalley, 270
perovskites, 167
photonic, 552
projected, 392
rocksalt, 164
spinels, 166
strained, 190
theory, 6
Bandwidth, 673, 683, 697, 885
Bardeen model, 584, 588
Barrier, 375, 419, 424
Coulomb, 474
finite, 421
height, 419, 603
reduction, 594
Schottky, 586
triangular, 376
Base, 42, 788
diatomic, 42, 130
monoatomic, 42
width, 798
Beam profile, 759, 779
Benzene, 30, 515
Biexciton, 483, 484
Binodal, 74
Bipolaron, 521
Birefringence, 292, 328, 449
Blackbody, 673
Bloch
function, *see* function, Bloch
oscillation, *see* oscillation, Bloch
Bloch's theorem, 154
Bohr radius, 211
Boltzmann
approximation, *see* approximation, Boltzmann
constant, *see* constant, Boltzmann
distribution, *see* distribution, Boltzmann
transport equation, 875
Bond
angle, 52
breaking, 150
covalent, 25
dangling, 52, 388
elastic, 111
electron pair, 26
homopolar, 34
ionic, 32
length, 34, 52, 77, 82, 136, 171
metallic, 36
mixed, 33
partially ionic, 129
sp^2, 29
sp^3, 26

strength, 168
strong, 409
tetrahedral, 28, 211
van-der-Waals, 37
Bose–Einstein
distribution, *see* distribution, Bose–Einstein
statistics, 860
Boson, 859
Bottleneck, 315
Boule, 401
Boundary
additional ∼ condition, 316
antiphase domain, 107
condition, 112, 379, 419, 461, 471, 548, 589, 620, 628, 790, 865
depletion layer, 699, 790
grain, *see* grain boundary
inversion domain, 107, 410
periodic ∼ condition, 112
sample, 457
Bow-tie, 223, 569
Bowing, 77
parameter, 170
Bragg mirror, 550, 750, 761, 780
Breakdown, 644, 680, 688, 797
voltage, *see* voltage, breakdown
Brightness, 727, 728
high, 738
perceived, 734
visible, 726
Brillouin zone, 70, 112, 113, 156, 386, 523
cubic
body-centered, 71
face centered, 71
simple, 71
hexagonally close packed, 71
orthorhombic, 71, 166
paths, 72
points, *see* point
size, 300
Broadening, 129
alloy, 358
inhomogeneous, 429, 457, 486
phonon, 313
Stark, 325
Buffer, 409
graded, 836
metamorphic, 722
relaxed, 836
Bulk modulus, 136
Burger's vector, 96, 147
Burstein–Moss shift, 332

C

Capacitance, 595, 603, 623, 632
insulator, 623
parasitic, 800
Capture, 464
barrier, 246
cross section, 371
Carrier
capture, 235, 370
concentration, 204
intrinsic, 207
density, 343, 449
equilibrium, 349
excess, 349, 379
nonequilibrium, 350
excess, 368, 464
excess ∼ profile, 379
free, 329
freeze-out, 216
hot, 714
injection, 234, 343
itinerant charge, 509
lifetime, *see* lifetime, carrier
majority, 583, 598
minority, 583, 626
release, 370
threshold ∼ density, 763
Catalysis, 385, 400
Catastrophical optical damage, 758, 767
Cathodoluminescence, 464
Cavity
empty, 562
external, 750
Fabry–Perot, 562, 750
hexagonal, 571
micro-, *see* microcavity
mirror, 562
mode, 562
optical, 749
resonance, 568
short, 761
Cell
central ∼ correction, 228, 324
elementary, 43
internal parameter, 56, 77, 490
photoelectrochemical, 670
primitive elementary, 43
primitive unit, 64
solar, *see* solar cell
tandem, 714
unit, 43, 489
Wigner–Seitz, 43, 70
Chalcogenide, 57, 831

atoms, 238
europium, 505
glass, 242
impurity, 239
lead, 163
Chalcopyrite structure, *see* structure
Channel, 512, 802
buried, 704
depth, 805
edge, 457
inversion, 814
isolation, 704
length, 804, 824
long, 812
n-type, 818
p-type, 818
short, 811
stop, 703
strain, 820
width, 806, 824
Charge
conservation, 372
deficit, 459
density, 424
effective, 35
elementary, 236
excess, 343
fixed, 211
image, 423, 593
inversion, 622, 698
ionic, 625
maximum, 701
neutrality, 620, 622
packet, 703
polarization, 489
relative shift, 489
retention, 830
sign, 449
signal, 699
state, 236
storage, 661
surface, 586
transfer, 424
Charge coupled device, 14, 698–706
Chemical shift, 324
Chromaticity, 727
Circuit
common base, 788, 796, 798
common collector, 796, 799
common emitter, 792, 798, 799
equivalent, 644, 658, 673, 683
feedback, 800
integrated, 10, 14, 818
millimeter-wave, 655
open, 286, 711
optoelectronic integrated, 558
read-out, 701, 706
short, 289, 711
Cleaving, 150, 387, 465, 761
Clock speed, 826
Cluster, 72
size, 508
Clustering, 72, 110
CMOS technology, 706, 818–829
Coating
antireflection, 691, 762, 785
high-reflection, 762
Coefficient
absorption, 296, 301, 304, 329, 380, 706, 852
Auger recombination, 368
bimolecular recombination, 348
correlation, 883
diffusion, 249, 282
distribution, 89, 94
elastic, 133
electron ionization, 273
gain compression, 775, 778
hole ionization, 273
impact ionization, 693
negative temperature, 803
Peltier, 286
sign, 289
Seebeck, 879
segregation, 89
stiffness, 134
temperature, 645, 647, 658
thermal expansion, 132, 142, 174, 413
transmission, 547
two-photon absorption, 323
Coincident site lattice, 106
Collector, 788
Collision, 876
Colloid, 478, 487
Color space, 730
Commensurability, 460
Compensation, 223, 247, 366
Composite fermion, 459
Compound
binary, 35
carbon, 29, 515
I–VII, 32
II–VI, 33
ionic, 33
Compressibility, 136
Condensation, 321

Bose–Einstein, 321
bosonic, 322
Conduction
band, *see* band, conduction
heat, 284–285
hole, 221
intrinsic, 207, 218
n-type, 221
ohmic, 208
p-type, 221
two-band, 444, 880
Conductivity, 203, 256–257, 521, 670, 699, 876
channel, 474, 802
complex, 275, 328
heat, 284
longitudinal, 457
metal, 257
one-dimensional, 457
type, 229
zero-field, 454
Configuration
atomic, 236
coordinate, 242
electron, *see* electron, configuration
Confinement
energy, *see* energy
potential, *see* potential
spatial, 302
Constant
Boltzmann, 862
damping, 275
dielectric, 131, 211, 337, 394, 423
high frequency, 131
negative, 332
static, 131
effective Richardson, 608
elastic, 112, 134
fine-structure, 456
force, 114, 117, 124, 125, 242
Fröhlich coupling, 276
lattice, *see* lattice, constant
normalization, 154
Poisson, 142
Richardson, 600, 603
spring, 112
static dielectric, 590
von-Klitzing, 455
Contact
back, 715
base, 790
intracavity, 780
lines, 696
metal–semiconductor, 7, 583–615
nonohmic, 802
Ohmic, 584, 611–613, 820
resistance, 584, 611
Schottky, 7, 62, 686
transparent, 691
Continuity equation, 284
Convolution, 429
Cooler
thermoelectric, 289
Coordination number, 43
Correlation, 883
Coulomb
blockade, 474
charging energy, 475
correlation, 311
gauge, 297
staircase, 474
Coupling, 791
antiferromagnetic, 506
capacitive, 473
ferromagnetic, 509
intervalley, 324
nanostructures, 461
phonon–plasmon, 340
strong, 562
weak, 562
Crack, 103, 413
Crescent, 463
Critical thickness, *see* thickness, critical
Crystal
class, 46
nobel gas, 37
structure, 13
Curie–Weiss law, 495
Current
amplification, 792
dark, 678, 683, 691, 699, 711
diffusion, 282, 681
divergence, 284
excess, 664, 666
leakage, 714
particle, 282
photo-, 672, 709
photogenerated, 681
recombination, *see* recombination, current
saturation, 600, 609, 807
short-circuit, 711
spreading, 767
transparency, 763
tunneling, 223, 608, 664

Curvature, 141, 159, 178, 182, 315, 373, 463, 647, 736
C–V spectroscopy, 595, 596, 603
Cyclotron
  frequency, 442, 449
  motion, 256, 441, 460
  orbit, 450, 460
  resonance, 181

D
De Broglie wavelength, 461
De Haas–van Alphén effect, 451
Debye length, 620, 631, 817
Decay
  hyperbolic, 350
  time, 350
Defect, 72, 81–110, 146
  acceptor, 220
  annihilation, 103
  antisite, 82, 238, 247
  area, 81
  density, 410
  diffusion, 88
  donor, 211
  double acceptor, 240
  double donor, 238
  EL2, 247
  electronic states, 203–253
  etching induced, 477
  Frenkel, 82
  interstitial, 82, 229, 237, 247, 252
  isoelectronic, 249
  line, 81, 96, 558
  metastable, 81
  nucleation, 415
  pair, 92, 223
  passivation, 251
  point, 81, 229, 373, 558
  symmetry, 241
  thermodynamics, 83
  vacancy, 82
Deformation
  volume, 170
Degeneracy, 85, 117, 220, 244, 451, 474
  holes, 420
  Kramer's, 159
  spin, 195, 214, 222, 456, 531
  valley, 200, 215, 452, 531
Delafossite structure, *see* structure
Demodulation, 655
Density of modes, 551, 567
Density of states, 195–201, 204, 420, 424, 450, 451, 461, 522
  amorphous semiconductor, 196
  band-edge, 372
  conduction-band edge, 205
  δ-like, 457
  joint, 299, 429
  surface, 584, 587
  two-dimensional, 200
  valence-band edge, 205
Depletion, 617
  deep, 698
Depletion layer, 208, 586, 680
  width, 591, 630
Depolarization, 489
Deposition
  chemical vapor, 400
  pulsed laser, 400
Detailed balance, 343, 347
Detectivity, 673, 677, 680
Deuterium, 510
Device
  charge coupled, *see* charge coupled device
  cooling, 256
  high-power, 256
  high-speed, 247
  optoelectronic, 291
  performance, 208
  photonic, 291
  reliability, 835
  two-terminal, 583
Diagram
  chromaticity, 727
Diamond structure, *see* structure
Dielectric constant, *see* constant, dielectric
Dielectric function, *see* function, dielectric
Diffraction, 293–296
  reflection high energy electron, 400
Diffusion, 255, 282–283, 464, 598, 607, 628, 699, 703, 788
  equation, 379
  lateral, 431, 674
  length, 379
  pair $\sim$ mechanism, 88
  point defects, 85
  surface, 400
Dimer, 388
Diode, 255, 583–650
  backward, 666
  bipolar, 583, 626–648
    I–V characteristics, 633–644
  fast-recovery, 660
  Gunn, 269, 667
  heterostructure, 626, 649, 661, 800

ideal characteristic, 600
laser, *see* laser
light-emitting, 399, 733–743
application, 734
white, 743
metal–insulator–semiconductor, 584, 615–626
metal–oxide–semiconductor, 452, 615
metal–semiconductor, 583
nonideal MIS, 624
one-sided, 636, 640
photo-, *see* photodiode
pin, 662
pn-junction, 583
Schottky, 484, 584
I–V characteristics, 598–610
step-recovery, 661
tunneling, 663
unipolar, 583
Zener, 658
Dipole moment, 439
Dirac particle, 536
Direction
growth ∼, 750
polar, 106
surface ∼, 778
Dislocation, 96, 401
60°, 96
α, β, 96
core, 96, 148
density, 148
edge, 96
half-loop, 147
line, 96
misfit, 96, 146, 412
partial, 98, 105
screw, 96
spacing, 106
threading, 147
Disorder, 51, 77, 110, 128, 171, 196, 302, 307, 451, 603
cation, 58
configurational, 84
isotope, 356, 358
Dispersion
free electron, 178
linear, 315
quadratic, 315
spatial, 292, 314
Dispersion relation, 177, 178, 331, 420, 548, 867
branch, 114, 118, 339
free electron, 156
hole, 184
lattice vibration, 111
linear chain, 113, 116
parabolic, 159
Displacement, 96, 112, 140, 246, 262, 491
atomic, 241
ion, 275
parameter, 58
Display, 727
application, 730
electroluminescence, 837
field-effect, 727
liquid crystal, 837
plasma, 727
Distortion
tetragonal, 145, 480
Distribution
binomial, 359
Boltzmann, 862
Bose–Einstein, 262, 347, 863
degenerate, 204, 332
Fermi–Dirac, 204, 428, 862
Gaussian, 429, 603
momentum, 875
nondegenerate, 204
Planck, 861
spectral power, 727
Domain
antiphase, 106, 408
high field, 667
inversion, 106
polarization, 496
Donor, 211
–acceptor pair, 366
binding energy, 211
deep, 247
double, 238
empty, 215
fine structure, 216
ionization energy, 211, 216
ionized, 214
neutral, 214, 366
populated, 215
shallow, 211
Dopant, 370, 425
Doping, 82, 209, 211–233, 329, 520, 703
concentration, 595
depth profile, 596
glass, 478
modulation, 266, 425, 451, 834
profile, 659, 811
Drain, 802
Drift, 255, 598, 684, 699

self-induced, 703
time, 691
Droop, 741, 746
Droplet, 478
Drude theory, 449, 453
DX center, 245, 834

E
Early effect, 796, 798
Ebers–Moll model, 794, 796
Effect
field, 375
Hall, *see* Hall, effect
Jahn–Teller, *see* Jahn–Teller effect
polaronic, 276
quantum Hall, *see* Hall, effect
Stark, *see* Stark effect
thermoelectric, 286
Effective mass, *see* mass, effective
Effective-mass
impurity, 213, 235
theory, 211, 873
Efficiency
conversion, 712, 727
differential, 766
emitter, 792
external, 766
external quantum, 735
internal quantum, 735, 766
light extraction, 714, 735
maximum solar cell, 712
packaging, 735
quantum, 477, 525, 672, 680
total quantum, 735, 767
wall-plug, 735, 767, 786
Einstein relation, 283
EL2 defect, *see* defect, EL2
Elasticity, 132
Electroabsorption, 613
Electroluminescence, 3, 837
Electromagnetic spectrum, 291
Electron
affinity, 210, 519, 586
conduction, 36
configuration, 26, 505
density, 204
dispersion, 177
distribution function, 204
equation of motion, 177
mass, *see* mass, electron
trap, 371
valence, 36, 153
wave packet, 177
Electron gas
free, 199
one-dimensional, 200
three-dimensional, 199, 449
two-dimensional, 200, 259, 424, 451, 460
Electron–hole droplet, 321
Electronegativity, 34, 338
Electrophotography, 674
Ellipsometry, 449
Emission
amplified spontaneous, 769
directional, 561
field, 598, 609
pattern, 570
phonon, 305
probability, 375
spontaneous, 344, 545
stimulated, 297, 347, 350
thermally activated, 375
thermionic, 598, 599, 677
thermionic field, 609
Emission rate, *see* rate, emission
Emitter, 788
follower, 800
Empty lattice, *see* lattice, empty
Energy
activation, 88, 253
barrier, 242
capture, 245
charging, 245
confinement, 419
conservation, 304
correlation, 320
Coulomb, 242
Coulomb charging, *see* Coulomb
defect formation, 81
density, 256
Dirac, 533, 535
dissipation, 457
elastic, 133
electrostatic, 473
exchange, 320
free, 236, 494
gap, 29, 30, 158
ionization, 211, 375, 519
kinetic, 157
loss, 269
Madelung, 32
optoelectronic, 439
parameter, 179
radiation, 562

Rydberg, 213, 309
strain, *see* strain, energy
surface, 387, 478
thermal, 262
zero-point, 37
Enthalpy, 84
formation, 84, 87, 237
free, 83, 237
migration, 87
mixing, 74
Entropy, 84, 244, 494
configurational, 74
disorder, 84
formation, 84
Envelope function, *see* function, envelope
Epitaxial relationship, 405
Epitaxy
liquid phase, 400
metalorganic vapor phase, 400
molecular beam, 400
thin film, 400
Equilibrium
crystal shape, 388
state, 343
stationary, 607
thermodynamic, 204, 207, 234, 235, 347–349, 368, 371, 372, 424, 584, 862
Etch, 100, 402
pit, 100
RCA, 402
Shiraki, 402
Etching, 477
anisotropic, 107, 462
plasma, 101
reactive ion, 101, 477, 761
thermal, 402
wet chemical, 477
Excitation
external, 343
neutral, 349
optical, 246
Exciton, 7, 296, 309–312, 523
binding energy, 309, 422
bound, 317
absorption, *see* absorption
recombination, *see* recombination
bright, 314
charged, 319, 482
correlation, 439
dark, 314
delocalized, 431
dynamics, 429
free, 351, 359
Frenkel, 523
interchain, 524
intrachain, 524
ionization, 439, 441
localized, 431
longitudinal, 314
ortho-, 313
oscillator strength, 310
para-, 313
polariton, 313
radius, 309
recombination, 351
scattering states, 311
self-trapped, 732
transverse, 314
volume, 359
Exclusion principle, 859
Exhaustion, 218
regime, 226
Eye pattern, 776

F

Facet
cleaved, 465, 750
etched, 761
laser, 762
side, 463
tilted, 778
Faceting, 392
Factor
base transport, 792
collector multiplication, 792
fractional filling, 458
ideality, 601, 604, 793
linewidth enhancement, 778
optical confinement, 695
quality, 559, 568
Sommerfeld, 311
spontaneous emission, 765
Fano resonance, 326
Far field, 759
Feedback
distributed, 770
loop, 400
optical, 749
Fermi
energy, 450
function, *see* function, Fermi
integral, 204, 205, 283
intrinsic ∼ level, 208, 222, 617
level, 216, 218, 222, 225, 333, 373, 424, 452, 485, 584, 614, 862

gradient, 282
liquid, 321
local quasi ∼ level, 235
quasi ∼ level, 234, 347, 350, 598, 698, 812
sphere, 256
surface, 450
vector, 450
Fermi–Dirac
distribution, *see* distribution, Fermi–Dirac
Fermi–Dirac statistics, 236, 237, 861
Fermion, 859
Fermi's golden rule, 297, 545, 567
Ferroelectricity, 489, 490
Ferromagnet
Heisenberg, 506
Fick's law, 282
Field
built-in, 613
crossed electric and magnetic, 784
crystal, 248
displacement, 131, 853
drift, 789
effect, *see* effect, field
electric, 177, 255, 273, 276, 282, 328, 437, 590
electromagnetic, 297
external, 437
fringing, 703
high magnetic, 453
homogeneous, 439
internal electric, 255
macroscopic electric, 129
magnetic, 177, 181, 256, 302, 325, 441
static, 328
piezoelectric, 262, 473
ring, 647
strength, 695
strong electric, 375
surface, 620
time dependent electric, 328
transverse electric, 441
Filling factor, 49, 711
Finesse, 559
Flip-flop, 10
Flow
heat, 284
Fluctuation, 75, 359, 429, 431, 881
amplitude, 778
phase, 778
quantum dot size, 486
vacuum, 545
Fluorescence, 522
Flux
luminous, 726
radiant, 725
Focal plane array, 677
Force, 178
dissipative, 275
image, 593
Lorentz, 441, 453, 876
restoring, 132, 492
van-der-Waals, 32
Force constant, *see* constant, force
Fourier
coefficient, 156, 157
series, 156
transform, 851
transformation, 66
Franck–Condon principle, 364
Franz–Keldysh oscillation, *see* oscillation, Franz–Keldysh
Freeze-out
regime, 218
Frequency
cutoff, 800
high, 800, 812, 836
mixing, 655
Fresnel formulas, 294, 550
Friction, 256
Fröhlich coupling constant, *see* constant, Fröhlich coupling
Function
auto correlation, 884
Bessel, 472
Bloch, 154, 179, 235, 297, 314, 417, 421, 869
Boltzmann, 430
color matching, 728
cross correlation, 883
dielectric, 131, 132, 297, 329, 563, 851, 854
distribution, 876
envelope, 417, 873
error-, 430
Fermi, 204, 347, 862
Hankel, 472
spherical harmonic, 471
static dielectric, 493
work, *see* work function

G

Gain, 678, 749, 769
differential, 776, 778

maximum, 772
Gate, 473, 802
charge, 830
contact, 512
voltage, 452, 474, 699
Gauge invariance, 457
Generation–recombination, 699
noise, *see* noise
g-factor, 441
Gibbs–Thomson effect, 478
Ginzburg–Landau theory, 493
Glide plane, 96
Glide reflection, 45
Grain
boundary, 41, 105, 252
boundary, small-angle, 105
size, 51, 720, 731, 837
Graphene, 16
Grating, 675
sampled, 773
Group
point, 43, 386, 847, 848
space, 45, 386, 405, 847, 849
theory, 41, 160
Growth
Czochralski, 401
Frank-van der Merwe, 404
kinetics, 228
methods, 400
mode, 403
pseudomorphic, 141
rate, *see* rate, growth
spiral, 96
Stranski-Krastanow, 404, 478
template, 477
Volmer-Weber, 404
Guard ring, 647
Gunn element, 667

H
Halbleiter, 4
Hall
bar, 455, 460
coefficient, 443, 455
sign, 445
constant, 444
current, 457
effect, 246, 441–447, 879
anomalous, 511
electrical, 449
fractional quantum, 458
integral quantized, 455
optical, 449
quantized, 14, 453
quantum, 399
factor, 447
plateau, 455, 458
quantum ∼ liquid, 457
resistivity, 457, 460
Hayne's rule, 353
Heat
capacity, 494
conduction, *see* conduction, heat
latent, 494
sink, 289, 738, 749, 767
transport, *see* transport, heat
Heavy metal, 525
Helmholtz equation, 546
Heteroepitaxy, 132, 137, 501
Heterointerface, 451
graded, 650
Heterojunction, 714, 833
Heterostructure, 14, 434, 455, 478, 565
type-I, 416
type-II, 416
Hexagonality index, 65
Hole, 183–190
capture, 373
concept, 183
density, 204, 222
dispersion, 183, 420
dispersion relation, *see* dispersion relation, hole
effective mass, 184
gas, 509
heavy, 184, 420
light, 184, 420
mass, *see* mass, hole
split-off, 184, 272, 302
HOMO, 30, 519
Hopping, 276
Hopping transport, *see* transport, hopping
Huang–Rhys parameter, *see* parameter, Huang–Rhys
Hue, 727
Hund's rule, 507
Hybridization
p–d, 509
sp^2, 30, 515, 529
sp^3, 26
Hydrogen, 211, 251, 356, 510
atom, 439
model, 309
molecule, 25, 319
problem, 213

Hydrogenation, 52
Hysteresis, 489, 496, 509

I
Ideality factor, *see* factor, ideality
Illuminance, 726
Illumination, 671, 691, 695, 704, 709
Image charge, *see* charge, image
Image sensor, 677, 701, 705
Impact ionization, 271, 368, 645, 692
Impedance, 655, 697
  amplifier, 799
  input, 803
Impurity, 82, 101, 203, 252, 324, 351, 370
  amphoteric, 228
  background, 223
  band, *see* band, impurity
  binding energy, 422
  charged, 325
  concentration, 203
  hardening, 101
  incorporation, 228
  isoelectronic, 318, 733
  isovalent, 82
  magnetic, 507
  shallow, 211
In-situ
  control, 400
  monitoring, 400
Inclusion, 59
Incompressible, 457
Index of refraction, 293, 329, 852
Insulator, 584, 615, 617, 622
  topological, 395
Interaction
  antiferromagnetic, 508
  atom–cavity, 562
  Coulomb, 32, 211, 309, 422, 458, 461
  dipole–dipole, 37
  double exchange, 509
  electron–phonon, 39, 174, 276, 297–299
  electrostatic, 32, 39
  exchange, 314, 321
  gap, 508
  hyperfine, 214, 236, 242
  indirect exchange, 509
  London, 37
  many-body, 459
  nearest-neighbor, 506
  parameter, 74
  Ruderman–Kittel–Kasuya–Yoshida, 509
  spin-orbit, 184, 441, 512, 525, 870
  superexchange, 506, 508
  van-der-Waals, 37, 516
Interconnect, 822
Interface, 146, 422, 490, 499, 617, 699
  flat, 431
  geometry, 399
  heterostructure, 584
  planar, 293, 399
  plane, 420, 598
  single hetero-, 417, 424
  state, 614
Interstitial, *see* defect
Inversion, 44, 106, 160, 617, 784
  center, 498
  charge, *see* charge, inversion
  strong, 621
  weak, 621
Inversion domain, *see* domain, inversion
Inverted opal, 556
Inverter, 820
Ionicity, 35, 136, 168, 338
Ionization, 324, 628
  exciton, *see* exciton
  photothermal, 324
Iron, 95, 107, 237, 249
Irradiance, 725
Irradiation, 373
Island growth, 478
Isomer shift, 228
Isotope, 110, 228, 285, 356
  effect on phonon, 120

J
Jahn–Teller effect, 82, 241, 242
  dynamic, 242
Joule heating, 286, 289
Junction
  abrupt, 626
  deep, 647
  graded, 767
  hyperabrupt, 659
  multiple, 714
  one-sided, 626, 632

K
Kane model, 870
Keating criterion, 135
Kick-out mechanism, 88
Kink, 390
Kirchhoff's law, 806
Klein paradox, 536

k · **p** theory, 179, 869–872
Kramer's degeneracy, *see* degeneracy, Kramer's
Kramers–Kronig relation, 292, 299, 851–852, 854
Kronig–Penney model, 156, 169, 424, 865

L
Lambert–Beer's law, 295, 684
Lamé's constant, 136
Landau level, *see* level, Landau
Large scale integration, 821
Laser, 11, 255, 399, 749–785
  cascade, 750
  condition
    thermodynamic, 350
  diode, 750
  double heterostructure, 11, 12
  edge emitting, 750
  horizontal cavity surface-emitting, 778
  hot hole, 784
  lead salt, 771
  modulation, 774
  monochromatic, 770
  monomode, 770
  multisection, 773
  optically pumped, 782
  output power, 766
  quantum cascade, 783
  surface-emitting, 750, 778
  tunable, 771, 780
  two-section, 773
  vertical-cavity surface-emitting, 562, 779
  zero-threshold, 546
Lattice
  1D Bravais, 112
  2D Bravais, 46
  3D Bravais, 46
  body-centered cubic, 48
  Bravais, 42, 46, 386
  constant, 77, 112, 117, 141, 168
  empty, 155
  expansion, 174
  face-centered cubic, 48
  hexagonally close packed, 49
  ionic, 275
  match, 172
  mismatch, 132, 141, 412
  period, 460
  point, 42, 43, 113
  reciprocal, 66–70, 154
  relaxation, 82, 241, 246
  simple cubic, 47
  site, 228
  temperature, 268, 322
  vibration, 111–132
Lattice matched, 412
Laughlin theory, 459
Layer
  active, 749
  amorphous, 406
  depletion, *see* depletion layer
  inversion, 452, 455, 803
  nucleation, 405
  semi-insulating, 236
  sequence, 417
  space-charge, 586
  spacer, 833
  wetting, 404
Level
  deep, 197, 211, 235, 326, 660
  Landau, 449, 451, 452, 455, 459, 508
  midgap, 373, 377
  occupancy, 242
  quantized, 419
Lifetime, 349, 374, 431
  carrier, 672
  minority carrier, 350, 373, 655
  phosphorescent, 525
  photon, 764
  triplet, 522
Lift-off, 143
Light-emitting diode, *see* diode
Linear chain
  diatomic, 115
  monoatomic, 112
Lithography, 477
Localization, 457
  light, 558
Löwdin perturbation theory, 870
Loss, 762
  internal, 762
  mirror, 762
LST relation, *see* Lyddane–Sachs–Teller relation
Luminance, 726
Luminescence, 358, 522
  decay, 568
  impurity, 733
  mechanism, 732
  negative, 348
Luminosity, 726
LUMO, 30, 519, 522
Luttinger Hamiltonian, 872

Luttinger parameter, *see* parameter, Luttinger
Lyddane–Sachs–Teller relation, 132, 337

M
Macfarlane–Roberts plot, 305
Madelung constant, 33
Magnetic moment, 507
Magnetoresistance, 451, 452
Magnetotransport, 441, 460
Mask, 410
Mass, 112, 117
  anisotropic, 324, 441
  carrier, 449
  density of states, 181, 200, 205
  effective, 159, 178, 182, 211, 259, 371, 419, 420
  effective conductivity, 257
  electron, 180
  hole, 185
  isotropic, 181
  longitudinal, 181
  nonparabolicity, 181
  polaron, 275
  reduced, 338, 437
  strain effect, 193
  transverse, 181, 420
Mass-action law, 206
Matrix element, 297, 298, 304, 347, 421, 477
  bulk momentum, 179
  dipole, 675
  k-dependence, 179
  momentum, 179, 298, 869
Matthiesen rule, 259
Maxwell's equations, 293, 554
Mean free path, 257, 460
Memory, 787, 821
  flash, 830
  nonvolatile, 830
Metalorganic, 400, 525
Microcavity, 322, 559, 568, 765, 780
Microdisc, 565
Micropillar, 568
Microscopy
  scanning tunneling, 219, 222
  secondary electron, 464
Midgap level, 323
Miller indices, 67
  wurtzite, 68
Miniband, 424
Minigap, 424
Mirror operation, 44
Miscibility gap, 72
Mobility, 258, 425, 449, 460, 598, 800, 837, 879
  channel, 835
  edge, 431, 457
  extremely high, 399
  high, 453
  high electron, 259
  hole, 259
  metal, 259
  negative differential, 809
  optical carrier, 448
  surface, 409
  temperature dependence, 264
Mobility edge, 280
Mode
  chaotic, 569
  defect, 558, 561
  evanescent, 569
  gap, 127
  hopping, 772
  localized vibrational, 84, 124–253
  longitudinal phonon plasmon, 340
  normal, 38
  optical, 116, 545
  out of plane, 529
  single longitudinal, 769
  soft phonon, 492
  spectrum, 769
  stretching, 522
  volume, 568
  whispering gallery, 566, 568, 571
Modulation
  large-signal, 775
  pulsewidth, 745
  small-signal, 777
Momentum
  angular, 302, 420, 471
  conservation, 298
  cyrstal, 177
  matrix element, *see* matrix element
  orbital, 324
  orbital angular, 184
  total angular, 184
Monolayer, 390, 400, 417, 431
Moore's law, 821
Mott transition, *see* transition, Mott
Multiexciton, 483
Multiferroics, 489
Multiplication
  region, 695

N
Nanobelt, 467
Nanohelix, 143
Nanolaser, 466
Nanoscroll, 143, 145
Nanostructure, 461–829
Nanotube, 143
  BN, 541
  carbon, 536
  metallic, 540
Nanowhisker, 465, 469
Negative-U center, 242
Neutrality, 207
  charge, *see* charge, neutrality
  condition, 208, 218, 224
  constraint, 236
Newton's law, 177
Nobel Prize, 12, 15, 399, 663
Noise, 881
  $1/f$, 886
  equivalent power, 673, 681
  excess, 693
  Flicker, 886
  generation–recombination, 673
  generation-recombination, 888
  shot, 683, 887
  source, 683
  thermal, 673, 683, 695, 885
  white, 886
Nonequilibrium, 234, 343
  thermodynamic, 235
Nonlinear optics, 323, 496, 546
Nonparabolicity, 439
Nonpolar, 262, 501
Nucleation, 405, 838
  dislocation, 147
Number
  atomic order, 82, 358
  complexion, 84
  order number, 54

O
Ohm's law, 257
Orbital, 28
  antibinding, 29
  antisymmetric, 26
  bonding, 29
  highest populated, 30
  lowest unoccupied, 30
  overlap, 111
  symmetric, 26
Order
  antiferromagnetic long-range, 508
  long-range, 51
  spontaneous magnetic, 505
  stacking, 56, 64
Ordering, 77
  CuAu, 80
  CuPt, 77
  in-plane, 481
Orientation, 253
  random in-plane, 406
Oscillation
  Bloch, 256
  Franz–Keldysh, 439
  Shubnikov–de Haas, 451, 452
  Weiss, 460
Oscillator, 10, 269
  harmonic, 38, 111, 449, 471, 475, 853
  local, 655
  master, 786
  strength, 127, 299, 314, 337, 675, 853–858
    exciton, *see* exciton
  voltage-controlled, 658
Ostwald ripening, 478
Overgrowth
  cleaved-edge, 465
  epitaxial lateral, 410
Overlap
  geometrical, 696
Oxide
  aperture, 780
  high-k dielectrics, 820
  transparent conductive, 4, 575–579

P
Parabola, 181
Paramagnetic ion, 505
Parameter
  s, 35, 105
  Heisenberg exchange, 506
  Huang–Rhys, 364
  Luttinger, 185
Partition sum, 859
Passivation, 52
Peak-to-valley ratio, 666
Peltier effect, 255, 289
Periodic system, 21, 26, 211
Phase, 96, 106, 117, 773
  Bloch, 549
  change, 831
  cubic, 490
  diagram, 74, 321

factor, 548, 865
ferroelectric, 490, 494
metastable, 65, 163
opposite, 129
ordered, 110
orthorhombic, 491, 496
paraelectric, 490, 493
paramagnetic, 508
rhombohedral, 496
separation, 74
shift, 459
spin glass, 508
tetragonal, 167, 490, 496
transition, 51, 56, 172, 321, 490, 493
first-order, 495
second-order, 493
trigonal, 491
Phonon, 39, 112, 123, 304
absorption, 376
acoustic, 117, 261, 315
alloy, 127
one-mode, 127
two-mode, 127
average temperature, 176
Bose–Einstein model, 176
broadening, 313
dispersion, 118, 176
emission, 368
energy, 433
infrared active, 296
LA, 117
LO, 117, 132, 262, 276
long wavelength, 132
optical, 117, 176, 337, 341
emission, 269
replica, 522
soft, 492
TA, 117
TO, 132
Phonon-drag, 289, 880
Phosphor, 727, 730, 743
Phosphorescence, 522
Photocatalysis, 385, 669
Photoconductivity, 527
persistent, 246
Photoconductor, 670, 674
Photodetector, 236, 255
FIR, 324
quantum well intersubband, 675
traveling wave, 695
Photodiode, 680–708
array, 705
avalanche, 680, 692–695, 706
bipolar, 681
metal–semiconductor–metal, 686
pin, 684
stacking, 706
Photometry, 725
Photon
counting, 645
Photoresistor, 583
Piezoelectricity, 489, 498
Pinch-off, 804, 806, 809, 814
Planar technology, 10, 583, 677, 779, 823
Planck's law, 347
Plane
high index, 390
Plasma, 727
frequency, 331, 341, 448
Plasmon, 332, 393
Plastic flow, 147
Pnictide, 57
Point
critical, 299, 300, 308
crossover, 245
defect, *see* defect, point
Dirac, 530, 534, 540
F, 166
Γ, 70, 114, 116, 117, 162, 167, 305
group, *see* group, point
K, 72
L, 72, 163, 172
M, 167
saddle, 300
X, 72, 117, 163, 172
Poisson
equation, 424, 589, 620, 628, 805
ratio, 136
statistics, 883
Polariton, 322, 339
lower branch, 315
surface plasmon, 393
Polarizability, 338
Polarization, 131, 294, 314, 421, 675, 852
circular, 302, 512
electric, 262, 489
ferroelectric, 62
light, 323
p, 294
s, 294
spin, 509
spontaneous, 77, 489
TE, 422
time-dependent, 129
TM, 422
Polaron, 275, 521, 732

small, 277
Poling, 496, 617
periodic, 496
Polyhedra, 43
Polymer, 30
chain, 519
Polytypism, 64
Poole–Frenkel effect, 375
Population, 371
inverted, 350
Position sensing detector, 686
Potential
asymmetric, 111, 675
atomic interaction, 111
built-in, 587
chemical, 237, 862
confinement, 419
Coulomb, 156
crystal, 156
distribution, 593
double well, 169
external, 591
fluctuation, 431
hard wall, 865
harmonic, 112
hydrostatic deformation, 191, 261
inversion surface, 815
ion core, 211
lateral ∼ well, 703
Lennard–Jones, 37
long-range Coulomb, 375
minimum, 429
optical deformation, 192
periodic, 6, 153, 211
piezoelectric, 503
pure Coulomb, 324
screened Coulomb, 260
short range, 211
triangular, 424
two-dimensional well, 468
well
three-dimensional, 471
Power
maximum, 712
noise, 885
output, 711
spectral, 727
thermoelectric, 286
total, 725
Pre-breakdown, 644
Precipitate, 226
Precursor, 400
Pressure, 84, 132
high, 14
hydrostatic, 56
partial, 228
vapor, 478
Process
activation, 241
causal, 851
Processing temperature, 576
Propagation
direction, 422
Punch-through, 798
Purcell effect, 562, 567
Purity, 203
Pyroelectricity, 489
Pyrolysis, 400

Q

Quadrupole, 503, 569
Quality factor, *see* factor, quality
Quantum
box, *see* quantum, dot
dot, 139, 201, 431, 461, 471–675
charge tunable, 484
cleaved-edge overgrowth, 465
cubic, 472
lens-shape, 477
pyramidal, 473
self-assembled, 476, 477
spherical, 472
stack, 480
efficiency, *see* efficiency, quantum
electrodynamics, 562
magnetic flux, 457, 458
well, 200, 322, 417, 439, 801
coupled, 424
energy level, 417
multiple, 417
sidewall, 463
vertical, 464
wire, 139, 200, 461, 462
cleaved-edge overgrowth, 465
T-shaped, 465
V-groove, 462
Quantum dot, *see* quantum, dot
Quantum Hall effect, *see* Hall, effect
Quantum statistics, 859–863
Quantum wire, *see* quantum, wire
Quarter-Wave stack, 550
Quasicrystal, 44
Quasi-Fermi level, *see* Fermi, level

R
Rabi frequency, 562, 563
Radiance, 725
Radiometry, 725
Radius
  self-limited, 463
Random bit pattern, 776
Random walk, 282
Rashba effect, 512
Rate
  Auger recombination, 368
  capture, 371
  emission, 371, 375
  escape, 677, 764
  generation, 284, 343, 671
  growth, 400
  net recombination, 347
  pulling, 401
  recombination, 284, 348
  thermal Auger generation, 368
  thermal generation, 348
  tunneling, 377
Recombination, 223, 343–381, 512, 522, 598, 623, 655, 669, 801
  Auger, 368
  band–band, 344
  band–impurity, 370
  bimolecular, 348
  bound-exciton, 351
  center, 373, 661
  current, 377, 735
  donor–acceptor pair, 366
  dynamics, 349
  excitons, 351
  free-exciton, 351
  lineshape, 429
  nonradiative, 105, 235, 793
  quantum well, 427
  radiative, 399, 523, 733
  rate, *see* rate, recombination
  spectrum, 483
  spontaneous, 344
  surface, 377
  velocity, 608
Rectification, 2, 7, 653
Rectifier
  metal–semiconductor, 7
  point contact, 7
Reflectance, 295, 338
Reflection, 291, 293–296, 773
  anisotropy spectroscopy, 400
  distributed, 770
  low, 685
  total, 294, 569, 735
Region
  space-charge, 586
Relaxation
  carrier, 343
  plastic, 146, 480
  time-constant, *see* time-constant, relaxation
Resharpening, 463
Resistance
  negative differential, 666, 667
  serial, 714
  shunt, 714
Resistivity, 257
  high, 247
  negative differential, 269
  transverse, 455
Resonator
  deformed, 567, 568
  microscopic, 565
  spiral, 569
Responsivity, 683
Reststrahlenbande, 338, 340
Richardson constant, *see* constant, Richardson
Rocksalt structure, *see* structure
Rotation, 44, 106, 261, 843
  general, 844
  improper, 44
Roughness, 403, 569

S
Saturation, 727
  electron density, 217
Scalar, 844
Scattering, 441, 456, 875, 880
  deformation potential, 261
  elastic ∼ process, 261
  grain boundary, 264
  hot-carrier, 784
  impurity, 834
  inelastic, 262
  intervalley, 811
  ionized impurity, 260, 425
  matrix, 548
  microscopic process, 259
  orbit, 460
  phonon, 328
  piezoelectric potential, 262
  polar optical, 262
  process, 256
  Rutherford, 260

spin, 512
theory, 546
time, 445
Schönfließ notation, 44
Schottky
barrier, *see* barrier, Schottky
Schottky effect, 593
Schottky–Mott model, 584
Schrödinger equation, 154, 157, 418, 424, 438, 461, 547, 869
Scintillation detector, 731
Scintillator, 727
Scrolling, 143
direction, 145
Second-harmonic generation, 496, 655
Seebeck effect, 255, 286
Selection rule, 302, 421
optical, 422
polarization, 675
Self-assembly, 478
Self-consistent, 424
Semiconductor, 4
alloy, 170
amorphous, 41, 51–52, 173, 196, 252, 308, 837
compound, 9, 56, 163
diluted magnetic, 505
doped, 207
elemental, 163, 168
ferroelectric, 492
history, 1
II–VI, 168, 262
III–V, 168
indirect, 305, 370
inhomogeneous, 255
intrinsic, 207
lead salts, 174
magnetic, 62, 505
nonpolar, 262
organic, 29, 515–613
oxide, 400
polarized, 489
polycrystalline, 41, 51
properties, 18
semi-insulating, 247
small band gap, 259
small-gap, 207
wide band gap, 208, 505, 509, 612
wide-gap, 207
Semipolar, 411, 501
Shell structure, *see* structure, shell
Shockley–Read–Hall kinetics, 348, 370
Shubnikov–de Haas effect, 452
Side-mode suppression ratio, 769
Signal-to-noise ratio, 673, 683, 695, 881
Singularity, 141, 195, 300
Fermi-edge, 428
van-Hove, 195, 299
Snapback time, 661
Snellius' law, 294
Sol-gel process, 478
Solar
cell, 2, 255, 707–723
spectrum, 708
Solid-state multiplier, 695
Solubility, 76, 93, 94, 233, 248
Source, 802
Space group, *see* group, space
Space-charge region, 589, 619, 626
capacitance, 595, 623
Spectroscopy
deep level transient, 246
Mössbauer, 228
Raman, 433
Spectrum
noise, 884
Spin, 160, 184, 302, 507
alignment, 512
glass, 508
LED, 512
nuclear, 214, 356
polarization, 397
rotation, 512
splitting, *see* splitting, spin
total, 313
transistor, 511, 512
Spinodal, 75
decomposition, 75
Spin-orbit interaction, 160, 396, 870
Spintronics, 505, 511
Splitting, 241, 356
crystal field, 188
Rabi, 563
spin, 160, 441, 455
valley-orbit, 356
Zeeman, 325, 356
Stacking, 49
vertical, 463, 480
Stacking fault, 98, 103
energy, 105
extrinsic, 103
intrinsic, 103
Stark effect
quantum confined, 439, 501, 774
second-order, 439
State

dark, 525
edge, 457
extended, 457
localized, 457
macroscopic quantum, 457
midgap surface, 588
triplet, 525
Step, 390
bunch, 390
monoatomic, 106
surface, 402
Stirling's formula, 85
Stoichiometry, 72, 229
Stokes shift, 430, 603
Stoney's formula, 143
Strain, 134, 262, 412, 675
bending, 144
biaxial, 137, 141, 192
compressive, 137, 192
distribution, 140, 473
energy, 133, 145, 149, 480
homogeneous, 190
hydrostatic, 191, 193
inhomogeneous, 190, 321, 872
in-plane, 501
large, 190
management, 143
microscopic, 58
misfit, 148
plastic relaxation, 96
random, 356
relaxation, 412, 480
shear, 137, 192, 498
small, 190
tensile, 137, 192
tensile surface, 480
tensor, 261
three-dimensional, 139, 473
Streaming motion, 785
Stress, 134, 253, 412
external, 489
–strain relation, 133, 134, 136
superposition, 140
uniaxial, 501
Structure
band, *see* band structure
chalcopyrite, 57
CsCl, 53
delafossite, 61
diamond, 54, 150, 171
dielectric, 545
field-ring, 647
fluorite, 61
interdigitated, 691
NiAs, 62
orthorhombic, 33
periodically poled, 496
perovskite, 61, 490
pseudomorphic, 412
rocksalt, 33, 53, 63, 172
shell, 475, 584
spinel, 59
tetragonal, 33
wurtzite, 56, 172
zincblende, 33, 55, 106, 150, 169, 172
Subband, 200, 451, 675
edge, 420, 421, 424
Sublattice, 55, 367, 505
anion, 56
cation, 56
Substrate
bending, 137, 141
compliant, 412
curved, 403
hetero-, 405
homo-, 405
patterned, 410
polished, 403
rotation, 400
transparent, 738
Sun, 708
Superconductivity, 39
Superlattice, 72, 256, 417, 420, 424–425, 750
buffer layer, 403
isotope, 433
Surface, 385–397
energy, 387
index, 588
passiviation, 385
phonon, 392
plasmon, 393
reconstruction, 388
resonance, 395
state, 395
vicinal, 390
Susceptibility, 852
electric, 853
magnetic, 451
nonlinear third-order electric dipole, 323
Switch, 236
Symmetry
inversion, 160
mirror, 439
reduction, 241
surface, 386

tetrahedral, 324
time reversal, 159, 396
trigonal, 245

T
Tail
carrier distribution, 591
exponential, 437
states, 451, 457
Urbach, 307, *see also* Urbach tail
Taylor series, 283
Temperature
blackbody, 673
characteristic, 769
Curie, 493, 505, 509
Curie–Weiss, 496
Debye, 262
difference, 289
electron, 268
gradient, 255, 284
lattice, 268, 886
local, 235
Tensor, 843, 845
conductivity, 453
dielectric
magneto-optic, 449
dielectric function, 292
effective-mass, 178, 213, 873
nonlinear third-order electric dipole susceptibility, 323
resistivity, 454
Terrace, 390
Theory
Drude, *see* Drude theory
effective mass, *see* effective mass, theory
Laughlin, *see* Laughlin theory
perturbation, 241
time-dependent perturbation, 297
Thermal instability, 644
Thermalization, 430, 603
incomplete, 431
Thermopower, 286, 879
Thickness
barrier, 424, 867
critical, 145–466
film, 148
oxide, 703
quantum well, 417, 429
Thomson heating, 289
Threshold, 763
Tilt, 105
Time constant, 343, 349, 595, 623, 703, 801
decay, 727
LO phonon emission, 269
RC, 696
relaxation, 256, 259
reorientation, 242
Transconductance, 799, 808, 816
differential, 799
Transistor, 7, 255, 787–841
bipolar, 10, 788–802
effect, 14
field effect, 5, 7, 802
heterobipolar, 399, 800
high electron mobility, 399, 833
JFET, 10
junction field effect, 7, 804
light-emitting, 801
MESFET, 7, 12, 804
MOSFET, 5, 11, 455, 812
organic, 839
planar, 11
point contact, 8
spin, *see* spin, transistor
thin film, 838
Transit time, 685, 811, 812
Transition
band–band, 299
dipole, 314
direct, 300
displacement, 492
donor–acceptor pair, 366, 733
forbidden, 311
indirect, 304
intersubband, 783
metal, 507
metal–insulator, 230, 233, 265, 280
Mott, 431
optical, 179, 188, 296, 324, 344, 482, 567
probability, 297
Transmission, 291
Transparency, 763
Transport, 255–289, 473
ballistic, 256, 598, 784
Boltzmann theory, 875
charge, 255
coupled heat and charge, 286
diode current, 598
heat, 284
heat energy, 255
high frequency, 275
high-field, 268
hopping, 278, 447, 521, 522
ionic, 281
low-Field, 258

magneto-, *see* magnetotransport
Trap, 375
filled, 371
multilevel, 374
surface, 587
Trion, 319, 483, 485
Tuning range, 773
Tunneling, 14, 419, 424, 429, 598, 608, 645, 648, 677
assisted, 376
current, 474
direct, 376
Fowler–Nordheim, 830
inelastic, 666
phonon-assisted, 377
photon-assisted, 377, 439
rate, *see* rate, tunneling
Zener, 7
Turn-on delay time, 775
Twin, 103
boundary, 103
lamella, 103
Twist, 105, 468
Two-electron satellite, 355
Two-photon process, 297

U
Umklapp process, 124
Unit cell, *see* cell, unit
volume, 179
Urbach tail, 307

V
Vacancy, 81, 229, 236, 244, 247
Vacuum, 465
level, 584
tube, 2, 5, 583, 787, 821
ultrahigh, 400
Valence band, *see* band, valence
Valley
current, 477, 664
L, 270
X, 324
Van-der-Pauw geometry, 442
Van-Hove singularity, *see* singularity, van-Hove
Varactor, 658
Variable range hopping, 278
Variance, 882
Varshi's formula, 175
Vector
antiphase, 106
Burger's, *see* Burger's vector
displacement, 133
in-plane wave, 421
line, 96
potential, 140, 297
reciprocal lattice, 156
translation, 42
wave, 158
Vegard's law, 77
Velocity
average carrier, 256
drift saturation, 269, 685, 807
effective diffusion, 608
group, 114, 177, 547
light, 329
match, 696
maximum, 257
maximum drift, 269
mismatch, 697
overshoot, 271
phase, 114
sound, 114, 262
surface recombination, 377, 380
thermal, 371, 600
Vernier effect, 773
Void, 82
Voltage
bias, 594
breakdown, 646, 656, 690, 801
built-in, 596, 626
diffusion, 587
flat-band, 625, 689
gain, 799
gate, *see* gate, voltage
maximum reverse, 645
open-circuit, 711
pinch-off, 804
reach-through, 688
reference, 658
regulator, 656
threshold, 816
turn-on, 801
Vortex, 459

W
Wafer, 400
bonding, 433, 714, 738, 740
breakage, 151
diameter, 402
edge, 151
epiready, 403
flat, 401

Wannier's theorem, 873
Warping, 185, 220
Wave
  acoustic, 114
  compression, 261
  electromagnetic, 328
  equation, 156, 328, 546
  evanescent, 394
  longitudinal, 112
  plane, 114, 154, 437
  shear, 261, 262
  sound, 114
  standing, 114, 158
  transverse, 112
  traveling, 695
  vector, 329
Wavefunction
  d, 223
  many-electron, 459
  overlap, 424, 461, 477
  strongly localized, 235
Waveguide, 552, 558, 561, 695, 758, 763, 771
  plasmon, 332
Weiss oscillation, *see* oscillation, Weiss
Well capacity, 701
Wiedemann–Franz law, 284
Wiener-Khintchine theorem, 884
Work function, 584, 587, 614, 616
Wurtzite structure, *see* structure

Y
Young's modulus, 136

Z
Zincblende, 160
Zincblende structure, *see* structure
Zone
  boundary, 116, 157, 256, 867
    vicinity, 159
  Brillouin, *see* Brillouin zone
  reduced scheme, 154
  scheme, 154
ZT-value, 289