SPRINGER NATURE
Reference

Karl W. Böer
Udo W. Pohl

Semiconductor Physics

Second Edition

Springer

Semiconductor Physics

Karl W. Böer • Udo W. Pohl

Semiconductor Physics

Second Edition

With 721 Figures and 100 Tables

Springer

Karl W. Böer
Naples, FL, USA

Udo W. Pohl
Institut für Festkörperphysik, EW5-1
Technische Universität Berlin
Berlin, Germany

ISBN 978-3-031-18285-3 ISBN 978-3-031-18286-0 (eBook)
https://doi.org/10.1007/978-3-031-18286-0

1st edition: © Springer International Publishing AG 2018
© Springer Nature Switzerland AG 2023
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, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

The authors would like to thank De Gruyter; John Wiley and Sons; the American Physical Society; Elsevier; AIP Publishing; the American Chemical Society; Plenum Press; Ames Laboratory and Department of Physics and Astronomy; AIP Publishing LLC; IOP Publishing; Taylor & Francis; Dover Publications; Nature; The Physical Society of Japan; Wiley-VCH Verlag GmbH & Co. KGaA; Chelsea Dielectric Press; Technical University of Berlin; SPIE publications; Annual Reviews Publishing; American Vacuum Society; Vieweg; World Scientific Publishing; The Minerals, Metals & Materials Society; the Royal Society of Chemistry; and the Optical Society of America, for giving the permission to use various figures from their publications.

Foreword

Materials donate their names to the great epochs of human civilizations. The Stone Age with its heavy, huge monuments, the Bronze Age with the beauty of carefully designed tools, finally the Iron Age with all the useful splendor of machinery – they all furnish the visible proof. Now, however, we live in the Age of Semiconductors. Strange and modest are their first two syllables: "semi" – not exactly convincing. And yet, these materials are essentially invisible, truly small, hidden, yet most important helpers in our activities every day.

Karl Boer was himself a major investigator and proponent for this new class of materials. He contributed strongly in a variety of topics for these novel members of condensed matter. He used to live in the western part of divided Berlin, but worked most efficiently in its eastern part, directed research, demonstrated applications, and actively assisted in spreading the knowledge of science and technology for all international communities. Later, in the United States, he became also one of the most successful pioneers of solar energy conversion – with semiconductors, of course.

His handbook documented Boer's broad range of personal knowledge and experience. All the readers had thus a wonderfully rich collection of the multitude of phenomena, very well explained and carefully presented. It is now a stroke of good fortune for all those users that Udo Pohl, who also became a Berliner, took it upon himself to amplify and modernize the handbook, to critically screen the contents. No question: the bulwark of semiconductor knowledge lives on!

Hans J. Queisser

Preface to the Second Edition

The first edition of *Semiconductor Physics* was published in spring 2018, in time for Karl W. Böer's 92nd birthday and only a month before he died. The book has been well received by the scientific community. Semiconductor devices are ubiquitous in our daily life, and today the global semiconductor industry is a key driver for economic growth. Recent discoveries in semiconductor physics established new and intriguing fields, which will open new scopes for applications in the future. One such field refers to the electronic states in a topological insulator, which recently led to a new classification scheme for electronic phases in condensed matter and in principle enables dissipationless surface conductance. Another field is quantum optics with the realization of photonic qubits based on excitonic states; rapid progress presently opens prospects for realizing quantum communication and quantum computing. Much of these developments are intimately related to the basic principles of semiconductors physics; some of the new topics are addressed in the second edition, keeping the style and the conciseness of the presentation. In addition, the initial text is thoroughly amended and updated.

Berlin *Udo W. Pohl*
January 2023

Preface to the First Edition

The roots of this book can be traced back to the 1950s when one of us (KWB) started giving lectures on a young and intriguing emerging field – the physics of semiconductors. Semiconductor technology was in its infancy at that time, and the other of us (UWP) was just born. The accumulated and steadily updated lecture notes provided a broad background on semiconductor physics, and eventually – in 1990 – led to the first edition of the *Survey of Semiconductor Physics*. The textbook became a helpful source for generations of students of physics and electrical engineering. This applies as well for the younger of us two, who appreciated the concise treatment of the new book for preparing his lectures held on semiconductor physics and epitaxy.

Now, a quarter of a century later, semiconductor physics is still young and intriguing. New fields and materials joined the established topics, such as structures employing quantum effects and designed light-matter interaction, or nitride and organic semiconductors. Still their study bases upon the principles learned in the early lectures.

Our common roots in Berlin and our pleasure on teaching brought us together – by the amicable encouragement of Claus Ascheron from Springer. In this book we like to keep the conciseness of the presentation also given in the *Survey* and organize new topics in the same style. We had to select the material for keeping the volume to a manageable size; furthermore, like any book covering a large spectrum of subjects, the treatment lacks the depth of more focused texts. We apologize to the reader who may not find his own specialized subject represented in sufficient detail. This applies also to the references, which, in a steadily developing field, may not remain best choices; they essentially represent examples, helping to familiarize with the field and find more recent literature.

In the course of writing Dieter Palme steadily supported us by furnishing hundreds of illustrations for this book; we are much obliged for his precise work and critical comments for graphical representations. We appreciate being granted permission by publishers and authors for using illustrations from other published source, which were adapted and redrawn for a consistent presentation. Last but not least, we gratefully acknowledge Hans-Joachim Queisser for writing his concise foreword.

Naples — *Karl W. Böer*
Berlin — *Udo W. Pohl*
January 2018

Contents

Volume 1

Part I Growth, Bonding and Structure 1

1 Properties and Growth of Semiconductors 3

2 Crystal Bonding 35

3 The Structure of Semiconductors 59

Part II Phonons 111

4 Elasticity and Phonons 113

5 Phonon-Induced Thermal Properties 157

Part III Energy Bands 191

6 The Origin of Band Structure 193

7 Quantum Mechanics of Electrons in Crystals 219

8 Bands and Bandgaps in Solids 257

9 Magnetic Semiconductors 319

Part IV Photons 367

10 Interaction of Light with Solids 369

11 Photon–Phonon Interaction 429

12 Photon–Free-Electron Interaction 467

13 Band-to-Band Transitions 499

14 Excitons 529

Volume 2

Part V Defects **593**

15 Crystal Defects 595

16 Crystal Interfaces 649

17 Optical Properties of Defects 703

18 Shallow-Level Centers 755

19 Deep-Level Centers 811

20 Defects in Amorphous and Organic Semiconductors 861

Part VI Transport **897**

21 Equilibrium Statistics of Carriers 899

22 Carrier-Transport Equations 933

23 Carrier Scattering at Low Electric Fields 985

24 Carrier Scattering at High Electric Fields 1023

25 Carriers in Magnetic Fields and Temperature Gradients 1055

26 Superconductivity 1087

27 Carrier Transport in Low-Dimensional Semiconductors 1113

28 Carrier Transport Induced and Controlled by Defects 1165

Part VII Generation - Recombination **1203**

29 Carrier Generation 1205

30 Carrier Recombination and Noise 1241

31 Photoconductivity 1299

32 Dynamic Processes 1325

Appendix: High-Field Domains 1381

Index 1403

About the Authors

Dr. mult. Karl Wolfgang Böer was distinguished professor of physics and solar energy, and fellow of AAAS, APS, ASES, and IEEE. Over five decades, he gave classes in solid state and semiconductor physics at the Humboldt University and the University of Delaware and gave keynote addresses at numerous international conferences. In his research, he was the first in measuring the Franz-Keldysh effect, discovered high-field domains, the condensation of excitons, and numerous properties of cadmium sulfide. He published 380 papers, 14 monographs, was editor of 23 international conference proceedings and obtained 18 patents. He created *Physica Status Solidi* in 1960 and the Series *Advances in Solar Energy* in 1978. He obtained highest honors from the University of Delaware, the American Solar Energy Society, and the International Solar Energy Society. The University of Delaware issued the prestigious solar energy award in his name that is given biannually to one of the world's leaders in solar energy research and industry. Karl Wolfgang Böer passed away aged 92 on April 18, 2018, shortly after the publication of the first edition of *Semiconductor Physics*.

Dr. rer. nat. Udo W. Pohl, professor of experimental physics, studied physics in Aachen and Berlin, Germany, and received his Ph.D. degree in 1988 from the Technical University of Berlin, where he was principal investigator in the Institute of Solid State Physics and retired end 2020. He is a member of the German Physical Society, participated in international conference committees, and has given keynote addresses at international conferences. Since the early 1990s, he has given lectures on semiconductor physics and epitaxy. In 2009,

he was appointed Adjunct Professor of Physics at the Technical University of Berlin. He is series editor of Materials Sciences, has authored over 200 journal articles and conference papers, the monograph *Epitaxy of Semiconductors*, 13 book contributions, and patents. His research interests include physics and epitaxy of semiconductor nanostructures and devices.

Karl W. Böer and Udo W. Pohl celebrating Karl's 91st birthday in March 2017 during the finalization of the first edition of *Semiconductor Physics*.

Part I

Growth, Bonding and Structure

Properties and Growth of Semiconductors

1

Karl W. Böer and Udo W. Pohl

Contents

1 Historic Development 4
2 Some General Properties of Semiconductors 6
2.1 Electrical Aspects 6
2.2 Structural Aspects 7
2.3 Chemical Aspects 8
3 Growth of Semiconductors 9
3.1 Driving Force and Nucleation 9
3.2 Growth of Bulk Single Crystals 18
3.3 Epitaxy of Layer Structures 24
4 Summary 31
References 31

Abstract

Semiconductor physics and devices have emerged from early studies on the conductivity of metal sulfides in the nineteenth century and experienced a strong progress since the middle of the twentieth century. This introductive chapter briefly highlights a couple of historic milestones and illustrates some general properties of semiconductors. Then the fabrication of semiconductors is described, pointing out the driving force of crystal growth, thermodynamics, and kinetics of nucleation and the occurrence of different growth modes. Various

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_1

methods for growing bulk single crystals from the liquid and the vapor phase are introduced, and the techniques of liquid-phase epitaxy, molecular-beam epitaxy, and metalorganic vapor-phase epitaxy for the fabrication of thin layers and sharp interfaces are pointed out.

Keywords

Bridgman growth · Crystal growth · Czochralski growth · Driving force of growth · Epitaxy · General properties of semiconductors · Growth kinetics · Growth modes · Nucleation · Liquid-phase epitaxy · Metalorganic vapor-phase epitaxy · Molecular-beam epitaxy · Nucleation · Semiconductor history · Supersaturation

1 Historic Development

A *semiconductor* is a solid with an electrical conductivity between that of a metal and an insulator. This conductivity is caused by *electronic particles* such as electrons, holes, or polarons, referred to later in this book as *carriers* (of charge) which are set free by ionization. Such ionization can be produced thermally, by light, other particles, or an electrical field. It involves only a small fraction of the total number of atoms, with a density of free electronic particles typically of an order of $10^{14} \ldots 10^{18}$ cm^{-3}, compared to atomic densities of 10^{22} cm^{-3}.

The *changes in electronic properties* controlled by external means – such as light, applied voltage, magnetic field, temperature, or mechanical pressure – make the semiconductor an interesting material for electronic and photonic devices, many of which have become familiar parts of our daily lives. Such devices include diodes, transistors, and integrated circuits in TV sets and computers: photosensors for imaging and the identification of goods in stores; a large variety of sensors for measuring temperature, weight, and magnetic fields; and all kinds of electro-optical displays. It is almost impossible to think of a modern appliance that does not contain a semiconducting device. Our cars, homes, and offices have become filled with such devices, ranging from smartphones and PCs for home banking, home office, Internet shopping, and mobile Internet to controls, copying machines, LEDs for general lighting, solar cells for energy harvesting, and imaging devices.

The first *semiconducting property* was reported by Seebeck for lead sulfide in 1822 and Faraday for silver sulfide in 1833. In 1851, Hittorf measured the semilogarithmic dependence of the conductivity on $1/T$ (temperature) in Ag_2S and Cu_2S. The first *application of a semiconductor as a device* was based on the observation of Braun (1874) that point contacts on some metal sulfides are rectifying; this became the well-known cat's whisker detector of the early twentieth century. Schuster noticed in 1874 that the contact of copper with copper oxide is rectifying; this observation became the basis for the copper-oxide rectifier introduced in 1926 by Grondahl. The discoveries of the photoconductivity of selenium by Smith (1873)

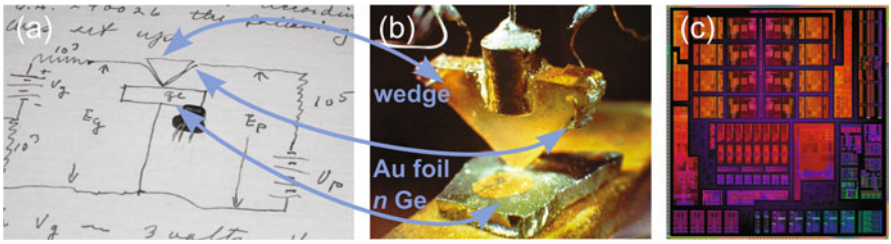


Fig. 1 (**a**) Principle and (**b**) realization of the first transistor, Bell Labs (1947). (**c**) AMD Ryzen 5000 desktop microprocessor chip comprising 10.7 billion 7nm transistors, AMD (2021)

and the photovoltaic effect in the same material by Adams and Day (1876) led to the first photocells as discussed by Bergmann (1931, 1934). The term *semiconductor*[1] was introduced much later by Königsberger and Weiss (1911). Only during the middle of the past century did other materials, such as Ge, Si, CdS, GaAs, and several other similar compounds gain great interest.

The invention of the transistor by Shockley, Bardeen, and Brattain in 1947 may be considered as the birth of modern electronics. This device consisted of a block of germanium with two very closely spaced gold contacts, realized by a gold foil attached over the point of a plastic triangle: the tip of the triangle sliced through the foil (Fig. 1a, b). The *n*-type Ge used had a thin *p*-type surface layer underneath the contacts, leading to a *pnp* transistor; a small change in the first (forward biased) contact current caused a greater change in the second (reversed biased) contact current. Nowadays, a billion transistors are integrated on a single advanced microprocessor chip as illustrated in Fig. 1c.

All initial studies involved crystalline materials; that is, their atoms are ordered in a three-dimensional periodic array. When the transistor was developed (ca. 1950), the first *amorphous semiconductors* (α-Se) became important for electrophotography (Xerox^{TM}). A few years later with the discovery of electronic switching (Ovshinsky 1968), another class of amorphous semiconductors, the chalcogenide glasses – and, more recently, with further development of solar cells, the amorphous Si alloys – gained substantial interest.

Some organic semiconductors and new artificial compounds, nanostructures and superlattices, more recently entered the vast inventory of semiconducting materials

[1]The word itself was rediscovered at this time. It was actually used much earlier (Ebert 1789) in approximately the correct context, and then again 62 years later by Bromme (1851). However, even after its more recent introduction, serious doubts were voiced as to whether even today's most prominent semiconductor, silicon, would not better be described as a metal (Wilson 1931).

of technical interest. Today, we experience that new compounds can be designed to exhibit specific desired properties, such as a large carrier mobility, appropriate optical absorption and emission, or higher-temperature superconductivity.

Semiconductors have sparked the beginning of a new material epoch.[2] Technology has evolved from the Stone Age into the age of semiconductors, materials, that are influencing culture and civilization to an unprecedented degree and enabled the advent of the information society.

2 Some General Properties of Semiconductors

2.1 Electrical Aspects

In several respects, the electrical aspects of semiconductors are unique and require a sophisticated knowledge of detail, which will be developed later in this book. Here, only a few general facts can be mentioned to emphasize some of the reasons which have sparked such enormous interest in this class of materials.

Semiconductors bridge a large gap of electrical conductivities between metals and insulators; however, more importantly, their electrical resistance[3] can change as a result of external forces, e.g., by applied voltage, magnetic field, light, mechanical stress, or a change in temperature. This is often accomplished by fabricating a *device* from such semiconductors when incorporating minute amounts of specific impurities (*dopants*) in a certain inhomogeneous pattern into the semiconductor. The electrical response of these devices makes them electronically *active*, as opposed to such passive elements as wires, insulators, or simple resistors made from thin metal layers.

Although the large field of semiconducting devices deals with the inhomogeneity aspect to a great extent, a thorough understanding of their operation requires a detailed analysis of the electronic and related properties of the homogeneous material. That is the topic of this book. The device aspects are dealt with in many reviews, most comprehensively by Sze and Ng (2007).

[2]This age was also termed the silicon age (Queisser 1985), in reference to the material now most widely used for semiconducting devices. Despite the great abundance in the earth's crust (27.5% surpassed only by oxygen with 50.5%, and followed by aluminum with 7.3% and iron with 3.4%), and its dominance as the material of choice in the semiconductor industry, other semiconductor materials (crystalline or amorphous) are now being identified which may show even greater potential in the future. The global semiconductor industry with a $440.4 billion market in 2020 and an expected growth to $551 billion in 2021 (source: World Semiconductor Trade Statistics, WSTS (2021)) is a key driver for economic growth, with an annual (long-term) average growth on the order of 10%.

[3]The resistance is used here rather than the material resistivity because of the inhomogeneity of the electronic transport through most of the devices.

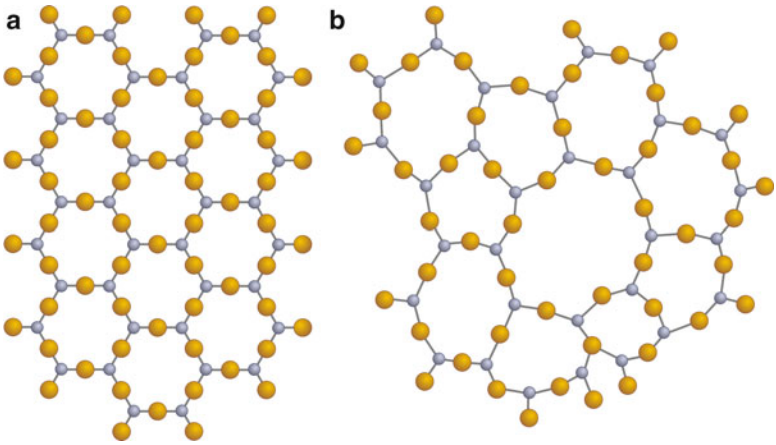


Fig. 2 (**a**) Crystalline and (**b**) amorphous states of an A_2B_3 compound – e.g., As_2S_3 (After Zachariasen 1932)

2.2 Structural Aspects

As are all crystals, semiconductors can be thought of as composed of *unit cells*, which describe the smallest atomic building blocks within an ordered crystalline lattice with translational symmetry in all directions. In a crystalline semiconductor, these unit cells are packed in a close-fitting, three-dimensional array that results in *long-range periodicity*.

In an amorphous semiconductor, there is a similar packing of statistically slightly *deformed unit cells* with variations in interatomic distance and bond angle. By constructing such a semiconductor from such deformed unit cells, long-range periodicity is lost, and a variation in the coordination number occurs. The *short-range order* of an amorphous semiconductor, however, is similar to the crystalline state of the same material. A two-dimensional illustration of maintaining short-range order while losing long-range order is given in Fig. 2.

The consequences of the microscopic structure for the physical properties are well known for crystalline semiconductors and comprise major parts of this book, while the understanding of amorphous semiconductors is less well developed. We have attempted to point out similarities and differences between the two states and to provide plausibility arguments where a rigorous treatment is still missing.

One of the reasons for a more manageable theoretical treatment of the crystalline state is its *periodicity*, which permits an analysis by replacing a many-body problem of astronomical extent (typically 10^{22} particles per cm^3) with a periodic repetition of one unit cell containing only a few relevant particles. Attempts to use a somewhat similar approach (e.g., a modified Bethe-lattice approach; see Sect. 3.2.1 in ▶ Chap. 3, "The Structure of Semiconductors") for the amorphous state have been relatively successful for theoretical prediction of some of the fundamental properties of amorphous semiconductors.

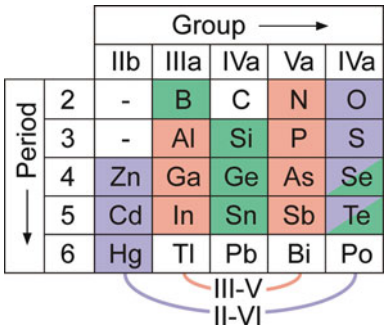


Fig. 3 Semiconductors in the periodic table. *Green* shadings signify elemental semiconductors and *red* and *blue* shadings mark III–V and II–VI compound semiconductors, respectively

Further discussion will be postponed until the elements necessary for a more sophisticated analysis are properly introduced.

2.3 Chemical Aspects

Only a small number of elements and simple compounds have semiconductive properties conducive for device development (cf. Fig. 3). Preferably, these materials are covalently bonded and are typically from group IV of the periodic system of elements (Si, Ge) or are compounds of group III with group V elements (GaAs, GaN) (see Mooser and Pearson (1956) and Sect. 1 in ▶ Chap. 2, "Crystal Bonding"). In these semiconductors, the atoms have a low number of nearest neighbors (*coordination number*). In tetrahedrally bonded crystals, the coordination number is 4. These semiconductors have *highly mobile carriers* which are desirable for many devices.

A number of important semiconducting properties of these devices are rather similar, regardless whether the material is in crystalline or amorphous form. This emphasizes the importance of the chemical aspect and justifies more attention to the subject, which will be provided in ▶ Chap. 2, "Crystal Bonding."

At closer inspection, however, almost every physical property of a semiconductor depends on the relative position of the atomic building blocks with respect to each other, i.e., on the *structure* of the semiconductor.

In the recently emerging *organic semiconductors*, the building blocks are *molecules* which are weakly bonded in organic crystals by van der Waals interaction (Forrest 2020). Organic crystals are usually not conductive. Conductivity occurs within organic molecules with a conjugated hydrocarbon structure, where single and double bonds between carbon atoms consecutively change; such a structure releases one electron per carbon atom being delocalized within the molecule. If the overlap of the electron wavefunctions of neighboring molecules is sufficiently large, band-like or hopping conductivity may occur on a macroscopic scale. Similar to their inorganic counterparts, the mobility of carriers in organic semiconductors depends sensitively on the relative arrangement of the molecular building blocks.

3 Growth of Semiconductors

Growth of a semiconductor requires nonequilibrium conditions to promote a transition from a more volatile to the intended solid phase. A process near equilibrium is usually described by thermodynamics in terms of macroscopic control parameters like temperature and pressure, while conditions far from equilibrium and on a short time scale may more appropriately be described by kinetic steps on an atomic scale with respective activation barriers. Basics of both approaches are pointed out in the following.

3.1 Driving Force and Nucleation

3.1.1 Equilibrium

We first consider a single-component system with an equilibrium of two phases like solid silicon in equilibrium with its own melt. The solid surface represents a boundary between the two phases, and the amount of substance dn leaving one phase equals the amount of substance entering the other phase, i.e., $-dn_{\mathrm{Si,liquid}} = +dn_{\mathrm{Si,solid}}$. In equilibrium, $dn_{\mathrm{Si,liquid}} = dn_{\mathrm{Si,solid}} = 0$. The *deviation* from equilibrium required to drive particles across the phase boundary toward the solid phase is obtained from the *chemical potential* μ. For the considered single-component system, $\mu(T,P) = G(T,P)/\,n$. The state function *Gibbs energy* $G(T,P)$ of the system is related to the internal energy $U(S,V)$ by[4] $G(T,P) = U + PV - TS$, and n is the amount of substance (usually expressed in units of mole). For two coexisting phases, μ is composed of the respective two parts, i.e., $\mu = \mu_{\mathrm{Si,liquid}} + \mu_{\mathrm{Si,solid}} = (G_{\mathrm{Si,liquid}} + G_{\mathrm{Si,solid}})/n$. At equilibrium, this sum is minimum, and $\mu_{\mathrm{Si,liquid}} = \mu_{\mathrm{Si,solid}}$. At a depart from equilibrium, $\mu_{\mathrm{Si,liquid}} \neq \mu_{\mathrm{Si,solid}}$ applies, and a tendency to restore the minimum exists.

In the *general case*, the system consists of N_c components instead of just one kind of particles. In such *multicomponent system*, more than two phases may coexist in equilibrium,[5] and the Gibbs energy depends also on the substance amounts $n_i, G = G(T, P, n_1, \ldots, n_{N_\mathrm{c}})$. The molar fraction of component i is $x_i = n_i/\,n$, and the sum of all N_c substance amounts yields n. The total chemical potential μ is composed of contributions μ_i of all components,

$$\mu = \frac{G}{n} = \mu_1 x_1 + \mu_2 x_2 + \ldots + \mu_{N_\mathrm{c}} x_{N_\mathrm{c}}, \tag{1}$$

where

[4] G is used instead of U, because control of temperature T and pressure P is more convenient than that of the parameters entropy S and volume V.

[5] The number of coexisting phases is specified by Gibbs phase rule.

$$\mu_i(T,P,x_1,\ldots,x_{N_c}) = \left(\frac{\partial G}{\partial n_i}\right)_{T,P,n_{j\neq i}}, \text{ and } x_i = \frac{n_i}{n}.$$

The minimum condition for G leads to the requirement that the chemical potentials μ_i of all N_c components are equal among each other in each of the N_p phases,

$$\mu_i(\sigma_1) = \mu_i(\sigma_2) = \ldots = \mu_i(N_p), i = 1 \ldots N_c. \text{ (equilibrium)} \quad (2)$$

The σ_j signify the coexisting phases (vapor, liquid, etc.) of component j. Equation 2 applies simultaneously for all N_c components of the system, yielding a set of N_c relations. The *compositions* of coexisting phases are generally *not* equal at equilibrium to fulfill these conditions, i.e., usually $x_1(\sigma_1) \neq x_1(\sigma_2) \neq \ldots \neq x_1(\sigma_{Np})$. In liquid-phase epitaxy of GaAs, e.g., a melt containing $x_{As,liquid} = 0.1$ (=10% of liquid substance moles) As dissolved in Ga may be in equilibrium with solid GaAs containing precisely $x_{As,solid} = 0.5$.

3.1.2 Driving Force

To induce a transition to the solid phase, the parameters temperature and pressure are controlled such that μ is larger in the nutrient phase. A measure for the deviation from equilibrium is the *supercooling* $\Delta T = T - T_e$ between the controlled temperature T and the (initial) equilibrium temperature T_e (melting temperature in a liquid–solid equilibrium). For a phase transition from the gas phase induced by a pressure variation, the *supersaturation* $\Delta P = P - P_e$ at constant temperature defines the depart from equilibrium, P_e being the saturation (or equilibrium) vapor pressure.

A phase transition induced by supercooling $\Delta T = T - T_e$ at constant pressure for a single-component system is illustrated in Fig. 4. Above the equilibrium temperature

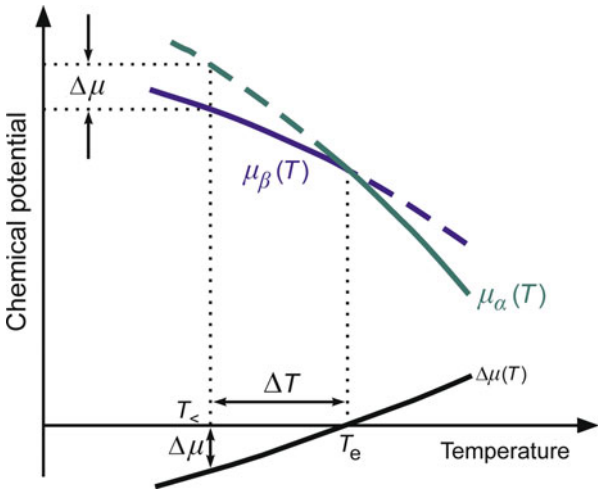


Fig. 4 Temperature dependence of the chemical potential μ of phases α and β at constant pressure. The growth affinity $\Delta\mu$ designates the difference of the chemical potentials μ_α and μ_β induced by a supercooling ΔT to a temperature $T_<$ below the equilibrium temperature T_e

T_e, the phase α is more stable than phase β, i.e., $\mu_\alpha(T) < \mu_\beta(T)$. Below T_e, however, $\mu_\beta(T) < \mu_\alpha(T)$ due to different slopes of the chemical potentials in the two phases and their intersection at T_e. Phases α and β exchange particles via the common phase boundary. Since $\mu_\beta(T) < \mu_\alpha(T)$ below T_e, the system can lower the total Gibbs energy $G = G_\alpha + G_\beta = n_\alpha \mu_\alpha + n_\beta \mu_\beta$ by transferring particles from phase α to phase β: the amount of substance n_β in the stable phase β increases on expense of the amount of substance n_α in the less stable phase α. The driving force for this phase transition is the difference in the chemical potential:

$$\Delta\mu \equiv \mu_\beta - \mu_\alpha \,(\text{nonequilibrium}). \tag{3}$$

$\Delta\,\mu$ is called *growth affinity* or *driving force for crystallization* if the stable phase β is a solid phase. $\Delta\,\mu$ is a negative quantity, albeit often chosen positive in literature. Figure 4 illustrates that an increase of supercooling $\Delta\,T$ raises the driving force $\Delta\,\mu$ and thereby enhances the rate of substance crossing the phase boundary.

The respective less stable phases indicated by dashed lines in Fig. 4 may sometimes be experimentally observed to some extent. Well-known examples are heating of a liquid above the boiling point (bumping) or supersaturated vapor in Wilson's cloud chamber. The effect is particularly pronounced for undercooled metal melts which may still be liquid some tens of degrees °C below the temperature of solidification.[6]

The driving force Eq. 3 may likewise be induced by a supersaturation $\Delta\,P = P - P_e$ at constant temperature. For solidification from the liquid phase, P must be increased if the liquid–solid boundary has a positive slope dP/dT and decreased otherwise. The driving force increases as $\Delta\,P$ is raised, similar to the dependence in supercooling discussed above.

3.1.3 Nucleation

Any growth starts from a nucleus or crystal seed. If the system was initially in a homogeneous phase, e.g., entirely liquid, the (new) solid phase is *not* spontaneously formed when nonequilibrium conditions are adjusted. Instead, the initial phase remains at first in a metastable state of undercooling (of a liquid below the freezing point) or supersaturation (of a vapor below the condensation point) – the states indicated by dashed lines in Fig. 4. The need to create an interface at the boundary to the new phase leads to a persistence of the no longer most stable phase. The formation of such an interface consumes some energy, related to the *surface energy* of each phase. The stable new phase is only formed at sufficient undercooling or supersaturation. *Homogeneous nucleation* occurs spontaneously and randomly in a homogeneous initial phase; this is generally not desired in crystal growth. Instead, *heterogeneous nucleation* induced by preferential *nucleation sites*, like a seed crystal or a substrate, is applied.

[6]The phenomenon should not be confused with the size effect of melting-point depression in nanoscale materials that originates from a large surface-to-volume ratio.

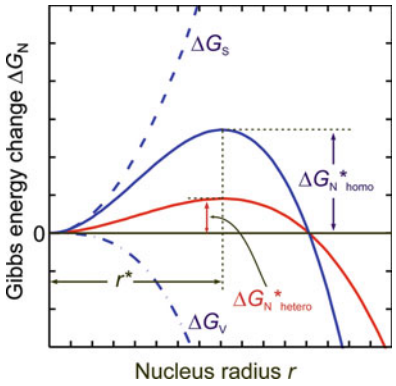


Fig. 5 Change of Gibbs energy ΔG_N (*solid blue line*) for creating an unstrained spherical nucleus with radius r in homogeneous nucleation. *Dashed* and *dash-dotted lines* are contributions ΔG_S of the surface and ΔG_V of the volume, respectively. r^* and ΔG_N^* are the critical nucleus radius and the critical work for nucleus creation, respectively. The *red line* refers to heterogeneous nucleation

In the classical capillary model developed by Gibbs (1874), Volmer and Weber (1926), Volmer (1939), and others, the onset of the phase transition in a small region called nucleus is connected to a change of Gibbs energy of the system ΔG_N. This quantity is composed of three contributions: at first, a favorable (i.e., negative) part ΔG_V due to the amount of substance entering the new stable phase. ΔG_V is proportional to the *volume* of the nucleus. Secondly, the interface between the new stable phase and the metastable surrounding (homogeneous) phase must be created, yielding a positive cost ΔG_S. This contribution of surface free-energy is proportional to the *area* of the interface. A third term ΔG_E arises if the nucleus is subjected to *elastic stress*. ΔG_E (>0) counteracts the favorable volume term ΔG_V and may even suppress nucleation if $\Delta G_E > |\Delta G_V|$. The total change of Gibbs energy upon creation of a nucleus is given by

$$\Delta G_N = \Delta G_V + \Delta G_S + \Delta G_E. \tag{4}$$

If we assume an *unstrained* spherical nucleus of radius r and an *isotropic* surface-free energy γ, we obtain $\Delta G_S = \gamma \Delta S = \gamma 4\,\pi r^2$ and $\Delta G_V = (\Delta g/\,v)\,(4/3)\,\pi r^3$. Here, r is the radius of the nucleus and ΔS its surface. v is the mole volume of the new phase and Δg (<0) is the change of Gibbs energy of the system by creating one mole of the new phase. Equation 4 then reads

$$\begin{aligned} \Delta G_N &= \Delta G_S + \Delta G_V \\ &= 4\pi \left(\gamma r^2 + (\Delta g/v) r^3/3\right) \text{ (unstrained spherical nucleus).} \end{aligned} \tag{5}$$

Gibbs energy change according to Eq. 5 drawn in Fig. 5 initially *increases* during formation of the stable phase due to the unfavorable creation of the boundary to the initial phase. After passing a maximum at some critical radius r^*, the volume term

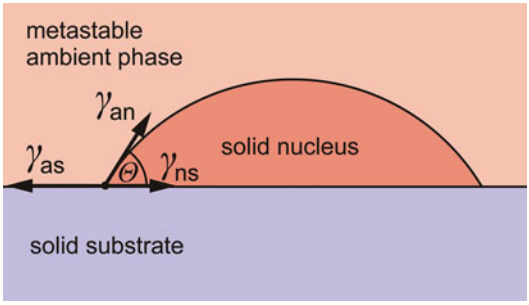


Fig. 6 Nucleus created on a structureless substrate. The balance of surface energies γ leads to a wetting angle Θ

($\propto r^3$) eventually prevails over the surface term ($\propto r^2$) and Gibbs energy change induced by the creation of the nucleus becomes negative. Such nuclei are stable and grow.

In *heterogeneous nucleation*, a solid phase preexists; the interface area to the ambient metastable phase is reduced, yielding a substantially smaller energy barrier $\Delta G_N{}^*_{hetero}$. Since a part of the preexisting solid (seed or substrate) is covered by the nucleus, a new interface is created between nucleus and substrate. If we assume an unstrained nucleus with the shape of a spherical cap with radius r on a substrate as depicted in Fig. 6, the balance of interface tensions at the line of contact between the three phases of metastable ambient (index a), nucleus (n), and substrate (s) is given by three quantities which represent the energies needed to create unit area of each of the three interfaces. From the figure, we read the *relation of Young* (1805) for the absolute values of tensions in balance $\gamma_{as} = \gamma_{ns} + \gamma_{an} \cos \Theta$, or

$$\cos \Theta = \frac{\gamma_{as} - \gamma_{ns}}{\gamma_{an}}. \tag{6}$$

The *wetting angle* Θ may vary between 0 and 180° depending on the degree of wetting, i.e., the affinity of nucleus and substrate materials. Θ determines the shape of the nucleus. The volume of a spherical cap is $(4/3)\,\pi\, r^3 \times f$, where $f(\Theta) < 1$ is a geometrical factor. This shape factor reduces the volume of the nucleus with respect to homogeneous nucleation according to $\Delta G_N{}^*_{hetero} = f \times \Delta G_N{}^*_{homo}$. Due to $f < 1$, homogeneous nucleation (i.e., random formation of solid nuclei in the bulk of the nutrient phase) can be suppressed and controlled seeded crystal growth can be performed.

The critical energy of nucleation $\Delta G_N{}^*$ represents an activation energy for the formation of viable nuclei. The rate of formation per unit area of such nuclei is described by the Arrhenius dependence

$$j = j_0 \exp\left(\frac{\Delta G_N^*}{kT}\right). \tag{7}$$

The prefactor j_0 follows from kinetic considerations of the nucleation process on an atomic scale.

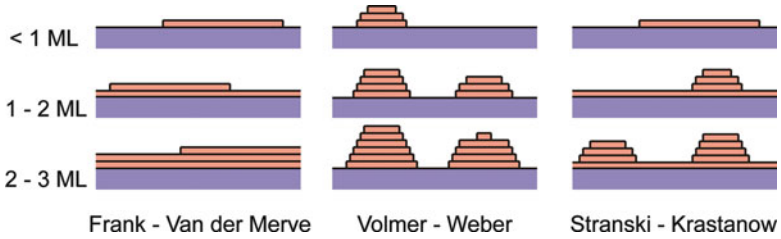


Fig. 7 Schematic of the three basic growth modes, illustrated as a function of approximately equal coverage given in units of monolayers (ML)

3.1.4 Growth Modes

The surface energies leading to Young's relation (Eq. 6) affect the initial stage of layer deposition on a substrate of different material. Growth with a smooth surface corresponds to a wetting angle of zero in Eq. 6, or $\gamma_{as} = \gamma_{ns} + \gamma_{an}$. If this condition applies or γ_{as} exceeds the sum of the two other interface energies, we obtain complete wetting of the layer on the substrate surface. Layer atoms are then more strongly attracted to the substrate than to themselves, and growth may proceed in an atomically flat layer-by-layer mode referred to as *Frank–van der Merwe growth mode*. Figure 7 illustrates the initial stages of such two-dimensional layer growth for different thickness of deposited material. Growth proceeds by the formation of *two-dimensional* islands and the advancement of surface steps as considered in the kinetic approach below.

A different surface morphology of the growth front is observed if the atoms of the growing solid are more strongly attracted to each other than to the substrate, as expressed in Eq. 6 by a wetting angle of π, or $\gamma_{ns} = \gamma_{as} + \gamma_{an}$. If this condition applies or γ_{ns} is even larger, then the deposit does not wet the substrate surface. The surface energy of deposit plus substrate is minimized if a maximum of substrate surface (with a low surface energy γ_{as}) is *not* covered by the growing material (with a large γ_{an}). This results in a three-dimensional growth of the deposit referred to as *Volmer–Weber growth mode*. Figure 7 illustrates that the growing solid is deposited in form of islands, which for thicker deposits eventually come into contact and coalesce.

Often, an intermediate case is found referred to as *Stranski–Krastanov growth mode* (also termed layer-plus-island growth). Here, the condition $\gamma_{as} \geq \gamma_{ns} + \gamma_{an}$ for Frank–van der Merwe growth applies solely for the first deposited monolayer (or the first few monolayers). After exceeding some critical thickness, the growth changes to the Volmer–Weber case where $\gamma_{ns} \geq \gamma_{as} + \gamma_{an}$ applies. Such a change may be induced by the gradual accumulation of strain in the deposit. The growing solid then resumes growth in form of three-dimensional islands, leaving a two-dimensional wetting layer underneath. Stranski–Krastanov growth has gained much advertence in recent years, because it may be employed for growth of defect-free quantum dots (see Sect. 2.2.3 in ► Chap. 3, "The Structure of Semiconductors").

The three categories of growth modes arising from a consideration of interface energies are sometimes complemented by further modes to account for observed

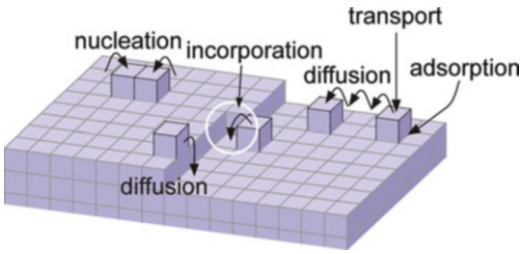


Fig. 8 Schematic illustration of basic kinetic processes in crystal growth. The *white circle* marks incorporation at a kink site located at a surface step

surface morphologies. Crystal faces which are slightly inclined with respect to faces with small Miller indices (vicinal surfaces) have persistent surface steps; under suitable conditions, these steps advance during growth without additional nucleation, leading to *step-flow growth*. Depending on growth conditions, also bunches of steps – so-called step bunching – may appear. In the presence of screw dislocations (Sect. 4.1 in ▶ Chap. 15, "Crystal Defects") which create persistent surface steps, *spiral growth* or *screw-island growth* is observed. Eventually, for materials with low surface mobility of adatoms, *columnar growth* may be obtained; similar to Volmer–Weber growth, islands nucleate, but they do not merge to continuous layers when growth proceeds. A mosaic crystal composed of numerous slightly tilted and twisted columnar crystallites is formed instead.

3.1.5 Kinetic Approach

The growth kinetics can usually be described by only a few categories of atomistic rate processes. Such processes strongly depend on the specific location of an atom on the surface as depicted in Fig. 8. Basically, four processes occur:

- Transport of atomic particles to the surface
- Diffusion of the particles across the surface
- Incorporation of each particle at a preferred surface site
- Dissipation of heat developed during growth

Each thermally activated kinetic growth process is controlled by characteristic parameters entering an Arrhenius dependence with an activation energy E, a characteristic time constant τ, and an attempt-rate constant ν_0:

$$\tau^{-1}(T) = \nu_0 e^{-\frac{E}{kT}}. \tag{8}$$

The time constant of a process can be altered by the experimental conditions.

Surface diffusion effects adatoms to find a proper lattice site for incorporation into the solid. A surface site of particular importance is the *kink site* or *half-crystal position* marked by a circle in Fig. 8. An atom at this site has one half of its bonds attached to neighboring atoms of the solid and the other half unsaturated. Crystal growth basically proceeds via the incorporation of atoms at this site. Occupation of a

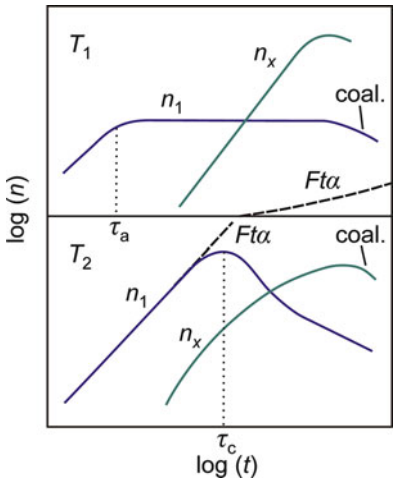


Fig. 9 Density of single atoms n_1, stable clusters n_x, and the total number $Ft\alpha$ of atoms condensed on the surface as a function of time t for a high temperature T_1 and a low temperatureT_2. τ_a and τ_c indicate reevaporation and capture times, respectively, and the label *coal.* marks a decrease of n_1 and n_x due to coalescence (After Venables et al. 1984)

kink site yields an adjacent kink site; thereby, the atom row at the step is gradually completed, leading to a lateral advancement of the step.[7]

The dynamics of growth is modeled in numerical Monte Carlo simulations using rate equations of a few basic processes. Only the rate-limiting steps, i.e., slower processes, are included in the calculations. Faster processes are accounted for in average by using *effective* kinetic parameters. The supersaturation $\Delta\mu$ is usually controlled by the experimental parameters T, the arrival rate of atoms F (flux, controlled by the pressure P), and the material parameters of the kinetic processes describing diffusion, reevaporation, and nucleation. The approach does not require a detailed knowledge of the atomic interactions and permits simulations including large time scales. Values for ν_0 and E (on the order of 10^{12} s^{-1} and eV, respectively) for each process are estimated from, e.g., molecular dynamics or are taken as parameters to fit experimental results.

The kinetic description yields the temperature-dependent density of single mobile adatoms, clusters of various sizes, and surface coverage. The critical radius r^* of the thermodynamic description corresponds to a *critical cluster size* comprising i atoms, where clusters with $j > i$ atoms are stable and likely to grow. The evolution of the surface coverage with *single* adatoms n_1, stable clusters n_x (=sum of all clusters n_j with $j > i$), and the total number of atoms on the surface, as resulting from the solution of kinetic rate equations, is illustrated in Fig. 9 for growth on a flat surface, starting at coverage $Z = 0$. At a *high temperature* T_1, nucleation is negligible; the single-atom density $n_1 \sim Ft$ initially rises for $t < \tau_a$, τ_a being the (short) residence time of a diffusing adatom from attachment on the surface (given by the flux rate F) and reevaporation. The constant value at longer time reflects the balance of arrival

[7]After completion of a step, a new kink must nucleate at the step for advancement, and after completion of an entire layer, a new (two-dimensional) nucleus with a step at its perimeter must be created. Particularly, the latter leads to a slow growth rate of flat surfaces.

and evaporation without nucleation or growth (high-temperature limit). The density of stable clusters n_x starts at a negligible value at $t = \tau_a$ and increases monotonously for $t > \tau_a$. Both n_1 and n_x decrease as coalescence sets in. The condensation coefficient $\alpha(t)$, which denotes the fraction of atomic dose impinging on the surface and being incorporated into the deposit, is initially very small in this high temperature, and so also the total deposit $Ft\alpha$.

At a *low temperature* T_2, no reevaporation occurs, i.e., $\alpha = 1$. In this temperature range, the single-atom density n_1 plotted in Fig. 9 increases linearly until capture by previously nucleated clusters sets in, causing n_1 to pass a maximum at the mean capture time τ_c and to decrease subsequently. The stable-cluster density n_x increases after a nucleation period and eventually decreases due to coalescence of clusters.

Growth of a flat surface proceeds basically by the attachment of adatoms at steps. The steps may originate either from nucleation as discussed above or from the terrace structure of a surface inclined by a small tilt angle with respect to a face with small Miller indices (vicinal surface) or from screw dislocations. The latter provide a source of persistent surface steps and may lead to efficient spiral growth of crystals.[8]

3.1.6 Growth Habit

When the transport to the surface is slower than the diffusion across the surface, the growth from a nucleus is ordered and a microcrystallite is formed. When such growth occurs from one nucleus (seed) only, a *single crystal* will grow. Growth from statistically formed and oriented nuclei results in *polycrystals*. The growth habit, i.e., the shape, of a single crystal often features flat faces with well-defined mutual angles. The reason is a pronounced dependence of the surface energy γ on the orientation of the surface. The *equilibrium crystal shape* is determined by the minimum of the sum of all products $\gamma_i \times S_i$, where the γ_i are the respective surface energies of the crystal faces with area S_i. According to *Wulff's theorem* (1901), the polyhedron of the crystal shape can be geometrically constructed by drawing vectors from the origin to all crystallographic directions with lengths proportional to the respective surface energies and placing a normal plane at the tip; the shape of the crystal is then given by the inner envelope of all such planes, i.e., by the planes with lowest surface energy.

The habit of real crystals often deviates significantly from the equilibrium shape: firstly, because growth occurs under nonequilibrium conditions altering the γ_i and secondly, due to the effect of imperfections like screw dislocations affecting the growth rate or due to spatial gradients in the driving force. Still, a simple rule holds: the growth habit is determined by the faces with slowest growth rate, because a slowly growing face with some angle ($<90°$) with respect to a fast growing face will gradually enlarge its surface on expense of the latter.

[8] A screw dislocation hitting a surface creates a steadily reproduced kink site at its core, enabling a spiral growth around the core that is much faster than growth on a planar surface. Under suitable conditions, a fine needle may form with an axial screw-dislocation line.

When the transport to the surface surpasses the diffusion velocity at the surface, i. e., when the viscosity of the material near the surface is too large, the atomic particles do not have time to find their proper, ordered position before more arriving atoms block their movement and freeze in a highly disordered state, creating an *amorphous solid*. Such growth occurs more easily when the atomic building blocks are larger molecules, or are formed by covalent forces, which, aside from space filling, must also fulfill well-defined angle relations.

3.2 Growth of Bulk Single Crystals

Single crystals can be grown from various ambient phases, such as vapors, solutions, melts, or even from a solid (e.g., amorphous) phase. The basics outlined above apply to all methods. There are numerous growth techniques as outlined in Table 1, and often, a specific material requires its own individually designed technique to obtain single crystals of the desired quality. Many reviews are available to describe this subject in detail, e.g., Buckley (1951), Laudise (1970), Goodman (1978), Pamplin (1980), Holden and Morrison (1982), Brice (1986), Scheel and Fukuda (2003), Markov (2017), and Handbooks, e.g., Hurle (1994) and Dhanaraj et al. (2010).

3.2.1 Growth from the Liquid Phase

Growth of bulk crystals from the liquid phase may occur by crystallization from either the molten material or a solution containing the dissolved material (the solute) in a suitable solvent.

Table 1 Techniques for growing bulk single crystals

Nutrient phase	Liquid		Vapor	Solid
	Melt	Solution		
Crystal growth process	Bridgman (vertical, horizontal)	Solvent evaporation	Physical vapor deposition	Recrystallization, strain annealing
	Gradient freeze (vertical (VGF), horizontal (HGF))	Slow cooling	Chemical vapor deposition	Precipitation from alloy
	Czochralski	Hydrothermal growth	Chemical vapor transport	Polymorphic phase change
	Kyropoulos	Diffusion and chemical reaction		
	Zone-melting, floating zone	Travelling solvent		
	Verneuil			

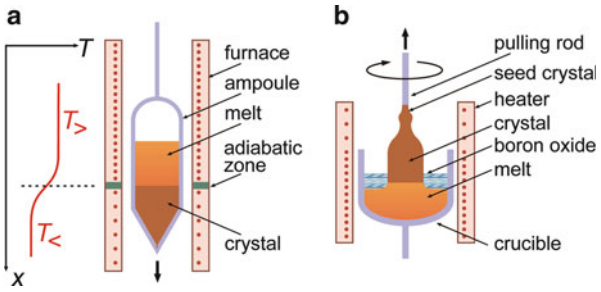


Fig. 10 Schematic for crystal growth from the melt. (**a**) Bridgman method with axial temperature profile of a furnace with two temperature zones. (**b**) Czochralski method with liquid encapsulation by boron oxide

Growth from the Melt

Melt growth is the most frequently applied method for growing large single crystals at high growth rates. Prerequisites for the application are a melting without decomposition (i.e., change of stoichiometry) and a solid state without polymorphic transitions (i.e., change of crystal structure without change of stoichiometry) below the melting point. Materials with a melting point below 1,800 °C are particularly made employing the Bridgman (1923) (also referred to as Bridgman Stockbarger) or the Czochralski method (1918), while high-melting materials are grown by zone-melting or the flame-fusion technique introduced by Verneuil (1902).

The *Bridgman method* illustrated in Fig. 10a employs a crucible containing the molten material which is moved relative to the axial temperature gradient of a vertical furnace. An additional adiabatic loss zone may be inserted between the high-temperature and the low-temperature zones of the furnace for a steep and better-adjusted temperature gradient. The speed of the directional crystallization is controlled by the speed of the temperature field; moving the furnace upward is favored over lowering the crucible to minimize agitation which may disturb crystallization. The coldest point of the crucible is the lower tip; here, crystallization of the melt commences, and a narrow part of the crucible supports the selection of a single crystal (with a proper, fast growing direction) from an initially polycrystalline material. Alternatively, a seed crystal may be placed at the lower tip of the crucible to induce growth along a specific orientation. The Bridgman method is also applied in a horizontal configuration.

A technique closely related to the Bridgman method is the *gradient freezing method* (see Gault et al. (1986), Asahi et al. (2003)): the crucible containing the melt is kept stationary, and the temperature gradient is moved along the melt by using a multiple-zone furnace with individually controlled temperature zones. The technique is implemented in vertical (*vertical gradient freeze* (VGF)) and horizontal (HGF) configurations.

Both Bridgman and VGF growths are widely employed for fabricating compound semiconductors (arsenides, phosphides, antimonides, II–VI). For growth of volatile or decomposing compounds with a high vapor pressure at melting point, an additional high pressure of an inert gas (Ar or N_2) is applied; since the diffusion coefficient in the vapor phase is inverse to the total pressure, the decomposition by volatile components is reduced. For binary III–V semiconductors, the growth rate is

in the range of 1 mm/h. Small temperature gradients (1–10 °C/cm) favor crystallization with a low dislocation density (<500 cm^{-2}).

The *Czochralski method* depicted in Fig. 10b is the most commonly used technique for growing Si and compound semiconductors. The method provides a high growth rate applicable to large crystals with a good perfection. The molten material is kept in a crucible at a temperature slightly above the melting point. A (cooler) seed crystal is dipped into the melt, leading to growth of the substance at the seed. The seed is then continuously pulled from the melt according to the speed of crystallization. The liquid–solid interface is kept slightly above the surface of the melt such that a rod-shaped crystal is formed.[9] The pulling speed depends on the geometry of crucible and crystal and on the balance of heat flow into and out of the crystal; it is on the order of some mm/h for semiconductors but may be as low as 0.1 mm/h for mixed crystals or as high as10 cm/h for elements. The pulling or heating rate must constantly be controlled to keep the growth front at the right position for maintaining a crystal with constant diameter. If either rate is too high, the crystallization front moves up, the crystal gets thinner, and removal of the heat of solidification gets worse; this leads to an even higher growth front. The stationary position of the crystallization front is thus labile and requires active control. Usually, the pulled crystal rotates at a rate of some to some ten rounds per minute to control also the convection of the melt.

For growth of volatile or decomposing compounds, a high pressure of an inert gas is applied, similar to the Bridgman technique outlined above. Particularly in the growth of GaAs, GaP, or InP, an additional molten boron oxide (B_2O_3) protection layer swimming on the melt is used (liquid encapsulation Czochralski (LEC); Metz et al. 1962). This chemically very stable *liquid encapsulation* has a low vapor pressure and a low density. Since it is transparent, the meniscus at the crystallization front can still be observed for controlling the pulling rate.

Growth from a Solution

This old technique is widely used for crystallizing materials which melt incongruently, i.e., decompose below the melting temperature. A further advantage particularly for low-temperature growth is the low density of equilibrium defects. Disadvantages are a relatively low growth rate (typically 1 mm/d) and a potential incorporation of solvent material. Supersaturation is achieved either by controlling the solution temperature, by evaporating the solvent, or by chemical reaction. The technique is called *hydrothermal method* if aqueous solvents are used, generally applied in an autoclave at high-pressure conditions to increase the solubility. For nonaqueous solvents, it is also termed *solvothermal method*. Hydrothermal growth is capable for growing very large single crystals, also from materials with a very low solubility at normal conditions, and was applied to fabricate a large variety of crystals like germinates, silicates, phosphates (KDP, $GaPO_4$), rare-earth (RE)

[9]This feature is different from the related *method of Kyropoulos* (1926, 1930), where crystallization proceeds by slowly cooling the melt, and the crystal grows into the melt.

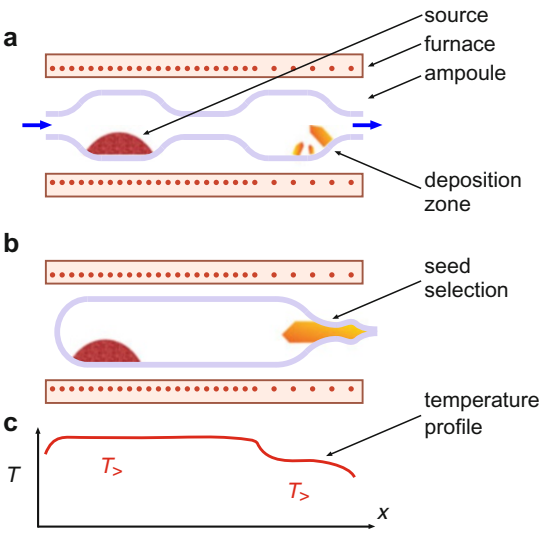


Fig. 11 Schematic for crystal growth from the vapor phase in (**a**) an open system and (**b**) a closed system. In (**b**), the ampoule has a bottleneck for selecting a seed. (**c**) Axial temperature profile of the furnace along the ampoule

tungstates ($NaRE(WO_4)_2$), III–V (GaN) and II–VI (ZnO) compounds, and organic compounds.

Growth from high-temperature solutions, also referred to as *flux method* (Wanklyn 1974), avoids the need of high pressure. Seed crystals for oriented growth may be used, and crystallization can be promoted by slow cooling or pulling of the seed. The method is particularly used for fabricating oxides (e.g., LBO (LiB_3O_5) and BBO (BaB_2O_4 or $Ba(BO_2)_2$)) and fluorides (BaF, PbF_2).

3.2.2 Growth from the Vapor Phase

In vapor-phase growth, the source material is transferred to the vapor phase by sublimation or chemical reaction at a high temperature $T_>$, transported to the growth region by diffusion or convection, and then crystallized at a lower temperature $T_<$. The sublimation–condensation process is also referred to as *physical vapor deposition* (PVD), while a chemically assisted process is often termed *chemical vapor deposition* (CVD) when performed in a closed configuration and *chemical vapor transport* (CVT) in an open configuration.[10] Recrystallization may occur at the ampoule walls by self-seeded growth or at a seed crystal. The basic setup is illustrated in Fig. 11.

[10]The differentiation between physical and chemical process is often not well defined, and CVT is also used for closed systems.

Physical Vapor Deposition

The equilibrium between a solid and a gaseous phase is described by the Clausius–Clapeyron relation $dP_e/dT = \Delta H/(T\Delta V)$. P_e is the equilibrium vapor pressure, ΔH the heat of sublimation, and ΔV the change of volume by the phase transition from solid to vapor. If ΔV is approximated by the (much larger) gas volume and the gas behaves approximately like an ideal gas, the Clausius–Clapeyron relation reads

$$\ln(P_{e2}/P_{e1}) \cong \frac{\Delta h(T_2 - T_1)}{R\ T_1 T_2}, \tag{9}$$

$\Delta h = \Delta H/n$ being the heat of sublimation of one mole of substance, R the universal molar gas constant, and P_{ei} the equilibrium vapor pressure at temperature T_i. Sublimation at a high temperature T_2 and condensation at a lower temperature T_1 then lead to a supersaturation

$$\Delta\mu = RT_1\ \ln(P_{e2}/P_{e1}). \tag{10}$$

The rate of material transport is proportional to $\Delta\mu$. Due to limits in equilibration and transport, Eq. 10 actually represents a maximum value. On the other hand, the transport rate must not be too large to avoid spontaneous parasitic nucleation and to ensure *crystalline* growth.

Chemical Vapor Transport

Application of a chemical transport agent C leads to a chemically assisted transition between solid and vapor phase (Binnewies et al. 2012). If we consider a compound AB with a gaseous component B (like GaN), the process can in a simple case be described by

$$AB^{\text{solid}} + C^{\text{gas}} \Leftrightarrow AC^{\text{gas}} + B^{\text{gas}}. \tag{11}$$

The equilibrium partial pressures are mutually related by the law of mass action,

$$P_{AC} \times P_B/P_C = K_P, \tag{12}$$

where K_P is the temperature-dependent equilibrium constant. K_P may have either sign and effects a transport from hot to cold for $K_P > 0$ (usual case) and reverse otherwise. Typical transport agents are hydrogen, halogens (chlorine, iodine, bromine), or hydrogen halides.

Growth from the vapor phase (both physical and chemical vapor deposition/transport and mixtures) is applied for a large variety of materials. It is the main

method to grow large boules of SiC (PVD and CVT), column-III nitrides (hydride/halide vapor-phase epitaxy), II-VI and I-III-VI_2 compounds (CVT), organic semiconductors (PVD), and many others.

3.2.3 Growth of Organic Crystals

The growth units of organic crystals are bulky molecules instead of the atoms illustrated schematically in Fig. 8, and their incorporation at a kink site of the surface requires some energy for proper orientation. The crystals have very weak intermolecular bonds and hence low melting temperatures (see Table 6 in ▶ Chap. 2, "Crystal Bonding"). Still the growth methods outlined above are basically also applied for fabricating organic semiconductors, but their vapor pressure, a low thermal stability of the growth units, and a limited solvent solubility often require specific modifications of conventional growth techniques; for a review, see Kloc et al. (2010).

Growth from the vapor phase is the most frequently used method for fabricating organic single crystals, particularly for smaller, more volatile molecules.[11] Generally, physical vapor transport is applied, using either an open tube or a sealed ampoule as illustrated in Fig. 11a, b. The organic source material is heated to sublimation temperature at one end of a quartz tube, vaporized, and either transported by an inert carrier gas like nitrogen or by diffusion in vacuum to the deposition zone held at decreased temperature. Organic semiconductors have usually a poor purity, but a comparable sensitivity like conventional semiconductors with respect to defects. This favors vapor-phase growth: heavy impurities may remain at the vaporization zone or already be deposited at higher temperatures before the crystallization zone, and lighter impurities behind this zone held at even lower temperatures. Crystals grown from the vapor phase have usually best purity and structural quality. They often show a preferred two-dimensional growth, yielding plate-type shapes with a very large aspect ratio of lateral dimensions versus thickness.

The vapor pressure, solubility, and stability of molecules decrease with increasing mass. This makes gas-phase growth difficult or even impossible for crystals composed of heavy molecules. A few materials like smaller acenes can be melted without decomposition. Such organic molecules can be crystallized from the melt by applying the Bridgman or the Czochralski method described above. For most materials with heavy molecules, *growth from a solution* is the only available method. Solution-phase growth can be performed by allowing crystals to precipitate from a supersaturated solution or by evaporating the solvent; the formation of larger crystals may be favored by a very slow cooling and slow solvent evaporation. The poor solubility may be increased by applying the solvothermal method with a high pressure and a moderately increased (still quite low) growth temperature. The solubility was also increased by modifying the molecules, e.g., by adding bulky

[11] Organic materials interesting for semiconductor applications are treated in Sect. 1.5 in ▶ Chap. 3, "The Structure of Semiconductors."

arms; however, such a substitution creates other molecules with altered physical properties. Growth can also be performed by coating techniques, e.g., by drop-casting a supersaturated solution of the dissolved semiconductor onto a substrate or by dip-coating the substrate into the solution. Such techniques lead to polycrystalline organic semiconductors, which are also widely used in organic electronics.

3.3 Epitaxy of Layer Structures

Advanced semiconductor devices often comprise quantum wells and superlattices with atomically abrupt interfaces, which can only be fabricated by epitaxial growth processes. The methods are essentially named after the nutrition phase. Physical vapor deposition techniques apply the vaporization of source material in vacuum, comprising thermal evaporation, laser ablation, or sputtering. Vapor-phase epitaxy (VPE) is often classified according to the chemistry of the source gases, such as metalorganic VPE (MOVPE or organometallic VPE (OMVPE), also termed metalorganic chemical vapor deposition (MOCVD)) and hydride VPE (HVPE). Liquid-phase epitaxy (LPE) is performed from a liquid solution or a melt. Epitaxy is treated in a number of textbooks, e.g., Herman et al. (2004), Ayers (2007), and Pohl (2020). Each epitaxy method has its strengths and weaknesses. The techniques of LPE, MBE, and MOVPE are outlined below; most electronic and optoelectronic devices are fabricated using one of these methods.

3.3.1 Liquid-Phase Epitaxy

In liquid-phase epitaxy (LPE), a crystalline layer grows from a supersaturated solution or melt on a substrate, similar to seeded bulk growth. The most unique features of this mature technology are:

- Layers with an extraordinary high structural perfection can be grown.
- High growth rates can be applied in the LPE growth process.

Growth conditions of LPE processes are close to thermodynamic equilibrium, enabling a very low density of point and dislocation defects. The carrier lifetime in such layers is consequently high. Highly efficient red GaAs-based LEDs and a majority of the world's production of high-performance HgCdTe γ-ray detectors and magneto-optic layers are presently fabricated using LPE. The technique has been applied to grow compound semiconductors (III–V, II–VI, IV–IV), magnetic or superconducting oxides, and many other materials. LPE is particularly useful for growing thick layers, also due to high deposition rates up to 1 μm/min. A restriction of LPE is the limitation to materials which are miscible at growth temperature and a quite limited ability for fabricating nanostructures like quantum wells due to interdiffusion at the interfaces. Reviews on LPE are given in, e.g., Giess and Ghez (1975), Astles (1990), Small et al. (1994), and Capper and Mauk (2007).

There exist several methods to bring the substrate into contact with the growth solution prepared for epitaxial growth and to separate them at the end of layer

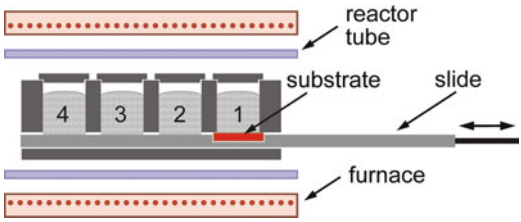


Fig. 12 Apparatus for liquid-phase epitaxy. Numbers denote solutions with different compositions

growth. Most versatile is the sliding-boat technique illustrated in Fig. 12, which allows for multiple-layer growth. The boat consists of two parts: a base carrying the substrate in a recess and wells in a block, which contain the solutions for growing successively different layers. Either the block with the wells or the base is movable, such that the solutions can be placed over the substrate; the surface tension of the liquids prevents leakage at the small clearance above the substrate. The boat is located in a furnace for precise temperature control and in a reactor tube for providing control of the gaseous ambient (usually purified H_2).

The LPE process is similar to that of seeded growth of bulk crystals from a solution or a melt. Growth from a solution is preferred due to low growth temperatures far below melting: the vapor pressure of volatile components is reduced, the interdiffusion of heterointerfaces is decreased, and detrimental effects of differences in the thermal expansion of substrate and epitaxial layers are reduced.

The supersaturation is created by either cooling or solvent evaporation of a saturated solution. A *diffusion boundary layer* of thickness δ with a concentration gradient and a temperature gradient is formed at the growth front. Growth species of the solute diffuse toward the liquid–solid interface, and solvent species diffuse contrariwise toward the bulk of the solution. The process within this layer is controlled by diffusion. LPE relies on well-established data of the liquid–solid phase equilibrium, and the development of a growth process for a new layer structure requires much more time than for MBE and MOVPE pointed out below. On the other hand, liquid-phase epitaxy bares the potential for fabricating epitaxial layers with the highest possible structural perfection and homogeneity.

3.3.2 Molecular-Beam Epitaxy

Molecular-beam epitaxy (MBE) is widely applied in research labs and industrial production. The technique is also termed metalorganic MBE (MOMBE), gas-source MBE (GSMBE), or chemical beam epitaxy (CBE) if different gas sources are employed. In early experiments, various beam techniques were used (cf. Miller and Bachmann (1958), Günther (1958)). Studies on the surface kinetics of GaAs epitaxy by Arthur (1968) provided an insight into the growth mechanisms, and soon later, layers with high quality were demonstrated by Cho (1971). Reviews on MBE are given in, e.g., Parker (1985), Joyce et al. (1994), and Asahi and Horikoshi (2019).

The characteristic feature of MBE is the mass transport in molecular or atomic beams, i.e., directed rays of neutral atoms or molecules in vacuum. An ultrahigh

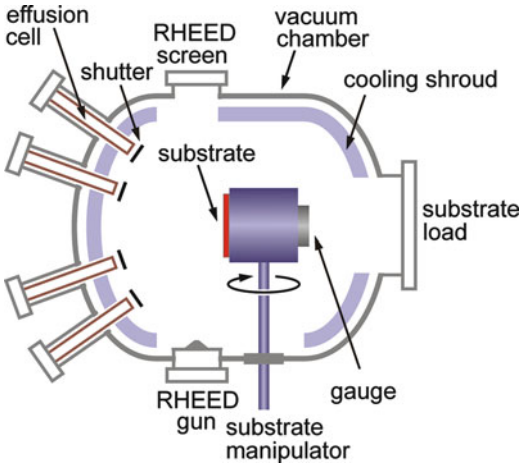


Fig. 13 Schematic of a molecular-beam epitaxy system. The *circular arrow* indicates the positioning of the gauge at the location of the substrate for calibrating the beam-equivalent pressure of the effusion cells

vacuum environment (UHV, $P < 10^{-7}$ Pa) is required to ensure a low impurity level in the layers originating from residual atoms of the background vapor; the background vapor provides a steady flux on the substrate that must be much smaller than that of the MBE sources. A schematic of an MBE system is given in Fig. 13. Beams of different species are produced by effusion cells mounted opposite to the substrate. The duration of the exposure is individually controlled by shutters for a rapid change of material composition. The beam-equivalent pressure (BEP) produced by the sources is measured by a gauge placed at the position of the substrate. Typical growth rates are below 1 μm/h. The vacuum environment enables in situ monitoring using electron diffraction. Usually, reflection high-energy electron diffraction (RHEED) with an electron beam nearly parallel to the growth surface is applied, yielding structural information on the surface crystallography during surface preparation and epitaxy.

Beam Sources

For the production of beams from *solid or liquid materials*, usually Knudsen cells (K cells) based on radiative heating are employed; for temperatures above 1,300 °C, mostly electron-beam evaporation is used. The effusion rate of thermal evaporation is described by the *Knudsen equation*

$$\Gamma = \frac{(P_e - P_{res})\,A}{\sqrt{2\pi\ m\ kT}} \quad \left(\frac{\text{particles}}{\text{s}}\right), \tag{13}$$

where the equilibrium vapor pressure $P_e(T)$ in the cell kept at temperature T can be expressed by the Clausius–Clapeyron relation, P_{res} is the residual gas pressure in the MBE chamber, and m is the mass of the effused atoms or molecules. The aperture A of the cell must be small for an ideal isothermal cell obeying Eq. 13, but often A is made large to increase the particle flux. The angular flux distribution is in the ideal case described by *Knudsen's cosine law* of effusion,

$$\Gamma(\vartheta)/\omega = \Gamma(0) \cos\vartheta, \tag{14}$$

the angle ϑ referring to the direction normal to the aperture and ω being the unit solid angle comprising the considered flux. Cells with a large aperture for enhanced effusion at moderate cell temperature have a flux distribution depending on the charging level of the cell; the effect originates from a gradually increasing collimation of the crucible side walls as the source material depletes.

G*aseous source materials* (supplied by external cylinders) are applied due to a quasi unlimited lifetime of the source and a precise control of the flux by pressure or mass-flow controllers which are also used in MOVPE. This allows also for a simple change of flux for, e.g., varying alloy composition. A gas source consists of the gas-control system and a cell (e.g., a heated nozzle) for the gas inlet into the MBE chamber. Thermally stable molecules are thermally dissociated in a cracking stage of the vacuum inlet cell to provide reactive species for growth.

Kinetic Aspects

Growth in MBE occurs usually far from thermodynamic equilibrium. Still both kinetic and thermodynamic processes are relevant. Since sources and substrate have different temperatures, no global equilibrium exists for the entire system. Effused particles have an energy distribution according to the temperature of the specific source. When they impinge on the substrate surface, they thermalize to the substrate temperature. Particles desorbing again from the surface were found to reflect an energy distribution according to the *substrate* temperature, indicating that the time for thermalization is much less than the mean time of surface diffusion. Such a finding justifies to assume an at least partial or local equilibrium on a time scale relevant for the growth process. The relevant equilibrium temperature is therefore that of the substrate.

The fraction of particles sticking to the surface and being incorporated during epitaxy N_{sticking} with respect to the total number of impinging particles of a considered species N_{total} is referred to as the *sticking coefficient s*:

$$s = \frac{N_{\text{sticking}}}{N_{\text{total}}} \quad . \tag{15}$$

The quantity s may also depend on the flux of other species[12] and may have any value between zero and unity. The sticking of particles on the surface is described by an adsorption energy as applied in Eq. 8. Usually, the terms *physisorption* and *chemisorption* are used to account for smaller and larger energies, respectively, although chemical interactions occur in both kinds of adsorptions and the two terms are not well defined. They are still useful for an overall description of surface processes to express different surface diffusivities of a considered adatom species.

[12]For example, in MBE of GaAs, the sticking coefficient of As_2,s_{As_2} , increases linearly with the (independent) Ga adsorption rate and reaches unity when flux(Ga) = 2 × flux(As_2).

For a simplified growth description, a two-step condensation process of the impinging particle is assumed implying two sticking coefficients. The high surface mobility often found for species arriving at the surface is then assigned to a physisorbed state with a lower adsorption energy and a larger desorption probability, and incorporation to the chemisorbed state.

3.3.3 Metalorganic Vapor-Phase Epitaxy

Metalorganic vapor-phase epitaxy (MOVPE) is presently applied for semiconductor device fabrication in large-scale reactors having a capacity for the simultaneous deposition on 30 four-inch wafers. Applications of MOVPE also include oxides, metals, and organic materials. The technique was first reported by Miederer et al. (1962), Manasevit and Simpson (1968) and Manasevit (1972). Due to a simple control of partial pressures, MOVPE is advantageous in realizing graded layers, or in applying source materials differing strongly in partial pressures, e.g., in growth of arsenide–phosphide alloys and nitride semiconductors. MOVPE is described in detail in, e.g., Jones and O'Brien (1997), Stringfellow (1999), and Irvine and Capper (2020).

Metalorganic Precursors

In MOVPE, the constituent elements are transported to the vapor–solid interface in form of volatile metalorganic compounds, typically by a carrier gas like hydrogen at 100 mbar total pressure. The gaseous species dissociate thermally at the heated substrate, thereby releasing the elements for layer growth. The dissociation is generally assisted by chemical reactions. The *net reaction* for the MOVPE of GaAs using the standard source compounds trimethylgallium and arsine reads

$$\mathrm{Ga\,(CH_3)_3 \;+\; AsH_3 \;\rightarrow\; GaAs \;+\; 3\,CH_4\uparrow.} \tag{16}$$

Actually, many reactions participate, and many successive steps and species are involved in the chemistry of deposition (see, e.g., Mountziaris and Jensen (1991)). Simple steps in the release of Ga from the trimethylgallium precursor in Eq. 16 are, e.g., given by the successive removal of the organic CH_3 groups:

$$\mathrm{Ga\,(CH_3)_3 \;\rightarrow\; Ga\,(CH_3)_2 \;+\; CH_3 \;\rightarrow\; Ga\,CH_3 \;+\; 2\,CH_3 \;\rightarrow\; Ga \;+\; 3\,CH_3.}$$

Basic requirements for the source compounds employed for MOVPE are a low stability to allow for decomposition in the process, but still a sufficient one for long-term storage. Furthermore, the volatility should be high, and a liquid state is favorable to provide a steady source flow by the carrier gas. Most source molecules have the form MR_n, where M denotes the source element (e.g., Ga) and R are alkyls like methyl CH_3, ethyl C_2H_5, or tertiary butyl $C(CH_3)_3$. The bond strength to a given element M can be selected by choosing a suitable organic ligand: the metal–carbon bond strength depends on the electronegativity of the metal M and the size and

configuration of the ligand R. As a thumb rule, the bond strength decreases as the number of carbon bonds to the central carbon in the alkyl is increased.

Besides metalorganic sources, also hydrides like arsine AsH_3 are employed as precursors. Their use is interesting since they release hydrogen radicals under decomposition that can assist removal of carbon-containing radicals from the surface, such as CH_3 forming stable CH_4 with H. A major obstacle is their high toxicity and their very high vapor pressure, requiring extensive safety precautions.

Precursor Supply

Most metalorganic sources are liquids which are stored in bubblers at a controlled temperature, yielding an equilibrium vapor pressure $P_{e\ MO}$. For transport, a carrier gas (usually hydrogen) with a flow Q_{MO} is introduced by a dip tube ending near the bottom of the bubbler, and the bubbles of the carrier gas saturate with precursor molecules. At the outlet of the bubbler, a pressure controller is installed, acting like a pressure-relief valve. This allows to define a fixed pressure P_B ($>P_{e\ MO}$) in the bubbler, thereby decoupling the bubbler pressure from the equilibrium vapor-pressure of the MO source (see Fig. 14). Also, the total pressure P_{tot} in the reactor is controlled independently. The partial pressure of a metalorganic source in the reactor P_{MO} results from the mentioned parameters by

$$P_{MO} = \frac{Q_{MO}}{Q_{tot}} \times \frac{P_{tot}}{P_B} \times P_{e\ MO}, \tag{17}$$

Q_{tot} denoting the total flow in the reactor. Both fractions in Eq. 17 are employed to control the partial pressure P_{MO} of the source in the reactor. Gaseous hydrides are directly controlled by their flow Q_{Hyd}, and Eq. 17 simplifies to $P_{Hyd} = Q_{Hyd} \times P_{tot}/Q_{tot}$.

The total flow in the reactor results from the sum of all component flows plus the additionally introduced flow of the carrier gas. This flow is generally much higher

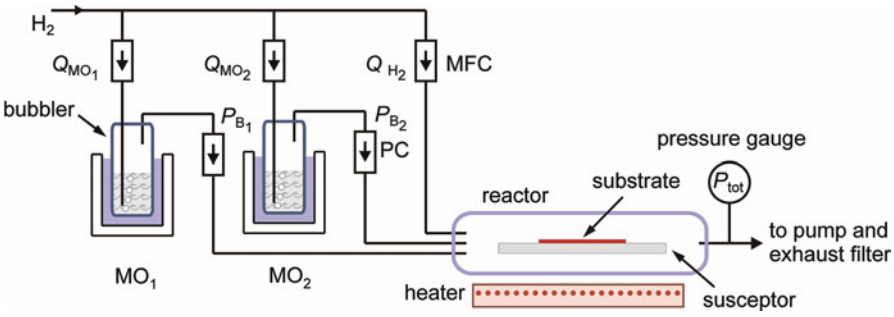


Fig. 14 Schematic of a metalorganic vapor-phase epitaxy apparatus. Hydrogen is used as carrier gas and introduced into the metalorganic sources MO_1 and MO_2. MFC and PC denote mass-flow and pressure controllers, respectively

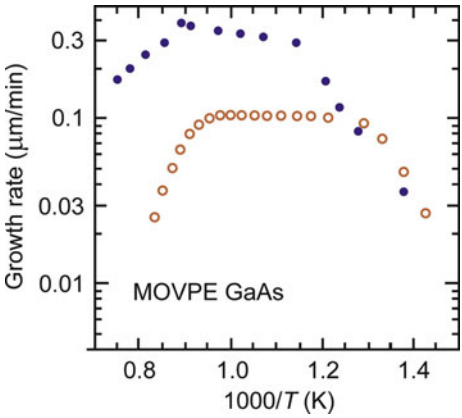


Fig. 15 Growth rate in the MOVPE of homoepitaxial GaAs layers as a function of reciprocal temperature. Trimethylgallium and arsine are used as precursors. Full and open *circles* represent measured data from Reep and Ghandhi (1983) and model predictions from Mountziaris and Jensen (1991), respectively

than that of all sources; consequently, the sum of all source partial pressures P_{MO} and P_{Hyd} is much smaller than the total pressure P_{tot}. The reactor pressure P_{tot} is controlled as an independent parameter by a control valve attached to an exhaust pump behind the reactor (Fig. 14).

Growth Regimes

The complete treatment of the MOVPE process involves numerous gas phase and surface reactions, in addition to hydrodynamic aspects. Numerical solutions of such a complex problem were developed for specific processes like the MOVPE of GaAs from trimethylgallium and arsine (Jensen 1994); the modeling of III-nitrides is reviewed by Dauelsberg and Talalaev (2020). Growth occurs generally far from equilibrium[13]; the growth rate is usually somewhat higher than in MBE (~µm/h) and generally limited by the slowest process. Without considering growth mechanisms in detail, limits by either transport or decomposition/surface kinetics can be well distinguished. Figure 15 shows the dependence of the GaAs growth rate from the substrate temperature. At low-temperature experiment and simulation show an exponential relation due to thermal activation of precursor decomposition and interface growth reactions. This regime is referred to as *kinetically limited growth*. As the temperature is increased, the growth rate becomes nearly independent on temperature. In this range called *transport-limited growth*, precursor decomposition and surface reactions are much faster than mass transport from the gas phase to the vapor–solid interface; the diffusion in the gas phase depends only weakly on temperature.[14] In the high-temperature range, growth rates decrease due to enhanced desorption and parasitic deposition at the reactor walls, leading to a depletion of the gas phase.

MOVPE is usually performed in the mid-temperature range of transport-limited growth, where variations of the substrate temperature have only a minor effect on the

[13]Similar to MBE

[14]The difference in the maximum growth rates in Fig. 15 originates from effects of the reactor geometry.

growth rate, the composition of alloys, and on the doping level. For arsenide and phosphide semiconductors, the range is typically between 500 °C and 800 °C and for nitrides with their strong bonds above 1,000 °C to ensure sufficient adatom surface mobility for crystalline growth.

4 Summary

Semiconductors have become basic elements of modern civilization, with an extremely large range of application, including high-speed communication, computation, and artificial intelligence. An increasing variety of materials, distinguished by their chemical composition and atomic structure, is used to produce semiconducting devices and to perform specific device functions with increasing efficiency in smaller dimensions and at lower costs.

The growth of a semiconductor requires a supersaturation or a supercooling of the nutrient phase. Growth commences from a nucleus or crystal seed and proceeds by incorporation of atoms at kink sites and the advancement of surface steps. Single-crystal growth from the melt achieves highest growth rates. Growth from the vapor phase may be performed by physical vapor deposition applying a sublimation–condensation process or by chemical vapor deposition implying chemical reactions.

Epitaxial growth of layers is mainly performed applying liquid-phase epitaxy (LPE), molecular-beam epitaxy (MBE), or metalorganic vapor-phase epitaxy (MOVPE). LPE yields highest growth rates and layer perfection, but not atomically sharp interfaces. MBE is a physical vapor deposition technique employing atomic or molecular beams in ultrahigh vacuum environment, enabling in situ monitoring by electron diffraction. MOVPE employs metalorganic source compounds which release the required elements by decomposition at growth temperature. Heterostructures with atomically abrupt interfaces like quantum wells can be fabricated by both MBE and MOVPE.

References

Adams WG, Day RE (1876) The action of light on selenium. Proc R Soc (Lond) 25:113

AMD (2021) image accessible at https://library.amd.com/media/?mediaId=DE133045-FEF1-4B57-A4741EC209817C03

Arthur JR Jr (1968) Interaction of Ga and As_2 molecular beams with GaAs surfaces. J Appl Phys 39:4032

Asahi H, Horikoshi Y (2019) Molecular beam epitaxy. Wiley, Hoboken

Asahi T, Kainosho K, Kohiro K, Noda A, Sato K, Oda O (2003) Chapter 15: Growth of III-V and II-VI single crystals. In: Scheel HJ, Fukuda T (eds) Crystal growth technology. Wiley, Chichester

Astles MG (1990) Liquid-phase epitaxial growth of III-V compound semiconductor materials and their device applications. Adam Hilger, Bristol

Ayers JE (2007) Heteroepitaxy of semiconductors: theory, growth, and characterization. CRC, Boca Raton

Bell Labs (1947) see: Bo Lojek (2007) History of semiconductor engineering. Springer, Berlin

Bergmann L (1931) Über eine neue Selen-Sperrschicht Photozelle. Phys Z 32:286 (On novel selenium-junction photo cells, in German)
Bergmann L (1934) Phys Z 35:450
Binnewies M, Glaum R, Schmidt M, Schmidt P (2012) Chemical vapor transport reactions. de Gruyter, Berlin/Boston
Braun F (1874) Über die Stromleitung durch Schwefelmetalle. Ann Phys Chem 153:556 (On the current conduction in sulfur metals, in German)
Brice JC (1986) Crystal growth processes. Halstead Press, New York
Bridgman PW (1923) The compressibility of thirty metals as a function of pressure and temperature. Proc Am Acad Arts Sci (Boston) 58:165; Ibid. 60:303(1925)
Bromme T (1851) Atlas der Physik der Welt. Krais & Hoffmann, Stuttgart (Physics atlas of the world, in German)
Buckley HE (1951) Crystal growth. Wiley, New York
Capper P, Mauk M (2007) Liquid phase epitaxy of electronic, optical and optoelectronic materials. Wiley, Chichester
Cho AY (1971) Film deposition by molecular-beam techniques. J Vac Sci Technol 8:S31
Czochralski J (1918) Ein neues Verfahren zur Messung der Kristallisationsgeschwindigkeit der Metalle. Z Phys Chemie 92:219 (New method for measuring the crystallization speed of metals, in German)
Dauelsberg M, Talalaev R (2020) Progress in modeling of III-nitride MOVPE. Prog Crystal Growth & Character Mater 66:100486
Dhanaraj G, Byrappa K, Prasad V, Dudley M (eds) (2010) Springer handbook of crystal growth. Springer, New York
Ebert JJ (1789) Unterweisung in den Anfangsgründen der Naturlehre. Chr. Gottlieb Hertel, Leipzig (Briefing in the elements of natural sciences, in German)
Faraday M (1833) Experimental researches in electricity, series IV. Bernard Quaritch, London, p 433
Forrest SR (2020) Organic electronics: foundations to applications. Oxford University Press, Oxford
Gault WA, Monberg EM, Clemans JE (1986) A novel application of the vertical gradient freeze method to the growth of high quality III–V crystals. J Cryst Growth 74:491
Gibbs JW (1874) On the equilibrium of heterogeneous substances. Trans Conn Acad Arts Sci 3:108–248, 343–524, (1874–1878). Reproduced in both The Scientific Papers (1906), pp 55–353 and The Collected Works of J. Willard Gibbs, vol 2, Longmans, Green and Co., New York (1928), p. 267
Giess EA, Ghez R (1975) Liquid-phase epitaxy. In: Matthews JW (ed) Epitaxial growth part B. Academic Press, New York, pp 183–213
Goodman CHL (1978) Crystal growth: theory and techniques. Plenum Press, New York
Grondahl LO (1926/1932) see: note on the discovery of the photoelectric effect in a copper-oxide rectifier. Phys Rev 40:635
Günther KG (1958) Aufdampfschichten aus halbleitenden III-V-Verbindungen. Z Naturforschg 13a:1081 (Vapor deposition of semiconducting III-V compound layers, in German)
Hermann MA, Richter W, Sitter H (2004) Epitaxy. Springer, Berlin
Hittorf JW (1851) Über das elektrische Leitvermögen des Schwefelsilbers und des Halbschwefelkupfers. Ann Phys Lpz 84:1 (On the electric conductivity of sulfur silver and semi-sulfur copper, in German)
Holden A, Morrison PS (1982) Crystals and crystal growing. MIT Press, Cambridge, MA
Hurle DTJ (1994) Handbook of crystal growth vol. 2a, bulk crystal growth, basic techniques. North Holland, Amsterdam
Irvine S, Capper P (eds) (2020) Metalorganic vapor phase epitaxy (MOVPE). Wiley, Hoboken
Jensen KF (1994) Transport phenomena in vapor phase epitaxy reactors. In: Hurle DRT (ed) Handbook of crystal growth. Elsevier, Amsterdam, pp 541–599
Jones AC, O'Brien P (1997) CVD of compound semiconductors. VCH, Weinheim
Joyce BA, Vvedenski DD, Foxon CT (1994) Growth mechanisms in MBE and CBE of III-V compounds. In: Mahajan S (ed) Handbook on semiconductors. Elsevier, Amsterdam

Kloc C, Siegrist T, Pflaum J (2010) Growth of single-crystal organic semiconductors. In: Dhanaraj G, Byrappa K, Prasad V, Dudley M (eds) Springer handbook of crystal growth. Springer, New York

Königsberger T, Weiss T (1911) Über die thermoelektrischen Effekte (Thermokräfte, Thomsonwärme) und die Wärmeleitung in einigen Elementen und Verbindungen und über die experimentelle Prüfung der Elektronentheorien. Ann Phys 35:1. (On the thermoelectrical effects and heat conductivity in some elements and compounds and on the experimental examination of the electron theory, in German)

Kyropoulos S (1926) Ein Verfahren zur Herstellung großer Kristalle. Z Anorg Allg Chemie 154:308 (A method for the fabrication of large crystals, in German)

Kyropoulos S (1930) Dielektrizitätskonstanten regulärer Kristalle. Z Phys 63:849 (Dielectric constants of normal crystals, in German)

Laudise RA (1970) The growth of single crystals. Prentice Hall, Englewood Cliffs

Manasevit HM, Simpson WI (1968) The use of metal-organics in the preparation of semiconductor materials on insulating substrates: I. Epitaxial III-V gallium compounds. J Electrochem Soc 12:66C

Manasevit HM (1972) The use of metalorganics in the preparation of semiconductor materials: Growth on insulating substrates. J Crystal Growth 13/14:306

Markov IV (2017) Crystal growth for beginners, 3rd edn. World Scientific, Singapore

Metz EAP, Miller RC, Mazelsky R (1962) A technique for pulling single crystals of volatile materials. J Appl Phys 33:2016

Miederer WG, Ziegler, Dötzer R (1962) Verfahren zum tiegelfreien Herstellen von Galliumarsenidstäben aus Galliumalkylen und Arsenverbindungen bei niedrigen Temperaturen. German Patent 1,176,102, filed 25.9.1962; and: Method of crucible-free production of gallium arsenide rods from alkyl galliums and arsenic compounds at low temperatures. US Patent 3,226,270, filed 24.9.1963

Miller RJ, Bachmann CH (1958) Production of cadmium sulfide crystals by coevaporation in a vacuum. J Appl Phys 29:1277

Mooser E, Pearson WB (1956) The chemical bond in semiconductors. J Electron 1:629

Mountziaris TJ, Jensen KF (1991) Gas-phase and surface reaction mechanisms in MOCVD of GaAs with trimethyl-gallium and arsine. J Electrochem Soc 138:2426

Ovshinsky SR (1968) Reversible electrical switching phenomena in disordered structures. Phys Rev Lett 21:1450

Pamplin B (1980) Crystal growth, 2nd edn. Pergamon, New York

Parker EHC (ed) (1985) The technology and physics of molecular beam epitaxy. Plenum Press, New York

Pohl UW (2020) Epitaxy of semiconductors, 2nd edn. Springer Nature, Switzerland

Queisser HJ (1985) Kristallene Krisen. Piper, München English: The Conquest of the Microchip. Harvard University Press, Cambridge MA

Reep DH, Ghandhi SK (1983) Deposition of GaAs epitaxial layers by organometallic CVD. J Electrochem Soc 130:675

Scheel HJ, Fukuda T (eds) (2003) Crystal growth technology. Wiley, Chichester

Schuster A (1874) On unilateral conductivity. Philos Mag 48:251

Seebeck TJ (1822) Magnetische Polarisation der Metalle und Erze durch Temperaturdifferenz. Abhandl Deut Akad Wiss. Berlin, p 265. (Magnetic polarization of metals and ore by temperature difference, in German)

Small MB, Giess EA, Ghez R (1994) Liquid-phase epitaxy. In: Hurle DTJ (ed) Handbook of crystal growth, vol 3. Elsevier, Amsterdam, pp 223–253

Smith W (1873) Effect of light on selenium during the passage of an electric current. Nature 7:303

Stringfellow GB (1999) Organometallic vapor-phase epitaxy, 2nd edn. Academic Press, New York

Sze SM, Ng KK (2007) Physics of semiconductor devices, 3rd edn. Wiley, Hoboken

Venables JA, Spiller GDT, Hanbrücken M (1984) Nucleation and growth of thin films. Rep Prog Phys 47:399

Verneuil AV (1902) Production artificielle du rubis par fusion (Artificial production of ruby by fusion, in French) C R Acad Sci Paris C 135:791; La synthese du rubis (Synthesis of ruby, in French) Ann Chim et Phys (Paris) 3:20 (1904)
Volmer M (1939) Kinetik der Phasenbildung. Theodor Steinkopf, Dresden (Kinetics of phase formation, in German)
Volmer M, Weber A (1926) Tröpfchenbildung in Dämpfen. Z Phys Chem 119:227 (Formation of droplets in vapor, in German)
Wanklyn BMR (1974) Practical aspects of flux growth by spontaneous nucleation. In: Pamplin BR (ed) Crystal growth, vol 1. Pergamon, Oxford, pp 217–288
Wilson AH (1931) The theory of electronic semi-conductors. Proc R Soc (Lond) Ser A 133:458
WSTS (2021) accessible at https://www.wsts.org/esraCMS/extension/media/f/WST/5145/WSTS_nr-2021_08.pdf
Wulff G (1901) Zur Frage der Geschwindigkeit des Wachstums und der Auflösung der Kristallflächen, Z. Kristallographie 34:449 (On the question of growth velocity and the decomposition of crystal faces, in German)
Young T (1805) An essay on the cohesion of fluids. Phil Trans R Soc Lond 95:65
Zachariasen WH (1932) The atomic arrangement in glass. J Am Chem Soc 54:3841

Crystal Bonding

2

Karl W. Böer and Udo W. Pohl

Contents

1 Ionic and Covalent Bonding 36
1.1 Ionic Bonding 37
1.2 Covalent Bonding 40
1.3 Mixed Bonding 42
2 Metallic Bonding 46
3 Further Types of Bonding in Solids 46
3.1 Atomic and Ionic Radii 48
3.2 Bond-Length Relaxation in Alloys 52
3.3 Bonding in Organic Crystals 54
4 Summary 56
References 56

Abstract

The bonding of atoms in semiconductors is accomplished by electrostatic forces – Coulomb forces between the electrons and atomic nuclei – and the tendency of atoms to fill their outer shells. Interatomic attraction is balanced by short-range repulsion due to strong resistance of atoms against interpenetration of core shells. Coulomb forces are the basis for ionic and hydrogen bonding forces but are also involved in metallic bonding and, as dipole–dipole interaction, in van der Waals bonding. In addition, strong quantum-mechanical effects, determining specific orbitals, and Pauli exclusion are major contributing factors in covalent and metallic bonding, respectively.

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_2

Keywords

Atomic radii · Buckingham potential · Bond-length relaxation · Bonding of atoms · Coulomb force · Covalent bonding · Electrostatic forces · Hydrogen bonding · Ionic bonding · Madelung constant · Metallic bonding · van der Waals bonding · Vegard's law

1 Ionic and Covalent Bonding

Interatomic Forces In this chapter, the different types of bonding of solids are reviewed, irrespective of whether the solids are crystalline or amorphous. This treatment is quite general and is not restricted to semiconductors.

The formation of solids is determined by the interatomic forces and the size of the atoms shaping the crystal lattice. The interatomic forces are composed of a far-reaching attractive and a short-range repulsive component, resulting in an *equilibrium distance* of vanishing forces at an interatomic distance r_e, at which the potential energy shows a minimum (Fig. 1). In binary compounds, this equilibrium distance, r_e, can be written as the sum of *atomic radii*,

$$r_e = r_A + r_B, \tag{1}$$

where r_A and r_B are characteristic for the two atoms A and B (Fig. 2) and can be used when other binary compounds are formed with the same bonding type, containing A or B. For a more detailed discussion on atomic radii, see Sect. 3.1.

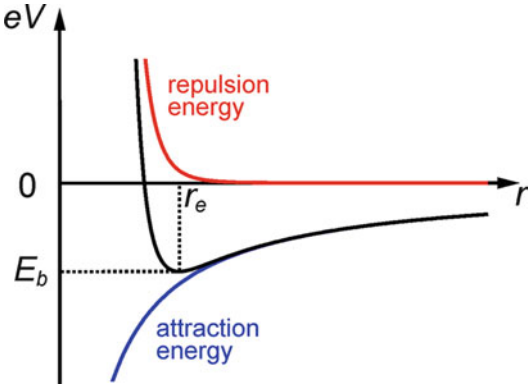


Fig. 1 Interaction potential eV between two atoms; r_e is the equilibrium distance; E_b is the bonding energy at $r = r_e$

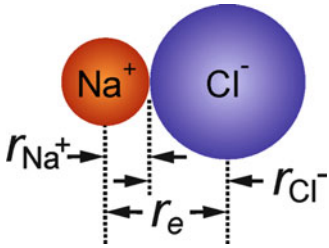


Fig. 2 Na^+ anion and Cl^- cation shown as hard spheres in actual ratio of radii

Attractive interatomic forces are predominantly electrostatic in character (e.g., in ionic, metallic, van der Waals, and hydrogen bonding) or are a consequence of sharing valence electrons of atoms to fill their outer shells, resulting in covalent bonding. Most materials show mixed bonding, i.e., at least two of these bond types contribute significantly to the interatomic interaction. In most technologically important compound semiconductors, these mixed bondings are more covalent and less ionic. In other semiconductors, one of the other types may contribute, e.g., van der Waals bonding in organic crystals and metallic bonding in highly conductive semiconductors. These different bonding types will be discussed in the following sections.

The *repulsive interatomic forces*, called Born forces (see Born and Huang 1954), are caused by a strong resistance of the electronic shells of atoms against interpenetration. The repulsive Born potential is usually modeled with a strong power law[1]

$$eV(r) = \beta / r^m, \quad m \approx 10 \ldots 12, \tag{3}$$

with β the force constant (see Eq. 1 in ▶ Chap. 4, "Elasticity and Phonons") and m an empirical exponent (see Sect. 1.1). For ionic crystals, the exponent m is somewhat smaller ($6 < m < 10$).

1.1 Ionic Bonding

Ionic bonding is caused by Coulomb attraction between ions. Such ions are formed by the tendency of atoms to complete their outer shells. This is most easily accomplished by compounds between elements of group I and group VII of the periodic system of elements; here, only one electron needs to be exchanged. For instance, in a NaCl crystal, the Cl atom captures one electron to form a negative Cl^- ion and the Na atom loses the single electron in its outer shell to become a positive Na^+ ion. The bonding is then described by isotropic (radial-symmetric) *nonsaturable* Coulomb forces attracting as many Na^+ ions as space permits around each Cl^- ion, and vice versa, while maintaining overall neutrality, i.e., an equal number of positive and negative ions. This results in a closely packed NaCl lattice with a coordination number 6 (=number of nearest neighbors).

The energy gain between two ions can be calculated from the potential equation

$$eV = -\frac{e^2}{4\pi\varepsilon_0 \, r} + \frac{\beta}{r^m} \quad \text{for } r = r_e, \tag{4}$$

[1] A better fit for the Born repulsion is obtained by the sum of a power and an exponential law:

$$eV_{\text{Born}} = \beta / r^m + \gamma \exp(-r/r_0) \, , \tag{2}$$

where r_0 is the softness parameter, listed for ions in Table 8. For more sophisticated repulsion potentials, see Shanker and Kumar (1987).

containing Coulomb attraction (first term) and Born repulsion (second term). For an equilibrium distance $r_e = r_{Na^+} + r_{Cl^-} = 2.8\text{Å}$, a minimum of the potential energy of $eV_{min} \approx -5$ eV results[2] for a typical value of $m = 9$. In a crystal, we must consider *all* neighbors. For example, in a NaCl lattice, six nearest neighbors exert Coulomb attraction in addition to 12 next-nearest neighbors of equal charge exerting Coulomb repulsion, etc. This alternating interaction results in a summation that can be expressed by a proportionality factor in the Coulomb term of Eq. 4, the *Madelung constant* (Madelung 1918). For a NaCl crystal structure, we have

$$A = \frac{6}{\sqrt{1}} - \frac{12}{\sqrt{2}} + \frac{8}{\sqrt{3}} - \frac{6}{\sqrt{4}} + \frac{25}{\sqrt{5}} - + \cdots, \tag{5}$$

where each term presents the number of equidistant neighbors in the numerator and the corresponding distance (in lattice units) in the denominator. This series is only slowly converging. Ewald's method (the theta-function method) is powerful and facilitates the numerical evaluation of A. For NaCl, we obtain from (Madelung 1918; Born and Landé 1918)

$$eV = -A\frac{e^2}{4\pi\varepsilon_0 r_e} + \frac{\beta'}{r_e^m} \tag{6}$$

with $A = 1.7476$, a lattice binding energy of $eV_{min}^{(A)} = H^0(\text{NaCl}) = 7.948$ eV, compared to an experimental value of 7.934 eV. Here, β' and m are empirically obtained from the observed lattice constant and compressibility. The Madelung constant is listed for several AB compounds in Table 1 (see Sherman 1932).

The Born–Haber cyclic process is an empirical way of obtaining the lattice energy, i.e., the binding energy per mole. The process starts with the solid metal and gaseous halogen and adds the heat of sublimation $W_{subl}(\text{Na})$ and the dissociation energy $(1/2)W_{diss}(Cl_2)$; it further adds the ionization energy $W_{ion}(\text{Na})$ and the electron affinity $W_{el\ aff}(\text{Cl})$ in order to obtain a diluted gas of Na^+ and Cl^- ions; all of

Table 1 Madelung constant for a number of crystal structures

Crystal structure	Madelung constant
NaCl	1.7476
CsCl	1.7627
Zincblende	1.6381
Wurtzite	1.6410
CaF_2	5.0388
Cu_2O	4.1155
TiO_2 (rutile)	4.8160

[2] β can be eliminated from the minimum condition $(dV/dr|_{r_e} = 0)$. One obtains $\beta = e^2 r_e^{m-1}/(4\pi\varepsilon_0 m)$ and as *cohesive energy* $eV_{min} = -e^2(m-1)/(4\pi\varepsilon_0 m\, r_e)$.

these energies can be obtained experimentally. These ions can be brought together from infinity to form the NaCl crystal by gaining the unknown lattice energy H^0(NaCl). This entire sum of processes must equal the heat of formation W^0(NaCl) which can be determined experimentally (Born 1919; Haber 1919):

$$W^0_{\text{solid}} = \left\{ W_{\text{subl}}(\text{Na}) + W_{\text{ion}}(\text{Na}) + \frac{1}{2} W_{\text{diss}}(\text{Cl}_2) + W_{\text{el aff}}(\text{Cl}_2) \right\} + H^0(\text{NaCl}). \quad (7)$$

In this equation, a minor correction of an isothermal compression of NaCl from $p = 0$ to $p = 1$ atm, heating from $T = 0$ K to room temperature, and an adiabatic expansion of the ion gases to $p = 0$ have been neglected. The corresponding energies almost cancel. The error is $<1\%$.

A listing of lattice constants (for definition, see ▶ Chap. 3, "The Structure of Semiconductors") of a number of predominantly ionic AB compounds is given in Table 2.

Table 2 Lattice constants a (in Å) and ratio of lattice constants c/a for simple AB compounds (After Weißmantel and Hamann 1979) (for explanation of the different crystal structures, see ▶ Chap. 3, "The Structure of Semiconductors")

NaCl structure				CsCl structure		Zincblende		Wurtzite		
	a		a		a		a		a	c/a
AgF	4.93	NaBr	5.973	BaS	6.363	AlP	5.431	AgI	4.589	1.63
AgCl	5.547	NaI	6.433	CsCl	4.118	AlAs	5.631	AlN	3.110	1.60
AgBr	5.775	PbS	5.935	CsBr	4.296	AlSb	6.142	BeO	2.700	1.63
BaO	5.534	PbSe	6.152	CsI	4.571	BeS	4.86	CdS	4.139	1.62
BaS	6.363	PbTe	6.353	TiI	4.206	BeSe	5.08	CdSe	4.309	1.63
BaSe	6.633	RbF	5.651	TlCl	3.842	BeTe	5.551	GaN	3.186	1.62
BaTe	7.000	RbCl	6.553	TlBr	3.978	CSi	4.357	InN	3.540	1.61
CaO	4.807	RbBr	6.868	TiI	4.198	CdS	5.832	MgTe	4.529	1.62
CaS	5.690	RbI	7.341	NH_4Cl	3.874	CdSe	6.052	MnS	3.984	1.62
CaSe	5.992	SnAs	5.692	NH_4Br	4.055	CdTe	6.423	MnSe	4.128	1.63
CaTe	6.358	SnTe	6.298	NH_4I	4.379	CuF	4.264	TaN	3.056	–
CdO	4.698	SrO	5.156	$TiNO_3$	4.31	CuCl	5.417	ZnO	3.249	1.60
KF	5.351	SrS	5.582	CsCN	4.25	CuBr	5.691	ZnS	3.819	1.64
KCl	6.283	SrSe	6.022			GaP	5.447	NH_4F	4.399	1.60
KBr	6.599	SrTc	6.483			GaAs	5.646			
KI	7.066	TaC	4.454			GaSb	6.130			
LiF	4.025	TiC	4.329			HgSe	6.082			
LiCl	5.130	TiN	4.244			HgTe	6.373			
LiBr	5.501	TiO	4.244			InAs	6.048			
LiI	6.012	VC	4.158			InSb	6.474			
MgO	4.211	VN	4.137			MnS	5.611			
MgS	5.200	VO	4.108			MnSe	5.832			
MgSe	5.462	ZrC	4.696			ZnS	5.423			
NaF	4.629	ZrN	4.619			ZnSe	5.661			
NaCl	5.693					ZnTe	6.082			

1.2 Covalent Bonding

Covalent bonding is caused by two electrons that are shared between two atoms: they form an electron bridge. The bridge formation can be understood quantum-mechanically by a nonspherical electron-density distribution that extends between the bonded atoms. Examples of such density distributions are shown schematically in Fig. 3 for a molecule formation with electrons in a 1*s* or 2*p* state, e.g., for H_2 or F_2, respectively.

If an atom approaching another atom of the same element has in its protruding part of the electron-density distribution an unpaired electron with antiparallel spin, both eigenfunctions may overlap; the Pauli principle is not violated. Their combined wavefunction ($\psi_+ = \psi_A + \psi_B$) yields an increased electron density $|\psi_+|^2$ in the overlap region (see Fig. 4a); the result is an *attractive force* between these two atoms in the direction of the overlapping eigenfunctions. This is the state of lowest energy of the two atoms, the *bonding state*. There is also a state of higher energy, the *antibonding state*, with $\psi_- = \psi_A - \psi_B$ in which the spin of both electrons is parallel. Here, the electrons are strongly repulsed because of the Pauli principle, and the electron clouds

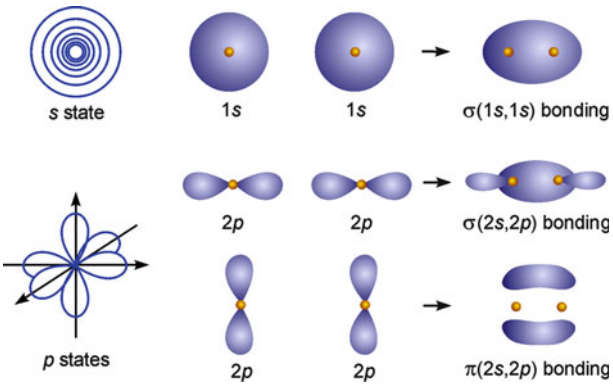


Fig. 3 Atomic and molecular electron-density distribution for $\sigma(s)$, $\sigma(p)$, and $\pi(p)$ bonding (After Weißmantel and Hamann 1979)

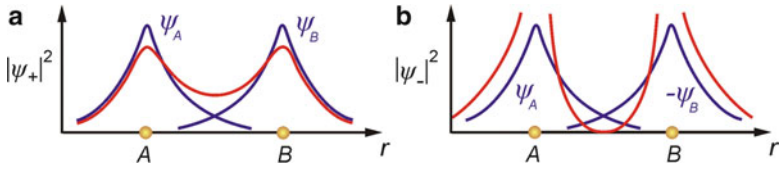


Fig. 4 Wavefunctions of one-electron states [*blue curves* – identical in (**a**) and (**b**)] and probability function to find one electron (*red curves*) in (**a**), a bonding state, and (**b**), an antibonding state, showing finite and vanishing electron density at the center between atoms *A* and *B* for these two states, respectively [observe the plotting of $-\psi_B$ in (**b**)]. The picture of these two one-electron states shown here shall not be confused with the two-electron potential given in Fig. 5

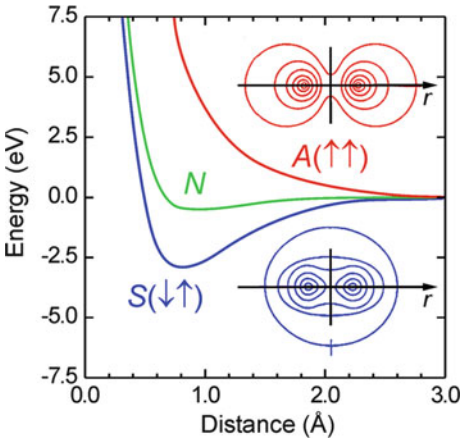


Fig. 5 Potential energy for the two valence electrons of two covalently bound hydrogen atoms approaching each other. *Upper curve*, antibonding state; *lower curve*, bonding state; *middle curve*, bonding potential from free atom charge distribution. Charge-density distributions shown in the *inset* are for the two covalent states (After Kittel 1996)

cannot penetrate each other; therefore, the electron density between both atoms vanishes (Fig. 4b). The resulting potential distribution as a function of the interatomic distance between two hydrogen atoms forming an H_2 molecule is given in Fig. 5. In this figure, the ground state (bonding) *S* and the excited state (antibonding) *A* are shown. The figure also contains as center curve the classical contribution (electrostatics) of two H atoms with a charge density corresponding to free atoms. Such bonding is small compared with the covalent bonding.

The bond length (center-to-center distance) for some semiconductors and molecules is listed in Table 3. In organic molecules, the bond length between C atoms decreases with increasing bonding valency as shown in the table.

With increasingly missing unpaired electrons in the outer shell, more than one atom of the same kind can be bound to each other. The number of bonded atoms is given by the following valency: monovalent atoms can form only diatomic molecules; divalent atoms, such as S or Se, can form chains; and trivalent atoms, such as As, can form two-dimensional (layered) lattices. Solids are formed from such elements by involving other bonding forces between the molecules, chains, or layers, e.g., van der Waals forces (see Sect. 3). Only tetravalent elements can form three-dimensional lattices which are covalently bound (e.g., Si).

Tetrahedrally Bound Elements Silicon has four electrons in its outer shell. In the ground state of an isolated atom, two of the electrons occupy the *s* state and two of them occupy *p* states, with a $2s^2 2p^2$ configuration. By investing a certain amount of promotion energy,[3] this s^2p^2 configuration is changed into an sp^3 configuration, in

[3]The promotion energy is 4.3, 3.5, and 3.3 eV for C, Si, and α-Sn, respectively. However, when forming bonds by establishing electron bridges to neighboring atoms, a substantially larger energy is gained, therefore resulting in net binding forces. Diamond has the highest cohesive energy in this series, despite the fact that its promotion energy is the largest, because its sp^3-sp^3 C–C bonds are the strongest (see Harrison 1980).

Table 3 Bond lengths relevant to organic molecules, α-Si, and related semiconductors (After Cotton and Wilkinson 1972)

Bond	Bond length (Å)	Bond	Bond length (Å)
C–C	1.54	Si–Si	2.35
C = C	1.38	Si–H	1.48
C = C	1.42 (graphite)	Ge–Ge	2.45
C ≡ C	1.21	Ge–H	1.55
C–H	1.09 (sp^3)	C–Si	1.87

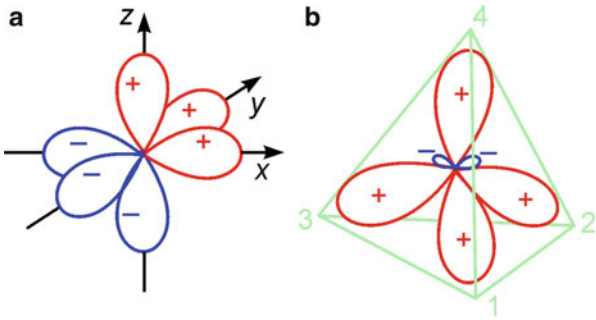


Fig. 6 Linear combination (hybridization) of a 1*s* function (spherical) with 3*p* functions (**a**) results in four sp^3 functions (**b**) which extend toward the four tetrahedra axes 1–4 and result in strongly directional bonding with a bond angle of 109.47°

which an unpaired electron sits in each one of the four singly occupied orbitals with tetrahedral geometry (see Fig. 6). From the *s* orbital and the three *p* orbitals, four linear combinations can be formed (depending upon the choice of signs: +++, +−−, −+−, −−+), represented as $\sigma_i = 1/2\ (\varphi_s \pm \varphi_{px} \pm \varphi_{py} \pm \varphi_{pz})$. This is referred to as *hybridization*, with σ_i as the *hybrid function* responsible for bonding. When we bring together a large number of Si atoms, they arrange themselves such that each of them has four neighbors in tetrahedral geometry as shown in Fig. 9 of ▶ Chap. 3, "The Structure of Semiconductors." Each atom then forms four electron bridges to its neighbors, in which each one is occupied with two electrons of opposite spin, as shown for the center atom in Fig. 7a. Such bridges become evident in a density profile within the (110) plane shown for two adjacent unit cells in Fig. 7b.

In contrast to the ionic bond, the covalent bond is angular dependent, since the protruding atomic eigenfunctions extend in well-defined directions. Covalent bonding is therefore a *directional* and *saturable bonding*; the corresponding force is known as a *chemical valence force*, and acts in exactly as many directions as the valency describes.

1.3 Mixed Bonding

Crystals that are bonded partially by ionic and partially by covalent forces are referred to as mixed-bond crystals. Most actual semiconductors have a fraction of covalent and ionic bonding components (see, e.g., Mooser and Pearson 1956).

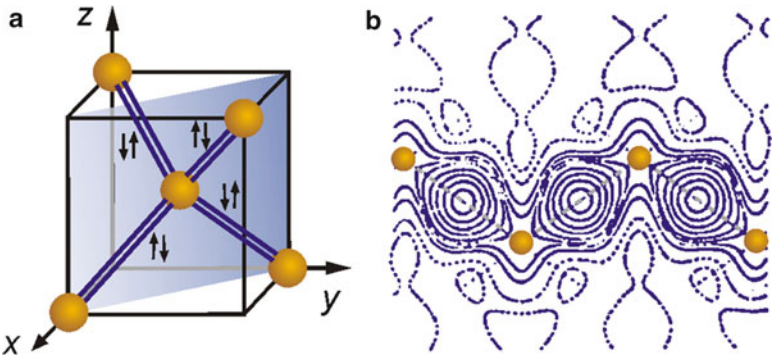


Fig. 7 (**a**) Unit cell of diamond with pairs of electrons indicated between adjacent atoms. (**b**) Electron-density profile within the (110) plane (After Dawson 1967)

Tetrahedrally Bonded Binaries Grimm and Sommerfeld predicted in 1926 that binary compounds with covalent character and (in average) 4 electrons per atom have tetrahedral coordination for both atoms. Based on this rule, Welker and Weiss (1954) predicted desirable semiconducting properties for III–V compounds.[4] Semiconducting III–V and II–VI compounds are bound in a *mixed bonding*, in which electron bridges exist, i.e., the bonding is directed, but the electron pair forming the bridge sits closer to the (more electronegative) anion. This *degree of ionicity* increases for these compounds with an increased difference in electronegativity (Fig. 8) from III–V to I–VII compounds and within one class of compounds, e.g., from RbI to LiF (see also Table 4).

The mixed bonding may be expressed as the sum of the wavefunctions describing covalent and ionic bonding:

$$\psi = a\psi_{\text{cov}} + b\psi_{\text{ion}}, \tag{8}$$

with the ratio b/a defining the *ionicity* of the bonding. This bonding can also be described as rapidly alternating between that of covalent and ionic. Over an average time period, a fraction of ionicity (b/a) results. The ionicity of the bonding can be described by a static *effective ion charge* e^*, as opposed to a dynamic effective ion charge (discussed in Sect. 1.1 in ▶ Chap. 11, "Photon–Phonon Interaction"), which is less by a fraction on the order of b/a than in a purely ionic compound with the charge given by the valency. For instance, in CdS, the divalent behavior of Cd and S could result in a doubly charged $Cd^{++}S^{--}$ lattice, while measurements of the electric dipole moment indicate an effective charge of 0.49 for CdS. The static effective charge for other II–VI and III–V compounds is given in Table 4.

The effective charge concept can be confusing if one does not clearly identify the ionic state of the system. For instance, in the case of CdS, a purely ionic state is $Cd^{2+}S^{2-}$, as opposed to the covalent state of $Cd^{2-}S^{2+}$ (which is equivalent to the

[4]Meaning compounds between one element of group III and one element of group V on the periodic system of elements (cf. Fig. 3 in ▶ Chap. 1, "Properties and Growth of Semiconductors")

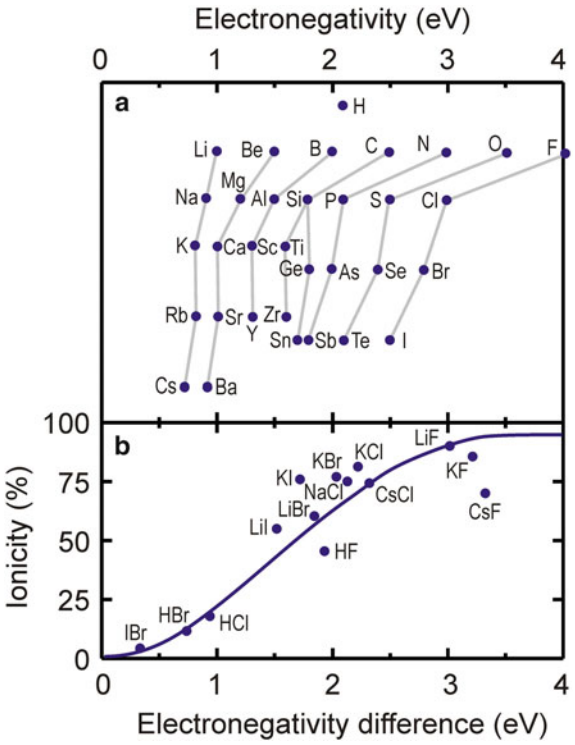


Fig. 8 (**a**) Electronegativity of the elements with groups from the periodic table of elements identified by interconnecting lines. (**b**) Ionicity of alkali halides and halide molecules as a function of the difference in electronegativity (After Pauling 1960)

Table 4 Static effective charges of partially covalent *AB* compounds (After Coulson et al. 1962)

Compound	e^*/e	Compound	e^*/e
ZnO	0.60	BN	0.43
		AlN	0.56
		GaN	0.55
		InN	0.58
ZnS	0.47	BP	0.32
CdS	0.49	AlP	0.46
HgS	0.46	GaP	0.45
		InP	0.49
ZnSe	0.47	AlAs	0.47
CdSe	0.49	GaAs	0.46
HgSe	0.46	InAs	0.49
ZnTe	0.45	AISb	0.44
CdTe	0.47	GaSb	0.43
HgTe	0.49	InSb	0.46

Si^0 Si^0 configuration). In other words, the covalent state is that in which both Cd and S have four valence electrons and are connected to each other by a double bond. This must not be confused with the neutral Cd^0S^0 configuration, which is a mixed-bonding state. The expression for the static effective charge (see Coulson et al. 1962) is

$$\frac{e*}{e} = \frac{N(a/b)^2 - (8-N)}{1 + (a/b)^2}, \tag{9}$$

with N as the valency. For $N = 2$, the effective charge vanishes when $a/b = \sqrt{3}$. For $N = 3$ in III–V compounds, e^* vanishes when $a/b = \sqrt{5/3}$, and for group IV semiconductors when $a = b$.

In crystals, low coordination numbers (typically 4) signify a considerable covalent contribution to the bonding.

The different degree of bridge formation in crystals with mixed bonding (Fig. 9) can be made visible by a Fourier analysis of x-ray diffraction from which the electron-density distribution around each atom can be obtained. This is shown for a mostly ionic crystal in Fig. 10a and for a mostly covalent crystal in Fig. 10b.

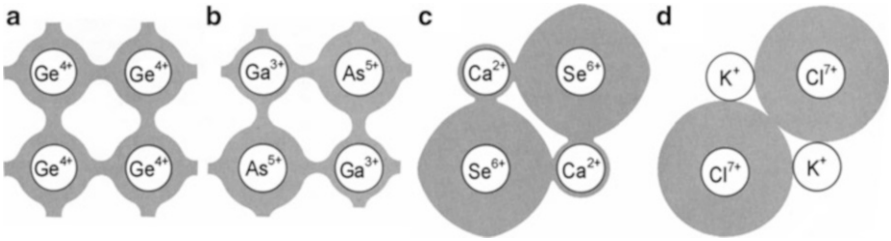


Fig. 9 Schematic sketch of mixed bonding from (**a**) nearly perfect covalent in Ge to (**d**) perfect ionic in KCl. It shows diminishing bridge formation and increasing cloud formation of electrons around anions with increasing ionicity (After Ashcroft and Mermin 1976)

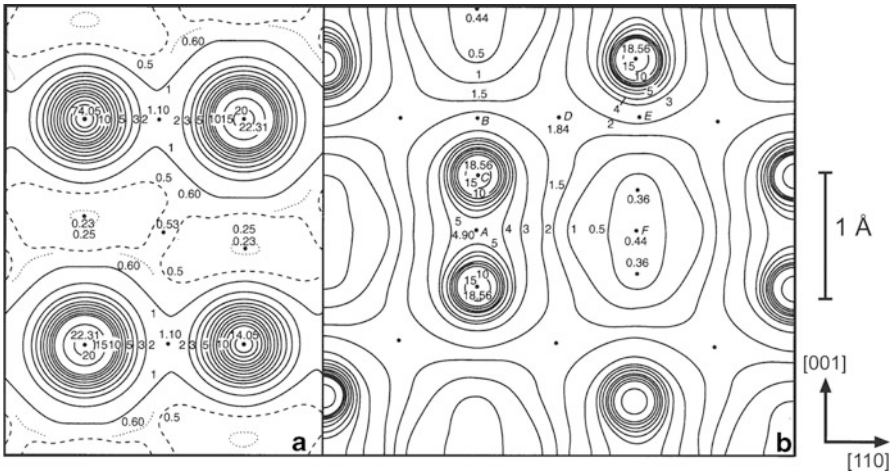


Fig. 10 Electron-density distribution obtained by Fourier analysis of the x-ray diffraction pattern of (**a**) NaCl, (110) plane, and (**b**) diamond, (110) plane (After Brill et al. 1939)

2 Metallic Bonding

Metallic bonding can be understood as a collective interaction of a mobile electron fluid with metal ions. Metallic bonding occurs when the number of valence electrons is only a small fraction of the coordination number; then, neither an ionic nor a covalent bond can be established. Metallic bonding of simple metals, e.g., alkali metals, can be modeled by assuming that each metal atom has given up its valence electron, forming a lattice of positively charged ions, submerged in a fluid of electrons. Between the repulsive electron–electron and ion–ion interactions and the attractive electron–ion interaction, a net attractive binding energy results, which is *nondirectional* and *not saturable*, and results in *close-packed structures* with high coordination numbers (8 or 12; Wigner and Seitz 1933), but relatively wide spacing between the submerged metal ions (Table 5). Such metals have low binding energies (~1 eV atom $^{-1}$) and high compressibility. They are mechanically soft, since the nondirectional lattice forces exert little resistance against plastic deformation. This makes metals attractive for forming and machining.

In other metals, such as transition-group elements, the bonding may be described as due to covalent bonds which rapidly hop from atom pair to atom pair. Again, free electrons are engaged in this resonance-type bonding. These metals have a higher binding energy of ~4 to 9 eV atom^{-1} and an interatomic distance that is closer to that given by the sum of ionic radii (Table 5). They are substantially harder when located in the middle of the transition metal row, e.g., Mo and W (Ashcroft and Mermin 1976).

In semiconductors with a very high density of free carriers, metallic binding forces may contribute a small fraction to the lattice bond, interfering with the predominant covalent bonding and usually weakening it, since these electrons are obtained by ionizing other bonds. Changes in the mechanical strength of the lattice can be observed in photoconductors in which a high density of free carriers can be created by light (Gorid'ko et al. 1961). For more information, see Ziman (1969) and Harrison (1966).

3 Further Types of Bonding in Solids

van der Waals Bonding Noble gas atoms or molecules with saturated covalent bonds can be bound to each other by dipole–dipole interaction (Debye force). The

Table 5 Ionic radii r_i and half the nearest-neighbor distances in metals r_m in Å (After Ashcroft and Mermin 1976)

Metal	r_i	r_m	r_m/r_i	Transition metal	r_i	r_m	r_m/r_i
Li	0.60	1.51	2.52	Cu	0.96	1.28	1.33
Na	0.95	1.83	1.93	Ag	1.26	1.45	1.15
K	1.33	2.26	1.70	Au	1.37	1.44	1.05
Rb	1.48	2.42	1.64				
Cs	1.69	2.62	1.55				

dipole is created between the nucleus (nuclei) of the atom (molecule) and the cloud of electrons moving around these nuclei and forms a fluctuating dipole moment even for a spherically symmetrical atom. The interaction creates very weak, nonsaturable attractive forces. The weakness of this van der Waals interaction results in low melting points (Table 6) and soft *molecule crystals*. The bonding potential $V(r)$ is complemented by a short-range repulsive component originating from the Coulomb repulsion of the core electrons and nuclei; this component can be approximated by an exponential term, yielding, according to Buckingham (1938), the potential energy

$$eV = -\frac{A}{r^6} + B\ e^{-\alpha r} \tag{10}$$

with empirical parameters A, B, and α. van der Waals forces (van der Waals 1873) are the main binding forces of organic semiconductors (see Sect. 3.3).

Hydrogen Bonding Hydrogen bonding (Fig. 11) is a type of ionic bonding in which the hydrogen atom has lost its electron to another atom of high electronegativity. The remaining proton establishes a strong Coulomb attraction. This force is not saturable. However, because of the small size of the proton, hydrogen bonding is strongly localized, and spatially no more than two ions have space to be attracted to it. When part of a molecule, the hydrogen bond – although ionic in nature – fixes the direction of the attached atom because of space consideration. It should not, however, be confused with the covalent bonding of hydrogen that occurs at dangling bonds (see Sect. 3.2.2 in ▶ Chap. 3, "The Structure of Semiconductors") in semiconductors, e.g., at the crystallite interfaces of polycrystalline Si or in amorphous Si:H.

Table 6 Melting points (m.p.) of crystals bonded by van der Waals interaction

Noble gas crystal		Organic crystal		
Atom	m.p. (K)	Molecule	Formula	m.p. (°C)
He	0.95 (25 bar)	Anthracene	$C_{14}H_{10}$	216
Ne	25	Tetracene	$C_{18}H_{12}$	357
Ar	84	Pentacene	$C_{22}H_{14}$	300
Kr	116	Quaterthiophene	$C_{16}H_{10}\ S_4$	~213
Xe	162	Quinquethiophene	$C_{20}H_{12}\ S_5$	253
Rn	202	Hexathiophene	$C_{24}H_{14}\ S_6$	290
		Rubrene	$C_{42}H_{28}$	315–330

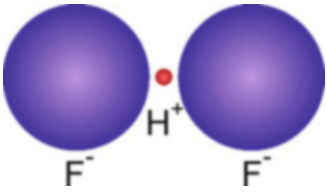


Fig. 11 Hydrogen bonding between a positive hydrogen ion (proton) and two ions (coordination number 2)

Intermediate Valence Bonding An interesting group of semiconductors are *transition-metal compounds*. The transition metals have partially filled *inner* 3*d*, 4*d*, 5*d*, or 4*f* shells and a filled outer shell that provides a shielding effect to the valence electrons. In these compounds, the crystal field has a reduced effect. Some of these compounds show *intermediate valence bonding*. The resulting unusual properties range from resonant valence-exchange transport in copper oxide compounds (Anderson et al. 1987) to giant magnetoresistance and very large magneto-optical effects in rare-earth semiconductors. For a review, see Holtzberg et al. (1980).

Other Bonding Considerations Other, more subtle bonding considerations have gained a great deal of interest because of their attractive properties. These are related to magnetic and special dielectric properties, to superconductivity, as well as to other exotic effects.

For instance, diluted ("semi-") magnetic semiconductors such as the alloy $Cd_{1-x}Mn_xTe$ (Furdyna 1982, 1986; Brandt and Moshchalkov 1984; Wei and Zunger 1986; Goede and Heimbrodt 1988) show interesting magneto-optical properties. They change from paramagnetic ($x < 0.17$) to antiferromagnetic ($0.6 < x$) and to the ferro- or antiferromagnetic behavior of MnTe (Sect. 2 in ▶ Chap. 9, "Magnetic Semiconductors"); they exhibit giant magneto-optical effects and bound magnetic polarons and offer opportunities for opto-electric devices that are tunable by magnetic fields.

These materials favor specific structures and permit the existence of certain quasiparticles, such as small polarons or Frenkel excitons. The discussion requires a substantial amount of understanding of the related physical effects and is therefore postponed to a more appropriate section of this book (see also Phillips 1973; Harrison 1980; Ehrenreich 1987).

3.1 Atomic and Ionic Radii

The equilibrium distances between atoms in a crystal define atomic radii when assuming hard-sphere atoms touching each other. In reality, however, these radii are soft with some variation of the electronic eigenfunctions and, for crystals with significant covalent fraction, with dependence on the angular atomic arrangement. However, for many crystals, the hard-sphere radii are very useful for most lattice estimates.

When comparing the lattice constants of chemically similar crystals, such as NaCl, NaBr, KCl, and KBr, one can determine the radii of the involved ions (Na^+, K^+, Cl^-, and Br^-) if at least one radius is known independently. Goldschmidt (1927) used the radii of F^- and 0^{--} for calibration. Consequently, listings of other ionic radii are therefore referred to as *Goldschmidt radii*. These radii are independent of the compound in which the atoms are incorporated as long as they exhibit the same type of bonding. One distinguishes atomic, ionic, metallic, and van der Waals radii. Ionic radii vary with changing valency.

A list of the most important ion and atomic radii is given in Table 7. The drastic change in radii with changing bonding force (Mooser and Pearson 1956) is best

Table 7 Covalent (effective ionic charge $e^* = 0$) and standard ionic (identified by $\pm e^*$) radii in Å

	+1	0		+2	0		+3	0		+4	0	−4		0	−3		0	−2
Li	0.68	1.34	Be	0.30	0.90	B	0.16	0.88	C		0.77	2.60	N	0.70	1.71	O	0.73	1.46
Na	0.98	1.54	Mg	0.65	1.30	Al	0.45	1.26	Si	0.38	1.17	2.71	P	1.10	2.12	S	1.04	1.90
K	1.33	1.96	Ca	0.94	1.74	Sc	0.68		Ti	0.60			As	1.18	2.22	Se	1.14	2.02
Cu	0.96		Zn	0.74	1.31	Ga	0.62	1.26	Ge	0.53	1.22	2.72						
Rb	1.48		Sr	1.10		Y	0.88		Zn	0.77			Sb	1.36	2.45	Te	1.32	2.22
Ag	1.26		Cd	0.97	1.48	In	0.81	1.44	Sn		1.40	2.94						
Cs	1.67		Ba	1.2		La	1.04		Ce	0.92			Bi	1.46		Po		2.30
Au	1.37		Hg	1.10	1.48	Tl	0.95	1.47	Pb	0.84	1.46							

demonstrated by comparing a few typical examples for some typical elements incorporated in semiconductors (Fig. 12). For more recent estimates of tetrahedral covalent radii, see van Vechten and Phillips (1970).

The deviation from strict rigidity, i.e., the softness of the ionic spheres, is conventionally considered by using a softness parameter r_0 in the exponential repulsion formula (Eq. 2). This parameter is listed for a number of ions in Table 8.

This softness also results in a change of the standard ionic radii as a function of the number of surrounding atoms. A small correction Δ_m in the interionic distance is listed in Table 9. This needs to be considered when crystals with different coordination numbers m, i.e., the number of surrounding atoms, are compared with each other (e.g., CsCl and NaCl).

With increasing atomic number, the atomic (or ionic) radius of homologous elements increases. The cohesive force therefore decreases with increasing atomic (ionic) radii. Thus, compounds formed by the same bonding forces, and crystallizing with similar crystal structure, show a decrease, for example, in hardness,[5] melting point, and band gap, but an increase in dielectric constant and carrier mobility (see the respective sections).

The ratio of ionic radii determines the preferred crystal structure of *ionic* compounds. This is caused by the fact that the energy gain of a crystal is increased with every additional atom that can be added per unit volume. When several possible atomic configurations are considered, the material crystallizes in a modification that *maximizes* the number of atoms in a given volume. This represents the state of lowest potential energy of the crystal, which is the most stable one. An elemental crystal with isotropic radial interatomic forces will therefore crystallize in a close-packed structure. In a binary crystal, the ratio of atomic radii will influence the possible crystal structure. For isotropic nonsaturable interatomic forces, the resulting stable lattices are shown in Table 11 for different ratios of the ion radii (see following sections).

When a substantial amount of covalent bonding forces are involved, the rules to select a stable crystal lattice for a given compound are more complex. Here, atomic bond length and bond angles must be considered. Both can now be determined from basic principal density-functions calculations (see Sect. 2 in ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"). We can then define atomic radii from the turning point of the electron-density distribution of each atom and obtain an angular-dependent internal energy scale from these calculations (Zunger and Cohen 1979). Using axes constructed from these radii, one obtains well-separated domains in which only one crystal structure is observed for binary compounds (Zunger and Cohen 1979; Villars and Calvert 1985; Yeh et al. 1992).

[5]This empirical quantity can be defined in several ways (e.g., as Mohs, Vickers, or Brinell *hardness*) and is a macroscopic mechanical representation of the cohesive strength of the lattice (Smith and Hashemi 2010). In Table 10, the often used Mohs hardness is listed, which orders the listed minerals according to the ability of the higher-numbered one to scratch the lower-numbered minerals. This simple scale remains relevant for metrology (Davis 2018) and can reasonably be predicted using machine learning (Garnett 2019).

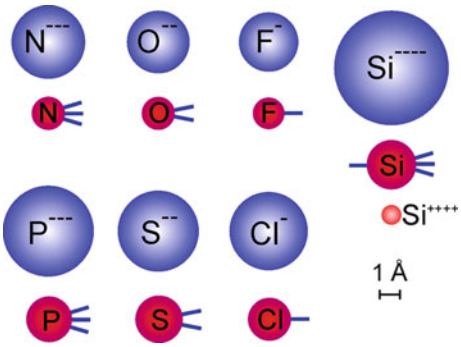


Fig. 12 Scale drawing of rigid sphere atoms with different bonding character [ionic or covalent, identified by the appropriate number of minus signs (*upper* row) or valence lines (*lower* row), respectively]

Table 8 Repulsion potential softness parameters (Eq. 2) in Å (After Shanker and Kumar 1987)

Ion	$r_0(th)$	$r_0(exp)$	Ion	$r_0(th)$	$r_0(exp)$
Li^-	0.069	0.042	F^-	0.179	0.215
Na^+	0.079	0.090	Cl^-	0.238	0.224
K^+	0.106	0.108	Br^-	0.258	0.254
Rb^+	0.115	0.089	I^-	0.289	0.315
Cs^+	0.130	0.100			

Table 9 Change of interatomic distance Δ_m (in Å) for compounds deviating from coordination number $m = 6$

m	Δ_m	m	Δ_m	m	Δ_m	Δ_m	Δ_m
1	−0.50	4	−0.11	7	+0.04	10	+0.14
2	−0.31	5	−0.05	8	+0.08	11	+0.17
3	−0.19	6	0	9	+0.11	12	+0.19

Table 10 Mohs hardness

Material	Chemistry	Lattice type	Hardness
Talc	$Mg_3H_2SiO_{12\text{-aq}}$	Layer lattice	1
Gypsum	$CaSO_4 \cdot H_2O$	Layer lattice	2
Iceland spar	$CaCO_3$	Layer lattice	3
Fluorite	CaF_2	Ion lattice	4
Apatite	$Ca_5F(PO_4)_3$	Ion lattice	5
Orthoclase	$KAlSi_3O_8$	SiO_4 frame	6
Quartz	SiO_2	SiO_4 frame	7
Topaz	$Al_2F_2SiO_4$	Mixed ion-valency lattice	8
Corundum	Al_2O_3	Valency lattice	9
Diamond	C	Valency lattice	10

Table 11 Preferred lattice structure for AB compounds with ionic binding forces (After Goldschmidt 1927)

r_A/r_B	Preferred stable lattice
<0.22	None
0.22. . .0.41	Zincblende or wurtzite
0.41. . .0.72	NaCl lattice
>0.72	CsCl lattice

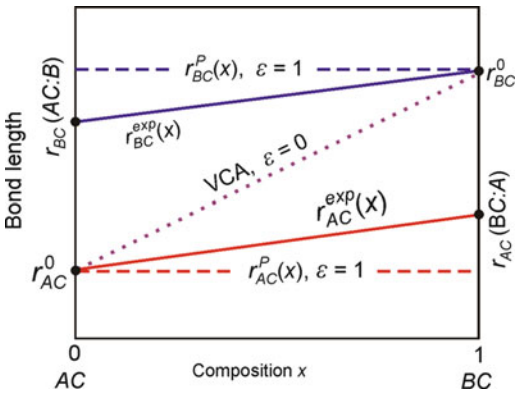


Fig. 13 Variation of bond length in an $A_{1-x}B_xC$ alloy for rigid atoms ($\varepsilon = 1$), virtual crystal approximation ($\varepsilon = 0$), and experimentally observed relaxation

3.2 Bond-Length Relaxation in Alloys

The lattice constant of alloys $A_{1-x}B_xC$ of binary compounds AC and BC with respective lattice constants a_{AC} and a_{BC} interpolates according to the concentration

$$a(x) = (1 - x)a_{AC} + xa_{BC} \tag{11}$$

when they crystallize with the same crystal structure (Vegard's rule, Vegard 1921). However, the bond length between any of the three pairs of atoms is neither a constant, as suggested from the use of constant atomic radii (Pauling 1960), nor a linear interpolation as shown by the dotted line in Fig. 13 for total relaxation of the bond of atom B in a different chemical environment AC (or of A in BC).

This nonrigidity of atoms is important when incorporating isovalent impurities into the lattice of a semiconductor (doping) and estimating the resulting deformation of the surrounding lattice. With the bond length r_{BC} within the AC lattice (see Table 12), one defines a relaxation parameter

$$\varepsilon = \frac{r_{BC}(AC:B) - r^0_{AC}}{r^0_{BC} - r^0_{AC}}. \tag{12}$$

The superscript 0 indicates the undisturbed pure crystal, the notation AC:B indicates B as doping element with a sufficiently small density incorporated in an AC compound, so that B–B interaction can be neglected.

The relaxation parameter can be estimated from the bond-stretching and bond-bending force constants α and β (see Table 13), according to Martins and Zunger (1984),

$$\varepsilon = \frac{1}{1 + \frac{1}{6}\frac{\alpha_{AC}}{\alpha_{BC}}\left(1 + 10\frac{\beta_{AC}}{\alpha_{AC}}\right)}, \tag{13}$$

yielding values of ε typically near 0.7 (see Table 12); that is, isovalent impurity atoms behave more like rigid atoms ($\varepsilon = 1$) than totally relaxed atoms ($\varepsilon = 0$) in a virtual crystal approximation (Eq. 11).

Table 12 Bond length of an isovalent impurity in a given host lattice and bond-length relaxation parameter (After Martins and Zunger 1984)

System	r_{BC} (AC:B) (Å)	ε	System	r_{BC} (AC:B) (Å)	ε
AlP:In	2.480	0.65	InP:Al	2.414	0.73
GaP:In	2.474	0.63	InP:Ga	2.409	0.73
AlAs:In	2.553	0.60	InAs:Al	2.495	0.74
GaAs:In	2.556	0.62	InAs:Ga	2.495	0.73
AlSb:In	2.746	0.61	InSb:Al	2.693	0.75
GaSb:In	2.739	0.60	InSb:Ga	2.683	0.74
AlP:As	2.422	0.65	AlAs:P	2.395	0.67
AlP:Sb	2.542	0.61	AlSb:P	2.444	0.73
AlAs:Sb	2.574	0.60	AlSb:As	2.510	0.71
GaP:As	2.414	0.62	GaAs:P	2.387	0.68
GaP:Sb	2.519	0.57	GaSb:P	2.436	0.73
GaAs:Sb	2.564	0.60	GaSb:As	2.505	0.70
InP:As	2.595	0.67	InAs:P	2.562	0.74
InP:Sb	2.700	0.60	InSb:P	2.597	0.79
InAs:Sb	2.739	0.64	InSb:As	2.667	0.75
ZnS:Se	2.420	0.70	ZnSe:S	2.367	0.78
ZnS:Te	2.539	0.67	ZnTe:S	2.407	0.78
ZnSe:Te	2.584	0.71	ZnTe:Se	2.502	0.74
β-HgS:Se	2.611	0.76	HgSe:S	2.553	0.80
β-HgS:Te	2.716	0.71	HgTe:S	2.579	0.82
HgSe:Te	2.748	0.74	HgTe:Se	2.665	0.80
ZnS:Hg	2.482	0.73	β-HgS:Zn	2.380	0.80
ZnSe:Hg	2.587	0.74	HgSe:Zn	2.494	0.78
ZnTe:Cd	2.755	0.70	CdTe:Zn	2.674	0.78
ZnTe:Hg	2.748	0.69	HgTe:Zn	2.673	0.78
γ-CuCl:Br	2.440	0.81	γ-CuBr:Cl	2.367	0.79
γ-CuCl:I	2.563	0.80	γ-CuI:Cl	2.407	0.76
γ-CuBr:I	2.585	0.79	γ-CuI:Br	2.500	0.76
C:Si	1.665	0.35	Si:C	2.009	0.74
Si:Ge	2.380	0.58	Ge:Si	2.419	0.63
Si:Sn	2.473	0.53	α-Sn:Si	2.645	0.70
Ge:Sn	2.549	0.55	α-Sn:Si	2.688	0.67

Table 13 Bond-length (d), bond-stretching (α), and bond-bending (β) force constants, calculated from elastic constants (After Martin 1970)

Crystal	d (Å)	α (N m^{-1})	β (N m^{-1})	Crystal	d (Å)	α (N m^{-1})	β (N m^{-1})
C	1.545	129.33	84.71	InP	2.541	43.04	6.24
Si	2.352	48.50	13.82	InAs	2.622	35.18	5.49
Ge	2.450	38.67	11.37	InSb	2.805	26.61	4.28
α-Sn	2.810	25.45	6.44	ZnS	2.342	44.92	4.81
SiC	1.888	88.	47.5	ZnSe	2.454	35.24	4.23
AlP	2.367	47.29	9.08	ZnTe	2.637	31.35	4.45
AlAs	2.451	43.05	9.86	CdTe	2.806	29.02	2.44
AlSb	2.656	35.35	6.79	*β-HgS*	2.534	41.33	2.56
GaP	2.360	47.32	10.46	HgSe	2.634	36.35	2.36
GaAs	2.448	41.19	8.94	HgTe	2.798	27.95	2.57
GaSb	2.640	33.16	7.23	γ-CuCl	2.341	22.9	1.01
				γ-CuBr	2.464	23.1	1.32
				γ-CuI	2.617	22.5	2.05

3.3 Bonding in Organic Crystals

Recently, semiconductors made from organic materials gained much advertence[6]; for a review, see Schwoerer and Wolf (2007). All organic semiconductors are solids comprising molecules with carbon atoms, which are bond by a system of *conjugated π electrons*. In such a system, two adjacent C atoms are not only bond by σ bonds (see Fig. 3), i.e., single bonds, but in addition by multiple (usually double) bonds. A simple example is the ethene molecule C_2H_4 illustrated in Fig. 14. Three of the four valence electrons ($2s^2$, $2p^2$) of each C atom form σ bonds from sp^2 hybrid orbitals: two to H atoms and one to the other C atom; all these bonds lie in one plane. The two remaining p_z electrons of the C atoms have their density distribution above and below this plane; they form an additional π bond, which is weaker than the strong σ bond because the overlap of the p_z wavefunctions of the adjacent C atoms is small.

The larger molecules of organic semiconductors have delocalized conjugated π electrons in alternating single and double bonds. The molecule may be linear or cyclic as illustrated in Fig. 15. Molecules of organic crystals are generally based on bonds with sp^2 hybridization, in contrast to the sp^3 hybridization of tetrahedrally bonded inorganic crystals discussed in Sect. 1.2.

The molecules are the building blocks of organic semiconductors. They may either be arranged in a regular order of a crystal as shown in Sect. 1.5 in ▶ Chap. 3, "The Structure of Semiconductors" or irregularly; while both forms are applied for organic electronics, generally best performance is obtained with crystalline modifications. The molecules of organic semiconductors do not have unsaturated bonds

[6]Two principal devices made of organic semiconductors recently entered the market: light-emitting diodes (OLEDs) and field-effect transistors (OFETs), processed as thin-film transistors (TFT). Prominent molecules used in organic (opto-) electronics are listed in Sect. 1.5 in ▶ Chap. 3, "The Structure of Semiconductors".

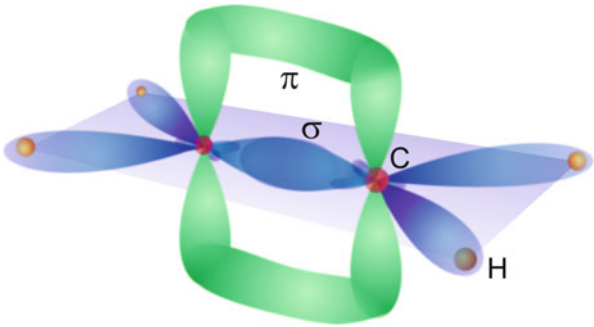


Fig. 14 σ (*blue*) and π (*green*) molecule orbitals of an ethene molecule C_2H_4. The *shaded area* illustrates that the carbon (C) and hydrogen (H) atoms lie in one plane

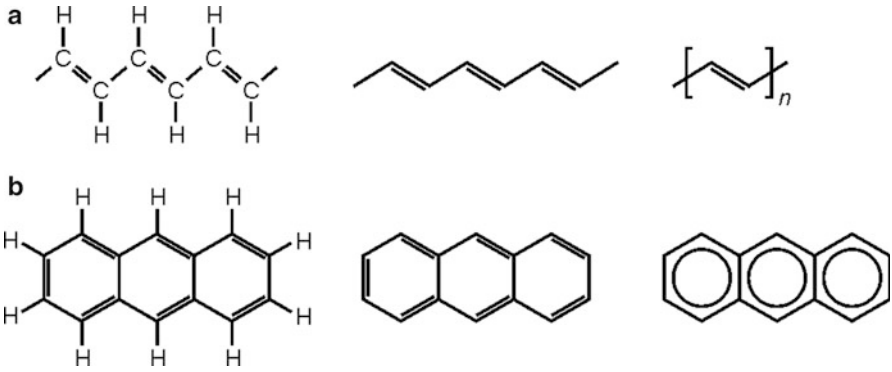


Fig. 15 Various representations of (**a**) polyethylene and (**b**) anthracene. Carbon atoms are generally left out and usually also the hydrogen atoms. The π electrons are indicated by double valence lines or, in cyclic molecules, also by a circle

and hence no free valences like the atoms of inorganic semiconductors. Intermolecular bonding forces of electrically neutral and nonpolar molecules are pure van der Waals interactions. If the molecules have a permanent dipole moment or polar substituents, static dipolar bonding and ionic or hydrogen bonding may be superimposed. In the presence of charges, a term $+ q_1q_2/r$ is added to the potential in Eq. 10 (see Starr and Williams 1977).

The bonding energy provided by van der Waals interaction is expressed by Eq. 10. However, in contrast to interatomic distances in noble gas crystals, the distance r in Eq. 10 is not well defined in organic solids: the distance between the molecules is of the same size as the extension of the molecules. A good description of experimental data is obtained, if the attractive potential between two neighboring molecules is calculated as the sum of all atom – atom potentials according to the Buckingham potential (Eq. 10), where r is the distance from an atom of one molecule to an atom of the other molecule (Kitaigorodskii 1966). Parameters of Eq. 10 for all three different pairs of atoms occurring in hydrocarbon molecules are listed in Table 14. The parameters were deduced from structural data and heat of sublimation of many organic molecules and apply in good approximation for all solids composed of hydrocarbon molecules

Table 14 Parameters of the Buckingham potential Eq. 10 for atom – atom pairs of neighboring hydrocarbon molecules (After Starr and Williams 1977)

Pair	A (kJ mol^{-1} Å^6)	B (kJ mol^{-1})	α (Å^{-1})
C–C	2,140	300,000	3.60
C–H	467	35,600	3.67
H–H	102	9,080	3.74

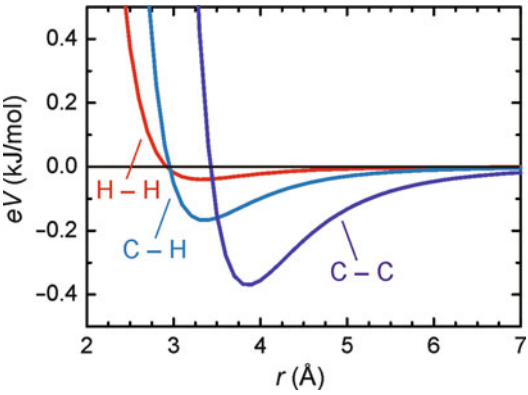


Fig. 16 Intermolecular potential of Eq. 10 for different atom – atom pairs of neighboring hydrocarbon molecules. The *curves* were computed using the parameters of Table 14

with conjugated π electrons. The intermolecular atom – atom potential calculated with these parameters is shown in Fig. 16. The intermolecular bond lengths and energies per atom pair are 3.88 Å, 4.1 meV for C–C; 3.30 Å, 2.1 meV for C–H; and 3.37 Å, 0.4 meV for H–H.

4 Summary

The interatomic forces responsible for crystal bonding are, to a large degree, electrostatic forces between the electrons and atomic nuclei. These Coulomb forces are the basic element for ionic and hydrogen bonding forces but are also involved in metallic bonding and, as dipole–dipole interaction, in van der Waals bonding. In addition, strong quantum-mechanical effects, determining specific orbitals, and Pauli exclusion are major contributing factors in covalent and metallic bonding, respectively. While the overlap of eigenfunctions of unpaired electrons with opposite spin provides the major contribution to the covalent attraction, the near impermeability of all other electronic orbitals determines the rigidity of atoms in close proximity to each other. This justifies the specification of atomic radii.

References

Anderson PW, Baskaran G, Zou Z, Hsu T (1987) Resonating–valence-bond theory of phase transitions and superconductivity in La_2CuO_4-based compounds. Phys Rev Lett 58:2790

Ashcroft NW, Mermin ND (1976) Solid state physics. Holt Reinhart and Winston, New York

Born M (1919) Die Elektronenaffinität der Halogenatome. Verh Dtsch Phys Ges 21:679 (The electron affinity of halogen atoms, in German)
Born M, Huang K (1954) Dynamical theory of crystal lattices. Oxford University Press, London
Born M, Landé A (1918) Über die Berechnung der Kompressibilität regulärer Kristalle aus der Gittertheorie. Verh Dtsch Phys Ges 20:210 (On computing the compressibility of normal crystals applying lattice theory, in German)
Brandt NB, Moshchalkov VV (1984) Semimagnetic semiconductors. Adv Phys 33:193
Brill R, Grimm HG, Hermann C, Peters CL (1939) Anwendung der röntgenographischen Fourieranalyse auf Fragen der chemischen, Bindung. Ann Physik, Lpz 34:26 (Application of x-ray Fourier analysis to problems of the chemical bond, in German)
Buckingham RA (1938) The classical equation of state of gaseous Helium, Neon and Argon. Proc R Soc Lond Ser A Math Phys Sci 168:264–283
Cotton FA, Wilkinson G (1972) Advanced inorganic chemistry: a comprehensive text, 3rd edn. Interscience, New York
Coulson CA, Redei LB, Stocker D (1962) The electronic properties of tetrahedral intermetallic compounds. I. Charge distribution. Proc R Soc Lond A 270:357
Davis R (2018) Why the Mohs scale remains relevant for metrology. IEEE Instrumentation & Measurement Magazine 21:49
Dawson B (1967) A general structure factor formalism for interpreting accurate x-ray and neutron diffraction data. Proc Roy Soc Lond A 298:255
Ehrenreich H (1987) Electronic theory for materials science. Science 235:1029
Furdyna JK (1982) Diluted magnetic semiconductors: an interface of semiconductor physics and magnetism. J Appl Phys 53:7637
Furdyna JK (1986) Diluted magnetic semiconductors: issues and opportunities. J Vac Sci Technol A 4:2002
Garnett JC (2019) Prediction of Mohs hardness with machine learning methods using compositional features. ACS Symposium Series 1326:23
Goede O, Heimbrodt W (1988) Optical Properties of (Zn, Mn) and (Cd, Mn) chalcogenide mixed crystals and superlattices. Phys Stat Sol B 146:11
Goldschmidt VM (1927) Geochemische Verteilungsgesetze der Elemente, Skrifter det Norske Videnskaps. Akad (Oslo), I Math Naturwiss Kl 1926:7 (Laws for the geometrical distribution of the elements, in German)
Gorid'ko NY, Kuzmenko PP, Novikov NN (1961) Change in mechanical properties of germanium caused by the change of concentration of current carriers. Fiz Tverd Tela 3:3650
Grimm HG, Sommerfeld A (1926) Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit den chemischen Valenzzahlen. Z Physik 36:36 (On the relation of the completion of electron groups in atoms to the chemical valence numbers, in German)
Haber F (1919) Theory of the heat of reaction. Verh Dtsch Phys Ges 21:750
Harrison WA (1966) Pseudopotentials in the theory of metals. W.A. Benjamin, New York
Harrison WA (1980) Electronic structure and the properties of solids: the physics of chemical bonds. Freeman, San Francisco
Holtzberg F, von Molnar S, Coey JMD (1980) Rare earth magnetic semiconductors. In: Moss TS, Keller SP (eds) Handbook of semiconductors, Vol. 3: materials properties and preparation. North Holland, Amsterdam
Kitaigorodskii AI (1966) Stacking of molecules in a crystal, interaction potential between atoms not linked by valence bonds, and calculation of molecular movements. J Chim Phys 63:9
Kittel C (1996) Introduction to solid state physics, 7th edn. Wiley, New York
Madelung E (1918) Das elektrische Feld in Systemen von regelmässig angeordneten Punktladungen. Phys Z 19:524 (The electric field of systems of regularly arranged point charges, in German)
Martin RM (1970) Elastic properties of ZnS structure semiconductors. Phys Rev B 1:4005
Martins JL, Zunger A (1984) Bond lengths around isovalent impurities and in semiconductor solid solutions. Phys Rev B 30:6217

Mooser E, Pearson WB (1956) The chemical bond in semiconductors. J Electron 1:629
Pauling L (1960) The nature of the chemical bond. Cornell University Press, Ithaca
Phillips JC (1973) Bonds and bands in semiconductors. Academic Press, New York
Schwoerer M, Wolf HC (2007) Organic molecular solids. Wiley-VCH, Weinheim
Shanker J, Kumar M (1987) Ion-dependent and crystal-independent interionic potentials. Phys Stat Sol B 142:325
Sherman J (1932) Crystal energies of ionic compounds and thermochemical applications. Chem Rev 11:93
Smith WF, Hashemi J (2010) Foundations of materials science and engineering, 5th edn. McGraw-Hill
Starr TL, Williams DE (1977) Coulombic nonbonded interatomic potential functions derived from crystal-lattice vibrational frequencies in hydrocarbons. Acta Cryst A 33:771
van der Waals JD (1873) Over de Continuiteit van den gasen vloeistof toestand. Sijthoff, Leiden (On the continuity of the gaseous and liquid state, in Dutch)
van Vechten JA, Phillips JC (1970) New set of tetrahedral covalent radii. Phys Rev B 2:2160
Vegard L (1921) Die Konstitution der Mischkristalle und die Raumfüllung der Atome. Z Phys 5:17 (The configuration of mixed crystals and space filling of atoms, in German)
Villars P, Calvert LD (1985) Pearson's handbook of crystallographic data for intermetallic phases. American Society for Metals, Metals Park
Wei S-H, Zunger A (1986) Total-energy and band-structure calculations for the semimagnetic $Cd_{1-x}Mn_xTe$ semiconductor alloy and its binary constituents. Phys Rev B 35:2340
Weißmantel C, Hamann C (1979) Grundlagen der Festkörperphysik. Springer, Berlin (Fundamentals of solid state physics, in German)
Welker H, Weiss H (1954) Zur transversalen magnetischen Widerstandsänderung von InSb. Z Phys 138:322 (On the change of the transversal magnetic resistance on InSb, in German)
Wigner EP, Seitz F (1933) On the constitution of metallic sodium. Phys Rev 43:804
Yeh C-Y, Lu ZW, Froyen S, Zunger A (1992) Zinc-blende–wurtzite polytypism in semiconductors. Phys Rev B 46:10086
Ziman JM (1969) The physics of metals, Vol. 1, electrons. Cambridge University Press, London
Zunger A, Cohen ML (1979) First-principles nonlocal-pseudopotential approach in the density-functional formalism. II. Application to electronic and structural properties of solids. Phys Rev B 20:4082

The Structure of Semiconductors

3

Karl W. Böer and Udo W. Pohl

Contents

1 Structure and Symmetry in Crystalline Solids 60
1.1 Crystal Systems and Bravais Lattices 61
1.2 Point Groups (Crystal Classes) and Space Groups 65
1.3 The Reciprocal Lattice 68
1.4 Relevance of Symmetry to Semiconductors 71
1.5 Structure of Organic Semiconductors 78
2 Superlattices and Quantum Structures 81
2.1 Superlattice Structures 81
2.2 Quantum Wells, Quantum Wires, and Quantum Dots 85
3 Amorphous Structures 91
3.1 Building Blocks and Short-Range Order 93
3.2 Network Structures and Matrix Glasses 98
4 Quasicrystals 100
4.1 Quasiperiodicity and Properties of Quasicrystals 101
4.2 Modeling Quasicrystals 103
5 Summary 105
References 105

Abstract

The bonding forces and atomic sizes determine the arrangement of the atoms in equilibrium in *crystals*. The crystal structure is determined by the tendency to fill a given space with the maximum number of atoms under the constraint of bonding forces and atomic radii. Crystal bonding and crystal structure are thus intimately related to each other and determine the intrinsic properties of

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_3

semiconductors. Nonequilibrium states can be frozen-in and determine the structure of *amorphous semiconductors*. In an amorphous structure the short-range order is much like that in a crystal, while long-range periodicity does not exist. *Quasicrystals* are solids with an order between crystalline and amorphous. These quasiperiodic crystals have no three-dimensional translational periodicity, but exhibit long-range order in a diffraction experiment. A quasicrystalline pattern continuously fills all available space; unlike regular crystals space filling requires an aperiodic repetition of (at least) two different unit cells.

Superlattices and low-dimensional structures like quantum wires and quantum dots, created by alternating thin depositions of different semiconductors, show material properties which can be engineered by designing size and chemical composition. This opens the feasibility for fabricating new and improved devices.

Keywords

Bonding forces · Bravais lattice · Brillouin zone · Crystal structure · Atomic radii · Crystal bonding · Miller indices · Organic semiconductors · Structure of amorphous semiconductors · Short-range order · Quasicrystals · Superlattices · Quantum wells · Quantum wires · Quantum dots · Reciprocal lattice · Unit cell

1 Structure and Symmetry in Crystalline Solids

Many physical properties of crystals depend on the *periodicity* and *symmetry of the lattice* that determines its *crystal structure*. A short summary of the basic elements of the crystal structure is presented in this chapter. For an extensive review, see Newnham (1975) and Barrett and Massalski (1980).

The easiest way to define the structure of a crystalline semiconductor (Fig. 1) is by its smallest three-dimensional building block, the *unit cell*. From these unit cells, the ideal crystal is constructed by three-dimensional repetition. The unit cell usually contains a small number of atoms, from one for a *primitive unit cell* to a few atoms for nonprimitive cells and compound crystals. In molecular crystals, this number can be much larger and is usually a small multiple of the number of atoms in the molecules forming the crystal. The three-dimensional periodic array of atoms is called the *crystal lattice*.

To define a unit cell, one introduces a three-dimensional *point lattice* and adds to this imaginary lattice an atomic *basis*, i.e., one, two, or more atoms in a specific arrangement for each point, in order to arrive at the crystal lattice (Klug and Alexander 1974; Buerger 1956). Figure 2 shows in two dimensions a crystal with a basis containing two atoms.

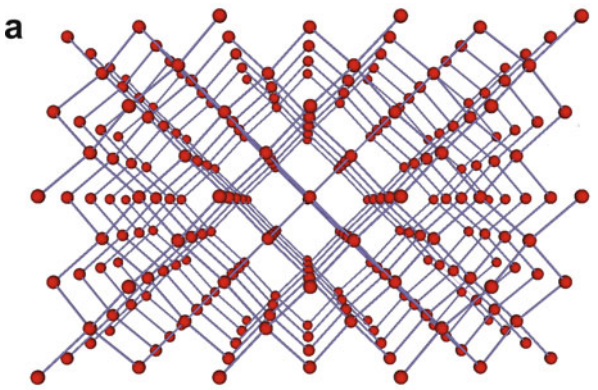
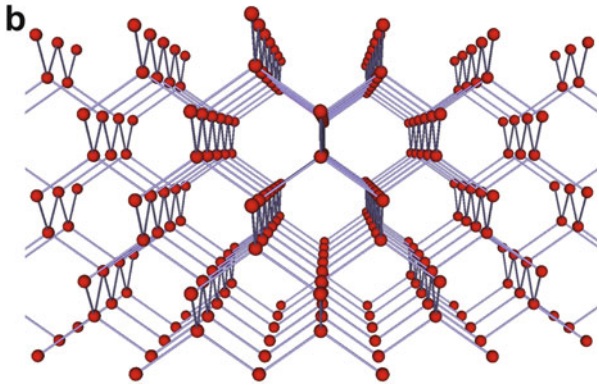


Fig. 1 Diamond structure viewed along a (**a**) ⟨100⟩ and (**b**) a ⟨110⟩ direction. (Courtesy D. Palme, Munich)



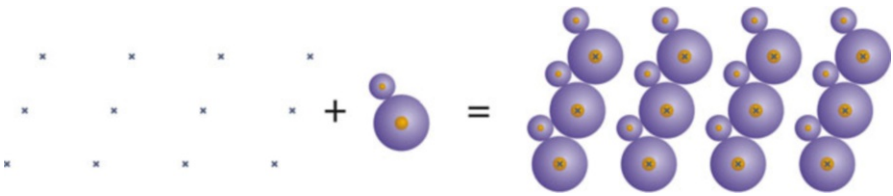


Fig. 2 The crystal lattice: point lattice plus basis with two atoms

1.1 Crystal Systems and Bravais Lattices

A *coordinate system* is introduced so that its origin lies at the center of an arbitrary atom (or basis) and its axes point through the centers of preferably adjacent atoms (or basis) while best representing the symmetry of the lattice[1] – see Fig. 5. A *lattice vector* points from the origin along each axis to the center of the next *equivalent*

[1]However, there is not always a unique way to define this coordinate system - see Sect. 1.1.3 below. For mathematical reasons, an orthogonal system is preferred when possible.

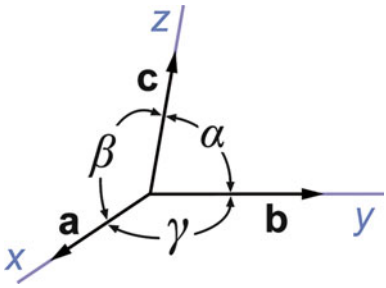


Fig. 3 Coordinate system and angles between lattice vectors within a crystal

atom[2] or from the center of one basis to the center of the next. The value of this vector is called the *lattice constant*.

1.1.1 Crystal Systems

All possible crystals can be ordered into seven *crystal systems* (i.e., different coordinate systems) according to the relative length of their lattice vectors and the angle between these vectors – see Fig. 3. These crystal systems are listed in Table 1, together with other properties identified in the following sections.

1.1.2 Bravais Lattices

There are several *symmetry operations* that transfer a crystal into itself. The simplest one is a *linear transformation*, which transfers the lattice point $\mathbf{t}_0$ into an equivalent lattice point $\mathbf{t}$:

$$\mathbf{t} = \mathbf{t}_0 + n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \tag{1}$$

with **a**, **b**, and **c** as the lattice vectors in *x*-, *y*-, and *z*-directions, respectively, and integers n_1, n_2, and n_3. This linear transformation shifts the entire lattice by an integer number of lattice constants and thereby reproduces the lattice. All lattices show linear transformation symmetry. A *unit cell* can now be defined as the smallest parallelepiped that forms the entire crystal when sequentially shifted by a linear transformation according to (1). There are only 14 different unit cells possible; they form 14 different lattices, called *Bravais lattices* (or translation lattices).

In each of the crystal systems, there is one lattice with a unit cell that contains only one lattice atom,[3] the *primitive unit cell* (P in Table 1). In some crystal systems, there exist lattices with unit cells containing more than one atom per cell. For example, in the orthorhombic system the extra atom(s) may sit in the center of the unit cell (*body centered*, I), in the center of the base [(**a** + **b**)/2, *base centered*, C], or

[2]For example, from Na to Na in a NaCl crystal, and not from Na to the next Cl ion. "Equivalent" refers to the neighborhood of this atom, which must be identical to the atom at the origin.

[3]Since each corner is shared by eight adjacent cells, only 1/8 of each corner atom belongs to each cell. Therefore, with eight corners one has $8 \times 1/8 = 1$ atom per primitive cell.

Table 1 Crystal systems, point groups, and Bravais lattices

Crystal systems	Lattice vector relation	Lattice angle relation	Crystal class		Bravais lattices
			Schönflies	Hermann-Mauguin	
Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$	C_1	1	P
			$\overline{C}_1$	$\overline{1}$	P
Monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90°$ $\beta > 90°$	C_2	2	P,C
			C_{1h} (C_2)	m	P,C
			C_{2h}	$\frac{2}{m}$	P,C
Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90°$	C_{2v}	2 mm	P,C,F,I
			D_2 (V)	222	P,C,F,I
			D_{2h} (V_h)	$\frac{2}{m}\frac{2}{m}\frac{2}{m}$	P,C,F,I
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90°$	C_4	4	P,I
			S_4	$\overline{4}$	P,I
			C_{4h}	$\frac{4}{m}$	P,I
			C_{4v}	4 mm	P,I
			D_{2d} (V_2)	$\overline{4}$2m	P,C,F,I
			D_4	422	P,I
			D_{4h}	$\frac{4}{m}\frac{2}{m}\frac{2}{m}$	P,I
Trigonal or rhombohedral	$a = b = c$	$\alpha = \beta = \gamma \neq 90°$	C_3	3	C,R
			S_6 (C_{3i})	$\overline{3}$	C,R
			C_{3v}	3 m	H,C,R
			D_3	32	H,C
			D_{3d}	$\overline{3}\frac{2}{m}$	H,C,R
Hexagonal	$a = b \neq c$	$\alpha = \gamma = 90°$ $\beta = 120°$	C_6	6	C
			C_{3h}	$\overline{6}$	C
			C_{6h}	$\frac{6}{m}$	C
			C_{6v}	6 mm	C
			D_{3h}	$\overline{6}$2m	C,H
			D_6	622	C
			D_{6h}	$\frac{6}{m}\frac{2}{m}\frac{2}{m}$	C
Cubic or isometric	$a = b = c$	$\alpha = \beta = \gamma = 90°$	T	23	P,F,I
			T_h	$\frac{2}{m}\overline{3}$	P,F,I
			T_d	$\overline{4}$3m	P,F,I
			O	432	P,F,I
			O_h	$\frac{4}{m}\overline{3}\frac{2}{m}$	P,F,I

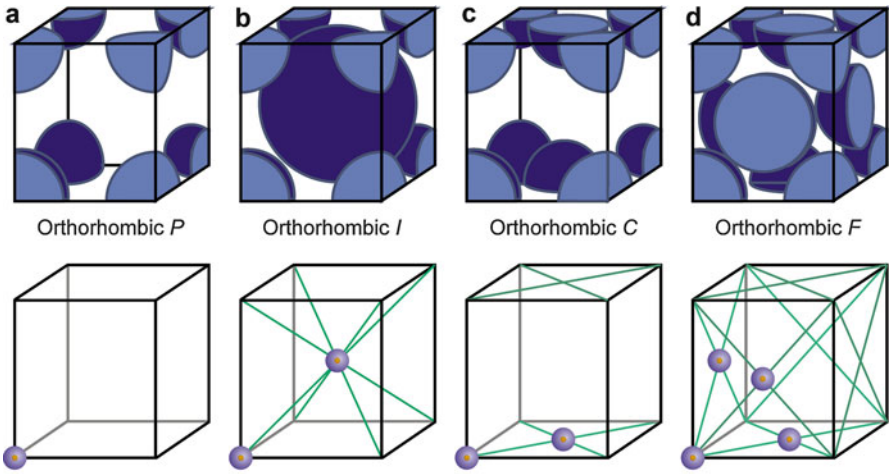


Fig. 4 Unit cells of the orthorhombic Bravais lattice. (**a**) Primitive *P*, (**b**) body-centered *I*, (**c**) base-centered *C*, and (**d**) face-centered *F*. *Upper row*: fractional atoms shown within each unit cell, *lower row*: number of atoms per unit cell indicated

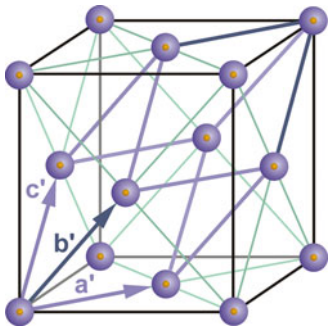


Fig. 5 Face-centered cubic unit cell with inscribed trigonal primitive cell (*blue lines*)

in the center of all faces[4] (*face centered*, F), as shown in Fig. 4. All Bravais lattices are listed in the last column of Table 1.

1.1.3 The Primitive Unit Cell

Occasionally one needs to describe the lattice as subdivided into primitive cells, while filling the entire space without voids. This can always be done; an example is presented in Fig. 5. The figure shows a face-centered cubic lattice with four lattice

[4]Each surface is shared by two neighbor cells; for example, with six surfaces, there are $6 \times 1/2 = 3$ surface atoms per unit cell.

atoms in its unit cell. If the orthogonal system of crystal axes is replaced with one connecting the corner atom to the nearest face-centered atom, the crystal structure becomes trigonal with

$$\alpha' = \beta' = \gamma' = 60^\circ \text{ and } a' = b' = c' = \frac{a}{\sqrt{2}}.$$

The cubic system is usually preferred because of a simpler mathematical description, but the trigonal representation is totally equivalent to it. This example shows that for a given crystal, the choice of a certain crystal system is not unique.

1.2 Point Groups (Crystal Classes) and Space Groups

1.2.1 Point Groups

The other symmetry operations, excluding any translation, are *rotation*, *reflection* (composed from rotation and inversion on a plane), and *inversion*, i.e., "reflection" at a point. Crystals that are distinguished by one or a combination of these can be divided into 32 different *crystal classes*. These symmetry operations are applied to the *basis* about a *point* of the Bravais lattice and therefore are also called *point groups*. The symmetry operations are usually identified by their *Schönflies* or *Hermann–Mauguin* symbol.

The Schönflies symbol identifies with capital letters C, D, T, and O the basic symmetry: *cyclic*, *dihedral*, *tetrahedral*, and *octahedral*. A subscript is used to identify the rotational symmetry, e.g., D_3 has threefold symmetry. Another index, *v*, *h*, *d* is used for further distinction – see, e.g., Brown and Forsyth (1973).

The Hermann–Mauguin nomenclature indicates the type of symmetry directly from the symbol. It is a combination of numbers (n) and the letter m: n indicates rotational symmetry (for $n = 2$, 3, 4, or 6, an n-fold symmetry) and $\overline{n}$ denotes either an inversion ($\bar{1}$) or a roto-inversion (with a $\overline{3}$-, $\overline{4}$-, or $\overline{6}$-fold symmetry); m indicates a mirror plane parallel to, and $\frac{n}{m}$ perpendicular to, the rotational axis with n-fold symmetry. Repetition of m or other symbols indicates the symmetry about the other orthogonal planes or axes – see, e.g., Hahn (1983).

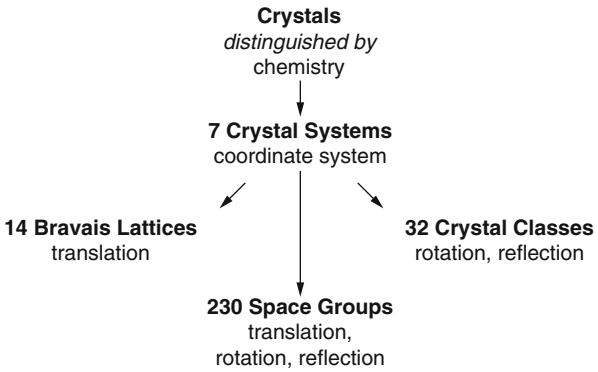
All possible combinations of rotation, reflection, and inversion are listed in Table 1, with both symbols to identify each of the 32 point groups.

1.2.2 Space Groups

Combining the symmetry operations leading to the point groups with nonprimitive translation yields a total of 230 *space groups* (Burns and Glazer 2013). Alternatively, there are 1421 space groups when the ordering of spins is also considered (Birss 1964). They include *screw axis* and *glide plane* operations; the former combines translation (shifting) with rotation; the latter combines translation with reflection.

The Schönflies symbol for space groups designates the different possibilities of combining the symmetry operations by a superscript referring to the point group symbol (e.g., O_h^7 for Si).

Table 2 Crystal classification



In the Hermann–Mauguin symbol, the Bravais lattice identifier is added: A, B, and C (identifying the specific base for face-centered symmetry)[5]; P (primitive); I, F, and R (rhombohedric); and H (hexagonal). In addition, small letters, *a*, *b*, *c*, *d*, or *n,* are appended to identify specific glide planes – namely, at *a*/2, *b*/2, and *c*/2, $\frac{r+s}{4}$, and $\frac{r+s}{2}$, for *a, b, c, d*, and *n*, respectively, with *r* and *s* standing for any *a*, *b*, or *c.*[6]

Typical element semiconductors have O_h symmetry, e.g., diamond O_h^7 (or F*d*3*m*) for Ge and Si. Other binary semiconductors have zincblende T_d^2 (or F43*m*) for GaAs, wurtzite C_{6v}^4 (P6$_3$*mc*) for GaN, or rock salt O_h^5 (or F*m*3*m*) for NaCl.

In summary, crystals are classified according to their lattice symmetry in four different ways, depending on the type of symmetry operation employed. This is shown in Table 2.

1.2.3 Crystallographic Notations

A *lattice point* is identified by the coefficients of the lattice vector pointing to it:

$$\mathbf{R}_n = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c}. \tag{2}$$

A lattice point is conventionally given by the three coefficients *without* brackets:

$$n_1\ n_2\ n_3.$$

A *lattice direction* is identified by a line pointing in this direction. When this line is shifted parallel so that it passes through the origin, the position of the nearest

[5]A is the face spun between **b** and **c**, B between **a** and **c**, and C between **a** and **b**.

[6]Thus, $(r + s)/4$ is a quarter of a face diagonal.

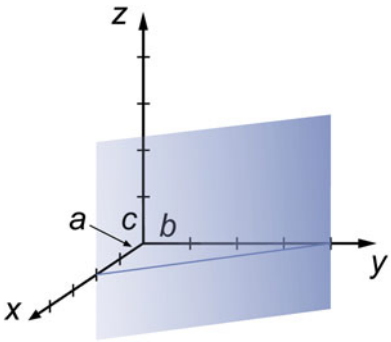


Fig. 6 Example of a (210) plane

lattice point on this line, identified by the coefficients of Eq. 2 and enclosed in square *brackets*, defines this direction:

$$[n_1\ n_2\ n_3].$$

Conveniently, one may reduce this notation by permitting simple fractions; for example, [221] may also be written as [1 1 1/2]. Negative coefficients are identified by a bar: $[00\bar{1}] = -[001]$ is a vector pointing downward.

Equivalent directions are directions which are crystallographically equivalent; for example, in a cube these are the directions [100], [010], [001], $[\bar{1}00]$, $[0\bar{1}0]$ and $[00\bar{1}]$. All of these are meant when one writes $\langle 100\rangle$; in general

$$\langle n_1\ n_2\ n_3\rangle.$$

A *lattice plane* is described by *Miller indices*. These are obtained by taking the three coefficients of the intercepts of this plane with the three axes n_1, n_2, and n_3; forming the reciprocals of these coefficients $1/n_1\ 1/n_2\ 1/n_3$; and clearing the fractions. For example, for a plane parallel to **c** and intersecting the *x*-axis at 2**a** (see Fig. 6) and the *y*-axis at 4**b**, the fractions are ½ ¼ $\frac{1}{\infty}$. Thus, the Miller indices are (210) and are enclosed in parentheses. The general form is

$$(h\ k\ l).$$

A *family of planes* which are crystallographically equivalent [such as (111), $(\bar{1}11)$, $(1\bar{1}1)$, $(11\bar{1})$, $(\bar{1}\,\bar{1}\,1)$, etc.] is identified by the Miller indices in curly parentheses. For this example the triple is {111}; in general, it is

$$\{h\ k\ l\}.$$

The Miller indices notation is a *reciprocal lattice* representation (see Sect. 1.3). It is quite useful for the discussion of interference phenomena, which requires the knowledge of distances between equivalent planes. The distance between the {*hkl*} planes is easily found; in a cubic system, it is simply

$$d_{hkl} = \frac{a}{\sqrt{h^2 + k^2 + l^2}}, \tag{3}$$

with a the lattice constant. In other crystal systems the expressions are slightly more complicated[7] (see Warren 1990; Zachariasen 2004; and James 1954 for more details).

The reciprocal lattice is a lattice in which each point relates to a corresponding point of the actual lattice by a reciprocity relation given below (Eqs. 6–10).

1.2.4 Morphology of Similar Crystals

When a specific chemical compound crystallizes in different crystal classes, it is called a *polymorph*. When crystals with the same structure are formed by compounds in which only one element is exchanged with a homologous element, they are referred to as *morphotrop*. When similar compounds crystallize in a similar crystal form, they are called *isomorph* when they also have other physical properties in common, such as similar cation to anion radii ratio and similar polarizability.

1.3 The Reciprocal Lattice

As indicated above, the introduction of a reciprocal lattice is advantageous when one needs to identify the distance between equivalent lattice planes. This is of help for all kinds of interference phenomena, such as x-ray diffraction, the behavior of electrons when taken as waves, or lattice oscillations themselves. In a quantitative description, the relevant waves are described by wave functions of the type[8]

$$\varphi(\mathbf{k}, \mathbf{r}) = A \exp\{i(\mathbf{k} \cdot \mathbf{r} - \omega t)\} \tag{4}$$

where A is the amplitude factor, $\mathbf{r}$ is a vector in real space, and $\mathbf{k}$ is a vector in reciprocal space. Here, $\mathbf{k}$ is referred to as the *wave vector*, or *wave number*, if only one relevant dimension is discussed; the wavevector is normal to the wave front and has the magnitude

[7]The general expression for the distance between two planes is given by

$$d_{hkl}^2 = \frac{\begin{vmatrix} 1 & \cos\gamma & \cos\beta \\ \cos\gamma & 1 & \cos\alpha \\ \cos\beta & \cos\alpha & 1 \end{vmatrix}}{\frac{h}{a}\begin{vmatrix} h/a & \cos\gamma & \cos\beta \\ k/b & 1 & \cos\alpha \\ l/c & \cos\alpha & 1 \end{vmatrix} + \frac{k}{b}\begin{vmatrix} 1 & h/a & \cos\beta \\ \cos\gamma & k/b & \cos\alpha \\ \cos\beta & l/c & 1 \end{vmatrix} + \frac{l}{c}\begin{vmatrix} 1 & \cos\gamma & h/a \\ \cos\gamma & 1 & k/b \\ \cos\beta & \cos\alpha & l/c \end{vmatrix}}$$

[8]This description is more convenient than an equivalent description, which in one direction reads $\phi(x) = A \exp\{2\pi i(x/\lambda - \nu t)\}$.

$$|\mathbf{k}| = 2\pi/\lambda \tag{5}$$

with λ the wavelength. Since $\mathbf{k} \cdot \mathbf{r}$ is dimensionless, $\mathbf{k}$ has the dimension of reciprocal length. Multiplied by $\hbar$, ($=h/2\pi$, where h is the Planck constant) $\hbar\mathbf{k}$ has the physical meaning of a momentum as will be shown in Sect. 2.1 of ▶ Chap. 6, "The Origin of Band Structure."

When $\mathbf{R}_n$ is a lattice vector [for ease of mathematical description, we now change from ($\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$) to ($\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$)]

$$\mathbf{R}_n = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \tag{6}$$

one obtains the corresponding vector $\mathbf{K}_m$ in reciprocal space with the three fundamental vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$:

$$\mathbf{K}_m = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3 \tag{7}$$

where both sets of unit vectors are related by the orthogonal relation

$$\mathbf{a}_i\,\mathbf{b}_j = 2\pi\,\delta_{ij} \quad \text{and} \quad i,j = 1,2,3, \tag{8}$$

where δ_{ij} is the *Kronecker delta symbol*

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}. \tag{9}$$

The orthogonal relation can also be expressed by

$$\mathbf{b}_1 = 2\pi\frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \times \mathbf{a}_2 \cdot \mathbf{a}_3}, \text{ etc. (cyclical)}, \tag{10}$$

that is, every vector in the reciprocal lattice is normal to the corresponding plane of the crystal lattice and its length is equal to the reciprocal distance between two neighboring corresponding lattice planes (see Kittel 2007). This definition is distinguished by a factor 2π from the definition of a reciprocal lattice found by crystallographers. This factor is included here to make the units of the reciprocal space identical to the wavevector units.

1.3.1 Wigner–Seitz Cells and Brillouin Zones

As knowledge about an entire crystal can be derived from the periodic repetition of its smallest unit, the *unit cell*, one can derive knowledge about the wave behavior from an equivalent cell in the reciprocal lattice. A convenient way to introduce this discussion is by examining the *Wigner–Seitz cell* rather than the unit cell itself.

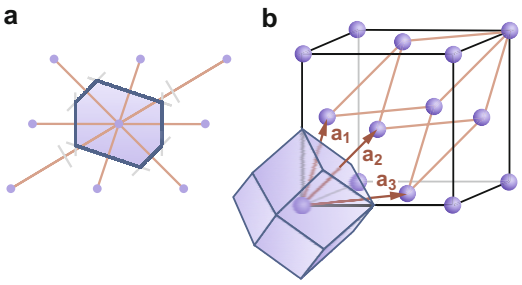


Fig. 7 (**a**) Illustration of the construction of a two-dimensional Wigner-Seitz cell (*blue shaded area*). (**b**) Face-centered cubic lattice (*blue* atoms on the *black* cube) with primitive parallelepiped (*red lines*) and from it the derived three-dimensional Wigner–Seitz cell in real space (*blue* polyhedron), which is equivalent to the Brillouin zone in reciprocal space

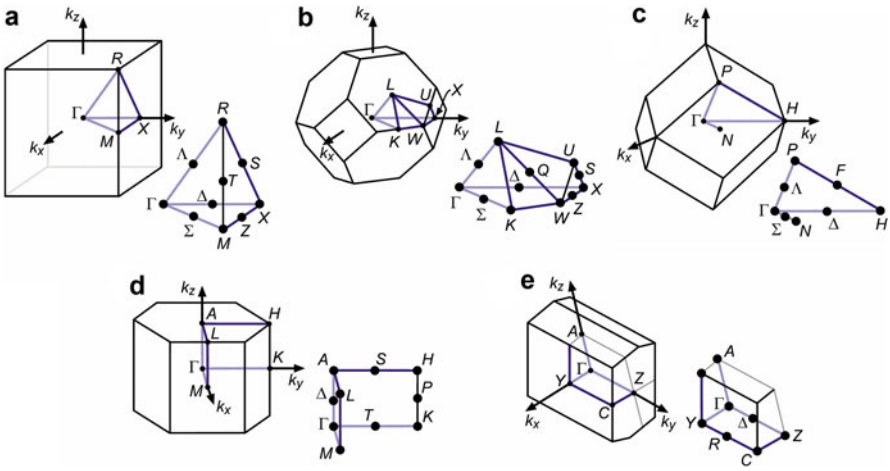


Fig. 8 Brillouin zones for the three cubic, hexagonal, and monoclinic lattices with important symmetry points and axes. (**a**) Primitive, (**b**) face-centered, and (**c**) body-centered cubic lattice, (**d**) primitive hexagonal, and (**e**) simple monoclinic lattice

A Wigner–Seitz cell is formed when a lattice point is connected with all equivalent neighbors, and planes are erected normal to and in the center of each of these interconnecting lines. The construction is illustrated for two dimensions in Fig. 7a. For the face-centered unit cell ($\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$), the Wigner–Seitz cell is constructed in Fig. 7b; the plane orthogonal to and intersecting the lattice vector $\mathbf{a}_2$ is visible.

When such a Wigner–Seitz cell is constructed from the unit cell of the reciprocal lattice, the resulting cell is called the *first Brillouin zone*. It is the basic unit for describing lattice oscillations and electronic phenomena.

Most semiconductors crystallize with cubic or hexagonal lattices; by contrast, *organic* semiconductors have low-symmetry – often monoclinic – unit cells. The first Brillouin zones of these lattices are given in Fig. 8 and will be referred to frequently later in the book.

In these discussions, lattice symmetry is of great importance, and points about which certain symmetry operations can reproduce the lattice are often cited. These symmetry points can also be transformed into the reciprocal lattice and are identified here by specific letters. The most important *symmetry points* with their conventional notations are identified in the different Brillouin zones of Fig. 8. Γ is always the center of the zone ($k_x = k_y = k_z = 0$), and in any of the cubic lattices, X is the intersection of the Brillouin zone surface with any of the main axes (k_x, k_y, or k_z); the points Δ, Λ, and Σ in face-centered cubic lattices lie halfway between Γ and X, Γ and L, and Γ and K, as shown in Fig. 8b. The positions of the other symmetry points (H, K, L, etc.) can be obtained directly from Fig. 8. In the hexagonal and monoclinic lattices other letters are used by convention as shown in Fig. 8.

The extent of the first Brillouin zone can easily be identified. For instance, in a primitive orthorhombic lattice with its unit cell extending to a, b, and c in the x-, y-, and z-directions, respectively, the first Brillouin zone extends from $-\frac{\pi}{a}$ to $\frac{\pi}{a}$ in k_x-, from $-\frac{\pi}{b}$ to $\frac{\pi}{b}$ in k_y-, and from $-\frac{\pi}{c}$ to $\frac{\pi}{c}$ in k_z-direction. Since the wave equation is periodic in **r** and **k**, all relevant information is contained within the first Brillouin zone.

1.4 Relevance of Symmetry to Semiconductors

Lattice periodicity is one of the major factors in determining the band structure of semiconductors – see ► Chap. 6, "The Origin of Band Structure." The symmetry elements of the lattice are reflected in the corresponding symmetry elements of the bands, from which important qualitative information about the electronic structure of a semiconductor is obtained. Therefore, the main features of the symmetry of some of the typical semiconductors are summarized below. A comprehensive review of element and compound structures is given by Wells (2012).

1.4.1 Elemental Semiconductors and Binary Semiconducting Compounds

Elemental Semiconductors

Most of the important crystalline semiconductors are elements (Ge, Si) or binary compounds (III–V or II–VI). They form crystals in which each atom is surrounded by four nearest neighbors,[9] i.e., they have a *coordination number* of 4. The connecting four atoms (*ligands*) surround each atom in the equidistant corners of a tetrahedron. The lattice is formed so that each of the surrounding atoms is again the center atom of an adjacent tetrahedron, as shown for two such tetrahedra in Fig. 9. Of

[9]There are other modifications possible. For example, seven for Si, of which four are stable at room temperature and ambient pressure (see Landoldt-Börnstein 1982, 1987). Only Si I and α-Si are included in this book. Si III is face-centered cubic and a semimetal; Si IV is hexagonal diamond and is a medium-gap semiconductor (see Besson et al. 1987).

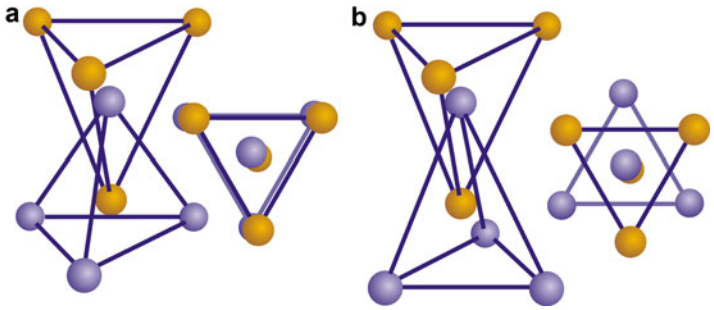


Fig. 9 Side and top view of two intertwined tetrahedra with (**a**) base triangles parallel (dihedral angle 0°) and (**b**) base triangles rotated by 60°

Fig. 10 Binary compounds with semiconducting properties

the two principal possibilities for arranging two tetrahedra, only one is realized in nature for elemental crystals: the *diamond lattice*, wherein the base triangles of the intertwined tetrahedra are rotated by 60°. Ge and Si are examples. In amorphous elemental semiconductors, however, both possibilities of arranging the tetrahedra are realized – see Sect. 3.1.

Binary Semiconducting Compounds

Binary III–V and II–VI compounds are formed by both tetrahedral arrangements which are dependent on relative atomic radii and preferred valence angles (see Sect. 1 in ▶ Chap. 2, "Crystal Bonding"), although with alternating atoms as nearest neighbors. These compounds can be thought of as an element (IV) semiconductor after replacing alternating atoms with an atom of the adjacent rows of elements (III and V). Similarly, II–VI compounds can be created by using elements from the next-to-adjacent rows – see Fig. 10 and Fig. 3 in ▶ Chap. 1, "Properties and Growth of Semiconductors."

Aside from these classical *AB* compounds, there are others that have interesting semiconducting (specifically thermoelectrical) properties. Examples include the II–V compounds (such as ZnSb, ZnAs, CdSb, or CdAs), which have orthorhombic structures. For a review, see Arushanov (1986).

The diamond lattice for *AB* compounds results in a *zincblende lattice* shown in Fig. 11a. Most III–V compounds, as, for instance, GaAs, are examples.

Unrotated interpenetrating tetrahedra, as shown in Fig. 9a, produce the *wurtzite lattice* (Fig. 11b) which can also be obtained for a number of *AB* compounds. Examples include ZnS, CdS, and GaN. The aforementioned semiconductors can also crystallize in a zincblende modification. Under certain conditions, alternating layers of wurtzite and zincblende, each several atomic layers thick, are observed.

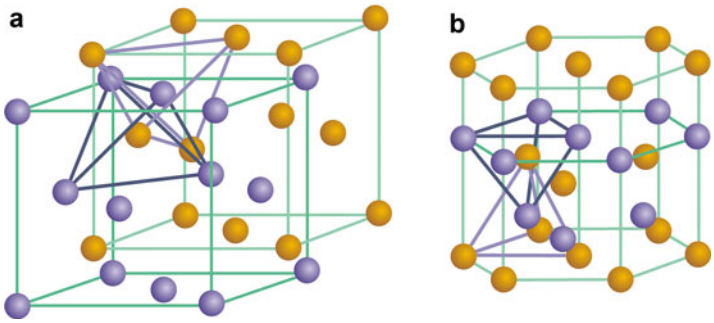


Fig. 11 (**a**) Zincblende lattice (GaAs) constructed from two interpenetrating face-centered cubic sublattices of Ga and As, with a displaced origin at $a/4$, $a/4$, $a/4$, with a the edge length of the elementary cube. (**b**) Wurtzite lattice (CdS, GaN) constructed from two intertwined hexagonal sublattices of Cd and S (or, Ga and N)

This is called a *polytype*. Often, the zincblende structure is more stable at lower temperatures and the wurtzite structure appears above a transition temperature (1053 °C in CdS). With rapid cooling the wurtzite structure can be frozen-in.

Other structures of binary semiconductors include:

- NaCl-type semiconductors, with PbTe as an example
- Cinnabar (deformed NaCl) structures, with HgS as an example
- Antifluorite silicide structures, with Mg_2Si as an example. These structures can be regarded as derived from the fcc lattice (Fig. 11a) with one of the two interstitial positions filled by the second metal atom, similar to the Nowotny–Juza compounds in Sect. 1.4.2 below, and
- $A_I{}^3B_{IV}$ structures, with Cs_3Si as an example. For a review, see Parthé (1964), Sommer (1968), and Abrikosov et al. (1969).

1.4.2 Ternary and Quaternary Semiconducting Compounds

There are several classes of ternary and quaternary compounds with known attractive semiconducting properties. All have tetrahedral structures: each atom is surrounded by four neighbors. Some examples are discussed in the following sections. For a review, see Zunger (1985).

One can conceptually form a wide variety of ternary, quaternary, or higher compounds which have desirable semiconducting properties by replacing within a tetrahedral lattice, subsequent to the original replacement shown in Fig. 10, certain atoms with those from adjacent rows, as given in Fig. 12. These examples represent a large number (~140) of such compounds and indicate the rules for this type of compound formation. For instance, a $II–III_2–VI_4$ compound can be formed by replacing 8 atoms of column IV first with 4 atoms each of columns II and VI and consequently the 4 atoms of column II with one vacancy (0), one atom of column II, and two atoms of column III.

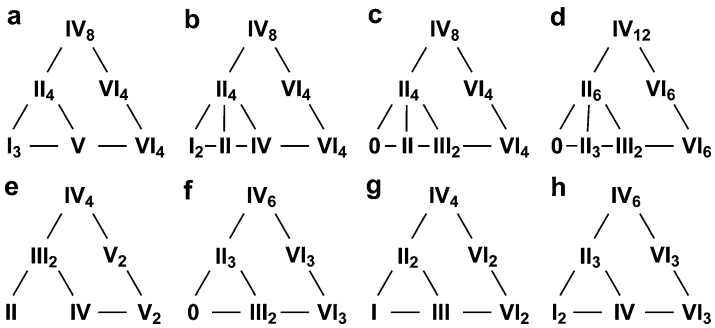


Fig. 12 Construction of pseudobinaries (**b**), ternaries (**a**, **c**, **d**, **e**) pseudoternaries (**g**, **h**), and quaternaries (**f**) from element (IV) semiconductors (0 represents a vacancy, i.e., a missing atom at a lattice position)

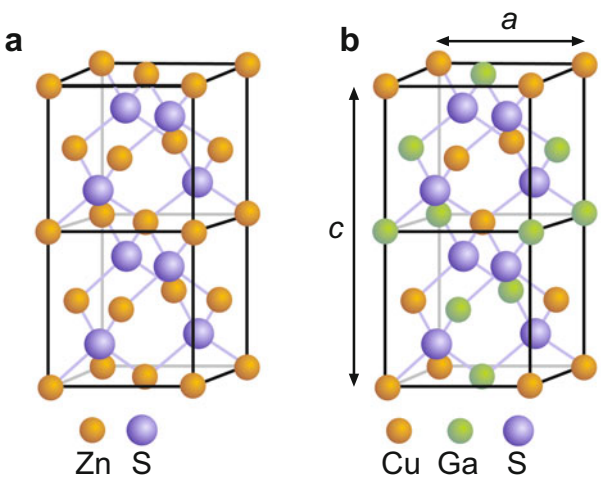


Fig. 13 (**a**) ZnS (or GaP) double unit cell; (**b**) $CuGaS_2$ (or $ZnGeP_2$) unit cell

Ternary Chalcopyrites

Best researched are the *ternary chalcopyrites* I–III–VI_2; they are constructed from two zincblende lattices in which the metal atoms are replaced by an atom from each of the adjacent columns. In a simple example one may think of the two Zn atoms from ZnS as transmuted into Cu and Ga:

$$^{30}\mathrm{ZnS} + {}^{30}\mathrm{ZnS} \rightarrow {}^{29}\mathrm{Cu}^{31}\mathrm{GaS}_2,$$

with some deformation of the zincblende lattice, since the Cu–S and Ga–S bonds have different strengths, and with a unit cell twice the size of that in the ZnS lattice (Fig. 13). For a review, see Miller et al. (1981).

Ternary Pnictides and ABC_2 Compounds

Other ternaries with good semiconducting properties are the *ternary pnictides* II–IV–V_2 (such as $ZnSiP_2$) which have the same chalcopyrite structure and, in a similar example, can be constructed from GaP by the transmutation

$$^{31}GaP + {}^{31}GaP \rightarrow {}^{30}Zn{}^{32}SiP_2.$$

Still another class with chalcopyrite structure is composed of the I–III–VI_2 compounds, of which $CuFeS_2$ is representative. (These structures are reviewed by Jaffe and Zunger 1984).

Also some Heusler compounds are semiconductors, e.g., Fe_2VAl (Anand et al. 2020) or CoTiSb. Their composition on a face-centered lattice corresponds to *ABC* (half Heuslers) or A_2BC (full Heuslers) with transition metals *A* and *B*. These compounds show interesting magnetic properties; for a review see Felser and Hirohata (2016).

Nowotny–Juza Compounds

Interesting variations of this tetrahedral structure (see Parthé 1972) are the *Nowotny–Juza compounds*, which are partially filled tetrahedral interstitial I–II–V compounds (e.g., LiZnN). Here the Li atom is inserted into exactly one half of the available interstitial sites of the zincblende lattice (e.g., on V_a or on V_c as shown in Fig. 14). A substantial preference for the Li atom to occur at the site closer to the N atom (rather than the site next to the Zn – the lattice energy of this structure is lower by about 1 eV) makes this compound an *ordered* crystal with good electronic properties (Carlson et al. 1985; Kuriyama and Nakamura 1987; Bacewicz and Ciszek 1988; Yu et al. 2004; Kalarasse and Bennecer 2006). It should be noted, however, that the Zn atom is fourfold coordinated with N atoms, while the N atom is fourfold coordinated with Zn and fourfold coordinated with Li; therefore, it has eight nearest neighbors.

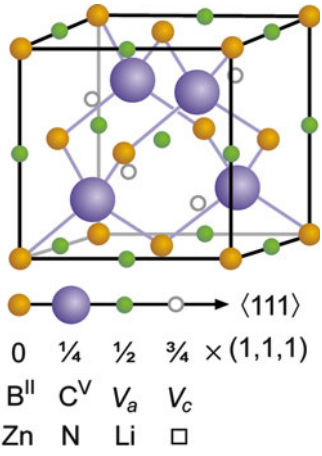


Fig. 14 Unit cell of the Nowotny–Juza compound

The Adamantine $A_nB_{4-n}C_4$ and Derived Vacancy Structures

Examples of this class of $A_nB_{4-n}C_4$ structures with $n = 1$ or 3, such as A_3BC_4 or AB_3C_4, are the famatinites (e.g., Cu_3SbS_4 or $InGa_3As_4$) or lazarevicites (e.g., Cu_3AsS_4). With $n = 2$ this class reduces to ABC_2 (e.g., $CuGaAs_2$ or $GaAlAs_2$), and with $n = 4$ it reduces to the zincblende (ZnS) lattice. The layered sublattices can be ordered (e.g., in $CuGaAs_2$) or disordered (alloyed) as in $GaAlAs_2$ and are discussed in the following section. All of these compounds follow the octet $(8 - N)$ rule (see Sect. 3.1.1); they are fourfold coordinated (each cation is surrounded by four anions and vice versa).

The $8 - N$ rule determines how many shared electrons are needed to satisfy perfect covalent bonding (Sect. 1 in ▶ Chap. 2, "Crystal Bonding") for any atom with N valency electrons, e.g., 1 for Cl with $N = 7$, 2 for S with $N = 6$, or 4 for Si with $N = 4$, requiring single, chain-like, or tetrahedral bondings, respectively.

Deviations from the $A_nB_{4-n}C_4$ composition may occur when including ordered vacancy compounds into this group, such as II–III_2VI_4 compounds (e.g., $CuIn_2Se_4$) in which one of the II or III atoms is removed in an ordered fashion, resulting in defect famatinites or defect stannites.

An instructive generic overview of the different structures of tetragonal ternaries or pseudotemaries is given by Bernard and Zunger (1988) (Fig. 15). See also Shay and Wernick (1974); Miller et al. (1981) and conference proceedings on ternary and multinary compounds.

Pseudoternary Compounds

Finally, one may consider *pseudoternary compounds* in which one of the components is replaced by an *alloy* of two homologous elements. For example, Ga replaced by a mixture of Al and Ga in GaAs yields $Al_xGa_{1-x}As$; replacement of As by P and As yields GaP_xAs_{1-x}. These pseudoternary compounds contain alloys of isovalent atoms in one of the sublattices.

When the two alloying elements are sufficiently different in size, preference for *ordering* exists for stoichiometric composition in the sublattice of this alloy. Substantial bandgap bowing (see Sect. 2.1 in ▶ Chap. 8, "Bands and Bandgaps in Solids") gives a helpful indication of predicting candidates for this ordering of stoichiometric compounds. Examples include $GaInP_2$, which shows strong bowing, where the Ga and In atoms are periodically ordered (Srivastava et al. 1985), Ga_3InP_4, or $GaIn_3P_4$ with similar chalcopyrite-type structures (see also Sect. 2.1 in ▶ Chap. 8, "Bands and Bandgaps in Solids"). Here again, the coordination number is four; each atom is surrounded by four nearest neighbors, although they are not necessarily of the same element.

A different class of such compounds is obtained when alloying with nonisovalent atoms, such as Si+GaAs.

The desire to obtain semiconductors with specific properties that are better suited for designing new and improved devices has focused major interest on synthesizing new semiconducting materials as discussed above, or using sophisticated growth methods to be discussed below, aided by theoretical analyses to predict potentially interesting target materials (see Ehrenreich 1987).

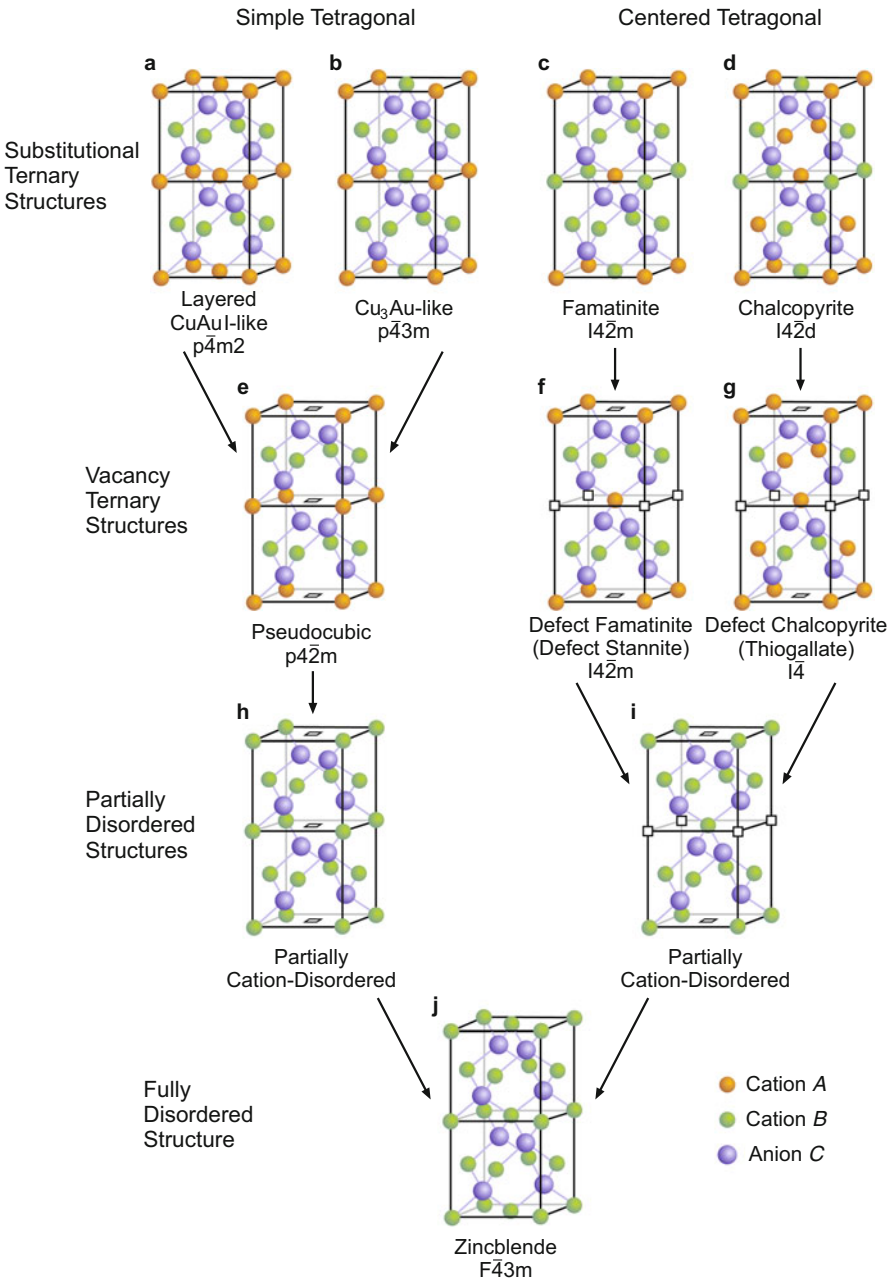


Fig. 15 Structure of $A_nB_{4-n}C_4$ (adamantine) compounds (**a**–**d**) and their derived, ordered vacancy structures (**e**–**g**). Also included are cation-disordered structures including ordered vacancies (**h**) and (**i**) and the parent zincblende (with ordered or disordered sublattice). Vacancies are shown as open rectangles (After Bernard and Zunger 1988)

1.5 Structure of Organic Semiconductors

The growth units in organic semiconductors are bulky molecules with a lower symmetry than single atoms, the growth units of inorganic semiconductors. Organic semiconductors therefore crystallize generally in low-symmetry unit cells. Consequently all physical properties have tensor character with often large anisotropies. The versatile ability for synthesizing organic molecules leads to a huge and steadily increasing number of organic crystals. Most of them are insulating, but quite a few show conductive or semiconducting properties; we focus on some important examples.

The structure of organic crystals is determined by their intermolecular forces. In *nonpolar molecules* these are van der Waals attractive and Born repulsive forces, combined described by the Buckingham potential Eq. 10 in ▶ Chap. 2, "Crystal Bonding." Nonpolar molecules comprise aliphatic[10] hydrocarbons like the alkanes $CH_3(CH_2)_nCH_3$ and aromates like the oligoacenes $C_{4n+2}H_{2n+4}$ listed in Table 3. Due to the weak attractive interaction, the molecules tend to crystallize in lattices with closest packing for maximizing the number of intermolecular contacts. The packing density is described by a coefficient (Kitaigorodskii 1973)

$$K = ZV_0/V_{uc}, \tag{11}$$

where V_{uc} is the volume of the unit cell and V_0 the volume of one of the Z molecules of the unit cell; V_0 can be computed from the molecule structure and the atomic radii. Stable crystals have packing coefficients between 0.65 and 0.80.

The mutual arrangement of the molecules follows the trend of close spacing: planar molecules prefer a parallel alignment. Furthermore, atoms tend to locate at interstices between atoms of the adjacent molecule. This favors a crystallization in a herringbone packing with an angle between adjacent columns of the planar molecules, observed, e.g., for oligoacenes and oligothiophenes[11]. The rule of thumb for interstitial alignment does not apply if the molecules have a permanent dipole moment or polar substituents; even small dipolar or ionic contributions to the intermolecular bonding have a respective long-ranging $1/r^3$ or $1/r$ dependence and thus a significant effect on the crystal structure.

Organic crystals often suffer for their limited perfection. Crystal growth is hampered by various factors: the orientational degree of freedom of their building blocks favors disorder-induced defects, crystal properties vary sensitively with the introduction of contaminants, and the rigid molecule structure combined with a weak intermolecular bonding make organic crystals fragile. Structural imperfections imply the frequently observed formation of polymorphs, which differ, e.g., in the herringbone angle or even in the number Z of molecules per unit cell.

[10] Aliphatic (from Greek *aleiphar*, "oil") designates organic compounds in which the carbon atoms are linked in open chains.

Table 3 Crystallographic data of some organic semiconductors. Z denotes the number of molecules per unit cell; values of vectors a, b, and c are given in Å, volume V of the unit cell in Å^3, and angles α, β, and γ in degrees. Most molecules form various polymorphs; only data of a few are given

Organic crystal	Formula	Crystal system	Z	a	b	c	α	β	γ	V
Anthracene	$C_{14}H_{10}$	Monoclinic	2	8.6	6.0	11.2	90	125	90	474
Tetracene	$C_{18}H_{12}$	Triclinic	2	7.9	6.0	13.5	100	113	86	583
Pentacene	$C_{22}H_{14}$	Triclinic	2	7.9	6.1	16.0	102	113	86	692
Rubrene	$C_{42}H_{28}$	Monoclinic	2	8.7	10.1	15.6	90	91	90	1,383
		Triclinic	1	7.0	8.5	11.9	93	106	96	684
		Orthorhombic	4	26.9	7.2	14.4	90	90	90	2,736
Perylene (α phase)	$C_{18}H_{12}$	Monoclinic	4	11.4	10.9	10.3	90	101	90	1,249
Perylene (β phase)	$C_{18}H_{12}$	Monoclinic	2	11.3	5.9	9.7	90	92	90	394
Quaterthiophene (α-4 T)	$C_{16}H_{10}S_4$	Monoclinic	4	30.5	7.9	6.1	90	92	90	1,471
Hexathiophene (α-6 T)	$C_{24}H_{14}S_6$	Monoclinic	4	44.7	7.9	6.0	90	91	90	2,117
CuPc (β phase)	$CuN_8C_{32}H_{16}$	Monoclinic	2	14.6	4.8	17.3	90	105	90	1,171
Alq3 (α phase)	Al $(C_9H_6NO)_3$	Triclinic		6.2	12.9	14.7	70	89	83	

Prominent organic semiconductors are listed in Table 3; small molecules (usually oligomers[11]) and polymers, both with conjugated π bonds, are used.[12] Crystals are generally formed from molecules; the structure of such molecules is shown in Fig. 16. The family of *acenes* is formed from polycyclic aromatic hydrocarbons fused in a linear chain of conjugated benzene rings (Fig. 16a). The polycyclic aromate *rubrene* (5,6,11,12-tetraphenylnaphthacene, the numbers indicate where four phenyl groups are attached to tetracene) is built on a tetracene backbone with four phenyl rings on the side that lie in a plane that is perpendicular to the plane of the backbone (Fig. 16d). The *perylene* molecule shown in Fig. 16c consists of two naphthalene molecules (similar to anthracene Fig. 16a with only *two* benzole rings, i.e., $n = 0$), connected by a carbon–carbon bond; all of the carbon atoms in perylene are sp^2 hybridized. The heterocyclic *thiophenes* Fig. 16b include a sulfur atom in their ring structure. Examples for more complex compounds used in organic devices are copper phthalocyanine (CuPc) shown in Fig. 16e and tris(8-hydroxyquinolinato)

[11]An oligomer (from Greek *oligos*, "a few," and *meros* "part") is a molecule consisting of a small number of the repeat units of a polymer; a polymer (from Greek *poly* "many" and *meros* "part") is a large molecule composed of many repeated subunits.

[12]The organic semiconductors listed in Table 3 are widely used particularly due to their high carrier mobility and stability. Highest hole mobilities at room temperature were reported for pentacene (35 cm^2/Vs; Jurchescu et al. 2004) and rubrene crystals (40 cm^2/Vs; Takeya et al. 2007); CuPc, known as blue dye in artificial organic pigments, is used in organic FETs, and Alq3 is commonly applied in organic LEDs.

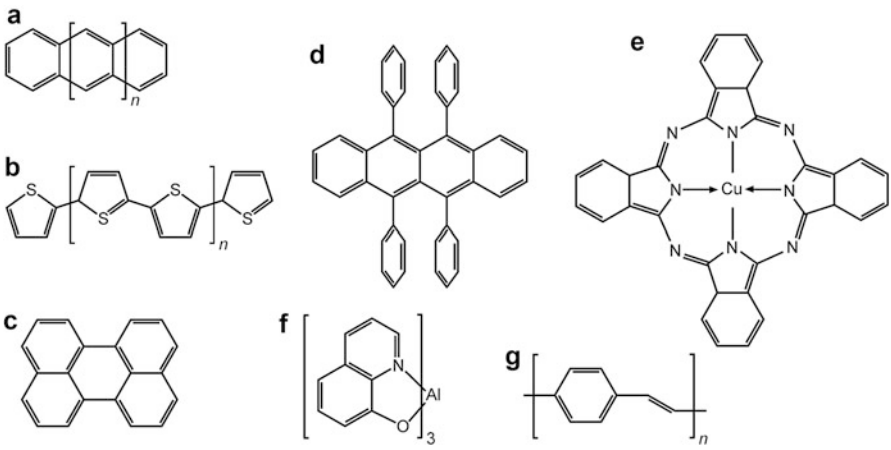


Fig. 16 Molecules of prominent organic semiconductors: (**a**) oligoacenes anthracene ($n = 1$), tetracene ($n = 2$), pentacene ($n = 3$); (**b**) oligothiophenes quaterthiophene ($n = 1$), hexathiophene ($n = 2$); (**c**) perylene, (**d**) rubrene, (**e**) copper phthalocyanine (CuPc), and (**f**) tris(8-hydroxyquinolinato)aluminum (Alq3). The small width of the side rings in (**d**) indicates a twist by 85° out of the plane of projection. (**g**) The repetition unit of the polymer poly(p-phenylene vinylene). For the representation of chemical structures, see Fig. 15 in ▶ Chap. 2, "Crystal Bonding"

aluminum (Alq3, Fig. 16f). There are numerous derivatives of all these compounds obtained from substituting one or several hydrogen atoms (which are not drawn in Fig. 16) for organic groups like methyl (CH_3), or a halogen like Cl, or a cyclic phenyl ring (C_6H_5) as those shown in rubrene Fig. 16d.

There are also *polymers* with conjugated π electrons used for semiconductor applications, in addition to organic crystals made of small molecules like those introduced above. A simple example is poly(p-phenylene vinylene), PPV, shown in Fig. 16g. Since this polymer does not dissolve in common solvents, more conveniently prepared derivatives of PPV are widely applied. Thin films of polymers are often formed by solution processing such as spin casting, resulting in polycrystalline or amorphous solids with entangled long polymer chains. These films are more robust than the crystalline films prepared from small molecules; their electrical properties are, however, inferior to crystalline solids.

Using such molecules as building blocks, organic crystals are formed with one or several molecules per unit cell. The frequently observed *herringbone alignment* of neighboring molecules is illustrated in Fig. 17a for a pentacene crystal. The nonpolar acene molecules are planar, and a similar crystalline arrangement of the molecules is found for the other family members, all with herringbone angles around 50°; the respective shapes of monoclinic and triclinic unit cells do not differ so much (except for the different molecule lengths and respective c values), as indicated by comparable angles α and γ near 90° listed in Table 3. The unit cell of an anthracene crystal is shown in Fig. 17b. The plane of the molecule does not coincide with a face of the unit cell. In pentacene the long axes of the two differently aligned molecules form respective angles of 22° and 20° to the c-axis, their short axes angles of 31° and 39° to the b-axis, and

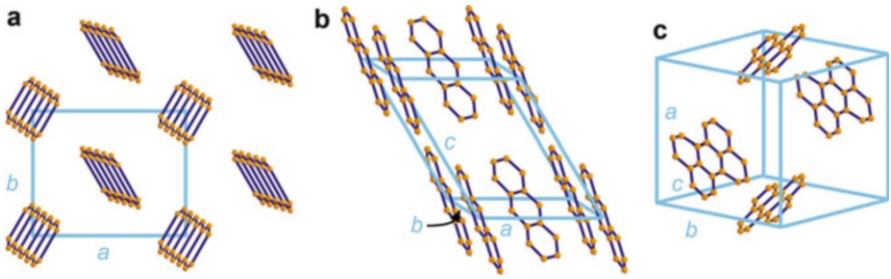


Fig. 17 Crystal structure of organic semiconductors; *orange circles* represent C atoms; H atoms are not shown. (**a**) The frequently observed herringbone packing of organic crystals, demonstrated for a top view on the *a*–*b* plane of a pentacene crystal. (**b**) Anthracene crystal with two anthracene molecules per unit cell. (**c**) Unit cell of the α phase of a perylene crystal comprising two pairs of perylene molecules

their normal axes angles of 27° and 32° to the *a*-axis of the unit cell; comparable values are found for the other acene crystals. It should be noted that the prevailing bulk structure differs from the structure predominately found in thin film growth.[13]

The herringbone packing is also realized in a variety of crystal structures of rubrene and perylene crystals. In the α phase of perylene shown in Fig. 17c, the pattern is built by molecule *pairs*, while it is formed by single molecules in the β phase (not shown). The pairing leads to a roughly doubled *b* value of the $\alpha -$ phase unit cell, while the other parameters are similar.

2 Superlattices and Quantum Structures

2.1 Superlattice Structures

Periodic alternation of one or a few monolayers of semiconductors *A* and *B* produces a composite semiconductor called a *superlattice*. Material *A* could stand for Ge or GaAs and *B* for Si or AlAs. A wide variety of other materials including alloys of such semiconductors and organic layers can also be used.

The width of each layer could be a few Angstroms in ultrathin superlattices to a few hundred Angstroms. In the first case, one may regard the resulting material as a new artificial compound (Isu et al. 1987); in the second case, the properties of the superlattice approach those of layers of the bulk material. Superlattices in the range between these extremes show interesting new properties. With epitaxial deposition techniques outlined in Sect. 3.3 in ▶ Chap. 1, "Properties and Growth of Semiconductors," one is able to deposit onto a planar substrate monolayer after monolayer of the same or a different material.

[13] In thin films the tilt of the long molecule axis with respect to the *a*–*b* plane is much smaller (3° instead of 22° for pentacene); see Ambrosch-Draxl et al. 2009.

2.1.1 Mini-Brillouin Zone

The introduction of a new superlattice periodicity has a profound influence on the structure of the Brillouin zones. In addition to the periodicity within each of the layers with lattice constant a, there is superlattice periodicity with lattice constant l. Consequently, within the first Brillouin zone of dimension π/a, a mini-Brillouin zone of dimension π/l will appear. Since l is usually much larger than a, e.g., $l = 20a$ for a periodic deposition of 10 monolayers of each material, the dimensions of the mini-Brillouin zone is only a small fraction (a/l) of the Brillouin zone and is located at its center with Γ coinciding. Such a mini-zone is of more than academic interest, since the superlattice is composed of alternating layers of different materials. Therefore, reflections of waves, e.g., excitons or electrons, can occur at the boundary between these materials. The related dispersion spectrum (discussed in Sect. 3.2 in ▶ Chap. 4, "Elasticity and Phonons," and Sect. 3.1.2 in ▶ Chap. 8, "Bands and Bandgaps in Solids") will become substantially modified, with important boundaries at the surface of such mini-zones. It is this mini-Brillouin zone structure that makes such superlattices especially interesting; this will become clearer in later discussions throughout the book. A more detailed discussion of the mini-zones is inherently coupled with corresponding new properties and is therefore postponed to the appropriate sections in this book.

2.1.2 Ultrathin Superlattices

Single or up to a few atomic layer sequential depositions can be accomplished (Gossard 1986; Petroff et al. 1979) even between materials with substantial lattice mismatch, e.g., Si and Ge, GaAs, and InAs (Fig. 18). The thickness of each layer must be thinner than the critical length beyond which dislocations (see Sect. 4 in ▶ Chap. 15, "Crystal Defects") can be created. This critical length decreases

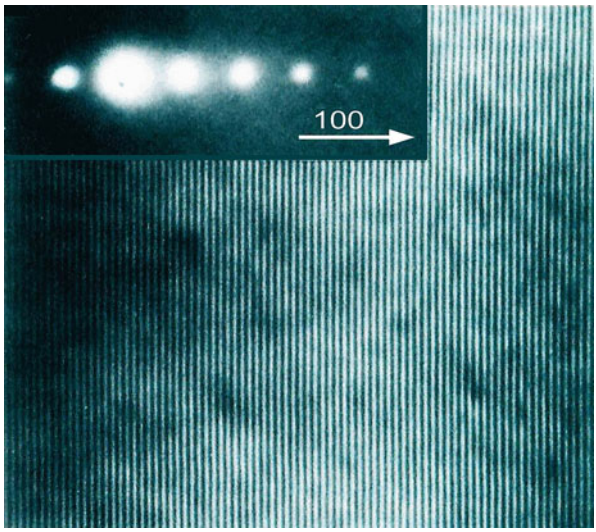


Fig. 18 Transmission-electron micrograph of an ultrathin superlattice of $(GaAs)_4$–$(AlAs)_4$ bilayers. The *inset* shows an electron-diffraction pattern (After Petroff et al. 1978)

(inversely) with increasing lattice mismatch and is on the order of 25 Å for a mismatch of 4%.

In ultrathin superlattices, the transition range between a true superlattice and an artificial new compound is reached. This opens an interesting field for synthesizing a large variety of compounds that may not otherwise grow by ordinary chemical reaction followed by conventional crystallization techniques.

Estimates as to whether or not such a spontaneous growth is possible have been carried out by estimating the enthalpy of formation of the ordered compound from the segregated phases. For instance, for a single-layer $(GaAs)_1$–$(AlAs)_1$ ultrathin superlattice, the formation enthalpy from the components GaAs and AlAs is given by

$$\Delta Q = E_{\mathrm{GaAlAs2}} - (E_{\mathrm{GaAs}} + E_{\mathrm{AlAs}}). \tag{12}$$

The formation enthalpy depends on the lattice mismatch. It is on the order of 10 meV for ultrathin superlattices with low mismatch (GaAs–AlAs) and about one order of magnitude larger for superlattices with large mismatch (such as GaAs–GaSb or GaP–InP), as shown in Table 4. The diatomic system Si–Ge, while having a large lattice mismatch, nevertheless shows a lower formation enthalpy, for reasons of lower constraint of the lattice.

The formation enthalpy also decreases with increasing thickness of each of the layers (Wood et al. 1988). Therefore, the ultrathin superlattices of isovalent semiconductors are chemically unstable with respect to the segregated compounds. These always have a lower formation enthalpy. Alloy formation does not require nucleation necessary for crystal growth of the segregated phases. Therefore, alloy formation of GaAs–AlAs is the dominant degradation mechanism. Recrystallization is usually frozen-in at room temperature.

Superlattices with low lattice mismatch, however, are also unstable with respect to alloy formation, e.g., to $Ga_{1-x}Al_xAs$, which has a formation enthalpy between that of the superlattice and the segregated phases. In contrast, the alloy formation energy of semiconductors with large mismatch lies above that for ultrathin superlattices. They are therefore more stable (Wood and Zunger 1988).

Several of these ultrathin superlattices can be grown under certain growth conditions *spontaneously* as an ordered compound, without artificially imposing layer-by-layer deposition, for instance, $(GaAs)_1$ $(AlAs)_1$ grown near 840 K by Petroff et al. (1978) and Kuan et al. (1985), $(InAs)_1$ $(GaAs)_1$ grown by Kuan et al. (1987), $(GaAs)_1$ $(GaSb)_1$ grown by Jen et al. (1986), $(InP)_n$ $(GaP)_n$ grown by Gomyo et al.

Table 4 Formation enthalpy for single-layer superlattices (After Wood et al. 1988; and Dandrea and Zunger 1991)

Superlattice	Lattice mismatch (%)	Formation enthalpy (meV/4 atoms)
$(GaAs)_1$–$(AlAs)_1$	0.1	11.5
$(GaP)_1$–$(InP)_1$	7.4	91
$(GaAs)_1$–$(AlSb)_1$	7.5	115

(1987), and $(InAs)_1$ $(GaAs)_3$ + $(InAs)_3$ $(GaAs)_1$ grown by Nakayama and Fujita (1985). All of these lattices grow as ordered compounds of the $A_nB_{4-n}C_4$ adamantine type (see Sect. 1.4.2).

2.1.3 Intercalated Compounds and Organic Superlattices

Intercalated Compounds

In crystals, such as graphite, which show a two-dimensional lattice structure, layers of other materials can be inserted between each single or multiple layer to form new compounds with unusual properties. This insertion of layers can be achieved easily by simply dipping graphite into molten metals, such as Li at 200–400 °C. After immersion, the *intercalation* starts at the edges and proceeds into the bulk by rapid diffusion. In graphite intercalation compounds may either occupy every graphite layer (stage 1 compounds) or every second layer (stage 2), such that two graphite layers alternate with a layer of intercalated material. Stage 1 binary graphite–metal intercalation has stoichiometry XC_8 for large metals (X = K, Rb, Cs) and XC_6 for small metals (X = Li, Sr, Ba, Eu, Yb, Ca). Intercalation changes the charge distribution and bonding in graphite; the compounds KC_8 or LiC_6, e.g., are transparent (yellow) and show anisotropic conductivity and low-temperature superconductivity. In the process of intercalation, the metal atom is ionized while the graphite layer becomes negatively charged. When immersed in an oxidizing liquid, the driving force to oxidize Li can be strong enough to reverse the reaction. This reversible process is attractive in the design of high-density rechargeable batteries when providing electrochemical driving forces.

Other layer-like lattices can also be intercalated easily. An example is TaS_2. Many of these compounds have extremely high diffusivity of the intercalating atoms. Some of them show a very large electrical anisotropy.

For reviews, see Whittingham and Jacobson (1982), Zabel and Solin (1990, 1992), Enoki et al. (2003), Emery et al. (2008); anion intercalation in perovskite oxides for supercapacitors is reviewed in Nan et al. (2019).

Organic Superlattices

Well known are the *Langmuir–Blodgett films* (Langmuir 1920; Blodgett 1935), which are monomolecular films of highly anisotropic organic molecules, such as alkanoic acids and their salts which form long hydrophobic chains. One end of the chain terminates in a hydrophobic acid group. Densely packed monomolecular layers can be obtained while floating on a water surface; by proper manipulation, these layers can be picked up, layer by layer (Fig. 19), onto an appropriate substrate, thereby producing a highly ordered superlattice structure; up to 10^3 such layers on top of each other have been produced. The ease in composing superlattices with a large variety of compositions makes these layers attractive for exploring a number of technical applications including electro-optical and microelectronic devices. For reviews, see Roberts (1985), Agarwal (1988), and Richardson (2000). More recent work also applied the Langmuir–Blodgett technique for fabricating well-ordered mesoscopic structural surfaces; see Chen et al. (2007).

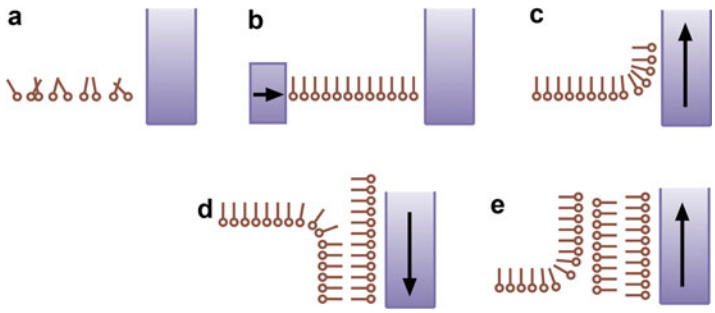


Fig. 19 Langmuir–Blodgett technique to produce multilayer films of amphiphilic, i.e., either hydrophilic or hydrophobic molecules from a water surface in a head-to-head and tail-to-tail mode. (**a**) Monolayer on top of water surface, (**b**) monolayer compressed and ordered, (**c**) monolayer picked up by glass slide moving upward, (**d**) second monolayer deposited by dipping of glass slide, and (**e**) third monolayer picked up by glass slide moving upward

Superlattices made of small molecules like those introduced in Sect. 1.5 have also been demonstrated, e.g., $H_2Pc/F_{16}CuPc$ (Zhu et al. 2009) or pentacene($C_{22}H_{14}$)/perfluoro-pentacene($C_{22}F_{14}$) (Kowarik et al. 2015); for a review see Yan et al. (2010).

2.2 Quantum Wells, Quantum Wires, and Quantum Dots

The reduction of the dimensions of a solid from three (3D) to 2D, 1D, or 0D leads to a modification of the electronic density-of-states (discussed in ▶ Chap. 8, "Bands and Bandgaps in Solids" Sect. 3.2). The effect of *size quantization* gets distinguishable if the motion of a quasi-free charge carrier with effective mass m^* (introduced in ▶ Chap. 6, "The Origin of Band Structure," Sect. 2.2) is confined to a length scale in the range of or below the de Broglie wavelength $\lambda = h/p = h/\sqrt{2m^*E}$. For a thermal energy $E = (3/2)\,kT = 26$ meV at room temperature and an effective mass of one tenth of the free electron mass, a typical length is in the 10 nm range. For excitons, i.e., correlated electron–hole pairs introduced in ▶ Chap. 14, "Excitons," the relevant length scale is the exciton Bohr radius given by

$$a_X = \frac{h^2 \varepsilon \varepsilon_0}{\pi \mu e_0^2}, \tag{13}$$

where ε, ε_0, μ, and e_0 are the relative permittivity of the solid, the permittivity of vacuum, the reduced mass of the exciton, and the electron charge, respectively. A typical length to observe size quantization for excitons is also in the 10 nm range. The Bohr radius of confined excitons is somewhat affected by a spatial localization (Bastard and Brum 1986).

Fabrication of such small semiconductor nanostructures usually employs self-organization phenomena during epitaxial growth, because patterning by etching or

implantation techniques inevitably introduce defects which deteriorate the electronic properties. Most approaches are based on an anisotropy of surface migration of supplied atoms originating from a nonuniform driving force like strain. Thereby structurally or compositionally nonuniform crystals with dimensions in the nanometer range may be coherently formed without structural defects.

2.2.1 Quantum Wells

A quantum well (QW) is made from a thin semiconductor layer with a smaller bandgap energy clad by semiconductors with a larger bandgap forming barriers. Usually the same material is used for lower and upper barrier, leading to a symmetrical square potential in one direction with a confinement given by the band offsets in the valence and conduction bands (Sect. 3.1 in ▶ Chap. 8, "Bands and Bandgaps in Solids"). Semiconductors with a small bandgap tend to have a large lattice constant; since coherent growth (without detrimental misfit dislocations, Sect. 1 in ▶ Chap. 16, "Crystal Interfaces") requires a low mismatch of lattice constants (typically below 1%), QWs or cladding barriers are usually alloyed by applying Vegard's rule (Eq. 11 in ▶ Chap. 2, "Crystal Bonding") to achieve matching. Still QWs are often coherently strained with an in-plane lattice parameter determined by the substrate material and a vertical lattice parameter resulting from Poisson's ratio of the QW material.[14] Even lattice matching at growth temperature may result in significant mismatch at room (or cryogenic) temperature due to differences in thermal expansion of QW and barriers or substrate materials (Sect. 2 in ▶ Chap. 5, "Phonon-Induced Thermal Properties").

Strain in a QW effects a splitting of confined carrier states. In addition, piezoelectric polarization is induced; the effect is particularly pronounced in semiconductors with wurtzite structure like column III nitrides or ZnO. The effect of strain on the bandgap is discussed in Sect. 2.2 in ▶ Chap. 8, "Bands and Bandgaps in Solids."

2.2.2 Quantum Wires

Fabrication of a one-dimensional quantum wire requires some patterning to define a lateral confinement in addition to the vertical cladding. The interface-to-volume ratio of 1D structures is larger than that of 2D quantum wells, so that interface fluctuations of thickness or composition on a length scale of the exciton Bohr radius easily lead to carrier localization referred to as zero-dimensional regime. Fabrication techniques of 1D wires with high optical quality imply epitaxial techniques like growth on V-groove substrates or corrugated substrates (for a review, see Wang and Voliotis 2006) and the approach of nanowire growth (see, e.g., Choi 2012).

[14]Poisson's ratio denotes the negative quotient of transverse strain/longitudinal strain for uniaxial stress, generally yielding a positive quantity (typically 0.25 . . . 0.3): transverse tensile strain leads to longitudinal compressive strain and vice versa; see also Sect. 1.1 in ▶ Chap. 4, "Elasticity and Phonons."

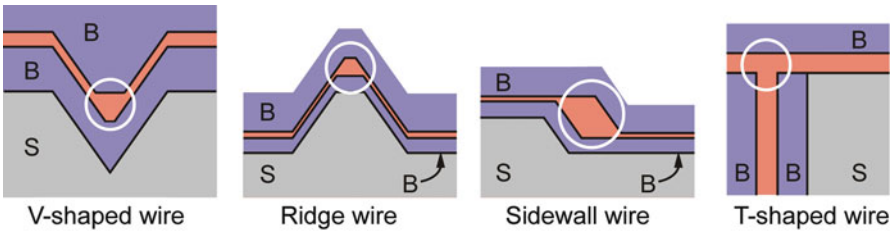


Fig. 20 Cross-section schemes of epitaxial quantum wires (*encircled*). B and S signify barrier and substrate materials, respectively

Epitaxial Quantum Wires

Most epitaxial techniques for fabricating 1D structures lead to complicate confinement potentials, and often an additional quantum well is coupled to the quantum wire. Successful approaches for epitaxial 1D quantum wires are V-shaped wires and T-shaped wires as illustrated in Fig. 20. The T-shaped wire depicted in Fig. 20 is formed from an overgrowth of the cleaved edge of a quantum-well structure (Wegscheider et al. 1993).

V-shaped ridge and sidewall wires are fabricated by employing the dependence of the growth rate on crystallographic orientation (Bhat et al. 1988). Column III controlled MBE of GaAs/AlGaAs superlattices yields a diffusion length of Ga adatoms according to $\lambda_{Ga} \propto \exp(-E_{eff}/(kT))$, with E_{eff} depending on the surface orientation. At 620 °C λ_{Ga} decreases in the order of GaAs surfaces related to (110), (111)A, $(\bar{1}\,\bar{1}\,\bar{1})$B, and (001) orientations. On a nonplanar GaAs surface, Ga adatoms migrate towards facets with minimum λ_{Ga} and are incorporated there. The growth rate of facets with a larger diffusion length is therefore decreased. Quantum wires were fabricated on (001)-oriented GaAs substrate with V-shaped grooves oriented along the $[\bar{1}10]$ direction and composed of two {111}A sidewalls (obtained by wet etching). During growth of a lower AlGaAs barrier layer, the adatom diffusion-length is quite short and does not show a pronounced facet dependence; the V-groove bottom therefore remains quite sharp. In the subsequent GaAs growth, Ga adatoms impinging on the {111}A sidewalls tend to migrate with a long diffusion length to facets with a short diffusion length. Thereby the growth rate is enhanced at the bottom of the V-groove, and a (001) facet is generated. Eventually the GaAs layer is capped by an upper AlGaAs barrier, leaving buried regions of an enhanced thickness which act as a quantum wire.

During growth of the upper AlGaAs layer, the diffusion length of Ga adatoms is again quite short. This leads to a sharpening of the V-groove bottom and allows for creating a vertical stack of quantum wires as shown in Fig. 21. The dark regions in the AlGaAs layers labeled VQW (vertical QW) represent Ga-rich material with a lower bandgap.

Nanowires

A different approach for creating a 1D nanowire is the *vapor–liquid–solid* (VLS) mechanism (Wagner and Ellis 1964). A metal catalyst (such as gold) forms at a high temperature liquid alloy droplets by adsorbing gaseous components of the material to be grown as illustrated in Fig. 22. At supersaturation the soluted components

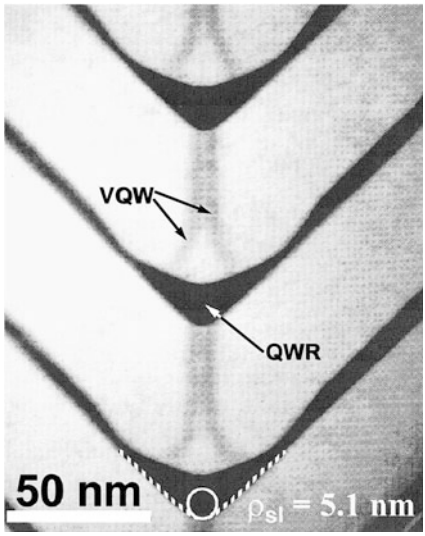


Fig. 21 Cross-section transmission-electron micrograph of vertically stacked GaAs/$Al_{0.42}Ga_{0.58}As$ quantum wires. AlGaAs appears bright; the *white circle* marks the radius of curvature at the *bottom* interface of the wire (After Gustafsson et al. 1995)

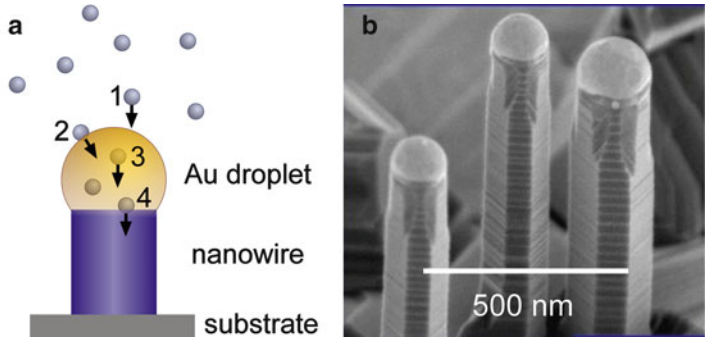


Fig. 22 (**a**) Successive steps in the VLS mechanism applied to create a nanowire. (**b**) Scanning electron micrographs of Si nanowires grown at 700 °C by the VLS mechanism on Si(111) (MPI Halle 2007)

precipitate at the liquid–solid interface (when growth commences precipitation starts at the interface to the substrate), leading to 1D whisker growth with typ. 0.1 μm diameter; the liquid droplet remains at the top of the growing nanowire. Axial heterostructures with a change of composition or doping along the wire axis are formed by changing the composition of the gas phase. Also radial heterostructures with interfaces along the wire axis can be formed; after completing the growth of a nanowire core, the growth conditions are altered to deposit a shell material. Multiple shell structures are produced by subsequent introduction of different materials.

2.2.3 Quantum Dots

A quantum dot (QD) is a zero-dimensional nanostructure providing fully quantized electron and hole states similar to discrete states in an atom. Interface perfection is

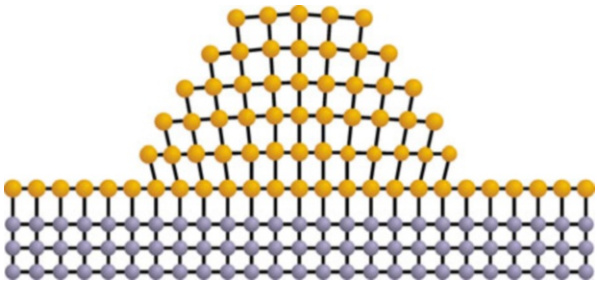


Fig. 23 Scheme illustrating elastic strain relaxation by forming three-dimensional islands of a layer material (*orange circles*) with large bond length, deposited on a substrate material with a smaller bond length (*blue circles*); note the fully strained wetting layer

crucial for 0D nanostructures due to a very high interface-to-volume ratio. Fabrication techniques comprise epitaxial QDs and colloidal QDs.

Epitaxial Quantum Dots

Epitaxial QDs are mostly fabricated self-organized by applying the Stranski–Krastanow growth mode introduced in Sect. 3.1.4 of ▶ Chap. 1, "Properties and Growth of Semiconductors." This growth mode may be induced by epitaxy of a highly strained layer, which initially grows two-dimensionally and subsequently transforms to three-dimensional islands due to elastic strain relaxation (Bimberg et al. 1999), cf. Fig. 23. Some part of the material remains as a two-dimensional wetting layer, due to a low surface free energy compared to the covered (barrier) material. The size of the three-dimensional islands lies for many semiconductors in the range required for quantum dots; the QDs are formed by capping such islands with an upper barrier material. The minimum diameter for a QD required to confine at least one bound state of a carrier is in the nanometer range, see section (Sect. 3.4 in ▶ Chap. 8, "Bands and Bandgaps in Solids").

The total energy gain for the formation of three-dimensional islands with respect to a two-dimensional layer is given by strain and surface-energy contributions of both the reorganized part of the material forming the islands and the part remaining in the wetting layer after the Stranski–Krastanow transition. The contributions sensitively depend on the shape of the islands; their sum is given in Fig. 24 for pyramidal InAs islands with {110} side facets and a (001) surface of the wetting layer, grown on (001)-oriented GaAs (Wang et al. 1999). The total energy density has an energy minimum for a particular island size (see arrows), creating a driving force towards a uniform size for an ensemble of islands.

Stranski–Krastanow growth induced by strain is found for both compressively and tensely strained layers in various materials systems and crystal structures. Table 5 and Fig. 25 give some examples. Usually substrate material is also employed for covering the islands after formation. The barrier material is then generally termed matrix. Often the island material is alloyed with matrix material to reduce the strain, yielding a parameter for controlling the transition energy of confined carriers. The shape of the islands is generally strongly modified during the capping process: the islands tend to become flat during cap layer deposition; often quantum dots with a shape of truncated pyramids are formed (Costantini et al. 2006).

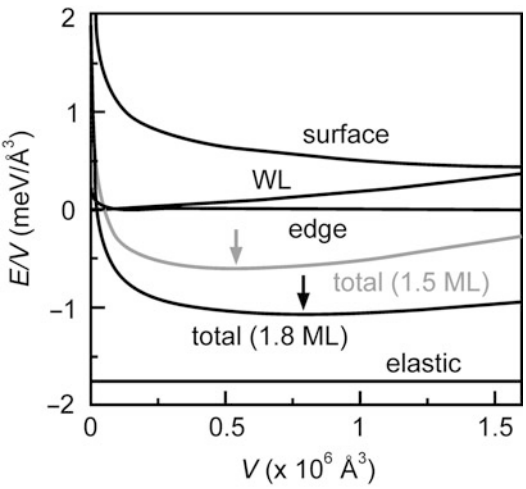


Fig. 24 Total energy gain calculated for the island formation of a 1.8 monolayers thick InAs layer on GaAs (*black lines*); an areal density of 10^{10} islands/cm^2 is assumed. The curves represent various contributions as indicated; WL signifies the wetting layer. The *gray line* refers to the total energy gain for 1.5 monolayers InAs. *Arrows* mark the minima of the total energy curves (After Wang et al. 1999)

Table 5 Some semiconductor materials used for strain-induced, self-organized Stranski–Krastanow formation of islands

Island/matrix	Ge/Si	InAs/GaAs	GaN/AlN	PbSe/PbTe
Structure	Diamond	Zincblende	Wurtzite	Sodium chloride
Orientation	(001)	(001)	(0001)	(111)
Mismatch	−3.6%	−7%	−2.5%	+5.5%

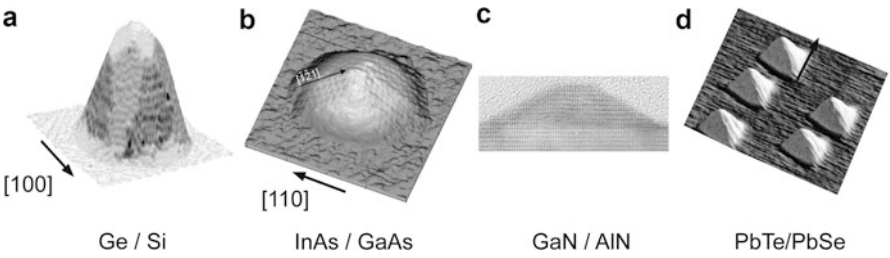


Fig. 25 Free-standing self-organized islands formed by Stranski–Krastanow growth in various strained heteroepitaxial materials: (**a**) Ge/Si(001) (After Rastelli et al. 2001), (**b**) InAs/GaAs(001) (After Márquez et al. 2001), (**c**) GaN/AlN(0001) (After Xu et al. 2007), (**d**) PbTe/PbSe(111) (After Pinczolits et al. 1998). The AFM images (**a**), (**b**), and (**d**) are vertically not to scale with respect to the lateral scale

Colloidal Quantum Dots

Colloidal quantum dots, also termed nanocrystals or nanocrystal QDs, are synthesized from precursor compounds dissolved in solutions (for a review see Murray et al. 2000). At high temperature the precursors chemically transform into monomers. Nanocrystal nucleation starts at sufficient supersaturation of dissolved

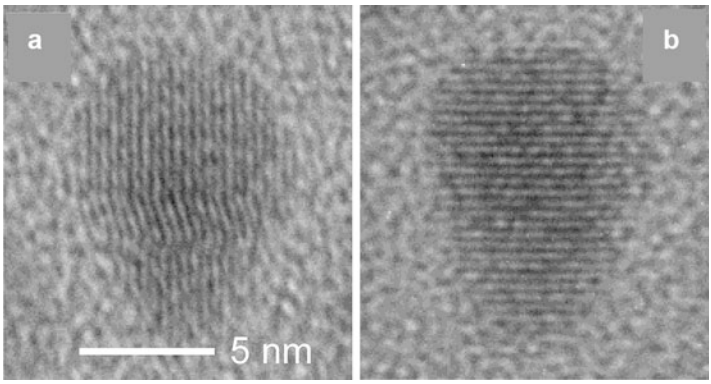


Fig. 26 High-resolution transmission-electron micrograph of nanocrystals with 3.5 nm CdSe core and 5 monolayers of CdS shell, elongated along the wurtzite *c*-axis (*vertical* in the figure); after Li et al. 2003

monomers. At high monomer concentration, the critical size where growth balances shrinkage is small; smaller nanocrystals grow faster than large ones (they need less atoms to grow), leading to a narrow size distribution of the ensemble.[15] Core/shell structures can be produced similar to nanowires; see Fig. 26.

Colloidal QDs can be embedded in glass matrices or in organic and related matrices; respective properties are described by Woggon 1997. Close-packed ordering of nanocrystals can be prepared by solvent evaporation, yielding nanocrystal solids with long-range order. Nanocrystals were also arranged in superlattices; reviews are given in Murray et al. (2000) and Hanrath (2012).

3 Amorphous Structures

Although there is no macroscopic structure[16] discernible in amorphous semiconductors (*glasses* for brevity), there is a well-determined *microscopic order* in atomic dimensions, which for nearest and next-nearest neighbors is usually nearly identical to the order in the crystalline state of the same material. The *long-range order*, however, is absent (see Phillips 1980 and Singh and Shimakawa 2003).

In many respects, the glass can be regarded as a supercooled liquid. When cooling down from a melt, glass-forming materials undergo two transition temperatures: T_f where it becomes possible to pull filaments (honey-like consistency) and T_g, where form elasticity is established, i.e., the glass can be formed into any arbitrary shape –

[15]Nanocrystals of 2–10 nm diameter (corresponding to 10–50 atom diameters) contain some 10^2–10^5 atoms.

[16]The surface of glasses, even at very high magnification, does not show any characteristic structure; after fracture, glasses show no preferred cleavage planes whatsoever.

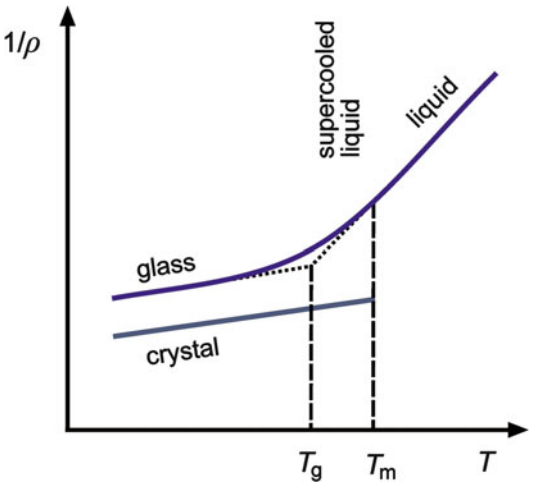


Fig. 27 Schematics of the dependence of the reciprocal density as a function of the temperature

its viscosity has reached $10^{15}p$, and its atomic rearrangement time is ~10^5 s. Only T_g is now used as the transition temperature and is identified in Fig. 27 (for a review see Jäckle 1986). When plotting certain properties of a semiconductor – such as its specific density (Fig. 27), the electrical conductivity, and many others as a function of the temperature – a jump and break in slopes are observed at the melting temperature T_m when crystallization occurs. Such a jump is absent when cooling proceeds sufficiently fast and an amorphous structure is frozen-in.

Fast cooling (quenching) for typical glasses is already achieved with a rate <1 deg/s, while many solids, including metals, become frozen-in liquids and remain amorphous at room temperature solely when this rate is ~10^7 deg/s, which can be achieved by *splat cooling* on fast rotating disks.

Near a transition temperature $T_g < T_m$, the slope gradually changes and, for $T < T_g$, the curves in Fig. 27 for a glass and a crystal of the same material run essentially parallel to each other.

Materials that have a large fraction of covalent bonding (see Sect. 1 in ▶ Chap. 2, "Crystal Bonding") show a tendency for glass formation. The liquid becomes significantly more viscous before crystallization takes place. The composition range for glass formation is shown for some ternary compounds in Fig. 28. In this range, while still liquid, cross-linking of many atoms has already taken place, and the principal building blocks (see below) of the glass are established; however, they cannot adjust with sufficient rigor to produce long-range periodicity. Nevertheless, all bonds tend to be satisfied by attachment to an appropriate neighbor.

The resulting structure is composed of principal building blocks that join each other with slight deviation from the preferred interatomic angle and distance (*frustration*) and consequent relaxation of these relations within the building blocks. In contrast, during crystallization, such building blocks easily break up so that a larger crystallite can grow by sequentially adding atoms rather than entire building blocks. During glass formation, there is usually little tendency to form *dangling bonds*, i.e., bonds not extending between two atoms – see *matrix glasses* in Sect. 3.2.2.

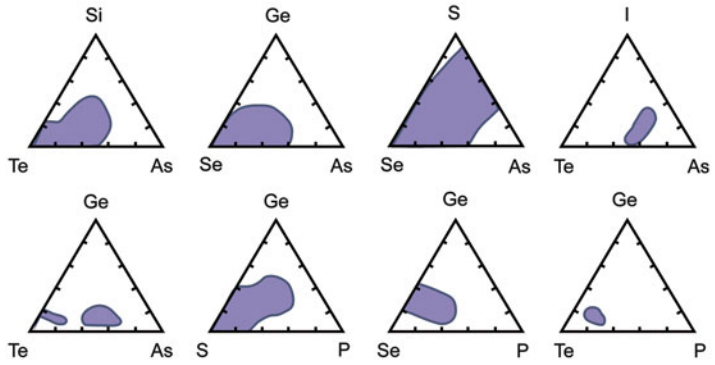


Fig. 28 Approximate glass-forming (*shaded*) regions in a few ternary semiconductor alloy systems – a point within this triangle represents an alloy of the three components of a composition given by the normal to each side of the triangle (After Mott and Davis 1979)

The structural analysis of an amorphous semiconductor can therefore be divided into two parts: the principal atomic building blocks and the arrangement of these blocks to form the glass.

3.1 Building Blocks and Short-Range Order

3.1.1 Building Blocks

Atomic semiconductors can be dealt with most easily since all of the neighbors are equivalent. Perfect crystalline order with tetrahedral binding (fourfold coordination) requires the formation of six-member rings – see Fig. 2. in ▶ Chap. 1, "Properties and Growth of Semiconductors." Polk (1971) introduced *odd-numbered rings* (5 or 7) and thereby formed glasses with an otherwise tetrahedrally coordinated arrangement of atoms around these building blocks. Such odd-numbered rings were later confirmed in α-Si (Pantelides 1987) and in α-C (Galli et al. 1988).

Comparing crystalline (c) and amorphous (α) structures of the same element (e.g., Ge), one sees that first- and second-neighbor distances (2.45 vs. 2.46 and 4.02 vs. 4.00 Å for c-Ge vs. α-Ge, respectively) are surprisingly similar, as is the average bond angle (109.5 vs. 108.5°). There is, however, a *spread* of ±10° in the bond angle for the amorphous structure, resulting in an *average coordination number* of 3.7 rather than 4 for c-Ge (Etherington et al. 1982). The lower effective coordination number indicates a principal building-block structure that is slightly less filled but without vacancies, which are ill-defined in amorphous structures (see Sect. 2 in ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors"). Hard-sphere models, which would assist in defining sufficient space between the spheres as vacancies, must be used with caution since covalent structures can relax interatomic lattice spacing when relaxing bond angles (Waire et al. 1971).

Binary compounds are more difficult to arrange in such a fashion, since odd-member rings cannot be formed in an *AB* sequence without requiring at least one *AA* or *BB* sequence. Random network models, however, can also be made with larger *even-numbered rings*. Zachariasen (1932) suggested the first one for SiO_2-type glasses, which was shown for a two-dimensional representation in Fig. 2 in ▶ Chap. 1, "Properties and Growth of Semiconductors."

Many covalent polyatomic binary compounds containing chalcogens easily form semiconducting glasses such as As_2S_3, As_2Se_3, or Ge_xTe_y. The principal building blocks obey the $8 - N$ *rule*, yielding the number of bonding partners. For example, As with $N = 5$ valence electrons is bonded to three Se atoms, while the Se with $N = 6$ in turn is surrounded by two As atoms in an As–Se–As configuration. Thereby both As and Se atoms obtain an octett configuration. Similarly, the Ge with $N = 4$ is surrounded by four Te atoms, while each Te atom with $N = 6$ has two Ge atoms as nearest neighbors in $GeTe_2$, similar to the SiO_2 configuration; for a discussion of the 8-*N* rule and its relation to dopant compensation see Robertson (2016).

In some of these amorphous chalcogen compounds, however, the interatomic nearest-neighbor distance is shorter and the coordination number is significantly lower than in the corresponding crystalline compounds (Bienenstock 1985). A chalcogen–chalcogen pairing [e.g., by including Ge–Te–Te–Ge or an ethane-like $Ge_2(Te_{l/2})_6$ formation[17]] can distort the building blocks. The large variety of possible Ge_xTe_y building blocks, still fulfilling the $8 - N$ rule, created by replacing Ge–Ge with Ge–Te or Te–Te bonds, is the reason that glasses of a continuous composition from pure Ge to pure Te can be formed (Boolchand 1985).

3.1.2 Short-Range Order

In crystalline covalent semiconductors, the coordination number m is given by the $8 - N$ rule, which is 4 for Si. For compounds one can define an *average coordination number*, drawing a shell in the atomic distribution function around an arbitrary atom and averaging (Fig. 29). These shells contain at nearest-neighbor distance a maximum of $m = 4$ atoms for GaAs, $m = 3$ for GeTe (also for As), and $m = 2$ for a linear lattice such as Se. The average coordination number is $\overline{m} = 2.7$ for $GeTe_2$ and $\overline{m} = 2.4$ for As_2Se_3. In a crystal there are $m/2$ *constraints* per atom with respect to bond length, since two atoms share a bond. This can easily be fulfilled if $m/2 \leq 3$, since each atom can shift with respect to its neighbor in three dimensions. There are also $m(m-1)/2$ constraints with respect to the bond angle, since it is defined by three atoms. Therefore, bond length and bond angle are constrained only if

$$\frac{\overline{m}}{2} + \frac{\overline{m}(\overline{m}-1)}{2} \leq 3, \quad \text{or} \quad \overline{m} \leq \sqrt{6} \cong 2.4. \tag{14}$$

[17]The "chemical formula" using $Te_{l/2}$ shows the symmetry of the Te binding and indicates that on the other side of each of the Te atoms another Ge atom is bound.

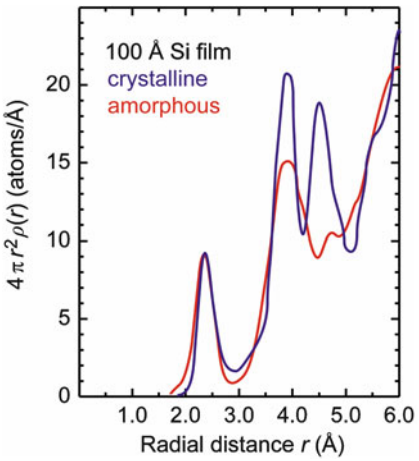


Fig. 29 Distance distribution (radial distribution function) of atoms in amorphous (*red*) and crystalline (*blue curve*) Si layers of 100 Å thickness (After Moss and Graczyk 1970)

Si- and GaAs-type semiconductors are *overconstrained*: a large internal strain prevents any significant deviation from its ordered, crystalline state (Li et al. 2019). In glasses, overconstrained conditions create stress-rigid networks with internal stress due to a competition among redundant constraints. Not so Se or As_2Se_3. The former, with $\overline{m} = 2$, is *underconstrained*: it provides a large amount of freedom for deviation from uniformity in bond length and angle; therefore, it easily forms amorphous structures. The latter, As_2Se_3, needs only minor alloying to cause $\overline{m}$ to drop below 2.4 and therefore also forms a glass easily. Since Eq. 14 shows a quadratic dependence on $\overline{m}$, a glass-forming tendency is rather sensitive to a lowering $\overline{m}$ (Ovshinsky 1976; Adler 1985; Phillips 1980).

The above-described principal building blocks are also described as *intermediate-range order*. These blocks are composed of subunits, identified by the *short-range order* of a few atoms, and characterized by bond lengths, bond angles (next-nearest-neighbor distances), and site geometry. Intermediate-range order describes *third-neighbor distances*, *dihedral angles*, *atomic ring structures*, and *local topology*. It distinguishes for tetra-, tri-, and divalent bonding truly three-dimensional (tetrahedral), two-dimensional (layer-like), and one-dimensional (chain-like) structures, respectively.

Intermediate-range order shows some interesting features that distinguish amorphous from crystalline states. For instance, monatomic column IV semiconductors crystallize only in the diamond lattice with a dihedral angle of 60°. Amorphous Ge, however, shows a dihedral angle of 0° (depicted in Fig. 9a). The interesting feature of this structure is the disappearance of the third-neighbor peak in diffraction analysis, which is observed at 4.7 Å for *c*-Ge. With a dihedral angle of 0° for *α*-Ge, this third-neighbor distance is only 4.02 Å; thus it is very close to, and nearly indistinguishable from, the second nearest neighbor at 4.0 Å (see Fig. 29).

The formation of amorphous solids may be illustrated by the ab initio random structure searching method (Pickard and Needs 2011). We consider small $(CdS)_n$ clusters made of *n* Cd-S pairs; for each pair number *n* the energy for a set of up to 300

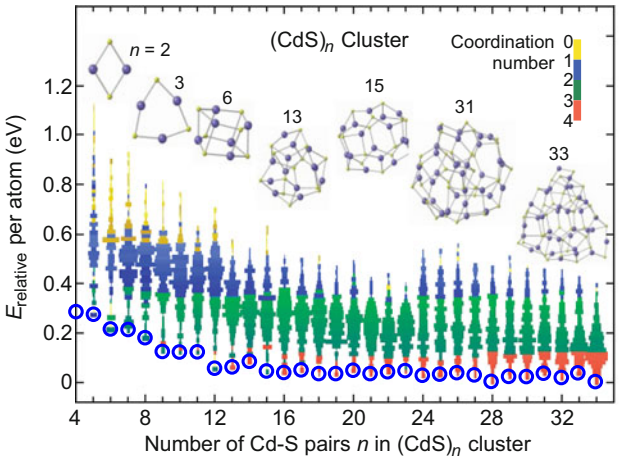


Fig. 30 Calculated relative energy per atom in $(CdS)_n$ nanoclusters formed by pairs of Cd and S atoms with various numbers n of Cd-S pairs. The *width of the bars* indicates the relative abundance, the *color of the bars* signifies the mean coordination number on a scale given at the top; *blue circles* mark the lowest-energy structures found for each cluster size. The *inset* shows minimum-energy clusters for some numbers of n. (After Tan et al. 2019)

stable random clusters was computed using using the density functional theory (Tan et al. 2019). Typical interatomic distances and a 1:1 stoichiometric ratio of the two atomic species were applied as constraints. Results given in Fig. 30 show an initital decrease in the relative energy of the cluster E_{relative} with increasing cluster size; the energy is defined by $E_{\text{relative}} = E(n)/n - E_{\text{min}}$, with $E(n)$ the total energy of of a cluster with n Cd-S pairs and E_{min} the minimum energy of all studied clusters. We note by the vertical spread of energies that there are many cluster configurations for each cluster size given by n. Their relative abundance is reflected by the width of the bars, indicating also different configurations with *similar* energy. Clusters with lower coordination have generally higher energy. Small clusters (n = 2 − 5) form rings with two-fold coordination as illustrated by the two clusters with lowest energy in the inset. Three-dimensional structures appear for $n \geq 6$; $(CdS)_6$ is formed from two stacks of hexagonal $(CdS)_3$ rings, yielding a simple cage structure. For the larger clusters $(CdS)_n$ with n = 10−20, cage structures with tri-coordinated atoms and cage −core structures consisting of a cage with some atom pairs in a core structure appear. Above a certain size (n = 19 for CdS) bulklike structures with an average coordination number >3 at interior sites are favored. The wider bars reflect the large number of cluster structures for a given size n found in the corresponding energy range. Lowest energy per atom is attained for largest clusters and four-fold coordination, indicating a decreasing effect of the surface; still a spread in coordination remains in amorphous structures.

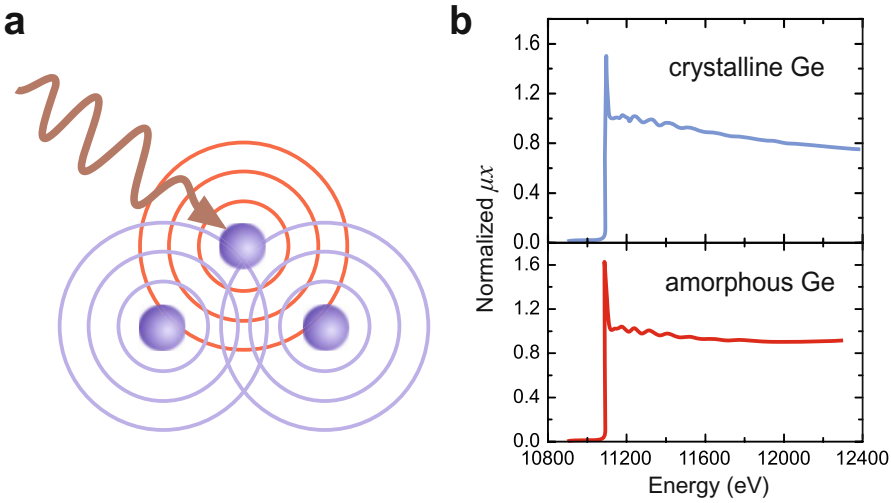


Fig. 31 (**a**) EXAFS representation with electron wave emitted from one atom (*red*) and scattered waves from adjacent atoms (*blue*). (**b**) EXAFS for crystalline Ge (**a**) and for amorphous Ge (**b**) (After Stern 1985)

EXAFS and NEXAFS

Information about the structure that surrounds specific types of atoms can be obtained from the extended x-ray absorption fine structure (EXAFS). With synchrotron radiation a continuous spectrum of x-rays is available for investigating absorption or luminescence spectra which show characteristic edges when an electron of a specific atom is excited from an inner shell into the continuum. Interference of such electrons with backscattered electrons from the surrounding atoms (Fig. 31a) results in a fine structure of the absorption beyond the edge (Fig. 31b). This results from interference between outgoing and reflected parts of the electron de Broglie wave, as indicated by red and blue rings in Fig. 31a. This fine structure, therefore, yields information about the distance to the surrounding atoms, as well as their number, and provides species identification of the neighbor atoms (Hayes and Boyce 1985; Bertrand et al. 2012).

EXAFS measurements do not require long-range periodicity and therefore are useful in analyzing amorphous short-range structures.

When measuring x-ray fluorescence rather than absorption, the surrounding of specific impurities of low density can also be analyzed, since such fluorescence has a much lower probability of overlapping with other emission in the same spectral range. In addition, near-edge x-ray fine structure (NEXAFS), within 30 eV of the edge, gives information from low-energy photoelectrons which undergo multiple scattering and provides information on the average coordination number, mass of neighbor atoms, average distance, and their variations with temperature. For special cases, it also yields information on the angular distribution of the surrounding atoms. For a review, see Bienenstock (1985), Stern (1978, 1985), Stöhr (1996), Koide et al. (2014).

3.2 Network Structures and Matrix Glasses

3.2.1 Network Structures

The entire glass can be composed of a *network-like structure* from elements of intermediate-range order, or as a *matrix-like structure*, which is preferable for elemental semiconductors and will be explained below (see *matrix glasses*).

Network glasses are constructed from principal building blocks with *long-range disorder* added. Such disorder can be introduced in several ways by statistical variation of interatomic distances and bond angles.

The results of a calculation of the radial density-distribution function of such random network structures are in satisfactory agreement with the experimental observation obtained from x-ray diffraction data and from EXAFS or NEXAFS, as shown for amorphous Ge and GaAs in Fig. 32; see also Roorda et al. (2012).

Requirements for creating such a network were given by Bell and Dean (1972) and applied to α-SiO_2, α-GeO_2, and α-BeF_2. When starting from a $Si(O_{1/2})_4$ unit, one proceeds with a covalent random network, connecting to it other $Si(O_{1/2})_4$ units with twofold oxygen coordination, while requiring that:

- The bond angle of Si atoms must not deviate more than $\pm 10°$ from the ideal value of 109.47°.
- All tetrahedra are corner-connected.
- The bond angles of O atoms may spread by $\pm 25°$ from the ideal value of 150°.
- There is equal probability for all dihedral angles.
- There is no correlation between bond angles at O atoms and dihedral angle.
- There is complete space filling.

Modifications of these instructions yield slightly different networks. The relation to the dihedral angle (e.g., assuming some correlation) is an example of such a modification.

A relatively simple infinite aperiodic network structure is called a *Bethe lattice* (Bethe 1935; see also Runnels 1967; Allan et al. 1982). Another kind of network

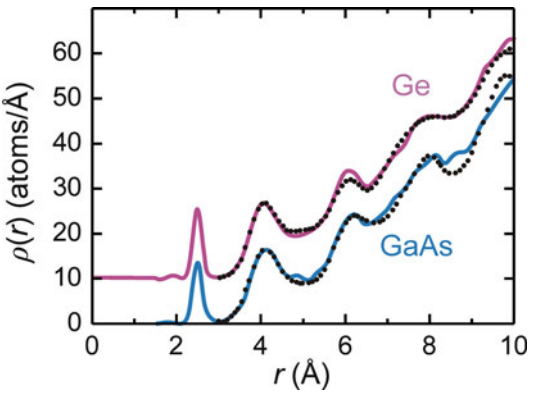


Fig. 32 Radial density function of α-Ge (Temkin et al. 1973) and α-GaAs (Temkin 1974) shown as *solid curves* and obtained from computation for a random network by Steinhardt et al. 1974 (*dots* at *upper curve*) and by Temkin 1974 (*dots* at *lower curve*)

structure is the *fractal structure*, in which void spaces between more densely filled regions can be identified (see Mandelbrot 1981).

3.2.2 Matrix Glasses, α-Si:H

By relaxing some above mentioned requirements or by including foreign atoms into an amporphous semiconductor (which acts as matrix) a large flexibility for forming new structures is achieved. Relaxing the requirements for a fourfold coordination of an atomic amorphous semiconductor creates dangling bonds. These bonds could attract monovalent elements such as H or F. Alternatively, the tetravalent host atom could be replaced with an element of lower valency such as N or O.

When foreign atoms are introduced in a density that is large enough so that their interaction can no longer be neglected, we call this process an *alloy formation*; as such, we may include homologous elements (e.g., C in α-Si). In all such cases, we then satisfy the $8 - N$ rule but achieve a greater degree of flexibility in constructing the amorphous *host matrix*, a reason why such alloys are easily formed.

In addition, we may include into such a host matrix more than one kind of atom and thereby create more complex alloys, e.g., forming α-Si:O:H or α-Si:N:H by also incorporating oxygen or nitrogen into α-Si:H. Examples of such clusters are shown in Fig. 33.

For a theoretical analysis of the various possibilities of incorporating foreign atoms (e.g., in Si–H, SiH_2, or SiH_3 configurations as shown in Fig. 34), it is useful to consider larger *atomic clusters* from this network. Such clusters contain nearest, next-nearest, and higher-order neighbors from the alloyed atom, as shown in Fig. 34.

Hydrogenated amorphous silicon (α-Si:H) is the most commonly used material for large-area optoelectronic devices because it absorbs light much more efficiently than crystalline silicon, in which crystal momentum conservation restricts optical transitions. Experimental analyses using a variety of techniques (neutron scattering, small-angle neutron scattering, EXAFS, etc.) indicate that α-Si:H contains hydrogen (about 10%) without substantially changing the bond length and the dihedral angle; see Fig. 35. It has a coordination number of 3.7 to 3.9 and may be described as a mixture of fourfold and threefold coordinated atoms. The hydrogenation is essential for eliminating the native dangling-bond defect concentration (to $\sim 10^{15}$ cm^{-3}), which produce a metastable light-induced degradation of the optoelectronic properties (*Staebler–Wronski effect*, Pantelides 1987, Fritzsche 2001). The defects are mostly dangling bonds on threefold coordinated Si atoms (SiH units; see Fig. 33,

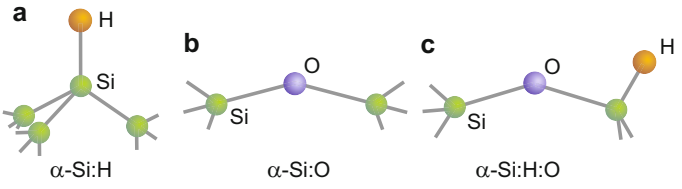


Fig. 33 Atomic clusters occurring in an α-Si:O:H alloy

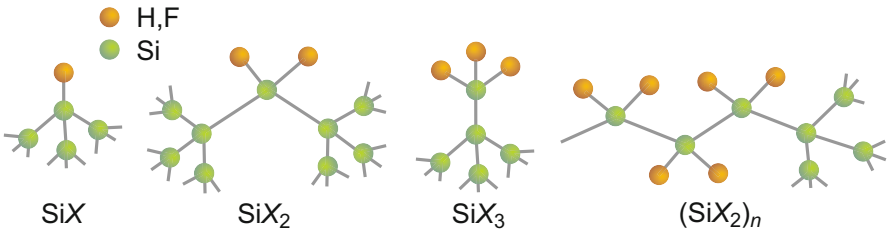


Fig. 34 Local bonding variation in an α-Si:H alloy with higher densities of hydrogen or fluorine X

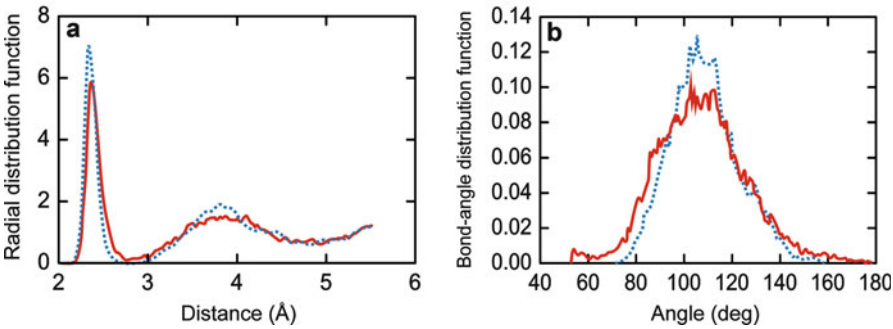


Fig. 35 (**a**) Si–Si radial distribution function and (**b**) Si–Si–Si bond-angle distribution of α-Si:H prepared by a fast (*solid red line*) or slow cooling rate (*blue dots*). The peaks of the distribution functions center near the values 2.35 Å, 3.84 Å, and 109.5° of crystalline Si and narrow with slower cooling; after Jarolimek et al. 2009

Menelle and Bellissent 1986, and Freysoldt et al. 2012); they give rise to amphoteric electronic states in the bandgap which may be occupied by up to two electrons and act as efficient recombination centers (Street 1991).

4 Quasicrystals

Quasicrystals are solids with an order between crystalline and amorphous. While a *crystal* is formed by a *periodic repetition of one* unit cell, *quasicrystals* can be assembled by an *aperiodic repetition of* (*at least*) *two different* unit cells. Such unit cells also may have fivefold symmetry as first shown by Shechtman et al. (1984), and Levine and Steinhardt (1984); this symmetry is forbidden for crystals since it cannot fill space without overlap or voids as illustrated in Fig. 36. A quasicrystal has no three-dimensional translational periodicity but still exhibits long-range order in a diffraction experiment. Furthermore, *orientational* order exists: bond angles between neighboring atoms have long-range correlations. There are two types of quasicrystals: *icosahedral quasicrystals*, which are aperiodic in all directions, and

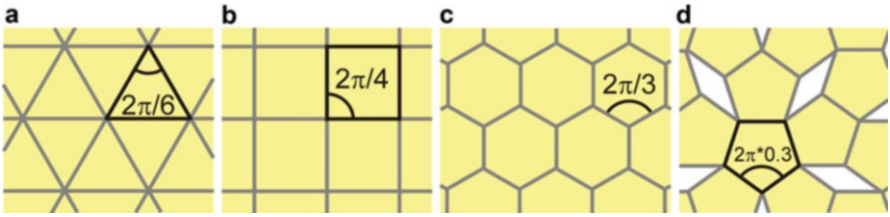


Fig. 36 Space filling in two dimensions using regular polygon tiles. For triangles, squares and hexagons (**a**–**c**) a single type of tile fills a surface; pentagons (**d**) leave gaps which can be filled with diamond-shaped tiles (Penrose 1974)

axial quasicrystals, which have an axis of 5-, 8-, 10-, or 12-fold symmetry (pentagonal, octagonal, decagonal, or dodecagonal quasicrystals, respectively); axial quasicrystals are periodic along their axis and quasiperiodic in planes normal to it (Bendersky 1985). Reviews are given in Janot (1994), Suck et al. (2002), Dubois (2005), Marciá-Barber (2021); structural aspects are discussed in Cahn et al. (1986), Steinhard (1987), Steurer and Deloudi (2009).

4.1 Quasiperiodicity and Properties of Quasicrystals

4.1.1 Quasiperiodicity

To illustrate the origin of a quasiperiodic structure, we consider the density of lattice points $\rho(x)$ on a one-dimensional lattice. A *periodic* structure of equally spaced lattice points with lattice parameter a is described by

$$\rho(x) = \sum_n \delta(x - na). \tag{15}$$

Superimposing a second periodic structure with a different lattice parameter $\alpha \times a$ yields the sum

$$\rho(x) = \sum_{n,m} \delta(x - n\ a) + \delta(x - m\ \alpha a). \tag{16}$$

This density of lattice points still has long-range order. It is, however, not periodic, if the ratio α of the two lattice parameters is not a rational number; a periodic spatial coincidence can only occur for rational numbers α. A prominent quasiperiodic chain is the one-dimensional Fibonacci sequence consisting of two spacings L (long) and S (short) with a surd ratio $L/S = \tau = \left(1 + \sqrt{5}\right)/2 = 1.618034\ldots$ (the golden ratio). The ratio τ appears in icosahedral symmetries, e.g., in the ratio of the diagonal and the edge length of a pentagonal plane, and τ is found in diffraction patterns of icosahedral quasicrystals; see Fig. 37a. The diffraction diagram also shows the self-similarity by scaling observed in

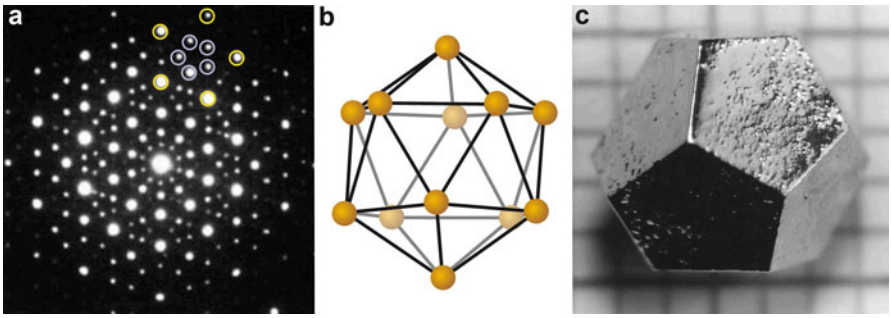


Fig. 37 (**a**) Electron diffraction pattern from an Al-14%Mn quasicrystal with icosahedral symmetry reported by Shechtman et al. 1984. *Yellow* and *blue circles* mark two pentagons scaled by the golden ratio τ. (**b**) Icosahedron viewed on one of the 20 triangular faces. (**c**) Icosahedral Zn–Mg–Ho quasicrystal (Ames Laboratory 2010)

quasicrystals. An experimental challenge of such measurements is the large variation of the intensity distribution over many orders of magnitude.

4.1.2 Quasicrystal Compounds

Quasicrystals exist in many metallic alloys, particularly in aluminum alloyed with transition metals like Mn, Co, or Cu, or with normal metals like Mg. Structures may be very complex with up to some 10^3 atoms per unit cell.[18] Both stable and metastable forms exist for quasiperiodic crystals. Stable quasicrystals synthesized using conventional metallurgy resist thermal treatment. Metastable quasicrystals are prepared by rapid cooling using melt spinning or by crystallization of the amorphous phase.[19] A few alloys forming quasicrystals are listed in Table 6.

Intermetallic quasicrystals are hard and brittle materials.[20] They have typically unusual transport properties and low surface energies. The electrical conductivity σ is low; a steady increase according to $\sigma = \sigma_{4K} + \Delta\sigma(T)$ was found as the temperature is raised with some general function $\Delta\sigma$ as reported by Mayou et al. 1993. The thermoelectric power at room temperature is large compared to Seebeck coefficients of crystalline and disordered metallic alloys; small deviations in the chemical composition give rise to large changes and even sign reversal, indicating a strong effect of the local order (Häussler et al. 2000). The thermal conductivity is very low, with a particularly small contribution of electrons. The unusual transport properties are related to the small number of free carriers, associated with a pronounced

[18]The largest unit cell experimentally found so far corresponds to the $Al_{60.3}Cu_{30.9}Fe_{9.7}$ compound comprising nearly 5,000 atoms.

[19]A natural quasicrystal, an $Al_{63}Cu_{24}Fe_{13}$ alloy termed icosahedrite, has been found in a meteorite (Bindi et al. 2011).

[20]For many quasicrystals a transition from brittle to ductile behavior was found at ~3/4 of the melting temperature (Dubois et al. 2000).

Table 6 Some metallic alloys with quasicrystal phases occurring at specific compositions

	Axial quasicrystals	
Icosahedral quasicrystals	Compound	Symmetry
$AlMn_{7.1}$, $YbCd_{5.7}$	Mn–Si, Cr–Ni–Si	Octagonal
Al–Pd–Mn, Al–Li–Cu, Al–Cu–Fe	Al–Fe–TM (TM = Cr,Mn,Ni,Pd)	Decagonal
$Zn_{60}Mg_{30}RE_{10}$ (RE = Y,Dy,Ho,Lu,Tb,Gd)	Al–Cu–TM (TM = Co,Ni,Rh)	
Al–Cu–TM (TM = Cr,Mn,Fe,Ru,Os)	Ni–V, Ni–Cr	Dodecagonal

pseudogap (Belin-Ferré 2004) around the Fermi level. Furthermore, extended electronic states like those found in periodic crystals cannot develop due to the lack of translational symmetry.

The unusual properties of quasicrystals are interesting for various applications like thermal insolation, reduced adhesion combined with hardness, reduced solid friction, resistance against corrosion, and catalysis (Maciá 2006; Dubois 2012).

4.2 Modeling Quasicrystals

Two methods are generally used for describing quasicrystals.

4.2.1 The Superspace Approach

Symmetries that are non-crystallographic in three dimensions (3D) may become crystallographic in higher-dimensional space (Hermann 1949). Icosahedral symmetry complies with translational symmetry in a 6-dimensional (6D) space, where each coordinate is perpendicular to a hyperplane spanned by the other 5 coordinates; the 6D space is composed of the physical 3D space (called parallel space) and a complementary 3D space (perpendicular space). A 3D icosahedron is then obtained as a projection of a 6D hypercube to three dimensions. The effect of such a transformation is illustrated in Fig. 38. The 2D square lattice can be projected to a 1D subspace as a straight line $R_{\|}$. If the slope $\Delta y/\Delta x$ of $R_{\|}$ is a rational number in units of a, a discrete periodic 1D set of sites is projected on $R_{\|}$. If the slope is irrational as depicted in the lower strip of the figure, the 1D sequence is aperiodic (in the figure the projection on $R_{\|}$ is restricted to the indicated strip). A rational slope close to an irrational slope creates an *approximant* with a periodic structure, labeled by the axial relations of the generating cut.[21] The 3/2 approximant depicted in Fig. 38 yields a periodic sequence *SLSLL*. The local atomic arrangement of approximant phases generated by rational cuts in superspace are expected to be similar to those of corresponding quasicrystals.[22]

[21]The relation is τ:1, τ:1, τ:1 for an icosahedral quasicrystal and, e.g., 2:1, 2:1, 2:1 for a cubic approximant (Pay Gómez and Lidin 2001).

[22]Since the theoretical treatment of quasicrystals is challenging, often large approximants which are found to have similar properties are modeled instead.

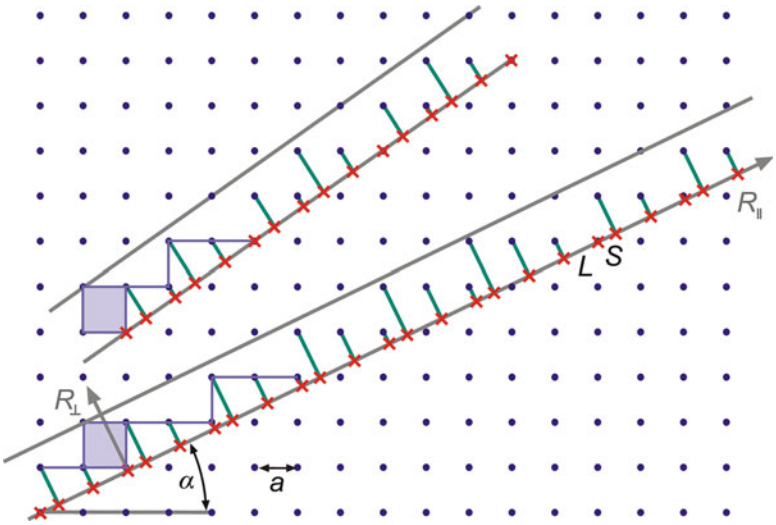


Fig. 38 Periodic (*upper strip*) and quasiperiodic sequence (*lower strip*) of one-dimensional tiles L and S, created by projecting points of a square lattice on a *straight line* $R_{\parallel}$. Using an irrational slope obeying $\cos\alpha/\sin\alpha = \tau$ (the golden mean) yields a Fibonacci sequence of L and S (*lower strip*). The *upper strip* is the 3/2 approximant of the golden mean with a rational slope

The strength of the superspace formalism is its straightforward mathematical concept and the ability to apply the mature diffraction theory developed for periodic 3D crystals also to quasiperiodic crystals. The description of *axial* quasicrystals requires a 5-dimensional space, which decomposes into the 2 orthogonal subspaces of the physical 3D space and a complementary 2D space.

4.2.2 Three-Dimensional Direct-Space Approach

This complementary method applies rules of tiling models like those depicted in Fig. 36d for a 2D Penrose tiling to atomic clusters of approximant phases. To create a quasiperiodic structure directly, the operations must transform the structure into itself within rescaling effects (Janot 1994) to ensure self-similarity. Thus, a 2D Penrose tiling can be created by iteratively applied deflation rules for two different rhombic unit cells (tiles).[23] An instructive example is the 1D quasiperiodic Fibonacci sequence of L and S segments shown in Fig. 39. Starting with a finite LS sequence and applying the substitution rules $L \rightarrow LS$ and $S \rightarrow L$, strings with increasing length are formed successively. The self-similarity of the chain is illustrated in the figure.

The advantage of the direct-space approach is its intuitive clearness and the ability for quantitative evaluation using microscopy and diffraction methods.

[23]To obtain a quasiperiodic instead of a just random tiling, matching rules for the arrangement of the two different tiles must be obeyed.

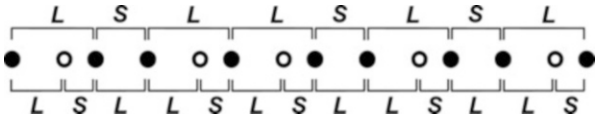


Fig. 39 Fibonacci sequence of *L* and *S* segments (*solid disks*) deflated to another Fibonacci sequence (*solid* and *open disks*) (After Janot 1994)

5 Summary

Crystal bonding and crystal structure are intimately related to each other. The crystal structure is determined by the tendency to fill a given space with the maximum number of atoms under the constraint of bonding forces and atomic radii. This tendency originates from the strong driving force of lowering the internal energy of a solid in thermal equilibrium and causes the number of atoms per unit volume to be maximized, leading to an ordered structure.

An amorphous structure results with no long-range order, when sufficient viscosity restricts atomic motion during fast enough cooling so that atomic building blocks cannot find their crystalline ordered position before further motion becomes more restricted by their rigidified surrounding. In an amorphous structure, the short-range order is much like that in a crystal, while long-range periodicity does not exist. This lack of periodicity has major consequences for the lack of interference effects, modifying substantially the theoretical analysis of photon, phonon, and electronic band-structure effects.

Quasicrystals are assembled by an *aperiodic* repetition of (at least) two unit cells and still exhibit long-range order in a diffraction experiment. A unit cell may have a symmetry not existing in crystals. There exist icosahedral quasicrystals with aperiodicity in all spatial directions and axial quasicrystals with a periodicity along one direction.

Superlattices and low-dimensional structures like quantum wells, quantum wires, and quantum dots enrich the variety of semiconducting materials. Their very attractive material properties can be engineered by designing size and chemical composition for creating new and improved devices.

References

Abrikosov NK, Bankina VF, Poretskaya LV, Shelimova LE, Skudnova EV (1969) Semiconducting II-VI, IV-VI, and V-VI compounds. Plenum Press, New York

Adler D (1985) Chemistry and physics of covalent amorphous semiconductors. In: Adler D, Schwartz BB, Steele MC (eds) Physical properties of amorphous materials. Plenum Press, New York, p 5–103.

Agarwal VK (1988) Langmuir-Blodgett films. Phys Today 41:40

Allan DC, Joannopoulos JD, Pollard WB (1982) Electronic states and total energies in hydrogenated amorphous silicon. Phys Rev B 25:1065

Ambrosch-Draxl C, Nabok D, Puschnig P, Meisenbichler C (2009) The role of polymorphism in organic thin films: oligoacenes investigated from first principles. New J Phys 11:125010
Ames Laboratory and Department of Physics and Astronomy, Iowa State University, Ames (2010) Image accessible at http://cmp.physics.iastate.edu/canfield/img/cbquasi1.jpg
Anand S, Gurunathan R, Soldi T, Borgsmiller L, Orenstein R, Snyder GJ (2020) Thermoelectric transport of semiconductor full-Heusler VFe_2Al. J Mater Chem C 8:10174
Arushanov EK (1986) Crystal growth, characterization and application of II V compounds. Prog Cryst Growth Charact 13:1
Bacewicz R, Ciszek TF (1988) Preparation and characterization of some $A^{I}B^{II}C^{V}$ type semiconductors. Appl Phys Lett 52:1150
Barrett CS, Massalski TB (1980) Structure of metals, 3rd revised edn. Pergamon Press, Oxford/New York
Bastard G, Brum JA (1986) Electronic states in semiconductor heterostructures. IEEE J Quantum Electron QE 22:1625
Belin-Ferré E (2004) Electronic structure of quasicrystalline compounds. J Non-Cryst Sol 334&335:323
Bell RJ, Dean P (1972) The structure of vitreous silica: validity of the random network theory. Philos Mag 25:1381
Bendersky L (1985) Quasicrystal with one-dimensional translational symmetry and a tenfold rotation axis. Phys Rev Lett 55:1461
Bernard JE, Zunger A (1988) Ordered-vacancy-compound semiconductors: pseudocubic $CdIn_2Se_4$. Phys Rev B 37:6835
Bertrand L, Cotte M, Stampanoni M, Thoury M, Marone F, Schöder S (2012) Development and trends in synchrotron studies of ancient and historical materials. Phys Rep 512:51
Besson JM, Mokhtari EH, Gonzalez J, Weill G (1987) Electrical properties of semimetallic silicon III and semiconductive silicon IV at ambient pressure. Phys Rev Lett 59:473
Bethe HA (1935) Statistical theory of superlattices. Proc Roy Soc A 150:552
Bhat R, Kapon E, Hwang DM, Koza MA, Yun CP (1988) Patterned quantum well heterostructures grown by OMCVD on non-planar substrates: applications to extremely narrow SQW lasers. J Cryst Growth 93:850
Bienenstock A (1985) Structural studies of amorphous materials. In: Adler D, Schwartz BB, Steele MC (eds) Physical properties of amorphous materials. Plenum Press, New York, pp 171–200.
Bimberg D, Grundmann M, Ledentsov NN (1999) Quantum dot heterostructures. Wiley, Chichester
Bindi L, Steinhardt PJ, Yao N, Lu PJ (2011) Icosahedrite, $Al_{63}Cu_{24}Fe_{13}$, the first natural quasicrystal. Am Mineral 96:928
Birss RR (1964) Symmetry and magnetism. North Holland, Amsterdam
Blodgett KB (1935) Films built by depositing successive monomolecular layers on a solid surface. J Am Chem Soc 57:1007
Boolchand P (1985) Mössbauer spectroscopy—a rewarding probe of morphological structure of semiconducting glasses. In: Adler D, Schwartz BB, Steele MC (eds) Physical properties of amorphous materials. Plenum Press, New York, pp 221–260.
Brown PJ, Forsyth JB (1973) The crystal structure of solids. Edwald Arnold, London
Buerger MJ (1956) Elementary crystallography: an introduction to the fundamental geometrical features of crystals. Wiley, New York
Burns G, Glazer AM (2013) Space groups for solid state scientists, 3rd edn. Academic Press, San Diego
Cahn JW, Shechtman D, Gratias D (1986) Indexing of icosahedral quasiperiodic crystals. J Mater Res 1:13
Carlson AE, Zunger A, Wood DM (1985) Electronic structure of LiZnN: interstitial insertion rule. Phys Rev B 32:1386
Chen X, Lenhert S, Hirtz M, Nan L, Fuchs H, Lifeng C (2007) Langmuir–Blodgett patterning: a bottom–up way to build mesostructures over large areas. Acc Chem Res 40:393

Choi HJ (2012) Vapor–liquid–solid growth of semiconductor nanowires. In: Yi G-C (ed) Semiconductor nanostructures for optoelectronic devices. Springer, Berlin
Costantini G, Rastelli A, Manzano C, Acosta-Diaz P, Songmuang R, Katsaros G, Schmidt OG, Kern K (2006) Interplay between thermodynamics and kinetics in the capping of InAs/GaAs(001) quantum dots. Phys Rev Lett 96:226106
Dandrea RG, Zunger A (1991) First-principles study of intervalley mixing: Ultrathin GaAs/GaP superlattices. Phys Rev B 43:8962
DiBenedetto AT (1967) The structure and properties of materials. McGraw-Hill, New York
Dubois JM, Brunet P, Belin-Ferré E (2000) Potential applications of quasicrystalline materials. In: Belin-Ferré E et al. (eds) Quasicrystals: current topics. World Scientific, Singapore, p 498.
Dubois J-M (2005) Useful quasicrystals. World Scientific, Singapore
Dubois J-M (2012) Properties and applications of quasicrystals and complex metallic alloys. Chem Soc Rev 41:6760
Ehrenreich H (1987) Electronic theory for materials science. Science 235:1029
Emery N, Hérold C, Marêché J-F, Lagrange P (2008) Synthesis and superconducting properties of CaC6. Sci Technol Adv Mater 9:044102
Enoki T, Suzuki M, Endo M (eds) (2003) Graphite intercalation compounds and applications. Oxford University Press, Oxford
Etherington G, Wright AC, Wenzel JT, Dore JC, Clarke JH, Sinclair RN (1982) A neutron diffraction study of the structure of evaporated amorphous germanium. J Non-Cryst Sol 48:265
Felser C, Hirohata A (eds) (2016) Heusler alloys: properties, growth, applications. Springer, Cham
Freysoldt C, Pfanner G, Neugebauer J (2012) The dangling-bond defect in amorphous silicon: statistical random versus kinetically driven defect geometries. J Non-Cryst Sol 358:2063
Fritzsche H (2001) Development in understanding and controlling the Staebler-Wronski effect in a-Si:H. Annu Rev Mater Res 31:47
Galli G, Martin RM, Car R, Parrinello M (1988) Structural and electronic properties of amorphous carbon. Phys Rev Lett 62:555
Gomyo A, Suzuki T, Kobayashi K, Kawata S, Hino I, Yuasa T (1987) Evidence for the existence of an ordered state in $Ga_{0.5}In_{0.5}P$ grown by metalorganic vapor phase epitaxy and its relation to band-gap energy. Appl Phys Lett 50:673
Gossard AC (1986) Growth of microstructures by molecular beam epitaxy. IEEE J Quantum Electron QE 22:1649
Gustafsson A, Reinhardt F, Biasiol G, Kapon E (1995) Low-pressure organometallic chemical vapor deposition of quantum wires on V-grooved substrates. Appl Phys Lett 67:3673
Hahn T (ed) (1983) International tables for crystallography vol. A. D. Reidel Publication, Dordrecht
Hanrath T (2012) Colloidal nanocrystal quantum dot assemblies as artificial solids. J Vac Sci Technol A 30:030802
Häussler P, Nowak H, Haberkern R (2000) From the disordered via the quasicrystalline to the crystalline state. Mater Sci Eng 294–296:283
Hayes TM, Boyce JC (1985) Extended X-ray absorption fine structure spectroscopy. Solid State Phys 37:173
Hermann C (1949) Kristallographie in Räumen beliebiger Dimensionszahl. 1. Die Symmetrieoperationen. Acta Crystallogr 2:139 (Crystallography in spaces of arbitrary dimension. 1. The symmetry operation, in German)
Li X, Song W, Smedskjaer MM, Mauro JC, Bauchy M (2019) Quantifying the internal stress in over-constrained glasses by molecular dynamics simulations. J Non-Crystalline Sol X 1:100013
Isu T, Jiang D-S, Ploog K (1987) Ultrathin-layer (AlAs)m (GaAs)m superlattices with m = 1,2,3 grown by molecular beam epitaxy. Appl Phys A 43:75
Jäckle J (1986) Models of the glass transition. Rep Prog Phys 49:171
Jaffe JE, Zunger A (1984) Electronic structure of the ternary pnictide semiconductors $ZnSiP_2$, $ZnGeP_2$, $ZnSnP_2$, $ZnSiAs_2$, and $MgSiP_2$. Phys Rev B 30:741
James RW (1954) The optical principles of the diffraction of X-Rays. G. Bell & Sons, London
Janot C (1994) Quasicrystals: a primer, 2nd edn. Oxford University Press, Oxford, UK, Reissued 2012

Jarolimek K, de Groot RA, de Wijs GA, Zeman M (2009) First-principles study of hydrogenated amorphous silicon. Phys Rev B 79:155206
Jen HR, Cherng MJ, Stringfellow GB (1986) Ordered structures in $GaAs_{0.5}Sb_{0.5}$ alloys grown by organometallic vapor phase epitaxy. Appl Phys Lett 48:1603
Jurchescu OD, Baas J, Palstra TTM (2004) Effect of impurities on the mobility of single crystal pentacene. Appl Phys Lett 84:3061
Kalarasse F, Bennecer B (2006) Optical properties of the filled tetrahedral semiconductors LiZnX (X = N, P, and As). J Phys Chem Sol 67:1850
Kitaigorodskii AI (1973) Molecular crystals and molecules. Academic Press, New York
Kittel C (2007) Introduction to solid state physics, 7th edn. Wiley, New York
Klug HP, Alexander LE (1974) X-ray diffraction procedures for polycrystalline and amorphous materials, 2nd edn. Wiley, New York
Koide A, Fujikawa T, Ichikuni N (2014) Recent progress in EXAFS/NEXAFS spectroscopy. J Electron Spectroscopy Related Phenomena 195:375
Kowarik S, Hinderhofer A, Wang C, Weber C, Gerlach A, Hexemer A, Leone SR, Schreiber F (2015) Identification of an organic semiconductor superlattice structure of pentacene and perfluoro-pentacene through resonant and non-resonant X-ray scattering. AIP Advances 5:117241
Kuan TS, Kuech TF, Wang WI, Wilkie EL (1985) Long-range order in $Al_xGa_{1-x}As$. Phys Rev Lett 54:201
Kuan TS, Wang WI, Wilkie EL (1987) Long-range order in $Al_xGa_{1-x}As$. Appl Phys Lett 51:51
Kuriyama K, Nakamura F (1987) Electrical transport properties and crystal structure of LiZnAs. Phys Rev B 36:4439
Landoldt-Börnstein (1982) New series, III. 17a and b. Madelung O, Schulz M, Weiss H (eds) Springer, Berlin
Landoldt-Börnstein (1987) New series, III, 22. Madelung O, Schulz M (eds) Springer, Berlin
Langmuir I (1920) The mechanism of the surface phenomena of flotation. Trans Faraday Soc 15:62
Levine D, Steinhardt PJ (1984) Quasicrystals: a new class of ordered structures. Phys Rev Lett 53:2477
Li JJ, Wang A, Ghuo W, Keay JC, Mishima TD, Johnson MB, Peng X (2003) Large-scale synthesis of nearly monodisperse CdSe/CdS core/shell nanocrystals using air-stable reagents via successive ion layer adsorption and reaction. J Am Chem Soc 125:12567
Li X, Song W, Smedskjaer MM, Mauro JC, Bauchy M (2019) Quantifying the internal stress in over-constrained glasses by molecular dynamics simulations. J Non-Crystalline Solids X 1:100013
Maciá E (2006) The role of aperiodic order in science and technology. Rep Prog Phys 69:397
Mandelbrot BB (1981) The fractal geometry of nature. Freeman, San Francisco
Marciá-Barber E (2021) Quasicrystals - Fundamentals and applications. CRC Press, Boca Raton FL
Márquez J, Geelhaar L, Jacobi K (2001) Atomically resolved structure of InAs quantum dots. Appl Phys Lett 78:2309
Mayou D, Berger C, Cyrot-Lackmann F, Klein T, Lanco P (1993) Evidence for unconventional electronic transport in quasicrystals. Phys Rev Lett 70:3915
Menelle A, Bellissent R (1986) EXAFS and neutron scattering determination of local order in a-Si:H. In: Engström O (ed) Proceedings of international conference on the physics of semiconductors. World Scientific, Stockholm, p 1049–1052
Miller A, MacKinnon A, Weaire D (1981) Beyond the binaries – the chalcopyrite and related semiconducting compounds. In: Ehrenreich H, Seitz F, Turnbull D (eds) Solid state physics, vol 36. Academic Press, New York
Moss SC, Graczyk JF (1970) Structure of amorphous silicon. In: Proceedings of 10th international conference on semiconductors, Washington DC, pp 658–662
Mott NF, Davis EA (1979) Electronic processes in non-crystalline materials, 2nd edn. Oxford University Press, Oxford

MPI Halle (2007): Max Planck Institute of Microstructure Physics, Halle, Germany. The image is accessible at http://www.mpi-halle.mpg.de/department2/research-areas/nanowires-nanoobjects/semiconductor-nanowires/abstract/si-nanowires-by-cvd-and-ebe/
Murray C, Kagan C, Bawendi M (2000) Synthesis and characterization of monodisperse nanocrystals and closed-packed nanocrystal assemblies. Ann Rev Mater Sci 30:545
Nakayama H, Fujita H (1985) Direct observation of an ordered phase in a disordered In1-xGaxAs alloy. In: Fujimoto M (ed) Gallium arsenide and related compounds 1985. Institute of physics conference series, vol 79. IOP, Adam Hilger, Boston, pp 289–294
Nan H-s, Hu X-y, Tian H-w (2019) Recent advances in perovskite oxides for anion-intercalation supercapacitor: a review. Mater Sci Semicond Processing 94:35
Newnham RE (1975) Structure–property relations. Springer, Berlin
Nobuhara K, Nakayama H, Nose M, Nakanishi S, Iba H (2013) First-principles study of alkali metal-graphite intercalation compounds. J Power Sources 243:585
Ovshinsky SR (1976) Lone-pair relationships and the origin of excited states in amorphous chalcogenides. AIP Conf Proc 31:31
Pantelides ST (1987) Defect dynamics and the Staebler-Wronski effect in hydrogenated amorphous silicon. Phys Rev B 36:3479
Parthé E (1964) Crystal chemistry of tetrahedral structures. Gordon & Breach, New York
Parthé E (1972) Cristallochimie des structures tétraédriques. Gordon & Breach, New York
Pay Gómez C, Lidin S (2001) Structure of $Ca_{13}Cd_{76}$: A novel approximant to the $MCd_{5.7}$ quasicrystals (M = Ca, Yb). Angewandte Chemie 40:4037
Penrose R (1974) Role of aesthetics in pure and applied research. Bull Inst Math Appl 10:266
Petroff PM, Gossard AC, Wiegmann W, Savage A (1978) Crystal growth kinetics in (GaAs)n − (AlAs)m superlattices deposited by molecular beam epitaxy: I. Growth on singular (100)GaAs substrates. J Cryst Growth 44:5
Petroff PM, Gossard AC, Savage A, Wiegmann W (1979) Molecular beam epitaxy of Ge and $Ga_{1-x}Al_xAs$ ultra thin film superlattices. J Cryst Growth 46:172
Phillips JC (1980) Morphology of amorphous semiconductors. Comments Solid State Phys 9:191
Pickard CJ, Needs RJ (2011) Ab initio random structure searching. J Phys Condens Matter 23:053201
Pinczolits M, Springholz G, Bauer G (1998) Direct formation of self-assembled quantum dots under tensile strain by heteroepitaxy of PbSe on PbTe (111). Appl Phys Lett 73:250
Polk DE (1971) Structural model for amorphous silicon and germanium. J Non-Cryst Sol 5:365
Rastelli A, Kummer M, von Känel H (2001) Reversible shape evolution of Ge islands on Si(001). Phys Rev Lett 87:256101
Richardson TH (ed) (2000) Functional organic and polymeric materials. Wiley, New York
Roberts GG (1985) An applied science perspective of Langmuir-Blodgett films. Adv Phys 34:475
Roorda S, Martin C, Droui M, Chicoine M, Kazimirov A, Kycia S (2012) Disentangling neighbors and extended range density oscillations in monatomic amorphous semiconductors. Phys Rev Lett 108:255501
Runnels LK (1967) Phase transition of a Bethe lattice gas of hard molecules. J Math Phys 8:2081
Shay L, Wernick JH (1974) Ternary chalcopyrite semiconductors: growth, electronic properties, and applications. Pergamon Press, Oxford
Shechtman D, Blech I, Gratias D, Cahn JW (1984) Metallic phase with long-range orientational order and no translational symmetry. Phys Rev Lett 53:1951
Singh J, Shimakawa K (2003) Advances in amorphous semiconductors. Taylor and Francis, London
Sommer AH (1968) Photoemissive materials: preparation, properties, and uses. Wiley, New York
Srivastava GP, Martins JL, Zunger A (1985) Atomic structure and ordering in semiconductor alloys. Phys Rev B 31:2561
Steinhardt P, Alben R, Weaire D (1974) Relaxed continuous random network models: (I). Structural characteristics. J Non-Cryst Sol 15:199
Steinhardt PJ (1987) Icosahedral solids: a new phase of matter? Science 238:1242
Stern EA (1978) Structure determination by X-ray absorption. Contemp Phys 19:289

Stern EA (1985) EXAFS of disordered systems. In: Adler D, Schwartz BB, Steele MC (eds) Physical properties of amorphous materials. Plenum Press, New York p 201–219
Steurer W, Deloudi S (2009) Crystallography of quasicrystals. Springer, Berlin
Stöhr J (1996) NEXAFS spectroscopy. Springer, Berlin
Street RA (1991) Hydrogenated amorphous silicon. Cambridge University Press, Cambridge, UK
Suck J-B, Schreiber M, Häussler P (eds) (2002) Quasicrystals: an introduction to structure, physical properties and applications. Springer, Berlin
Takeya J, Yamagishi M, Tominari Y, Hirahara R, Nakazawa Y, Nishikawa T, Kawase T, Shimoda T, Ogawa S (2007) Very high-mobility organic single-crystal transistors with in-crystal conduction channels. Appl Phys Lett 90:102120
Tan L, Pickard CJ, Yu K, Sapelkin A, Misquitta AJ, Dove MT (2019) Structures of CdSe and CdS nanoclusters from ab initio random structure searching. J Phys Chem C 123:29370
Temkin RJ, Paul W, Connell GAN (1973) Amorphous germanium II. Structural properties. Adv Phys 22:581
Temkin RJ (1974) Comparison of the structure of amorphous Ge and GaAs. Sol State Commun 15:1325
Wagner RS, Ellis WC (1964) Vapor–liquid-solid mechanism of single crystal growth. Appl Phys Lett 4:89
Waire D, Ashby MF, Logan J, Weins MJ (1971) On the use of pair potentials to calculate the properties of amorphous metals. Acta Metall 19:779
Wang LG, Kratzer P, Scheffler M, Moll N (1999) Formation and stability of self-assembled coherent islands in highly mismatched heteroepitaxy. Phys Rev Lett 82:4042
Wang X-L, Voliotis V (2006) Epitaxial growth and optical properties of semiconductor quantum wires. J Appl Phys 99:121301
Warren BE (1990) X-ray diffraction. Dover, New York
Wegscheider W, Pfeiffer LN, Dignam MM, Pinczuk A, West KW, McCall SL, Hull R (1993) Appl Phys Lett 71:4071
Wells AF (2012) Structural inorganic chemistry, 5th edn. Oxford University Press, Oxford
Whittingham MS, Jacobson AJ (eds) (1982) Intercalation chemistry. Academic Press, New York
Woggon U (1997) Optical properties of semiconductor quantum dots. Springer, Berlin
Wood DM, Zunger A (1988) Epitaxial effects on coherent phase diagrams of alloys. Phys Rev Lett 61:1501
Wood DM, Wei S-H, Zunger A (1988) Stability and electronic structure of ultrathin [001] $(GaAs)_m(AlAs)_m$ superlattices. Phys Rev B 37:1342
Xu T, Zhou L, Wang Y, Özcan AS, Ludwig KF (2007) GaN quantum dot superlattices grown by molecular beam epitaxy at high temperature. J Appl Phys 102:073517
Yan D, Wang H, Du B (2010) Introduction to organic semiconductor heterojunctions. Wiley, Singapore
Yu LH, Yao KL, Liu ZL (2004) Electronic structures of filled tetrahedral semiconductors LiMgN and LiZnN: conduction band distortion. Phys B 353:278
Zabel H, Solin SA (eds) (1990) Graphite intercalation compounds I. Springer, Berlin
Zabel H, Solin SA (eds) (1992) Graphite intercalation compounds II. Springer, Berlin
Zachariasen WH (1932) The atomic arrangement in glass. J Am Chem Soc 54:3841
Zachariasen WH (2004) Theory of X-ray diffraction in crystals. Dover, New York
Zhu F, Lou K, Huang L, Yang J, Zhang J, Wang H, Geng Y, Yana D (2009) Crystalline organic superlattice. Appl Phys Lett 95:203106
Zunger A (1985) Ternary semiconductors and ordered pseudobinary alloys: electronic structure and predictions of new materials. Int J Quant Chem 28:629

Part II

Phonons

Elasticity and Phonons

4

Karl W. Böer and Udo W. Pohl

Contents

1 Elastic Properties 114
1.1 Stress–Strain Relations 115
1.2 Elastic Stiffness Constants 117
2 Elastic Waves 122
2.1 Sound Waves in Crystals 123
3 Phonon Spectra 128
3.1 Oscillations of One-Dimensional Lattices 129
3.2 Phonons in a Three-Dimensional Lattice 137
3.3 Phonons in Superlattices, at Surfaces, in Organic Semiconductors, and in Amorphous Semiconductors 145
4 Summary 152
References 152

Abstract

Springlike interatomic forces allow macroscopic elastic deformations of the semiconductor and coupled microscopic oscillations of each atom. The strain occurring as a response to external stress is conventionally described by elastic stiffness constants. When the strain exceeds the range in which the harmonic approximation of the interatomic potential is valid, higher-order stiffness constants are used. The symmetry of crystals strongly reduces the number of independent constants. Elastic properties can be measured by static deformations or kinetically by sound wave propagation.

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_4

Different modes of sound waves propagate with different velocities from which all stiffness constants can be determined. Each mode of such collective oscillations is equivalent to a harmonic oscillator which can be quantized as a phonon. Phonons are one of the most important quasiparticles in solids. They are responsible for all thermal properties and, when interacting with other quasiparticles, for the damping of their motion.

Keywords

Bulk modulus · Elastic compliance · Elastic properties · Elastic stiffness constants · Elastic moduli · Harmonic approximation · Higher-order stiffness constants · Hooke's law · Phonon · Phonon density of states · Phonon dispersion · Shear modulus · Sound wave propagation · Strain · Stress · Stress-strain relation · Young's modulus

1 Elastic Properties

Elastic properties of solids are determined by interatomic forces and the crystal structure of solids. These properties determine not only the macroscopic elastic behavior of a solid but also provide the basic elements for all lattice oscillations.

The interatomic forces can be described in a first approximation as spring forces. This means that the force F is proportional to an atomic displacement $u = r - r_e$:

$$F = \beta u, \tag{1}$$

with β the (spring) force constant. This holds for small displacements from the equilibrium position of each atom as long as the potential can be approximated by a parabola (see Fig. 1 for values of r close to the equilibrium distance r_e). For the harmonic potential $V = V_0 + \alpha(r - r_e)^2$ and $F = dV/dr$, we obtain the linear relation (1) with $\beta = 2\alpha$ for the spring forces. For nonlinear effects, see Sect. 1.2 and in ▶ Chap. 10, "Interaction of Light with Solids," Sect. 3.

In a three-dimensional lattice with interatomic forces restraining changes in the *bond length and bond angle*, a more complicated relation between the force and the

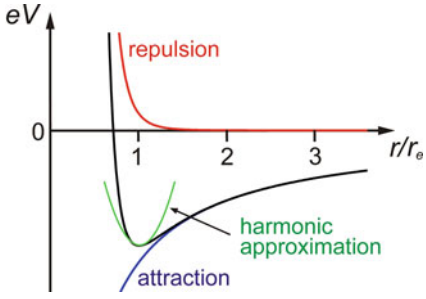


Fig. 1 Interatomic potential energy eV (*black line*) and parabolic, i.e., harmonic, approximation (*green line*). r_e is the equilibrium distance

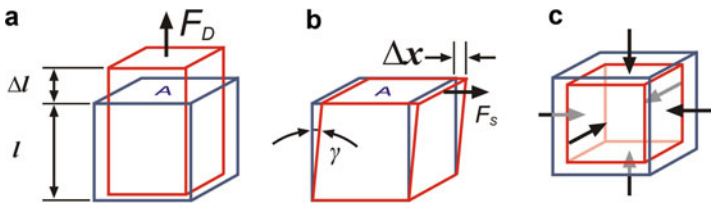


Fig. 2 Three basic types of mechanical deformation of a cube: (**a**) stretching (in one dimension), (**b**) shearing, and (**c**) volume compression

atomic displacement in such a network must be considered. Since various force constants can be distinguished in such a three-dimensional lattice, the general relation between forces and lattice deformation will be analyzed first.

Microscopic parameters can be derived from macroscopic experiments by exposing a sample to specific forces (*stress*) and observing the resulting deformation (*strain*). The following section will deal briefly with *stationary* deformation of the semiconductor described as a macroscopic continuum. In Sect. 2, we will then discuss additional information which can be obtained from *dynamic* deformation.

1.1 Stress–Strain Relations

The basis for deriving the elastic constants is a simple application of *Hooke's law* (Eq. 1) to the different ways of applying mechanical forces to a semiconductor, which are summarized for isotropic systems and homogeneous strains in Fig. 2.

There are three macroscopic *elastic moduli* which can be expressed for a homogeneous medium by the following relations:

- The *linear elasticity modulus* (*Young's modulus* or stretching modulus) *E:*

$$E = \frac{F_D/A}{\Delta l/l} \tag{2}$$

- The *shear modulus* (or rigidity modulus) *G*:

$$G = \frac{F_S/A}{\Delta x/l} \tag{3}$$

- The *bulk modulus* (or compression modulus) *B* (=1/κ with κ the compressibility):

$$B = -V\frac{\Delta P}{\Delta V}, \tag{4}$$

where l, A, and V are, respectively, the length, the face area, and the volume of the undeformed cube; F_D is the pull force; F_S is the shear force; and P is the pressure.[1] In a typical semiconductor, these moduli depend on the crystal orientation. From this mechanical approach, we can deduce important consequences for the different types of waves of interest in a semiconductor.

The mechanical deformation depends on the type of stress applied as illustrated in Fig. 2. A parallelepiped (e.g., a crystallite) can be deformed by changing the lengths of its edges and the angle between them or, equivalently, by changing its unit cell accordingly. During such a deformation, an arbitrary point, described by a vector $\mathbf{r}$ from the origin, is displaced to $\mathbf{r}'$, which in turn is given in component form as

$$\begin{aligned} x' &= \left(1+\frac{\partial u_1}{\partial x}\right) x + \left(1+\frac{\partial u_1}{\partial y}\right) y + \left(1+\frac{\partial u_1}{\partial z}\right) z \\ y' &= \left(1+\frac{\partial u_2}{\partial x}\right) x + \left(1+\frac{\partial u_2}{\partial y}\right) y + \left(1+\frac{\partial u_2}{\partial z}\right) z \\ z' &= \left(1+\frac{\partial u_3}{\partial x}\right) x + \left(1+\frac{\partial u_3}{\partial y}\right) y + \left(1+\frac{\partial u_3}{\partial z}\right) z \end{aligned} \tag{5}$$

where u_1, u_2, and u_3 are the components of the displacement vector and assuming that the dimensionless strain quantities ($\partial u_i/\partial x$, $\partial u_i/\partial y$, $\partial u_i/\partial z$) < 1. When expressed in tensor form, we have

$$\mathbf{r}' = \mathbf{r} + \underline{\boldsymbol{\varepsilon}}\cdot\mathbf{r} \tag{6}$$

with the components of the tensor $\underline{\boldsymbol{\varepsilon}}$ given by the derivatives in Eq. 5. A pure deformation without translation or rotation of the crystal is described by the symmetric part of this tensor

$$\underline{\boldsymbol{\varepsilon}}_{\mathrm{s}} = \begin{pmatrix} \dfrac{\partial u_1}{\partial x} & \dfrac{1}{2}\left(\dfrac{\partial u_1}{\partial y}+\dfrac{\partial u_2}{\partial x}\right) & \dfrac{1}{2}\left(\dfrac{\partial u_1}{\partial z}+\dfrac{\partial u_3}{\partial x}\right) \\ \dfrac{1}{2}\left(\dfrac{\partial u_1}{\partial y}+\dfrac{\partial u_2}{\partial x}\right) & \dfrac{\partial u_2}{\partial y} & \dfrac{1}{2}\left(\dfrac{\partial u_2}{\partial z}+\dfrac{\partial u_3}{\partial y}\right) \\ \dfrac{1}{2}\left(\dfrac{\partial u_1}{\partial z}+\dfrac{\partial u_3}{\partial x}\right) & \dfrac{1}{2}\left(\dfrac{\partial u_2}{\partial z}+\dfrac{\partial u_3}{\partial y}\right) & \dfrac{\partial u_3}{\partial z} \end{pmatrix} \tag{7}$$

defining a set of strain coefficients ε_{ik} with which the crystallite deformation is easily described. For instance, after application of hydrostatic pressure, the relative volume change and the new volume are given from $V' = \mathbf{a}' \cdot \mathbf{b}' \times \mathbf{c}'$, when retaining only first-order terms by

[1] The ratios $\sigma_D = F_D/A$ and $\sigma_S = F_S/A$ are a normal stress (uniaxial pressure) and a shear stress, respectively, and P is the hydrostatic pressure. Stresses are the origin of deformations and have the unit of a force per area.

$$\begin{aligned} \Delta V/V = (V - V')/V = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} \quad \text{and} \\ V' = V\big(1 - \big(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}\big)\big). \end{aligned} \tag{8}$$

(For more detail, see Joos (1945) and Nye (1972).) The strain coefficients ε_{ik} relate to the six independent components of the stresses (σ_{xx}, σ_{yy}, σ_{zz} and σ_{yz}, σ_{zx}, σ_{xy}) in a tensor relationship

$$\underline{\boldsymbol{\varepsilon}}_{\mathrm{s}} = \underline{\underline{\mathbf{S}}}\ \underline{\boldsymbol{\sigma}} \tag{9}$$

with the compliance tensor $\underline{\underline{\mathbf{S}}}$ comprising 81 *elastic compliance constants* S_{ijkl} (with $i, j, k, l = 1, \ldots 3$); usually a reduced notation with 36 constants S_{ik} (with $i, k = 1, \ldots 6$) is applied.[2]

Often the inverse relationship is used:

$$\underline{\boldsymbol{\sigma}} = \underline{\underline{\mathbf{C}}}\ \underline{\boldsymbol{\varepsilon}}_{\mathrm{s}} \quad \text{and} \quad \underline{\underline{\mathbf{C}}} = \underline{\underline{\mathbf{S}}}^{-1} \tag{10}$$

with the *elastic stiffness constants* C_{ijkl} (with $i, j, k, l = 1, \ldots 3$) usually reduced to 36 constants C_{ik} (with $i, k = 1, \ldots 6$). Equation 10 is also known as the generalized Hooke's law. Since $\sigma_{yz} = \sigma_{zy}$, etc., and $\varepsilon_{yz} = \varepsilon_{zy}$, etc., both second-rank tensors can be reduced in a symbolic notation to vectors $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ with six components. This permits us to write a matrix relation between $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ where the fourth-rank tensors $\underline{\underline{\mathbf{C}}}$ and $\underline{\underline{\mathbf{S}}}$ are replaced by 6×6 matrices $\mathbf{C}$ and $\mathbf{S}$. The stress components can then be written as

$$\sigma_i = \sum\nolimits_{k=1}^{6} C_{ik}\ \varepsilon_k \tag{11}$$

with the convention $k = 1 \ldots 6$ equivalent to xx, yy, zz, yz, zx, and xy, respectively, for a single index (*Voigt notation*).[3]

1.2 Elastic Stiffness Constants

The number of elastic stiffness constants can be reduced from 36 because of two considerations:

[2] $\underline{\underline{\mathbf{S}}}$ is a fourth-rank tensor with $3^4 = 81$ components S_{ijkl}. The tensor (S_{ijkl}) is usually reduced to a 6×6 matrix (S_{ik}) by the index substitution given in the text, thereby simplifying the notation but losing the tensor transformation properties; for details, see Dunstan (1997).

[3] The full tensor notation reads $\sigma_{ij} = \sum_{k=1}^{3} \sum_{l=1}^{3} C_{ijkl}\ \varepsilon_{kl}$. This form yields correctly transformed stress–strain relations for arbitrary stress directions, while the reduced notation applies only for stresses along the principal axes, when the generally listed constants C_{mn} are used.

1. The parabolicity of the lattice potential (in harmonic approximation) renders the C_{ik} symmetrical: $C_{ik} = C_{ki}$, leaving 21 independent constants.
2. Lattice symmetry further reduces this number.

For instance, in a *cubic lattice* with cyclic interchange of x, y, and z, the same state is reproduced. This yields

$$C_{12} = C_{13} = C_{23}, C_{11} = C_{22} = C_{33} \text{ and } C_{44} = C_{55} = C_{66}$$

$$C_{14} = C_{15} = C_{16} = C_{61} = C_{62} = C_{63} = C_{64} = C_{65} = 0$$

and an exceedingly simple matrix results:

$$\mathbf{C}_{\text{cubic}} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{pmatrix} \tag{12}$$

with only three independent elastic stiffness constants C_{11}, C_{12}, and C_{44}. The hexagonal lattice has a lower symmetry and requires five independent elastic stiffness constants; the respective matrix reads

$$\mathbf{C}_{\text{hex}} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{13} & 0 & 0 & 0 \\ C_{13} & C_{13} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & \dfrac{C_{11} - C_{12}}{2} \end{pmatrix} \tag{13}$$

The independent elastic stiffness constants for the seven crystal systems and isotropic solids are given in Table 1.

The values of these elastic stiffness constants for the more important semiconductors are listed in Table 2. These stiffness constants depend on the lattice temperature (see below); the room temperature values are given when not otherwise stated.

1.2.1 Third-Order Elastic Constants

In the previous discussion, only the harmonic part of the interatomic potential was included; this limits the amplitude of deformation to the validity of Hooke's law. For nonlinear effects, which include thermal expansion and phonon scattering, the *third-order elastic constants* must be included (see Brugger 1964). These can be obtained from the lattice potential with higher terms:

Table 1 Elastic stiffness constants for primitive lattices and isotropic solids

Crystal system	Independent elastic stiffness constants	
	Number	Constants
Triclinic	21	$C_{11}C_{12}C_{13}C_{14}C_{15}C_{16}C_{22}C_{23}C_{24}C_{25}C_{26}$
		$C_{33}C_{34}C_{35}C_{36}C_{44}C_{45}C_{46}C_{55}C_{56}C_{66}$
Monoclinic (standard orientation)	13	$C_{11}C_{12}C_{13}C_{15}C_{22}C_{23}C_{25}$
		$C_{33}C_{35}C_{44}C_{46}C_{55}C_{66}$
Trigonal	6	$C_{11}C_{12}C_{13}C_{14}$ $C_{33}C_{44}C_{55}$
Hexagonal	5	$C_{11}C_{12}C_{13}$ $C_{33}C_{44}$
Orthorhombic	9	$C_{11}C_{12}C_{13}C_{22}C_{23}C_{33}C_{44}C_{55}C_{66}$
Tetragonal	6	$C_{11}C_{12}C_{13}$ $C_{33}C_{44}$ C_{66}
Cubic	3	$C_{11}C_{12}$ C_{44}
Isotropic	2	$C_{11}C_{12}$

Table 2 Elastic stiffness constants for some important semiconductors in GPa

Crystal	Structure	C_{11}	C_{12}	C_{44}	C_{13}	C_{33}
Diamond[a]	D	1,076.4	125.2	557.4		
Si[a]	D	165.8	63.9	79.6		
Ge[a]	D	124.0	41.3	68.3		
α-Sn[a]	D	69.0	29.3	36.2		
SiC[b]	ZB	390	142	256		
AlP[c]	ZB	133.0	63.0	61.5		
AlAs[c]	ZB	125.0	53.4	54.2		
AlSb[c]	ZB	87.7	43.4	40.8		
β-GaN[d]	ZB	293	159	155		
GaP[c]	ZB	140.5	62.0	70.3		
GaAs[c]	ZB	122.1	56.6	60.0		
GaSb[c]	ZB	88.4	40.3	43.2		
InP[c]	ZB	101.1	56.1	45.6		
InAs[c]	ZB	83.3	45.3	39.6		
InSb[c]	ZB	68.5	37.4	31.1		
ZnS[e]	ZB	98.1	62.7	44.8		
ZnSe[e]	ZB	87.2	52.4	39.2		
ZnTe[e]	ZB	71.3	40.7	31.2		
CdTe[a]	ZB	53.8	37.4	20.2		
HgSe[a]	ZB	62.2	46.4	22.7		
HgTe[a]	ZB	53.6	36.6	21.1		
PbS	NaCl	113.9	28.9	27.2		
PbSe	NaCl	123.7	19.3	15.9		
PbTe	NaCl	107.2	7.7	1.3		
SiC[e]	6H	501	111	163	53	553
AlN[d]	W	396	137	116	108	373
GaN[d]	W	390	145	105	106	398

(continued)

Table 2 (continued)

Crystal	Structure	C_{11}	C_{12}	C_{44}	C_{13}	C_{33}
InN[d]	W	223	115	48	92	224
ZnO[a]	W	206	118	44	118	211
CdS[b]	W	86.5	54.0	15.0	47.3	94.4
CdSe[a]	W	74.1	45.2	13.4	39	84.3

[a]Martienssen and Warlimont (2005)
[b]Adachi (2005)
[c]Vurgaftman et al. (2001)
[d]Vurgaftman and Meyer (2003)
[e]Ayers (2007)

Table 3 Third-order elastic constants at room temperature in GPa

Crystal	Structure	C_{111}	C_{112}	C_{123}	C_{144}	C_{166}	C_{456}
C (diamond)[a]	D	−7603	−1909	+835	+1438	−3938	−2316
Si[b]	D	−834	−531	−2	−95	−296	−7
Ge[b]	D	−823	−512	−234	−34	−288	−39
GaP[c]	ZB	−737	−474	−131	−107	−234	+62
GaAs[d]	ZB	−628	−387	−90	+24	−269	−44
GaSb[e]	ZB	−475	−308	−44	+50	−216	−25
InP[f]	ZB	−860	−185	−510	−650	+160	−4
InAs[f]	ZB	−518	−225	−239	−190	−18	−7
InSb[g]	ZB	−356	−266	−100	+16	−139	−0.4
ZnSe[f]	ZB	−827	−136	−511	+222	−265	−278
HgTe[f]	ZB	−260	−170	−77	−17	−57	−1

[a]Lang and Gupta (2011)
[b]Sahasrabudhe and Lambade (1998)
[c]Yoğurtçu et al. (1981)
[d]Joharapurkar et al. (1992)
[e]Raja and Reddy (1976)
[f]Adachi (2005)
[g]Sundararaja and Reddy (1977)

$$V = V_0 + \alpha_1 (r - r_e)^2 + \alpha_2 (r - r_e)^3 + \ldots \tag{14}$$

with the strain energy

$$E_S = E_{S_0} + \frac{1}{2} \sum_{i,k} C_{ik} \ \varepsilon_i \varepsilon_k + \frac{1}{6} \sum_{i,k,l} C_{ikl} \ \varepsilon_i \varepsilon_k \varepsilon_l \tag{15}$$

defining the third-order elastic constants. These constants can be obtained from sound-wave propagation (see Sect. 2.1) in a sufficiently prestressed semiconductor; more recently nonlinear elasticity was also calculated from first principles based on density functional theory (Łopuszyński and Majewski 2007, Tanner et al. 2019). In Table 3,

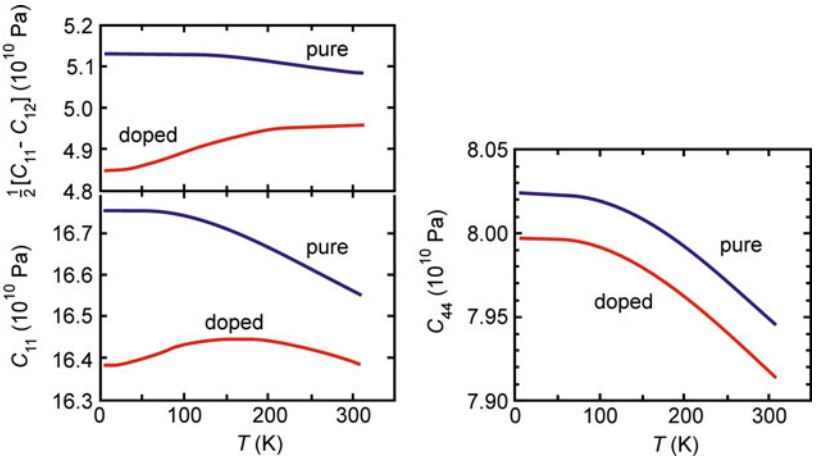


Fig. 3 Temperature dependence of the elastic stiffness constants for Si (After Hall 1967)

the set of third-order elastic constants for a few semiconductors is given. (See McSkimmin and Andreatch (1964, 1967); for a general review, see Hiki (1981).)

1.2.2 Temperature Dependence

The expansion of the lattice with increased temperature leads to a reduction of interatomic forces and thus a reduction in the value of the elastic stiffness constants (Garber and Granato 1975). Typical behavior is shown for Si in Fig. 3. Near 0 K this dependence disappears, shown as a horizontal tangent in Fig. 3 – as required by the third law of thermodynamics. The figure also shows that doping (Sects. 1.2 in ▶ Chap. 15, "Crystal Defects," and 1 in ▶ Chap. 18, "Shallow-Level Centers") usually weakens the lattice and thereby reduces the elastic stiffness constants.

1.2.3 Information from Elastic Stiffness Constants

The elastic stiffness constants provide a wealth of information about the interatomic forces. Their magnitude reflects the strength of the bond and, for a similar bonding type, shows the decreasing strength of the bond with increasing atomic number (atomic radius), as seen in Table 2 by comparing compounds in which only one of the homologous elements is varied, e.g., InP, InAs, and InSb.

Most direct information can be derived by expressing the compressibility in terms of C_{ik}. For instance, the compressibility in cubic crystals is given by[4]

[4]For hexagonal crystals, the compressibility is given by $\kappa_{\text{hex}} = \frac{C_{11}+C_{12}-4C_{13}+2C_{33}}{(C_{11}+C_{12})C_{33}-2C_{13}^2}$.

$$\kappa = \frac{3}{C_{11} + 2C_{12}}. \tag{16}$$

When binding forces are radially symmetric (ionic bond), one expects the *Cauchy relation* (Cauchy 1828; see Zener 1947; Musgrave 1970) to hold

$$C_{44} = \frac{C_{11} - C_{12}}{2}. \tag{17}$$

Deviation from this relation indicates an elastic anisotropy and is presented as anisotropy parameter $a = C_{11} - C_{12}/ (2C_{44})$. Such anisotropy is given by *Keating's relation* (Keating 1966) for monatomic cubic semiconductors:

$$C_{44} = \frac{C_{11} - C_{12}}{2} \cdot \frac{C_{11} + 3C_{12}}{C_{11} + C_{12}}. \tag{18}$$

Other information can be obtained about the influence of lattice defects (Mason and Bateman 1964) and about the influence of free electrons (Bruner and Keyes 1961) which reduce the strength of the lattice bonds. A decrease in all elastic stiffness constants is shown by Hall (1967) as illustrated in Fig. 3.

2 Elastic Waves

Acoustic oscillations correspond to the macroscopic behavior of a solid responding to sound waves. At long wavelengths, this behavior can be described (Brown 1967; Musgrave 1970) by a continuum model in which the displacement u_i (in the harmonic approximation) follows the equation of motion:

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \sum_k \frac{\partial \sigma_{ik}}{\partial x_k} = \sum_j \sum_k \sum_l C_{ijkl} \frac{\partial^2 u_i}{\partial x_j \partial x_k}, \tag{19}$$

where u_1, u_2, and u_3 are displacements in the x-, y-, and z-directions, respectively; ρ is the mass density, and σ_{ik} is the stress tensor (see Sect. 1.1). The C_{ik} are related to the components of the C_{ijkl} tensor. This relation is given by[5]

[5]The factors are necessary due to a simplification in Voigt notation; the full tensor notation comprises, e.g., off-diagonal summands $C_{ijkl}\varepsilon_{kl} + C_{ijlk}\varepsilon_{lk} (k \neq l)$, while in the shorter matrix notation, the corresponding summands $C_{mn}\varepsilon_n$ ($n = 4, 5, 6$) only appear once.

$$C_{\alpha\beta} = C_{ijkl} \text{ for } 1 \leq \alpha, \beta \leq 3$$
$$C_{\alpha\beta} = \frac{1}{4} C_{ijkl} \text{ for } 4 \leq \alpha, \beta \leq 6$$
$$C_{\alpha\beta} = \frac{1}{2} C_{ijkl} \text{ for } 1 \leq \alpha \leq 3, \ 4 \leq \beta \leq 6, \text{ or } 1 \leq \beta \leq 3, \ 4 \leq \alpha \leq 6$$

and $\alpha = (ij), \beta = (kl)$ with i, j, k, l permutating for x, y, z, following the convention for (α or β) $= 1 \ldots 6$, equivalent to xx, yy, zz, yz, zx, and xy, respectively. Equation 19 is the wave equation.

In a cubic crystal, the relation (19) can be expressed as

$$\rho \frac{\partial^2 u_1}{\partial t^2} = C_{11} \frac{\partial^2 u_1}{\partial x^2} + C_{44} \left(\frac{\partial^2 u_1}{\partial y^2} + \frac{\partial^2 u_1}{\partial z^2} \right) + (C_{12} + C_{44}) \left(\frac{\partial^2 u_2}{\partial x \partial y} + \frac{\partial^2 u_3}{\partial x \partial z} \right), \quad (20)$$

with similar equations for u_2 and u_3 obtained by cyclic exchange of (u_1, u_2, u_3) and (x, y, z). Their solutions can be written for the displacement vector in plane wave form:

$$\mathbf{u}(\mathbf{r}, t) = \mathbf{A} \exp[i(\mathbf{q}\mathbf{r} - \omega t)], \quad (21)$$

with $|\mathbf{q}| = 2\pi/\lambda$ and $\mathbf{q}$ the wavevector for lattice oscillations. Here $\mathbf{q}$ will be used consistently to set it apart from $\mathbf{k}$, the electron wavevector (see Sect. 1.2 in ▶ Chap. 6, "The Origin of Band Structure"). In one dimension, q is referred to as the *wavenumber*. Introducing this *ansatz* (German for "trial solution") into Eq. 20 yields the set of equations that connect ω and $\mathbf{q}$, called the *dispersion equations* (see Sect. 3.1).

Ultrasound Measurement of Elastic Constants The sound velocities can be measured by using ultrasound transducers coupled to properly cut semiconductor platelets and measuring the time delay between an emitted pulse and its reflected echo. For a review, see Truell et al. (1969), Bogdanov and Sapozhnikov (2009). This method can be used to determine experimentally the elastic stiffness constants from the sound velocity for longitudinal and transverse acoustic waves in different crystallographic directions.

2.1 Sound Waves in Crystals

2.1.1 Cubic Crystals

We now determine the crystallographic direction in which the sound wave propagates and whether it is induced as a compression (longitudinal) wave or as a shear (transverse) wave. Only in special directions (see below) are these waves purely longitudinal or transverse; otherwise, they have components of each. In either case, the resulting dispersion relations are easily obtained. For instance, when a longitudinal sound wave propagates in the ⟨100⟩ direction, one obtains from Eq. 19

$$\rho \frac{\partial^2 u_1}{\partial t^2} = C_{11} \frac{\partial^2 u_1}{\partial x^2} \tag{22}$$

which yields with the ansatz $u_1 = A \exp[i(q_x x - \omega t)]$ the dispersion relation

$$\rho \omega^2 = C_{11} q_x^2. \tag{23}$$

Rewriting Eq. 22 as

$$\frac{\partial^2 u_1}{\partial t^2} = v_s^2 \frac{\partial^2 u_1}{\partial x^2}, \tag{24}$$

one obtains the sound velocity (see also Sect. 3.1):

$$v_s = \frac{\omega}{q} = \sqrt{\frac{C_{11}}{\rho}}. \tag{25}$$

For longitudinal and transverse acoustic waves, one has the simple relations

$$\text{(a)}\ \ v_{s,l} = \sqrt{\frac{C_{11}}{\rho}} \quad \text{and} \quad \text{(b)}\ \ v_{s,t} = \sqrt{\frac{C_{44}}{\rho}}, \tag{26}$$

obtained from an equation similar to Eq. 22. In the ⟨110⟩ direction, the relationship is a bit more involved, since two transverse modes must be distinguished:

$$\begin{aligned} &\text{(a)}\ \ v_{s,l} = \sqrt{\frac{C_{11} + C_{12} + 2C_{44}}{2\rho}}\ , \quad \text{(b)}\ \ v_{s,t_1} = \sqrt{\frac{C_{44}}{\rho}}, \\ &\text{(c)}\ \ v_{s,t_2} = \sqrt{\frac{C_{11} - C_{12}}{2\rho}}. \end{aligned} \tag{27}$$

Finally, in the ⟨111⟩ direction, one obtains

$$\text{(a)}\ \ v_{s,l} = \sqrt{\frac{C_{11} + 2C_{12} + 4C_{44}}{3\rho}} \quad \text{and} \quad \text{(b)}\ \ v_{s,t} = \sqrt{\frac{C_{11} - C_{12} + C_{44}}{3\rho}}. \tag{28}$$

Later (Sect. 2.2 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields"), when electron scattering with longitudinal acoustic phonons is considered, C_l is used as an abbreviation for a longitudinal elastic constant, which equals C_{11} in the ⟨100⟩ direction and $(C_{11} + C_{12} + 2C_{44})/2$ in the ⟨110⟩ direction.

These velocities are conventionally labeled v_1, v_3, and v_6 for longitudinal waves in the ⟨100⟩, ⟨110⟩, and ⟨111⟩ directions, respectively. The corresponding transverse

Table 4 Sound velocities in some cubic semiconductors (structure diamond (D) or zincblende (ZB)) given in 10^3 m/s. Expressions with square brackets (top) denote the propagation direction; below is the direction [] or plane () of polarization; the last line of the header gives the reference to the respective equation in the text

		v_1	v_2	v_3	v_4	v_5	v_6	v_7
		LA	TA_1, TA_2	LA	TA_1	TA_2	LA	TA_1, TA_2
		$\frac{[001]}{[001]}$	$\frac{[001]}{(001)}$	$\frac{[110]}{[110]}$	$\frac{[110]}{(001)}$	$\frac{[110]}{(110)}$	$\frac{[111]}{[111]}$	$\frac{[111]}{(111)}$
Crystal	Structure	Eq. 26a	Eq. 26b	Eq. 27a	Eq. 27b	Eq. 27c	Eq. 28a	Eq. 28b
Diamond	D	17.52	12.82	18.32	12.82	11.66	18.58	12.06
Si	D	8.43	5.84	9.13	5.84	4.67	9.36	5.09
Ge	D	4.87	3.57	5.36	3.57	2.77	5.51	3.06
SiC	ZB	11.0	8.92	12.7	8.92	6.21	13.3	7.23
β-GaN	ZB	6.9	5.02	7.87	5.02	3.3	8.17	3.96
GaP	ZB	5.83	4.12	6.43	4.12	3.08	6.63	3.46
GaAs	ZB	4.73	3.35	5.24	3.35	2.48	5.4	2.8
GaSb	ZB	3.97	2.77	4.38	2.77	2.07	4.50	2.33
InP	ZB	4.58	3.08	5.08	3.08	2.16	5.23	2.51
InAs	ZB	3.83	2.64	4.28	2.64	1.83	4.41	2.13
ZnSe	ZB	4.04	2.77	4.55	2.77	1.82	4.70	2.19
HgTe	ZB	2.57	1.60	2.85	1.60	1.01	2.94	1.24

waves are labeled v_2; v_4, v_5; and v_7, v_8 for the $\langle 100 \rangle$ (one wave), $\langle 110 \rangle$ (two waves), and $\langle 111 \rangle$ (two waves) directions, respectively.[6]

Sound velocities v_i are given for a number of semiconductors in Tables 4 and 5. Since they are related to the elastic stiffness constants, the sound velocities decrease with increasing temperature, as shown in Fig. 4. The change in crystal volume with temperature has the opposite influence on the sound velocity, but usually is a smaller effect. With application of uniaxial stress, the lattice symmetry is lowered, and new branches appear for the sound velocities, which split proportional to the applied stress.

2.1.2 Hexagonal Crystals

In hexagonal crystals, sound propagation along and perpendicular to the *c*-axis is distinguished. Values for some semiconductors are given in Table 5; the last line in the header gives the relations of v to the elastic stiffness constants. For sound velocities and second-order elastic constants of hexagonal crystals calculated from first principles see Sarasamak et al. (2010).

2.1.3 Anharmonic Effects

The above given relations $v_s = (C/\rho)$ with appropriate stiffness constant(s) C apply only for the harmonic approximation presumed in Eq. 19. If we include nonlinear

[6] Also the following labels are used (see also Eqs. 26, 27, and 28): v_L (corresponding to v_1 or $v_{s,l}^{100}$), v_T (v_2, $v_{s,t}^{100}$), v_l (v_3, $v_{s,l}^{100}$), $v_{t\|}$ (v_4, v_{t_1}), $v_{t\perp}$ (v_5, v_{t_2}), $v_{l'}$ (v_6, $v_{s,l}^{111}$), $v_{t'}$ (v_7, $v_{s,t}^{111}$).

Table 5 Sound velocities in hexagonal semiconductors (wurtzite (W) or 6H) at room temperature given in 10^3 m/s. Expressions with square brackets (top) denote the propagation direction; below is the direction [] or plane () of polarization (After Adachi 2005)

Crystal	Structure	v_1	v_2	v_3	v_4	v_5
		LA	TA_1,TA_2	TA_1	TA_2	LA
		$\frac{[00.1]}{[00.1]}$	$\frac{[00.1]}{(00.1)}$	$\frac{[01.0]}{[00.1]}$	$\frac{[01.0]}{[21.0]}$	$\frac{[01.0]}{[01.0]}$
		$\sqrt{\frac{C_{33}}{\rho}}$	$\sqrt{\frac{C_{44}}{\rho}}$	$\sqrt{\frac{C_{44}}{\rho}}$	$\sqrt{\frac{C_{11}-C_{12}}{\rho}}$	$\sqrt{\frac{C_{11}}{\rho}}$
SiC	6H	13.1	7.12	7.12	7.79	12.5
AlN	W	10.9	6.07	6.07	7.80	11.2
GaN	W	7.97	3.93	3.93	4.37	7.83
InN	W	5.17	1.21	1.21	2.51	5.28
ZnO	W	6.20	2.79	2.79	2.80	6.07
CdS	W	4.43	1.76	1.76	1.84	4.24
CdSe	W	3.86	1.54	1.54	1.60	3.62

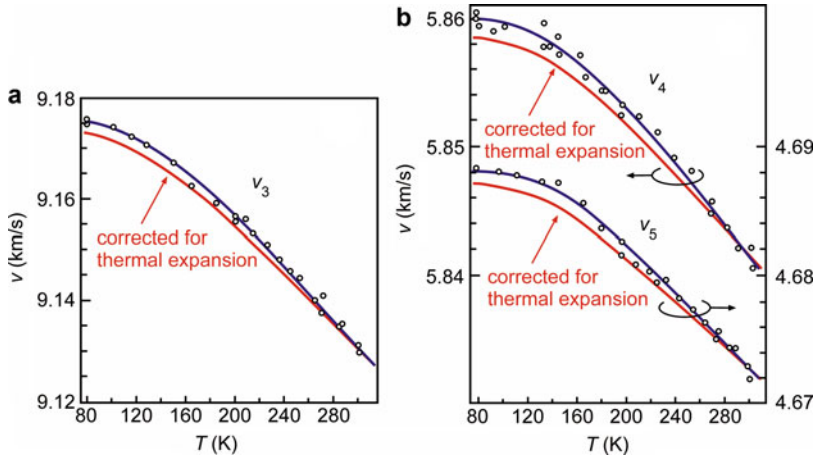


Fig. 4 Sound velocities of shear waves in Si versus temperature for propagation along ⟨110⟩: (**a**) longitudinal wave and (**b**) shear waves for particle motion parallel to ⟨001⟩ direction (v_4) and ⟨110⟩ direction (v_5)

terms in the equation of motion, no pure transverse modes exist.[7] The *nonlinear* wave equation of cubic crystals for propagation of *longitudinal* waves along the principal directions assumes the simple form (Joharapurkar and Breazeale 1990)

[7]There exist still pure longitudinal modes along the principal directions <100>, <110>, and <111>.

$$\rho \frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} \left(K_2 + (3K_2 + K_3) \frac{\partial u}{\partial x} + \text{higher terms} \right). \tag{29}$$

The quantities of K_2 for the respective directions contain the second-order stiffness constants and are given by $K_2 = \rho v_{s,l}^2$, while K_3 contains third-order stiffness constants. We obtain

$$\text{for } \langle 100 \rangle : K_2 = C_{11}, K_3 = C_{111}$$

$$\text{for } \langle 110 \rangle : K_2 = \frac{1}{2}(C_{11} + C_{12} + 2C_{44}), K_3 = \frac{1}{4}(C_{111} + 3C_{112} + 12C_{166})$$

$$\text{for } \langle 111 \rangle : K_2 = \frac{1}{3}(C_{11} + 2C_{12} + 4C_{44}),$$

$$K_3 = \frac{1}{9}(C_{111} + 6C_{112} + 12C_{144} + 24C_{166} + 2C_{123} + 16C_{456}).$$

Sound Beam Mixing At sufficient amplitudes, the nonlinear part of the interatomic potential becomes important in sound waves; this permits two sound beams of the same or different frequency to mix, and results in the creation of a third beam with sum or difference frequency (Hiki and Mukai 1973). The amplitude of the generated second harmonic is proportional to $(3\ K_2 + K_3)/K_3$.

Sound Damping and Crystal Defects Sound echos in good semiconducting samples can be observed after many reflections at the surfaces. Damping of the sound occurs because of inelastic scattering with phonons (see Sect. 3 for definition) and crystal defects. For interaction with phonons, considering the anharmonic part of the lattice potential, there are two extreme regimes: the Akhieser regime (Akhieser 1939) with $\omega_s \tau < 1$ and the Landau–Rumer regime with $\omega_s \tau > 1$ (ω_s = sound frequency, τ = phonon relaxation time) see Sect. 1 in ▶ Chap. 11, "Photon–Phonon Interaction." In the former, the sound attenuation is given by (Woodruff and Ehrenreich 1961)

$$\gamma_A = \delta^2 \omega_s^2 T \tag{30}$$

and in the second by (Landau and Rumer 1937)

$$\gamma_{L-R} = \delta^2 \omega_s T^4, \tag{31}$$

where δ is a quantity[8] proportional to the lattice anharmonicity and T is the temperature. A similar attenuation is important for phonon–phonon interaction, as

[8]The constant δ can be expressed as $\Gamma^2 C_v/(3\rho v_s^3)$ with Γ the Grüneisen parameter (Eq. 26 of ▶ Chap. 5, "Phonon-Induced Thermal Properties"), C_v the specific heat, ρ the density, and v_s the average sound velocity.

it determines heat conductivity (see Sect. 3 in ▶ Chap. 5, "Phonon-Induced Thermal Properties"). For a consistent treatment in the entire frequency range, see Guyer (1966).

There is a wide variety of crystal defects (see ▶ Chap. 15, "Crystal Defects") which cause further sound attenuation. The detectability of such defects depends on the wavelength of the sound: small defects (point defects and dislocations) require wavelengths on the order of the lattice constant; larger defects, such as crystallite boundaries, precipitates, and small cavities, can be detected with conventional ultrasound. In general, the wavelength should be on the order of the defect dimension to cause detectable sound attenuation.

In a continuum model, the elastic (acoustic) waves with a long wavelength are only one type of possible oscillations. The description will be extended below to more general lattice oscillations.

3 Phonon Spectra

When an atom is coupled to another atom of the same mass in a diatomic molecule, it can oscillate in a vibrational mode with a well-defined eigenfrequency

$$\omega_0 = \sqrt{\frac{\beta}{M}} \tag{32}$$

with β as the spring (force) constant and M as its mass. For Coulomb interaction, the spring constant can be given easily, relating the force to the spatial derivative of the Coulomb potential and yielding, near the equilibrium distance a, for $\beta \cong e^2/(4\pi\varepsilon_0 a)^2$.

When this same atom is imbedded in a monatomic crystal, a broad spectrum of oscillation is possible. In the following sections, we will analyze this vibration *spectrum*. Here, such an oscillatory state may be envisaged by holding a lattice of steel balls interconnected by springs, as shown in a two-dimensional model in Fig. 5, and by wiggling it, setting the balls in a jiggling oscillatory motion. The motion of each individual ball, although oscillatory in nature, is a complicated one, with components in all directions. It can be decomposed into a sum of many harmonic parts of a large number of frequencies. A systematic discussion in the following pages will show, however, that order can be brought into this complex picture.

The coupled lattice atoms in such a state of complex oscillation can be thought of as performing collective oscillations. Each mode into which it can be decomposed belongs to the entire crystal. Such modes can be regarded as quanta of elementary excitations, $\hbar\omega$, and are called *phonons*.[9] They are not localized; together, they belong to the entire crystal, until a specific event (a *scattering event*) occurs where,

[9] An elastic wave can thus be regarded as a stream of phonons, in analogy to an electromagnetic wave which can be described as a stream of photons. Both quasiparticles are not conserved; phonons or photons can be created by simply increasing the temperature or the electromagnetic field.

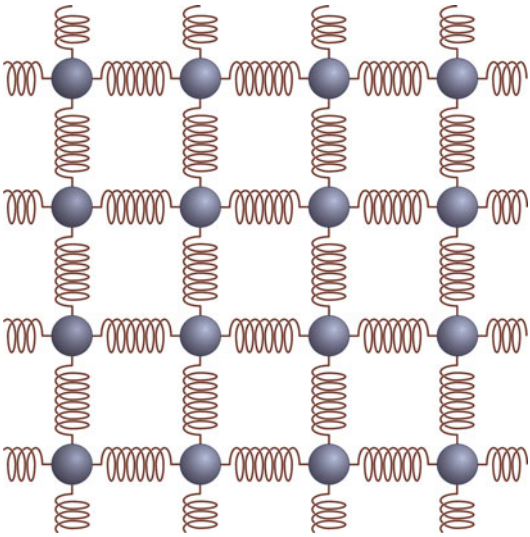


Fig. 5 Two-dimensional representation of a lattice of steel balls held in place by interconnecting springs

temporarily, such localization takes place. In our steel-ball model, it could be demonstrated by hitting one of the balls with a small hammer, temporarily making it oscillate more orderly in its resonance frequency, until, by interaction with its neighbors, the more erratic, jiggling motion is restored.

3.1 Oscillations of One-Dimensional Lattices

In order to understand this concept better, an analysis of the classical equation of motion for a simple one-dimensional lattice is given first. In such a lattice, *longitudinal* and *transverse modes of oscillations* are possible. In pure form, these oscillations entail oscillations in the direction of, or perpendicular to, the atomic chain, respectively. In general, therefore, one can describe an arbitrary state of oscillations by a superposition of one longitudinal and two (perpendicular to each other) transverse branches of oscillations. Pure longitudinal oscillations are described first.

3.1.1 Longitudinal Lattice Oscillation

In a *longitudinal* mode, the atoms in a monatomic chain oscillate, thereby changing the distance between each other, although remaining entirely along the x-axis; see the lower row of Fig. 6. For small displacements, assuming a harmonic lattice potential,[10] the force F is proportional to the relative atomic displacement u (Hooke's law).

[10] In the harmonic approximation, there is no exchange of energy between different modes of oscillation. Such exchange, necessary to return to equilibrium after any perturbation, requires anharmonicity.

For an arbitrary atom (index n), the forces acting on it are transmitted from its two neighbors ($n - 1$ and $n + 1$) and are given by (see Fig. 6)

$$F = \beta(u_{n-1} - u_n) - \beta(u_n - u_{n+1}) \tag{33}$$

with β the spring constant. Using Newton's law, the equation for motion of this atom is

$$M\frac{d^2u_n}{dt^2} = F = \beta(u_{n+1} + u_{n-1} - 2u_n). \tag{34}$$

The solution of this equation must describe atomic oscillations, which are given by

$$u_n = A\exp(i(q_n a - \omega t)) \tag{35}$$

with q as the wavenumber and na as the position x of the n^{th} atom. Equation 35 is used as a trial solution and is introduced into Eq. 34 for u_n, u_{n+1}, and u_{n-1}. We then obtain the relationship between ω and q,

$$M\omega^2 = -\beta(\exp(iqa) + \exp(-iqa) - 2) = 2\beta(1 - \cos(qa)), \tag{36}$$

from which the *dispersion equation* is obtained:

$$\omega(q) = \pm 2\sqrt{\frac{\beta}{M}}\left|\sin\left(\frac{qa}{2}\right)\right|. \tag{37}$$

The $\pm$ sign refers to waves traveling to the left or to the right. Standing waves, as required for the boundary conditions (Eq. 38), are obtained by superposition of the running waves with opposite propagation and equal amplitude.

The dispersion relation is pictured in Fig. 7.[11] The frequency increases linearly with q for $(qa/2) \ll 1$, as is expected for elastic waves in a homogeneous elastic material. However, $\omega(q)$ then levels off and reaches a saturation value $\left(\sin\frac{\pi}{2} = 1\right)$ when $q \to \frac{\pi}{a}$, i.e., when the wavelength of the oscillations ($\lambda = 2\pi/q$) approaches twice the interatomic distance; the atomic character of the medium becomes apparent

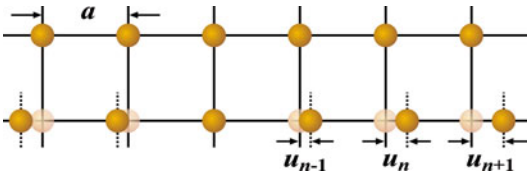


Fig. 6 Linear chains of atoms with the *upper* chain at rest and the *lower* vibrating in a longitudinal mode

[11] To point out the similarity to the $E(k)$ diagram for electrons in a periodic potential (see Sect. 2 in ▶ Chap. 6, "The Origin of Band Structure"), $\hbar\omega$ is plotted rather than ω as a function of q throughout the book. The limit of the wavenumber between $-\pi/a$ and $+\pi/a$ indicates the boundaries of the first

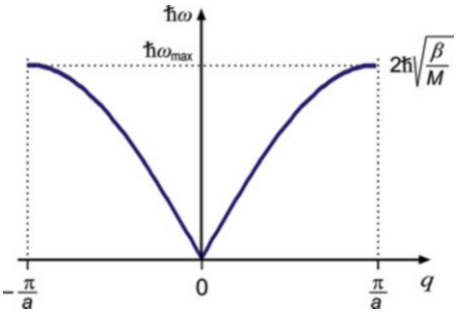


Fig. 7 Dispersion relation (here for the energy $\hbar\omega$) as given by Eq. 37 for a linear chain of the same type of atoms with an interatomic distance a. The boundaries in q are those of the first Brillouin zone

here (see below). The maximum frequencies $\left(\omega_{\max} = 2\sqrt{\beta/M}\right)$ are typically on the order of 10^{13} Hz; the related maximum phonon energy $\hbar\omega_{\max}$ is on the order of 30 meV (240 cm^{-1}) for typical semiconductors.

3.1.2 Transverse Lattice Oscillation

The same dispersion relation is obtained for transverse waves in a linear chain as long as the amplitude remains small and the force constants can be described by the same β. For these waves, however, the oscillations occur in a plane perpendicular to the x-axis. Since there are two of these planes orthogonal to each other, we distinguish *two transverse polarizations* of these oscillations. Figure 8 presents some examples of such *modes* of transverse vibrations in a very short chain.

As a boundary condition, we can require that no energy is transferred from or to the outside; this is the adiabatic boundary condition and requires nodes at the outer "surfaces" ($x = 0$ and $x = l$). With this condition, q can only attain discrete values q_n, with

$$q = q_n = \frac{2\pi n_q}{l} \quad n_q = 1, 2, \ldots (N = l/a), \tag{38}$$

where N is the number of atoms in the chain and l is the length of the chain. The upper limit $n_{q,\max}$ is given by the fact that the minimum wavelength is a when fulfilling the boundary condition of nodes at both ends of the chain. Smaller wavelength representations, although possible, describe no new type of atomic oscillation and are disregarded. Therefore, a chain of N atoms has exactly N different modes for the given polarization of such oscillations.[12] Or, an arbitrary state of motion of such a chain can be decomposed by the superposition of $2N$ normal modes. The factor of two stems from the two polarizations of transverse

Brillouin zone (Sect. 1.3 in ▶ Chap. 3, "The Structure of Semiconductors"). Extending the diagram beyond this interval provides no new information.

[12]Strictly speaking, there are only $N - 2$ different modes when requiring that the surface atoms remain at rest. By bending the (long) linear chain into a circle, one can obtain an equivalent condition by introducing cyclic boundary conditions [requiring $u(x = 0) = u(x = l)$] to get around this "$N - 2$" peculiarity (Born and von Karman 1912). For large N, however, one always has $N - 2 \cong N$. Such a cyclic boundary condition is also necessary to permit propagating waves.

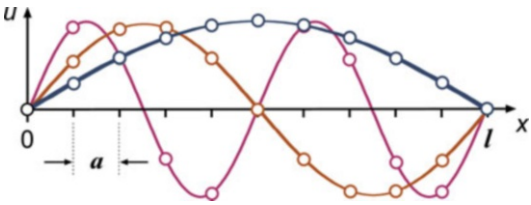


Fig. 8 Three modes of transverse oscillation in a short one-dimensional chain of length l

oscillations. Depending on the total energy (i.e., thermal energy) in the chain, these different modes can be excited to a greater or lesser degree; however, with oscillations of quantized energy steps

$$E_n = \hbar\omega_n \quad \text{and} \quad \omega_n = \omega(q_n), \tag{39}$$

each one of these steps represents a *phonon*. For a single oscillator, a larger excitation causes a larger amplitude at the same eigenfrequency, i.e., a larger energy and therefore a larger number of phonons, each one of them having the same energy. The occupation of a specific mode is then given by the total energy within this mode divided by $\hbar\omega$:

$$\frac{E_n^{(m)}}{E_n} = n + \frac{1}{2} \tag{40}$$

with m phonons residing in this mode. The term ½ stems from the *zero-point energy*, i.e., a remaining fraction of energy at $T = 0$ K.

In thermal equilibrium, the many different modes are occupied with phonons according to the Bose–Einstein distribution function: the higher the temperature, the more phonons that appear.

With one longitudinal and two transverse branches, we have a total of $3N$ phonon modes in a linear lattice. This also holds true for two- or three-dimensional *monatomic* lattices, where N is the total number of lattice atoms (i.e., there is also a total of $3N$ phonon modes).

Comparing the relative displacement of atoms in the three different modes shown in Fig. 8, we see that the displacement of adjacent atoms with respect to each other becomes larger with decreased wavelength, i.e., with increased q. Therefore, more energy is contained in modes with higher ω. This relation between phonon energy and wavelength $\lambda = 2\pi/q$ is given by the dispersion relations (Eqs. 37 and 45).

3.1.3 Transverse Oscillation in a Diatomic Lattice

In a lattice with a basis (see Sect. 1 in ▶ Chap. 3, "The Structure of Semiconductors") (e.g., a diatomic linear chain with alternating masses), the equation of motion can be split into a set of two oscillatory equations:

$$M_1 \frac{d^2 u_{2n}}{dt^2} = \beta(u_{2n+1} + u_{2n-1} - 2u_{2n}) \tag{41}$$

and

$$M_2 \frac{d^2 u_{2n+1}}{dt^2} = \beta(u_{2n+2} + u_{2n} - 2u_{2n+1}). \tag{42}$$

These require two wave equations for the displacements of light and heavy atoms with indices 1 and 2, respectively, and a' the nearest neighbor distance $= a/2$ with a the lattice constant:

$$u_{2n} = A_1 \exp(i(2nqa' - \omega t)) \tag{43}$$

and

$$u_{2n+1} = A_2 \exp(i([2n+1]qa' - \omega t)). \tag{44}$$

with amplitudes A_1 and A_2. Introducing Eqs. 43 and 44 into Eqs. 41 and 42 and solving the ensuing secular equations yields the *dispersion equation* for a diatomic chain

$$\omega_{\pm}^2 = \frac{\beta(M_1 + M_2)}{M_1 M_2} \left(1 \pm \sqrt{1 - \frac{4M_1 M_2 \sin^2(qa')}{(M_1 + M_2)^2}}\right). \tag{45}$$

It has two solutions depending on the sign of the square root. The two corresponding $\hbar\omega(q)$ branches are shown in Fig. 9.

The upper branch is referred to as the *optical branch*, the lower as the *acoustic branch*. The amplitudes of the different sublattice oscillations near $q = 0$ for the same wavelength are the same for the *acoustic* branch:

$$\frac{A_1}{A_2} \cong \begin{cases} 1 & \text{for} \quad q \to 0 \\ \infty & \text{for} \quad q \to \dfrac{\pi}{a} \end{cases}. \tag{46}$$

For $q = \pi/a$, the heavy atom is at rest, as shown in Fig. 10b (top), with amplitude $A_2 = 0$. For the *optical* branch,

$$\frac{A_1}{A_2} \cong \begin{cases} -\dfrac{M_2}{M_1} & \text{for} \quad q \to 0 \\ 0 & \text{for} \quad q \to \dfrac{\pi}{a} \end{cases}. \tag{47}$$

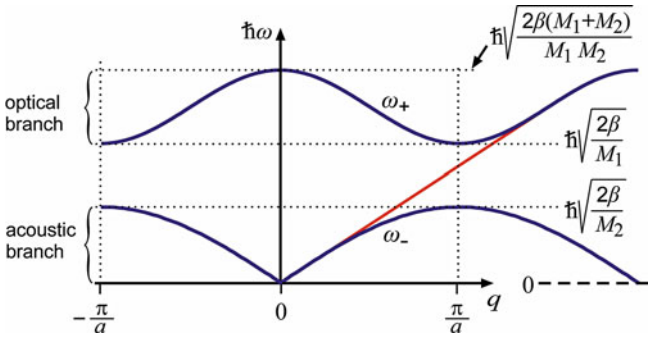


Fig. 9 Dispersion relations for a diatomic chain of light (M_1) and heavy (M_2) atoms in alternating order, indicating acoustic and optical branches with a forbidden frequency range in between. The energies $\hbar\omega$ at $q = 0$ and $q = \pm \pi/a$ are given at the right-hand side. The *red curve part* indicates the one branch when $M_1 = M_2$

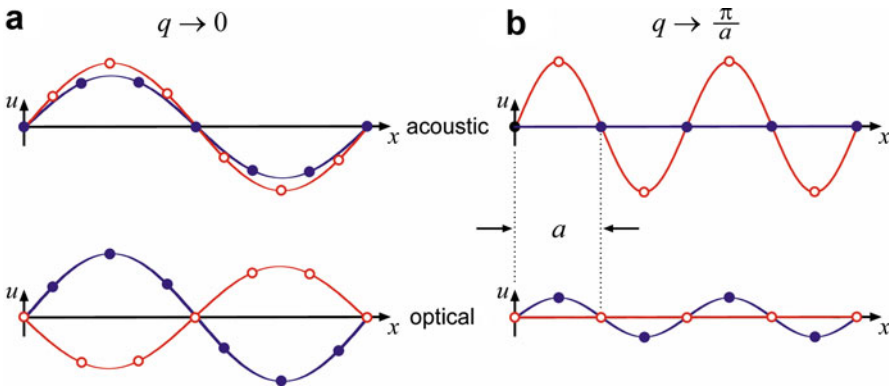


Fig. 10 Acoustic and optical branches of transverse oscillations with the same wavelength in a diatomic lattice for (**a**) long-wavelength limits and for (**b**) short-wavelength limits ($\lambda = 2a$) with heavy (*blue*) and light atoms (*red*) at rest for acoustic and optical modes, respectively

Here, for $q = \pi/a$, the light atom is at rest with amplitude $A_1 = 0$. Near $q = 0$, the lattices oscillate opposite to each other, while the center of mass remains stationary; therefore, the lighter mass oscillates with larger amplitude, as indicated in Fig. 10.

If the lattice binding force is (partially) ionic, the optical branch of oscillation will show a large dipole moment interacting with electromagnetic radiation (photons)

– hence the name *optical phonons*. The acoustic branch has a much smaller dipole moment since adjacent atoms oscillate with each other (Fig. 10a); therefore, optical stimulation is less effective for acoustic mode excitation. This mode is stimulated preferably by mechanical means such as sound waves.

At short wavelengths ($q \cong \pi/a$), the heavy atoms are oscillating while the light atoms are at rest in the optical mode; the oscillation energy is large. At the same wavelength in the acoustic mode, only the light atoms are oscillating with larger amplitude but with lesser energy (see Fig. 11).

3.1.4 Phonon Velocity

A phonon with energy $\hbar\omega$ has been introduced as related to lattice oscillation after localization with a certain frequency ω, i.e., a certain mode of a collective state of oscillations. A phonon may also be seen as a quasiparticle when described as a wave packet, similar to the description of an electron which is to be discussed in Sect. 2.1 in ▶ Chap. 6, "The Origin of Band Structure." The velocity of such a wave packet is given by the *group velocity*[13] or *phonon velocity*:

$$v_g = \frac{\partial}{\partial \mathbf{q}} \omega(\mathbf{q}), \tag{48}$$

that is, by the slope of $\omega(\mathbf{q})$. In a one-dimensional lattice and isotropic solids, the gradient $\partial\omega(\mathbf{q})/\partial\mathbf{q}$ is replaced by the derivative $d\omega(q)/dq$.

[13]The group velocity can be defined when at least two waves of slightly different frequencies interact and form a wave train with an envelope forming beats (Fig. 12). Since energy cannot flow past a node, one readily sees that the velocity with which energy is transmitted must equal the velocity with which the nodes move. Adding two waves with ω, q and $\omega + d\omega$, $q + dq$, one obtains for the superposition

$$u_1 + u_2 = (A_1 + A_2) \cos\left(\frac{t}{2} d\omega - \frac{x}{2} dq\right). \tag{49}$$

Thus the motion of the zero-phase point of the envelope

$$\frac{t}{2} d\omega - \frac{x}{2} dq = 0 \tag{50}$$

yields for the velocity of the groups of waves:

$$v_g = \frac{x}{t} = \frac{d\omega}{dq} \tag{51}$$

or, more generally, Eq. 48.

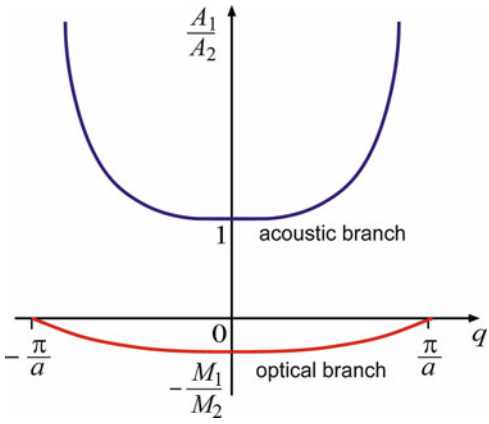


Fig. 11 Amplitude ratio for light and heavy atoms as a function of the wavenumber (After Weißmantel and Hamann 1979)

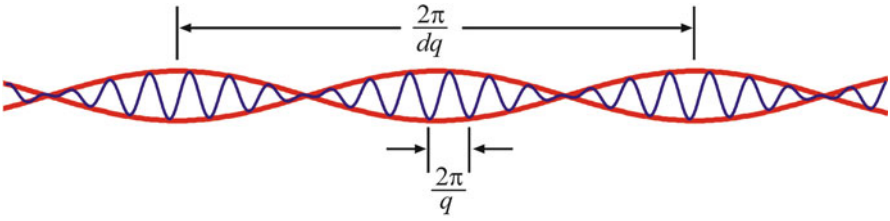


Fig. 12 Wave train (*blue curve*) made by superposition of two waves with slightly different frequencies, resulting in beats of the envelope function (*red curve*)

There is another important velocity in solids, the *phase velocity*,[14] or *sound velocity*, which is given by

$$v_s = \frac{\omega}{q}. \tag{52}$$

Both group and phase velocities are the same in the acoustic branch at low values of q, i.e., at long wavelengths where $\omega \propto q$ (see Figs. 7, 9, and three-dimensional phonon dispersion-curves near the Γ point, i.e., at $q \to 0$); however, they become substantially different where dispersion occurs.

With decreasing wavelength (increasing q), the phase velocity decreases but remains on the same order of magnitude as the sound velocity near the Γ point. At the surface of the Brillouin zone (e.g., at the X point of a cubic semiconductor), the

[14]The velocity in which the phase of a single wave moves; for a node, it is given by

$$\omega t - qx = 0, \tag{53}$$

resulting in the phase velocity given by Eq. 53; this is the velocity with which energy is transported in such a wave.

sound velocity is decreased slightly to $v_s(q = \pi/a) = \omega(q = \pi/a)/(\pi/a)$. An estimate for the sound velocity is obtained by extrapolating the acoustic branch $\omega = v_s q$ to $q = \pi/a$ with $\hbar\omega_n(q = \pi/a) \approx k\theta$, yielding

$$v_s = \frac{k}{\hbar}\frac{a\theta}{\pi} = 4.168 \;\; \theta(\mathrm{K}) \;\; a(\text{Å}) \quad (\mathrm{m/s}); \tag{54}$$

with the Debye temperature θ (see Sect. 1.1.2 in ▶ Chap. 5, "Phonon-Induced Thermal Properties") on the order of 300 K, the average phonon velocity is on the order of 10^3 m/s.

In contrast, the group velocity, given by the slope of $\omega(q)$, vanishes at $q = \pi/a$; at the surface of the Brillouin zone, the waves have become standing waves, composed of two components moving in opposite directions with the same amplitude and velocity. This is the condition of Bragg reflection and will be discussed later in more detail for a similar problem dealing with electrons (Sect. 2.2 in ▶ Chap. 6, "The Origin of Band Structure").

In summary, in the diatomic one-dimensional chain, we have for the acoustic branch

$$v_g = v_s = a\sqrt{\frac{\beta}{2(M_1 + M_2)}} \quad \text{for} \quad q \ll \frac{\pi}{a} \tag{55}$$

and

$$\left.\begin{aligned} v_g &= 0 \\ v_s &= a\sqrt{\frac{2\beta}{\pi^2 M_2}} \end{aligned}\right\} \quad \text{for} \quad q \cong \frac{\pi}{a}; \tag{56}$$

and for the optical branch

$$\left.\begin{aligned} v_g &= 0 \\ v_s &: \text{meaningless} \end{aligned}\right\} \quad \text{for} \quad q \ll \frac{\pi}{a} \quad \text{and for} \quad q \cong \frac{\pi}{a}. \tag{57}$$

In a three-dimensional lattice, similar results are obtained except for the anisotropic velocities (see Sect. 2.1).

3.2 Phonons in a Three-Dimensional Lattice

In a three-dimensional lattice, the dispersion equation $\omega(\mathbf{q})$ is given by a hypersurface. A cut through such a surface in a specific crystallographic direction is conventionally presented and is shown for the [100] direction in Fig. 13 for two typical semiconductors. There are $3p$ branches of the phonon dispersion, where p is the number of atoms in the basis. In three branches, $\hbar\omega$ goes to zero for vanishing q; these are the *acoustic branches* (1 longitudinal, LA, and 2 transversal, TA). All

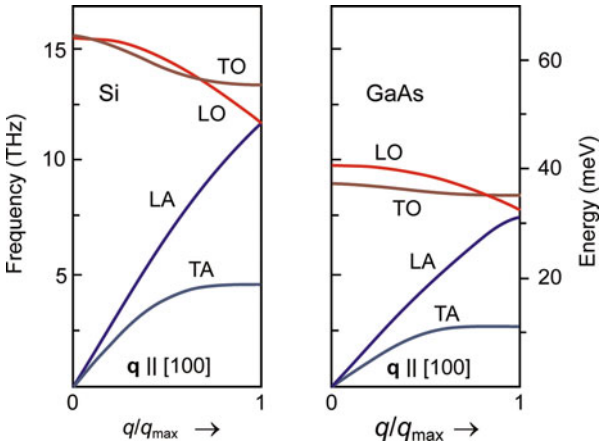


Fig. 13 Phonon spectrum for Si and GaAs in the [100] direction (details of Figs. 14 and 15a, respectively)

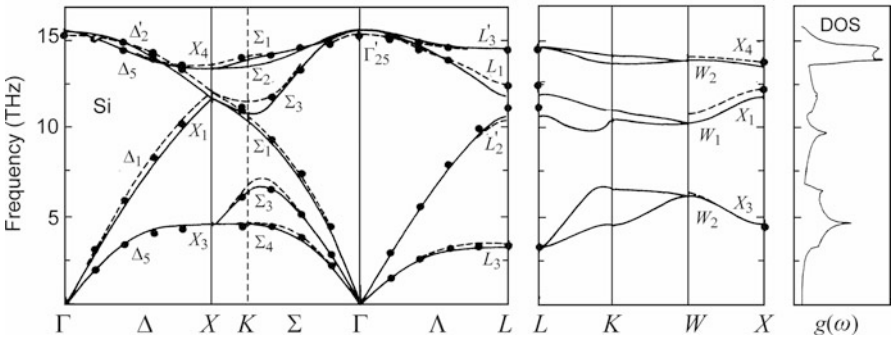


Fig. 14 Phonon dispersion curves for Si. *Solid curves* are calculated by Weber (1977); *filled circles* are experimental values. The *right panel* shows the density of states (DOS) of the phonon modes

$3p - 3$ others are *optical branches*.[15] Their energy $\hbar\omega$ does not vanish in the long-wavelength limit at $q \to 0$. Monatomic lattices also show optical branches when they have more than one atom per unit cell (e.g., Si); here we must distinguish between oscillations in which the atoms within the *unit cell* vibrate about their center of mass (optical), and oscillations in which the *center* of the unit cell vibrates (acoustic).[16] In diatomic lattices, there is a gap between optical and acoustic branches (GaAs), while

[15]For instance, in a crystal with four atoms in the basis (primitive unit cell), one has three acoustic and nine optical branches; see the phonon dispersion of wurtzite GaN shown in Fig. 15b.

[16]Some materials undergo a phase transition between diatomic and monatomic unit cells (e.g., *hcp* → *fcc* transition), whereby the optical branch disappears during the transition. For example, Ca shows such a phase transition at 450 °C.

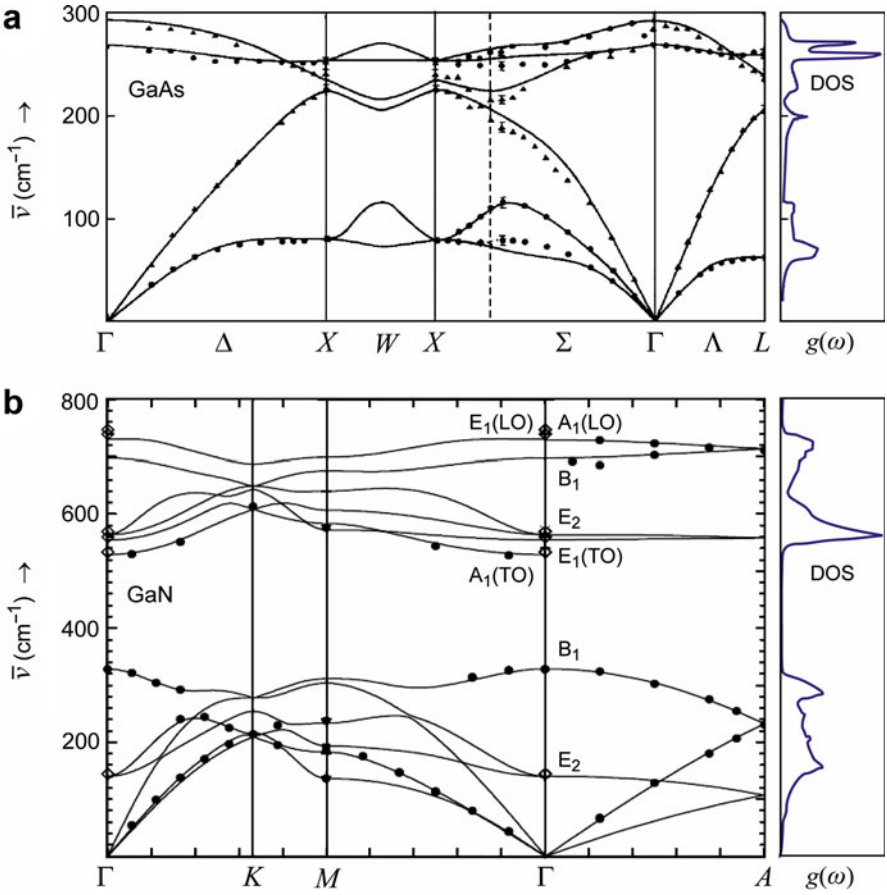


Fig. 15 (**a**) Phonon dispersion curves of GaAs as measured from neutron diffraction and one-phonon density of states as a function of frequency calculated from the rigid ion model (After Patel et al. 1984). (**b**) Same for GaN measured from inelastic x-ray scattering (After Ruf et al. 2001; DOS after Davydov et al. 1998)

in monatomic lattices, there is none (Si). The phonon energies are higher in lattices with larger binding energies (compare GaAs and GaN in Fig. 15).

Optical phonon branches show a small $\omega(\mathbf{q})$ dependence near $\mathbf{q} = 0$, with a rather well-defined phonon energy $\hbar\omega_0 \cong \hbar\omega(\mathbf{q} = 0)$. The transverse optical branch near $q = 0$ is important for optical absorption, while the longitudinal branch is more important for scattering with electrons; see Sect. 2.3 in ▶ Chap. 23, "Carrier Scattering at Low Electric Fields," for more detail. Acoustic phonon branches show a linear dispersion $\omega(\mathbf{q}) \propto q$ near $\mathbf{q} = 0$, with a derivative reflecting the sound velocity discussed in Sect. 3.1. In anisotropic lattices, one has to distinguish more complicated modes, such as involving bond stretching and bond bending between the different lattice atoms.

The conventional diagram within the first Brillouin zone shown in Figs. 14 and 15 is an intersection of the $\omega(\mathbf{q})$ hypersurface with planar surfaces connecting the listed symmetry points (see Sect. 1.3 and Fig. 8 in ▶ Chap. 3, "The Structure of Semiconductors") and changing direction at each vertical line (recent data for Si is reported in Kim et al. 2020). A similar representation is conventionally used in the $E(\mathbf{k})$ diagram for the dispersion relation of electrons in the semiconductor; see Sect. 4 in ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals." Optical branches are distinguished as A_1 and B (both nondegenerate) and E_1 and E_2 (both twofold degenerate). Only A_1 and E_1 are dipole active, i.e., they show up in absorption or reflection. A_1, E_1, and E_2 are Raman active (see Sect. 3.3 in ▶ Chap. 11, "Photon–Phonon Interaction"); B is not optically active. A listing of the characteristic phonon energies is given in Table 6.

3.2.1 Phonon Density of States

The phonon density distribution $N(\omega)d\omega$ (see Eq. 13 in ▶ Chap. 5, "Phonon-Induced Thermal Properties") determines many properties of solids including mechanical (thermal expansion), thermal (energy content), electrical, and optical (phonon scattering) properties. The density of states $g(\omega)$, contained in $N(\omega)$, can be obtained by counting the number of *modes* occurring within a frequency interval $(\omega, \omega + d\omega)$.

It is instructive to derive the density of states analytically for the linear chain of atoms. Here, the dispersion relation (37) has discrete normal modes for $q = q_n$ given in Eq. 38:

$$\omega(q_n) = \omega_0 \sin\left(\frac{q_n a}{2}\right) = \omega_0 \sin\left(\frac{\pi n_q a}{l}\right). \tag{58}$$

The density of states is given by the modes per frequency interval, which can be obtained by differentiation of Eq. 58, replacing cos $(\pi n_q/l)$, and using $\sin^2 + \cos^2 = 1$:

$$\frac{d\omega}{dn_q} = \frac{\pi a}{l}\omega_0 \cos\left(\frac{\pi n_q a}{l}\right) = \frac{\pi a}{l}\omega_0 \sqrt{1 - \left(\frac{\omega}{\omega_0}\right)^2}. \tag{59}$$

Consequently, one has for the density of states of the linear chain the reciprocal of Eq. 59:

$$dn_q = g(\omega)d\omega = \frac{l}{\pi a \omega_0} \frac{d\omega}{\sqrt{1 - \left(\frac{\omega}{\omega_0}\right)^2}}. \tag{60}$$

In a three-dimensional lattice, and in more general terms, this density of states can be expressed by

$$g(\omega)d\omega = V \int_{S_\omega} \frac{dS_\omega}{\left|\frac{\partial}{\partial \mathbf{q}}\omega\right|} d\omega, \tag{61}$$

Table 6 Phonon energies of cubic semiconductors at room temperature in meV, given for characteristic points of the Brillouin zone (quantities are often also given in cm^{-1} or THz units; the conversion factors are 0.124 meV/cm^{-1} and 4.141 meV/THz)

		$\hbar\omega_{TO}$			$\hbar\omega_{LO}$			$\hbar\omega_{TA}$			$\hbar\omega_{LA}$		
Crystal	Structure	$\Gamma_{25'}$	$L_{3'}$	X_4	$\Gamma_{25'}$	L_1	X_1	L_3	X_3	X_1	$\Gamma_{25'}$	$L_{2'}$	X_1
Diamond	D	165	150	133	165	155 ($L_{2'}$)	145	70	100	147	165	125 (L_1)	147
Si	D	64.2	60.7	57.5	64.2	52.1		14.2	18.6	51		46.8	51
Ge	D	37.3	35.4	33.8	37.3	30.1	29.5	7.7	9.9			27.4	29.5
α-Sn	D	24.8	23.7	22.8	24.8	20.2	5.2 (X_3)	4.1	5.3	19.3		17.1	19.3
		Γ	L	X	Γ	L	X		L	X		L	X
SiC	ZB	98	94.9	94.0	102.8	103.9	102.8		33.0	45.8		75.6	78.7
GaP	ZB	45.3	43.8	44.0	50.7	50.3	45.4		10.3	12.9		26.7	31
InP	ZB	38.0	39.3	40.1	42.7	42.2	41.1		6.8	8.5		20.7	24.0
AlAs	ZB	44.9			50.1								
GaAs	ZB	33.0	22.3	31.4	35.2	29.4	29.8		7.7	9.8		25.8	28.1
InAs	ZB	26.9	27.6	26.7	29.5	25.1	25.1		5.4	6.6		17.2	19.8
ZnSe	ZB	25.4	25.6	25.6	31.0	27.8	27.7		7.2	8.7		22.5	23.6
ZnTe	ZB	21.9	21.5	21.5	25.6	22.3	22.8		5.2	6.7		16.9	17.7
CdTe	ZB	17.4	17.9		21.0	17.9	18.4		3.7	4.4		13.5	

where V is the volume, S_ω is the surface of constant energy,[17] dS_ω is the surface element of S_ω, and $\partial\omega/\partial\mathbf{q}$ is the gradient of ω in $\mathbf{q}$ space. In the nondispersive range, where $\omega \propto q$ for the acoustic branch (see Eq. 9 in ▶ Chap. 5, "Phonon-Induced Thermal Properties"), we can show that

$$g(\omega_{\text{acoustic}}) \propto \omega^2. \tag{62}$$

For higher values of q, and in optical branches, peaks occur in the density distribution when this gradient tends to zero (see also Fig. 2 in ▶ Chap. 5, "Phonon-Induced Thermal Properties"). These are the *critical points* identified by capital letters. The different branches are numbered in the $g(\omega)$ diagram starting from the acoustic branch. Most important are critical points in the center, or at the boundary, of the Brillouin zone (see Figs. 14 and 15): at Γ (for optical branches) and at X, K, and L (for identification, see Fig. 8 in ▶ Chap. 3, "The Structure of Semiconductors"): the corresponding density of state function has spikes; many more modes per $d\omega$ interval occur, where the branch $\omega(q)$ is flat rather than in a range where ω varies steeply with q, as, for example, in the acoustic branch at the Γ point.

3.2.2 Local Phonon Modes

In the neighborhood of lattice defects (see ▶ Chap. 15, "Crystal Defects"), the binding forces and the mass for extrinsic defects are altered; thus the oscillatory behavior is locally modified. The eigenfrequency of such a defect with mass M_d and force constant of bond β_d is different from the most abundant lattice frequencies. In the simple isotropic case, it is given by

$$\omega_{0,d} = 2\sqrt{\frac{\beta_d}{M_d}}. \tag{63}$$

When the density of a specific defect is large enough and its eigenfrequency lies outside the allowed ranges of the intrinsic phonon spectrum,[18] a *localized vibrational mode* (LVM) can be experimentally distinguished (e.g., its IR absorption can be observed), providing information about the defect center. For instance, the absorption spectrum of substitutional ^{12}C on an arsenic site in GaAs is shown in Fig. 16. The frequency is sensitive to the isotope masses of the four tetrahedrally surrounding Ga next neighbors, being ^{69}Ga or ^{71}Ga with a natural abundance ratio of 61%/39%. Pure tetrahedral ^{71}Ga and ^{69}Ga surrounding yield lowest and highest

[17] This surface is similar to the Fermi surface in the Brillouin zone discussed in Sect. 1.1 in ▶ Chap. 8, "Bands and Bandgaps in Solids." For phonons, however, the energy surface lies at much lower energies.

[18] For instance, if the mass of the defect atom is much smaller, the eigenfrequency can be higher than the highest lattice phonon frequency (*local mode*); for heavier atoms, its optical phonon may lie within the gap between the optical and acoustic branches (*gap mode*). In a *resonant mode*, the eigenfrequency of the foreign atom lies within the band of a phonon branch of the host lattice. This mode shows a greatly enhanced amplitude.

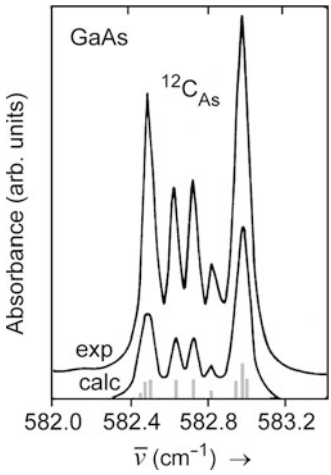


Fig. 16 Local vibrational modes of ^{12}C on As site in GaAs. Calculated energies and oscillator strengths of transitions related to different configurations of tetrahedrally coordinated ^{69}Ga and ^{71}Ga next neighbors are indicated by *gray bars*; calculated Lorentzians are artificially broadened (After Newman (1993))

energy, respectively; bars in Fig. 16 indicate positions and oscillator strength of theoretical transitions of various surroundings.

Often, the lattice oscillation surrounding a defect is described by radially expanding and contracting oscillations, i.e., by a *breathing mode*. For reviews, see Barker and Sievers (1975) and Mitra and Massa (1982). Further detail is given in Sect. 3.3 of ▶ Chap. 11, "Photon–Phonon Interaction." For local modes in highly disordered semiconductors, see also Sect. 3.3 of ▶ Chap. 5, "Phonon-Induced Thermal Properties."

3.2.3 Phonon Modes in Mixed Crystals

Phonon frequencies in alloyed crystals of the type $A_xB_{1-x}C$ (or, analogously, AB_xC_{1-x}) depend on the composition parameter x.[19] For *acoustic phonons* in the long-wavelength limit, the sound velocity, and correspondingly the phonon dispersion ($q \to 0$), is determined by the weighted average of the alloyed atoms. In the case $q \to \pi/a$, however, only one sublattice vibrates (see Fig. 10b). The phonon spectrum then may be broadened due to different masses and force constants of the alloyed atoms. The *optical phonons* exhibit an either one-mode or two-mode behavior (see Barker and Sievers 1975). In *one-mode behavior*, the mode frequencies of the pure *BC* crystal evolve continuously (and roughly linearly) with increasing x toward the mode frequencies of the pure *AC* crystal. In the frequently observed *two-mode behavior*, the LO–TO multiplet of the pure *BC* crystal ($x = 0$) evolves with increasing x approximately linearly toward a localized mode of *B* atoms in an *AC* crystal ($x = 1$); simultaneously a localized mode of *A* atoms in *BC* evolves to the LO–TO multiplet in the pure *AC* crystal. The two-mode behavior of $Al_xGa_{1-x}As$ mixed

[19]For simplicity, the frequently applied alloying on either the anion or the cation sublattice of a compound semiconductor is considered.

crystal is shown in Fig. 17. At small composition parameter x, the LO–TO multiplet is GaAs-like; at large x, it is AlAs-like.

Some mixed crystals exhibit a *mixed-mode behavior*, i.e., both kinds of behavior in different concentration ranges (see Deych et al. 2000).

3.2.4 Pressure Dependence of Phonons

With hydrostatic pressure, the interatomic distance is reduced, resulting in an increase of ion–ion interaction and a reduction of electronic screening. This causes an increase of the optical phonon energy at the Γ point of $E(\mathbf{q})$ with increasing pressure. However, the pressure coefficient for TA (transversal acoustic) modes at the Brillouin zone boundary (at X and L) is negative, which seems to be an intrinsic property of shear distortion in predominately covalent semiconductors that are involved in such modes (Martinez 1980). The corresponding pressure coefficients are listed in Table 7.

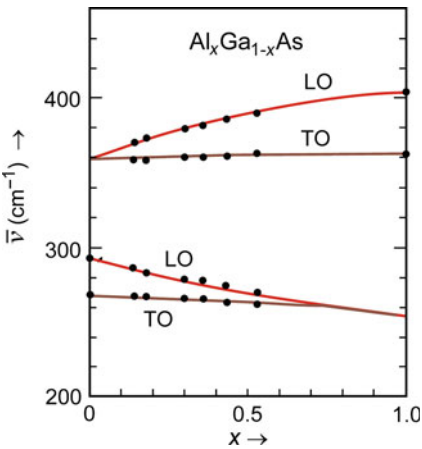


Fig. 17 Optical phonon branches for $Al_xGa_{1-x}As$ mixed crystals as a function of the alloying parameter x (After Adachi 1985)

Table 7 Pressure coefficient of phonon modes at room temperatures in meV/kbar (After Martinez 1980)

Crystal	Structure	$d\omega_{TO}/dp$	$d\omega_{LO}/dp$	$d\omega_{TA(X)}/dp$	$d\omega_{TA(L)}/dp$
C	D	0.445	0.1		
Si	D	0.064	−0.027	−0.019	
Ge	D	0.057		−0.004	
GaP	ZB	0.054	0.056	−0.011	−0.009
GaAs	ZB	0.053	0.054		
GaSb	ZB	0.062	0.062		
ZnS	ZB	0.077	0.053	−0.021	−0.18
ZnSe	ZB	0.063	0.052	−0.019	−0.17
ZnTe	ZB	0.070	0.056	−0.015	−0.01

3.2.5 Microscopic Force Models

More sophisticated models than the harmonic approach outlined in Sect. 3.1 consider the interacting potential $V(R)$ between the lattice atoms, from which the force constants $\beta_{\alpha i}$ can be determined by forming the spatial derivative of these potentials with respect to the deformation. When these force constants are introduced into the set of equations of motion similar to Eqs. 41 and 42, and a planar wave *ansatz* is introduced, the resulting secular equations have as their solution the dispersion equation. It should be noted that all of these calculations use an empirical potential with several (usually 4–6) adjustable parameters.

Depending on the interacting potential, we must distinguish the *rigid ion model* – which includes only ion–ion interaction, i.e., central forces except for nearest neighbor interaction, which may be viewed as noncentral (Rajagopal and Srinivasan 1960) – and several more refined models, which include the influence of the electron shell (bond angle influence). For a review, see Bilz and Kress (1979); an example is given in Fig. 15a for GaAs. *Shell models*, which include valence forces into the interaction potential, are used for covalent semiconductors. Turbino et al. (1972) applied such a potential to achieve good agreement with measured $\omega(q)$ curves for the group IV elements. Another shell model proposed by Jasval (1975) successfully calculated the dispersion relation for III–V and II–VI compounds (Jasval 1977). Best agreement between the model computation and experiment has been obtained for III–V compounds by Weber (1977) with a *bond-charge model* and is shown for Si in Fig. 13.

More recently, ab initio calculations are applied by tracing back lattice dynamics to electron–electron and electron–ion interaction, which together describe the interatomic forces. An example for the excellent agreement of calculations and experiment is given in Fig. 15b for GaN. (See also Sect. 2 in ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals.") Current ab-initio calculations of lattice dynamics are often based on a supercell approach as reported by Souvatzis et al. (2008) and applied by Lindsay et al. (2013). Latest developments employ a stochastic implementation of this ansatz (Monacelli et al. 2018).

3.3 Phonons in Superlattices, at Surfaces, in Organic Semiconductors, and in Amorphous Semiconductors

3.3.1 Phonons in Superlattices

A remarkable result of the periodic structure of a superlattice of alternating layers A and B with a superlattice constant l (=thickness of layers $A + B$) is its reduced dispersion relation with a boundary at $\pm \pi/l$ for the first superlattice Brillouin zone (*mini-zone* – see Sect. 2.1 in ▶ Chap. 3, "The Structure of Semiconductors" – and superlattices in Sect. 3.1.2 in ▶ Chap. 8,"Bands and Bandgaps in Solids") rather than at $\pm \pi/a$ for the bulk lattice. Consequently, the phonon spectrum is *folded* at the mini-zone boundary rather than continued to the main Brillouin zone boundary

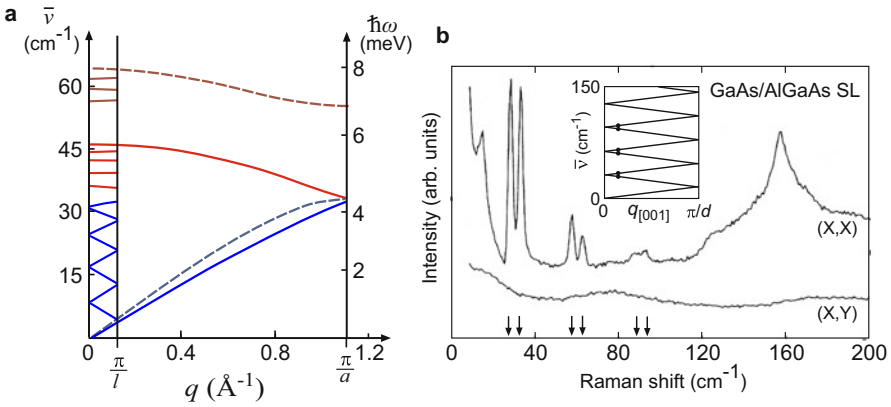


Fig. 18 (**a**) Phonon dispersion curves for longitudinal modes, calculated for a linear chain in GaAs (*solid curve*) and for AlAs (*dashed curve*), and shown folded for a mini-zone of a symmetrical superlattice with a periodicity of $l = 6a$. (**b**) Raman spectra of a (42 Å GaAs/ 8 Å $Al_{0.3}Ga_{0.7}As$) superlattice. *Vertical arrows* indicate folded-phonon doublets expected from the dispersion in the inset (After Colvard et al. 1985)

(Fig. 18). Since the alternating materials in a superlattice have different atomic masses and force constants, there is, as in an *AB* compound, a gap between each of the branches of the folded dispersion curve.

In comparison with the dispersion curve of an alloyed bulk lattice, several folded branches evolve from the *acoustic branch*. The dispersion entering the backfolding is an average of those of the two superlattice materials, weighted by their relative layer thickness; only the first folded branch is referred to as an acoustic branch of the superlattice. In the optical branches, two alternatives may appear. If the phonon energies of the two materials overlap, folded optical branches similar to the acoustic phonons appear. Otherwise, an optical oscillation in one material does not find a resonance in the other material and gets localized in the respective layer. The phonon dispersion spectrum is thus modified to a large degree from the bulk phonon behavior, as is expected considering the coupling of atom groups and the reflection at the boundaries between the superlattice layers.

The multibranch structure can be directly observed from phonon scattering experiments (see Sect. 3.3 in ▶ Chap. 11, "Photon–Phonon Interaction," and Colvard (1987)). For reviews, see Klein (1986) and Ruf (1998).

3.3.2 Surface Phonons

Atoms at and near the surface are bound with lesser strength because of missing external neighbors. Consequently, there is a modification of the phonon spectrum near the surface that can be excited through surface waves (see below). These surface phonons have a lower frequency caused by a lower force constant than the optical phonons in the crystal bulk and can be observed in the forbidden frequency range

(Wallis 1994). Surface phonons are localized at the surface, i.e., the vibrational amplitude decays (basically exponentially) away from the surface.[20]

A qualitative understanding of a phonon state in the forbidden frequency range can be obtained from the diatomic lattice discussed in Sect. 3.1.3. The acoustic and optical branches illustrated in Fig. 9 extend between $0 < \omega_{ac} < \omega_{ac,max}$ and $\omega_{opt,min} < \omega_{opt} < \omega_{opt,max}$, respectively, given by

$$\begin{aligned} \omega_{ac,max} &= \sqrt{2\beta/M_2}, \omega_{opt,min} = \sqrt{2\beta/M_1}, \\ \omega_{opt,max} &= \sqrt{2\beta/(M_1+M_2)/(M_1M_2)} \end{aligned} \tag{64}$$

For "bulk" modes there is a forbidden band (stop band) between $\omega_{ac,max}$ and $\omega_{opt,min}$, and another one above $\omega_{opt,max}$. Surface atoms are less bound, expressed by a smaller effective spring constant β. Thereby the vibration frequency of the lighter atom with mass M_1 drops from the optical branch into the stop band. If we designate the displacement of the "surface" atom by u_0, we have the equation of motion

$$M_1^2 \frac{d^2u_0}{dx^2} = \beta(u_1 - u_0). \tag{65}$$

For all other atoms the bulk equations apply. The solution (with no periodic boundary condition) yields a new eigenfrequency in the stop band

$$\omega_{stop} = \sqrt{\beta/(M_1+M_2)/(M_1M_2)}. \tag{66}$$

Since in our diatomic chain only a *single* spring ("bond") is attached to the "surface" atom (instead of two) there is no factor of two in the root of Eq. 66. The surface modes of the *heavy* atom with mass M_2 are degenerate with the bulk modes.

The spectrum of surface phonons is illustrated in Fig. 19. The energy $\hbar\omega$ of surface phonons depends uniquely on the wavevector component $\mathbf{q}_\parallel$ which lies in the surface plane. The spectrum of bulk phonons is indicated as a continuum in the presentation $\hbar\omega(\mathbf{q}_\parallel)$, since the component of the wavevector perpendicular to the surface is arbitrary.

Surface phonons can only persist if no bulk modes of the same symmetry exist at the same energy and wavevector component $\mathbf{q}_\parallel$. Another kind of surface vibrations are surface-phonon resonances; these resonances are bulk-phonon modes with a large amplitude at the surface.

In the acoustic branch, elastic surface waves propagate along the surface within a thin layer with a thickness approximately equals the wavelength (*Rayleigh waves*).

[20]The same applies for phonons localized at an interface. Both a surface and an interface break the three-dimensional translation symmetry of a solid, giving rise to such localization of the solutions of the equation of motion.

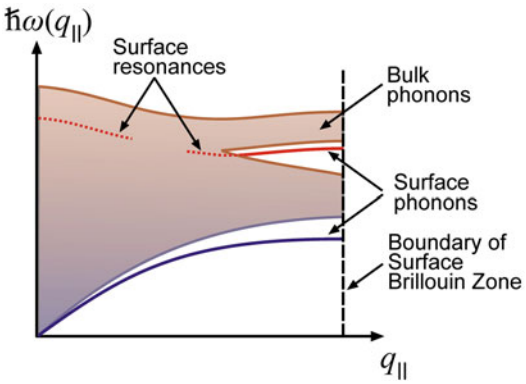


Fig. 19 Spectrum of surface phonons (*black curves*). The *shaded area* represents the continuum of bulk modes, and *dotted lines* are surface resonances

The frequency of this surface wave can be derived from Maxwell's equation, yielding

$$\omega_s = \omega_{\mathrm{TO}} \sqrt{\frac{\varepsilon_{\mathrm{st}} + 1}{\varepsilon_{\mathrm{opt}} + 1}} \tag{67}$$

where $\varepsilon_{\mathrm{st}}$ is the static and $\varepsilon_{\mathrm{opt}}$ is the optical dielectric constant (see Sect. 1.3 in ▶ Chap. 10, "Interaction of Light with Solids"). This frequency lies between the longitudinal and transverse branches (compare with Eq. 50 in ▶ Chap. 11, "Photon–Phonon Interaction") and can be observed in crystal plates of sufficient width (for cubic semiconductors, see Ludwig 1974).

3.3.3 Phonons in Organic Semiconductors

There are two different kinds of vibrational modes in organic crystals: intramolecular vibrations (*internal modes*) and intermolecular vibrations (*external* modes). Both have a strong effect on the electrical and optical properties of organic semiconductors.

Internal modes represent vibrations of individual molecules. Atoms of the molecule oscillate about their equilibrium position; the center of gravity of the molecule remains at rest. A free molecule with N atoms has $3N - 6$ internal eigenmodes. Due to the weak intermolecular interaction of organic crystals (Sect. 3.3 in ▶ Chap. 2, "Crystal Bonding"), the frequencies of internal modes do hardly differ from the eigenmodes of the free molecule. Furthermore, internal modes show only weak dispersion. Since intramolecular forces are strong, internal eigenmodes have usually a high energy and their thermal excitation is weak at room temperature. In some cases, the weak intermolecular interaction can be observed in the spectrum of soft low-frequency internal eigenmodes: the coupling to neighboring molecules leads to a – generally small – splitting of the resonance frequency (typically below 0.1 meV). This effect is analogous to the Davydov

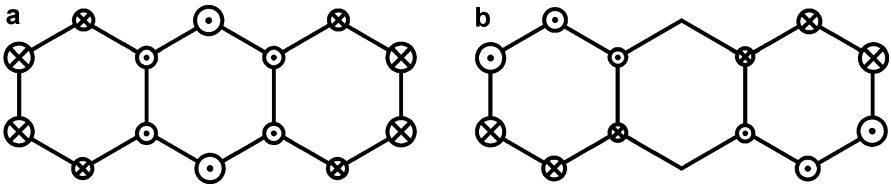


Fig. 20 Top view on a planar anthracene molecule showing two of the 66 internal modes: (**a**) a butterfly distortion and (**b**) a twisting distortion. The $\odot$ and $\otimes$ symbols indicate an atomic movement out of and into the drawing plane, and their size indicates the magnitude of the displacements; the two central carbon atoms in (**b**) remain at rest

splitting observed in excitonic spectra of organic molecules with $Z > 1$ molecules per unit cell (Davydov 1964).

The anthracene molecule consists of 14 C atoms and 10 H atoms (Fig. 15b in ▶ Chap. 2, "Crystal Bonding"), yielding 66 internal modes. Two out-of-plane modes with a low energy are illustrated in Fig. 20: a butterfly distortion (symmetry B_{3u}) and a twisting distortion (symmetry A_u); half a period after this snapshot, all amplitudes changed sign.

External modes are coupled oscillations of molecules, i.e., phonons like those found in inorganic solids. In addition to translational motions, molecules also perform oscillating rotations, so-called *librations*, about their equilibrium position. In external modes, the molecules can be assumed rigid. An organic crystal hence has $f = 6$ degrees of freedom (three axes of translation and three of rotation). A crystal with Z molecules per unit cell (Sect. 1.5 in ▶ Chap. 3, "The Structure of Semiconductors") has consequently $f \times Z$ external eigenmodes and therefore $f \times Z$ phonon branches.

The restoring forces of the external modes are given by the weak intermolecular interactions, i.e., basically by van der Waals forces. Since these interactions are much weaker than the intramolecular covalent interactions, frequencies of external modes are usually much lower than those of internal modes; this holds particularly for organic crystals composed of light molecules. The low phonon energies of organic crystals lead to a high occupation of phonon states at room temperature; consequently, phonons strongly affect interactions with carriers.

The phonon dispersion $\omega(\mathbf{q})$ of external modes is similar to that of inorganic crystals; in particular, there are acoustic branches, where $\hbar\omega(\mathbf{q} = 0) \to 0$ applies, and optical branches with a finite energy $\hbar\omega(\mathbf{q} = 0)$. Phonon energies of organic crystals are generally smaller than those of inorganic crystals.

Phonon dispersion curves for a deuterated d_{10}-anthracene[21] crystal are shown in Fig. 21. Anthracene has $Z = 2$ molecules in a unit cell (Table 3 in ▶ Chap. 3, "The Structure of Semiconductors"), yielding 12 phonon branches: 3 acoustic and 9 optical branches. There exist one symmetric and one antisymmetric set of modes with

[21] In d_{10}-anthracene, all 10 hydrogen atoms are substituted by deuterium atoms to increase the cross section for coherent inelastic neutron scattering.

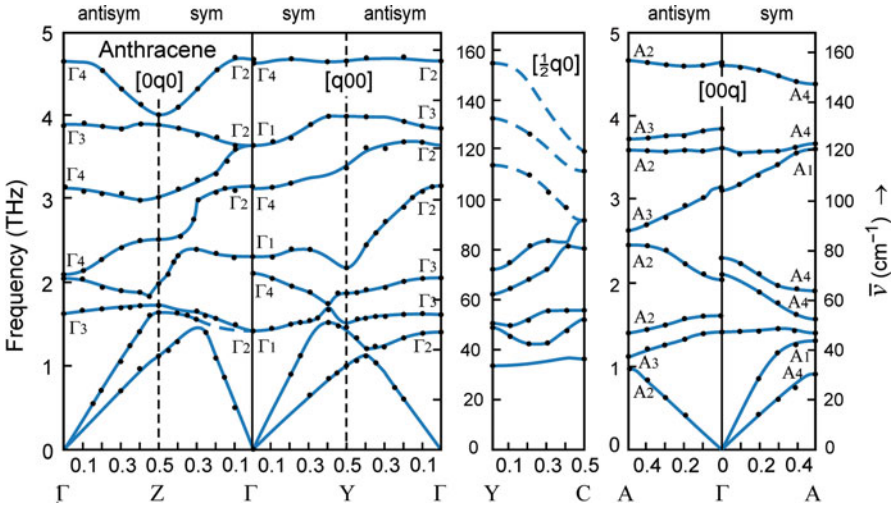


Fig. 21 Phonon dispersion $\omega(\mathbf{q})$ of the 12 external and the 4 low-energy internal modes of a deuterated anthracene crystal measured at 12 K using inelastic neutron scattering. In the presented extended-zone scheme, branches must not cross. The given directions of the phonon wavevector **q** refer to the monoclinic Brillouin zone shown in Fig. 8 in ▶ Chap. 3, "The Structure of Semiconductors." (After Dorner et al. 1982)

six modes each for every principal symmetry direction, originating from a twofold screw axis along the [0q0] direction; the additional two branches in each set found in the figure stem from internal modes. The sets are presented separately in the extended-zone scheme of Fig. 21, so as to avoid branch crossing (Dorner et al. 1982). The assignments of the modes are complicated by mixing for $q \neq 0$. Only at the Γ point at $q = 0$ all modes are distinguished by their symmetry: Γ_1 and Γ_3 contain librations, and Γ_2 and Γ_4 contain the external translations; still all vibrations may comprise contributions from internal modes. Model calculations reported by Dorner et al. (1972) indicated weak contributions of internal modes to dispersion curves with $\hbar\omega > 2$ THz and strong contributions above 3 THz; the uppermost branches have almost pure internal character composed of the modes illustrated in Fig. 20.

3.3.4 Phonons in Amorphous Semiconductors

The vibrational spectrum of amorphous semiconductors is an important means for obtaining information about their structure. However, the description used previously in an $\hbar\omega$ (q) representation is no longer appropriate because of nonperiodicity: q is not a good quantum number for phonons in a glass. Long-wavelength acoustic phonons, on the other hand, behave much like phonons in crystals. Instead, *local mode phonons* are used for exploring the atomic structure. From the phonon spectrum, we obtain valuable information about the specific bonding character, topology, and local symmetry, thus describing the phonon spectrum as a *local* (vibrational) *density of states* (LDOS).

The different local vibrational modes can be calculated using a molecular cluster model (Lucovsky et al. 1983), wherein the frequencies of specific modes of atomic vibrations are calculated. We must distinguish alloy atoms, e.g., H in an a-Si host, or foreign atoms (impurities) with a lower density so that impurity–impurity interaction can be neglected in the latter material. In many respects, these local modes are similar to those in a crystalline host; however, the LDOS features in a glass are broader since the local environment is deformed to a larger extent and variety.

3.3.5 Measurement of Phonon Spectra

There are numerous methods for obtaining information about the phonon spectra. Phonons near the center of the Brillouin zone can be excited by optical techniques due to the small photon momentum. The measurements can be divided into methods which reveal the structure of optical modes and those that deal with acoustic modes. The optical mode can be detected by optical absorption or reflection measurements. Its spectral distribution yields the most direct information about these lattice oscillations. In addition, optical scattering experiments provide information about the most abundant optical and acoustic phonons. Both techniques will be discussed in detail in Sects. 2 and 3 in ▶ Chap. 11, "Photon–Phonon Interaction," respectively.

Because of the large phonon momentum near the boundaries of the zone, only heavy particles can interact sufficiently. As such, neutrons are almost ideal probes for investigating the entire phonon spectrum.[22] Slow neutrons (mass M_n) have an energy $(3/2)kT$ and a momentum (wavenumber)

$$q = \frac{M_n v}{\hbar} = \frac{\sqrt{3M_n kT}}{\hbar} = 2.5\sqrt{T} \times 10^7 \left(\mathrm{cm}^{-1}\right), \tag{68}$$

which, at low temperatures, is comparable to the total extent of the Brillouin zone. In scattering with phonons while conserving total energy and momentum, such neutrons are therefore able to probe the entire phonon spectrum. *Slow* neutrons can be obtained by cooling neutrons from a nuclear reactor using a moderating material with a large scattering cross-section, such as water or liquid hydrogen. Monoenergetic neutrons are obtained by diffraction from a single crystal. A collimated beam of these monochromatic neutrons is scattered by phonons within a semiconductor. From the angular distribution of the energy (i.e., the incident neutron energy plus phonon energy), we then obtain a complete $\omega(\mathbf{q})$ dispersion curve of all active branches of the phonon spectrum (see reviews by Dolling (1974), Bührer and Iqbal (1984)). High-energy branches are inactive at temperatures much below the Debye temperature of the semiconductor.

[22] Except for compounds in which thermal neutrons react strongly with the nucleus of one of the elements, e.g., ^{10}B or ^{113}Cd.

4 Summary

Elastic deformations of solids are described by elastic stiffness constants accounting for the interatomic forces. The symmetry of crystals reduces their number in the harmonic approximation from 21 (triclinic) to 3 (cubic). Coupled acoustic oscillations of the atoms lead to different modes of sound waves. The excitation of each mode is quantized in multiples of elementary excitations called phonons. The number of branches in the phonon dispersion is determined by the number of atoms p in the basis; there are three acoustic branches (one longitudinal, two transversal) and $3p - 3$ optical branches.

Phonons are responsible for the thermal properties and cause thermalization of a wide variety of events when interacting with other particles, such as electrons and holes. They supply the necessary damping (i.e., dissipation of energy into the thermal reservoir), thereby providing for optical absorption, electric energy dissipation, and solidification of liquids, to mention just a few. The phonon spectrum is part of the fingerprint of the specific semiconductor. In mixed crystals (random alloys), the phonon frequencies depend on the composition parameter; one-mode and two-mode behaviors occur. In superlattices, the phonon dispersion is folded at the boundary of the mini-zone, if the phonon energies in the constituent materials overlap; otherwise, phonon modes are localized in one material.

The phonon spectrum can be divided into an intrinsic spectrum determined by the bulk of the semiconductor and additional features due to surfaces, impurities, and other defects. In organic semiconductors, internal modes of the molecules and librations, i.e., oscillating rotations of the molecules, occur in addition to translational modes. Phonon spectra can be determined by optical absorption or by scattering experiments and can be calculated, e.g., by ab initio methods from basic principles. From the dispersion relation, we can obtain directly the density of states of the different modes of oscillation, the knowledge of which is essential for the identification of the energy density in the most important phonon branches.

References

Adachi S (1985) GaAs, AlAs, and $Al_xGa_{1-x}As$: material parameters for use in research and device applications. J Appl Phys 58:R1

Adachi S (2005) Properties of group-IV, III-V and II-VI semiconductors. Wiley, Chichester

Akhieser A (1939) On the absorption of sound in solids. J Phys (USSR) 1:277

Ayers JE (2007) Heteroepitaxy of semiconductors: theory, growth, and characterization. CRC, Boca Raton

Bogdanov SV, Sapozhnikov VK (2009) Acoustooptic methods for measuring sound velocity. Optoelectron Instrument Proc 45:554

Barker AS Jr, Sievers AJ (1975) Optical studies of the vibrational properties of disordered solids. Rev Mod Phys 47:S1

Bilz H, Kress W (1979) Phonon dispersion relations in insulators. Springer, Berlin

Brown FC (1967) The physics of solids. W.A. Benjamin, New York

Brugger K (1964) Thermodynamic definition of higher order elastic coefficients. Phys Rev 133:1611

Bruner LJ, Keyes RW (1961) Electronic effect in the elastic constants of Germanium. Phys Rev Lett 7:55

Bührer W, Iqbal Z (1984) Neutron scattering studies of structural phase transitions. In: Iqbal Z, Owens FJ (eds) Vibrational spectroscopy of phase transitions. Academic Press, New York

Cauchy A-L (1828) Lehrbuch der Algebraischen Analysis (trans: Huzler CLB, Textbook on algebraic analysis, in German). Bornträger, Königsberg

Colvard C (1987) Raman scattering characterization of quantum wells and superlattices. Proc SPIE 794:209

Colvard C, Gant TA, Klein MV, Merlin R, Fisher R, Morkoç H, Gossard LA (1985) Folded acoustic and quantized optic phonons in (GaAl)As superlattices. Phys Rev B 31:2080

Davydov AS (1964) The theory of molecular excitons. Sov Phys Usp (English trans) 7:145

Davydov VY, Kitaev YE, Goncharuk IN, Smirnov AN, Graul J, Semchinova O, Uffmann D, Smirnov MB, Mirgorodsky AP, Evarestov RA (1998) Phonon dispersion and Raman scattering in hexagonal GaN and AlN. Phys Rev B 58:12899

Deych LI, Yamilov A, Lisyansky AA (2000) Concept of local polaritons and optical properties of mixed polar crystals. Phys Rev B 62:6301

Dolling G (1974) In: Norton GR, Maradudin AA (eds) Dynamic properties of solids. North Holland, Amsterdam

Dorner B, Bokhenkov EL, Chaplot SL, Kalus J, Natkaniec I, Pawley GS, Schmelzer U, Sheka EF (1982) The 12 external and the 4 lowest internal phonon dispersion branches in d_{10}-anthracene at 12 K. J Phys C Solid State Phys 15:2353

Dunstan DJ (1997) Strain and strain relaxation in semiconductors. J Mater Sci: Materials in Electronics 8:337

Garber JA, Granato AV (1975) Theory of the temperature dependence of second-order elastic constants in cubic materials. Phys Rev B 11:3990; Fourth-order elastic constants and the temperature dependence of second-order elastic constants in cubic materials. Phys Rev B 11:3998

Guyer RA (1966) Acoustic attenuation in dielectric solids. Phys Rev 148:789

Hall JJ (1967) Electronic effects in the elastic constants of *n*-type silicon. Phys Rev 161:756

Hiki Y (1981) Higher order elastic constants of solids. Ann Rev Mater Sci 11:51

Hiki Y, Mukai K (1973) Ultrasonic three-phonon process in copper crystal. J Phys Soc Jpn 34:454

Jasval SS (1975) Microscopic theory and deformation-dipole model of lattice dynamics. Phys Rev Lett 35:1600

Jasval SS (1977) In: Balkanski M (ed) Lattice dynamics. Flammarion, Paris

Joharapurkar DN, Breazeale MA (1990) Nonlinearity parameter, nonlinearity constant, and frequency dependence of ultrasonic attenuation in GaAs. J Appl Phys 67:76

Joharapurkar D, Gerlich D, Breazealec MA (1992) Temperature dependence of elastic nonlinearities in single-crystal gallium arsenide. J Appl Phys 72:2202

Joos G (1945) Lehrbuch der Theoretischen Physik (Textbook on theoretical physics, in German). Akademie Verlag Gesellschaft, Leipzig

Keating PN (1966) Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure. Phys Rev 145:637

Kim DS, Hellman O, Shulumba N, Saunders CN, Lin JYY, Smith HL, Herriman JE, Niedziela JL, Abernathy DL, Li CW, Fultz B (2020) Temperature-dependent phonon lifetimes and thermal conductivity of silicon by inelastic neutron scattering and ab initio calculations. Phys Rev B 102:174311

Klein MV (1986) Phonons in semiconductor superlattices. IEEE J Quantum Electron QE 22:1760

Landau LD, Rumer G (1937) Absorption of sound in solids. Phys Z Sowjetunion 11:18

Lang JM Jr, Gupta YM (2011) Experimental determination of third-order elastic constants of diamond. Phys Rev Lett 106:125502

Lindsay L, Broido DA, Reinecke TL (2013) *Ab initio* thermal transport in compound semiconductors. Phys Rev B 87:165201

Łopuszyński M, Majewski JA (2007) *Ab initio* calculations of third-order elastic constants and related properties for selected semiconductors. Phys Rev B 76:045202

Lucovsky G, Wong CK, Pollard WB (1983) Vibrational properties of glasses: intermediate range order. J Non Cryst Solid 59–60:839
Ludwig WEW (1974) Theory of surface phonons in diamond- and zincblende-lattices. Jpn J Appl Phys 2(Supplement 2-2):879
Martienssen W, Warlimont H (2005) Handbook of condensed matter and materials data. Springer, Berlin/Heidelberg/New York
Martinez G (1980) Optical properties of semiconductors under pressure. In: Moss TS, Balkanski M (eds) Handbook on semiconductors, vol 2: Optical properties of solids. North Holland, Amsterdam, pp 181–222
Mason WP, Bateman TB (1964) Ultrasonic wave propagation in doped *n*-germanium and *p*-silicon. Phys Rev 134:A1387
McSkimmin HJ, Andreatch P Jr (1964) Measurement of third-order moduli of silicon and germanium. J Appl Phys 35:3312
McSkimmin HJ, Andreatch P Jr (1967) Third-order elastic moduli of gallium arsenide. J Appl Phys 38:2610
Mitra SS, Massa ME (1982) Lattice vibrations in semiconductors. In: Moss TS, Paul W (eds) Handbook on semiconductors, vol 1: Band theory and transport properties. North Holland, Amsterdam, pp 81–192
Monacelli L, Errea I, Calandra M, Mauri F (2018) Pressure and stress tensor of complex anharmonic crystals within the stochastic self-consistent harmonic approximation. Phys Rev B 98:024106
Musgrave MJP (1970) Crystal acoustics. Holden-Day, San Francisco
Newman RC (1993) Vibrational mode spectroscopy of defects in III/V compounds. Semicond Semimet 38:117
Nye JF (1972) Physical properties of crystals – their representation by tensors and matrices. Oxford University Press, London. First published 1957, reprinted from corrected sheets
Patel C, Parker TJ, Jamshidi H, Sherman W (1984) Phonon frequencies in GaAs. Phys Stat Sol (B) 122:461
Raja VS, Reddy PJ (1976) Measurement of third-order elastic constants of GaSb. Phys Lett A 56:215
Rajagopal AK, Srinivasan R (1960) Lattice vibrations and specific heat of zinc blende. Z Phys 158:471
Ruf T (1998) Phonon raman scattering in semiconductors, quantum wells and superlattices. Springer, Berlin
Ruf T, Serrano J, Cardona M, Pavone P, Pabst M, Krisch M, D'Astuto M, Suski T, Grzegory I, Leszczynski M (2001) Phonon dispersion curves in wurtzite-structure GaN determined by inelastic X-Ray scattering. Phys Rev Lett 86:906
Sarasamak K, Limpijumnong S, Lambrecht WRL (2010) Pressure-dependent elastic constants and sound velocities of wurtzite SiC, GaN, InN, ZnO, and CdSe, and their relation to the high-pressure phase transition: A first-principles study. Phys Rev B 82:035201
Sahasrabudhe GG, Lambade SD (1998) Study of elastic and acoustic nonlinearities in solids at room temperature. J Phys Chem Sol 59:789
Souvatzis P, Eriksson O, Katsnelson MI, Rudin SP (2008) Entropy driven stabilization of energetically unstable crystal structures explained from first principles theory. Phys Rev Lett 100:095901
Sundararaja V, Reddy PJ (1977) Third-order elastic constants of indium antimonide at 300° anf 80° K. Solid State Commun 21:701
Tanner DSP, Caro MA, Schulz S, O'Reilly EP (2019) Hybrid functional study of nonlinear elasticity and internal strain in zinc-blende III-V materials. Phys Rev Materials 3:013604
Truell R, Elbaum C, Crick BB (1969) Ultrasonic methods in solid state physics. Academic Press, New York
Turbino R, Piseri L, Zerbi G (1972) Lattice dynamics and spectroscopic properties by a valence force potential of diamondlike crystals: Si, Ge, and Sn. J Chem Phys 56:1022

Vurgaftman I, Meyer JR (2003) Band parameters for III-V nitrogen-containing semiconductors. J Appl Phys 94:3675
Vurgaftman I, Meyer JR, Ram-Mohan LR (2001) Band parameters for III-V compound semiconductors and their alloys. J Appl Phys 89:5815
Wallis RF (1994) Surface phonons: theoretical developments. Surf Sci 299/300:612
Weber W (1977) Adiabatic bond charge model for the phonons in diamond, Si, Ge, and α-Sn. Phys Rev B 15:4789
Weißmantel C, Hamann C (1979) Grundlagen der Festkörperphysik (Fundamentals of Solid State Physics, in German). Springer, Berlin/Heidelberg/New York
Woodruff TO, Ehrenreich H (1961) Absorption of sound in insulators. Phys Rev 123:1553
Yoğurtçu YK, Miller AJ, Saunders GA (1981) Pressure dependence of elastic behaviour and force constants of GaP. J Phys Chem Sol 42:49
Zener C (1947) A defense of the cauchy relations. Phys Rev 71:323

Phonon-Induced Thermal Properties

5

Karl W. Böer and Udo W. Pohl

Contents

1 Heat Capacity 158
1.1 Classical Models 159
1.2 General Phonon-Distribution Function and Phase Changes 162
1.3 Specific Heat of Amorphous Semiconductors 164
2 Thermal Expansion 165
2.1 Phenomenological Description 165
2.2 Lattice Dynamic Consideration 166
3 Thermal Conductivity 170
3.1 Diffusive Thermal Transport 170
3.2 Phonon-Scattering Mechanisms 172
3.3 Phonon Scattering in Crystalline and Amorphous Semiconductors 177
4 Thermal Transport in Nanostructures 179
4.1 Scattering at Crystal Boundaries 179
4.2 Mean Free Path 180
4.3 Thermal Transport Parallel to Boundaries 182
4.4 Thermal Transport Normal to Boundaries 183
5 Summary 185
References 186

Abstract

Phonons are responsible for all thermal properties of a solid, such as its heat content and transport. The anharmonic part of lattice oscillations causes thermal expansion. All these are integral contributions of the phonon spectrum; only at low temperatures, where part of the spectrum can be frozen-out, do they become

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_5

partially spectrum selective. Thermal conductivity is determined by various mechanisms of phonon scattering. Most important are phonon-phonon and disordered-induced scattering. In nanostructures scattering at boundaries is dominant.

Keywords

Debye specific heat · Dulong-Petit law · Einstein specific heat · Heat capacity · Phonon scattering · Specific heat · Thermal conductivity · Thermal expansion · Thermal transport

1 Heat Capacity

There are three macroscopic thermal properties originating from phonons:

- Heat capacity
- Thermal expansion
- Thermal conductivity

neglecting electronic contributions here. Since these are easily observable effects, they were studied early, and rather general, often semiempirical descriptions were given. These properties are of interest with regard to the interaction of electrons with phonons and the dissipation of energy, e.g., in optical excitation. The first set of processes is discussed in ▶ Chaps. 23, "Carrier Scattering at Low Electric Fields," and ▶ 24, "Carrier Scattering at High Electric Fields." The latter processes are kinetic in nature and are discussed in ▶ Chap. 32, "Dynamic Processes." In this chapter we will provide the essential basic information.

Heat capacity is a measure of the content of thermal energy, which is stored in all active oscillations of atoms in the solid. The *specific heat* (or *specific heat capacity*) expresses the change of the total energy U of all phonons in the solid for a given change of temperature T; it is hence defined, for constant volume[1], as

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V . \tag{2}$$

[1] Although the specific heat is measured more easily for constant pressure C_P, the difference between C_V and C_P is very small for solids and is given by

$$C_P - C_V = \frac{9\alpha^2 T V_m}{\kappa} \tag{1}$$

where α is the thermal expansion coefficient, κ is the isothermal compressibility, and V_m is the molar volume.

In a monatomic semiconductor, each oscillating lattice atom represents a basically harmonic oscillator with an average energy of $kT/2$ per degree of freedom. Thus, with three degrees of freedom per atom, times two (for the kinetic and potential energies of the oscillator), a solid with N_A atoms per mole yields a total energy $U = 3N_A kT$. From Eq. 2 we then obtain for the specific heat the *Dulong–Petit law* (1819):

$$C_V = 3N_A k = 3R \cong 25 \ \mathrm{J\,mol^{-1}K^{-1}} \tag{3}$$

where R is the gas constant.[2] This equation is approximately fulfilled at sufficiently high temperatures when all modes of oscillation are excited.

1.1 Classical Models

1.1.1 Einstein Model

The Dulong–Petit law Eq. 3 states a *constant* specific heat for all solids and does not account for the generally observed decrease of C_V at lower temperatures. At low temperatures the phonons *freeze out*, i.e., they can no longer be thermally excited. A realistic description of the total energy U of all phonons therefore considers the energy of all phonon modes $\hbar\omega_i$ and their respective occupation in thermal equilibrium. Phonons are *bosons*, i.e., particles with integer spin, and follow Bose–Einstein statistics. The Einstein model assumes one mode of oscillation with one corresponding eigenfrequency ω_0 only. At sufficiently high temperatures, most of the oscillatory energy is indeed present in one kind of phonon, the transverse optical phonons, which – due to their flat dispersion curves – have nearly the same frequency in the entire Brillouin zone. Therefore, there is some justification for identifying the oscillatory lattice energy with $U_0 = 3N_A\hbar\omega_0$ for a one-atomic semiconductor with N_A atoms and three oscillators per atom (in the three lattice coordinates). The occupation of the states is given by multiplication of U_0 with the Bose–Einstein distribution $f_{BE} = (\exp(\hbar\omega_0/(kT)) - 1)^{-1}$, yielding

$$U = U_0 f_{BE} = \frac{3N_A\,\hbar\omega_0}{\exp(\hbar\omega_0/(kT)) - 1} \tag{4}$$

and therefore from Eqs. 2 and 3

$$C_V = 3R\left(\frac{\hbar\omega_0}{kT}\right)^2 \frac{\exp\left(\frac{\hbar\omega_0}{kT}\right)}{\left(\exp\left(\frac{\hbar\omega_0}{kT}\right) - 1\right)^2}, \tag{5}$$

which represents the Einstein model of the specific heat (Einstein 1907). Equation 5 shows the observed decrease of C_V with decreasing temperature as part of a quantum-mechanical phenomenon, the successive freeze-out of phonons, rather

[2] The specific heat is given here in Ws/(mol K) and should be distinguished from the values often given in tables in units of Ws/(cm^{3}K) or Ws/(gK).

than a continuous decrease of oscillatory amplitudes. The Einstein model describes the trend of $C_V(T)$ quite well for values exceeding about 3/2 R; the simplifying assumption of a *single* phonon mode leads, however, to an exponential dependence of C_V for $T \to 0$ that is experimentally not observed.

1.1.2 Debye Model

Debye (1912) approached the problem from a different point of view, having recognized that *various* acoustic oscillatory modes are dominant at low temperature. With a *distribution function* $g(\omega)$ of these modes, the phonons are distributed over these modes according to the Bose–Einstein statistics. Neglecting optical modes, Debye obtained for the thermal energy content

$$U = \int_0^{\omega_D} \hbar\omega\, g(\omega) f_{BE}(\hbar\omega)\, d\omega = \int_0^{\omega_D} \frac{\hbar\omega\, g(\omega)\, d\omega}{\exp\left(\frac{\hbar\omega}{kT}\right) - 1}. \tag{6}$$

The integration requires an upper limit ω_D of phonon frequencies, given by the fact that the total number of modes cannot exceed $3N$:

$$3N = \int_0^{\omega_D} g(\omega)\, d\omega. \tag{7}$$

Permitting all standing wave-type oscillations (or all running waves fulfilling cyclic boundary conditions), we obtain the distribution function by mode counting (see Sect. 3.2 in ▶ Chap. 4, "Elasticity and Phonons"), similar to that done for electron waves:

$$g(\omega)\, d\omega = \frac{V}{8\pi^3} 4\pi q^2 dq. \tag{8}$$

With the group velocity $v_g = \partial\omega/\partial q$, and assuming a *linear dispersion relation* ($\omega = v_g q$) for a macroscopic continuous solid, we obtain

$$g(\omega)\, d\omega = \frac{V}{2\pi^2} \frac{\omega^2}{v_g^3} d\omega = 9N \frac{\omega^2}{\omega_D^3} d\omega; \tag{9}$$

the limiting *Debye frequency* ω_D can be replaced by the so-called *Debye temperature* θ, according to

$$\hbar\omega_D = v_s \sqrt[3]{\frac{6\pi^2\hbar^2 N}{V}} \equiv k\theta. \tag{10}$$

This yields the well-known Debye expression for the specific heat [from Eq. 2 using Eqs. 6, 9, and 10]:

$$C_V = 9R\left(\frac{T}{\theta}\right)^3 \int_0^{\theta/T} \frac{x^4 \exp x}{(\exp x - 1)^2} dx, \tag{11}$$

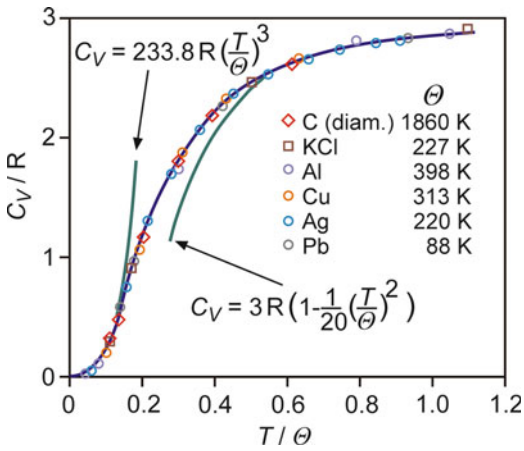


Fig. 1 Molar specific heat of crystals as a function of the normalized temperature and the two branches of the approximation (Eq. 12) (After Weißmantel and Hamann 1995)

Table 1 Debye temperatures for various solids at 300 K (largely after Martienssen and Warlimont 2005)

Material	θ (K)	Material	θ (K)	Material	θ (K)
$C_{Diamond}$	1860	ZnO	440	AlAs	417
Si	645	ZnS	351	AlSb	292
Ge	374	ZnSe	339	GaN	~600
α-Sn	230	ZnTe	180	GaP	445
LiF	730	CdO	255	GaAs	344
LiCl	422	CdS	219	GaSb	266
NaF	491	CdSe	182	InN	~660
NaCl	321	CdTe	140	InP	321
KF	336	AlN	~1150	InAs	247
KCl	233	AlP	588	InSb	206

which can be approximated by the expansion of the exponent (see Joos 1945 for detail) as

$$C_V = \begin{cases} 233.78R\left(\frac{T}{\theta}\right)^3 & \text{for } T \ll \theta \\ 3R\left(1 - 0.05\left(\frac{T}{\theta}\right)^2\right) & \text{for } T \gtrsim 0.5\theta. \end{cases} \tag{12}$$

A better approximation includes, in addition to the cubic term at low T, a linear component if free electron contributions are significant.

A quite good fit with the experiment can be obtained for some crystals by choosing the appropriate θ – see Fig. 1. The Debye temperatures for a variety of solids are listed in Table 1. For non-monatomic semiconductors (e.g., diatomic), the agreement is usually less satisfactory. We then evaluate the measured $C_V(T)$ with

Eq. 11 but permit a temperature-dependent Debye temperature. The deviation from $\theta = \text{const}$ is a measure for the departure of the actual phonon spectrum from the assumed distribution – see Fig. 3.

1.1.3 Validity Range of the Approximations

The specific heat is reasonably well described by the Dulong–Petit law at temperatures well above the Debye temperature. With decreasing temperature, the freezing-out of the dominant phonon mode makes the major contribution. Here the Einstein approximation is useful, while at very low temperatures (where $C_V < 3/2\ R$), contributions of the acoustic phonon branch dominate and the Debye approximation fits the experimental results remarkably well. However, the rapid change of the Debye temperature with the actual temperature, even at low temperatures (see Fig. 3), indicates the need for a more detailed analysis which will be discussed in the following section.

1.2 General Phonon-Distribution Function and Phase Changes

There are several branches of the phonon spectrum. In the simplest case, following Eq. 9, these can be taken into account for the acoustic phonons by setting $1/v_g^3 = \frac{1}{3}\left(1/v_l^3 + 2/v_t^3\right)$ while still assuming a parabolic distribution. However, the actual phonon-distribution function is much more complex; it can be measured using inelastic neutron scattering (Kim et al. 2015) or x-ray scattering (Holt et al. 1999). Figure 2 shows the difference between the simple quadratic distribution function and one that is obtained from neutron scattering. We observe a rough agreement for Cu, but an only poor description for Si with its complex phonon spectrum. Introducing this experimentally obtained distribution function into Eq. 6, the derived specific heat yields a temperature-dependent Debye temperature. Examples of such dependence for some II–VI semiconductors are

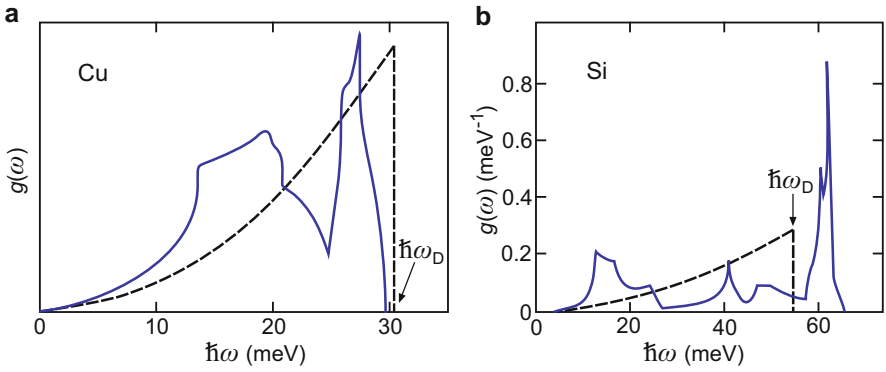


Fig. 2 Density of states for phonons as obtained from the Debye approximation (*dashed curves*) and experimental data for (**a**) copper (After Svenson et al. 1967) and (**b**) silicon obtained from neutron scattering (*solid curves*), scaled to yield equal area under both curves

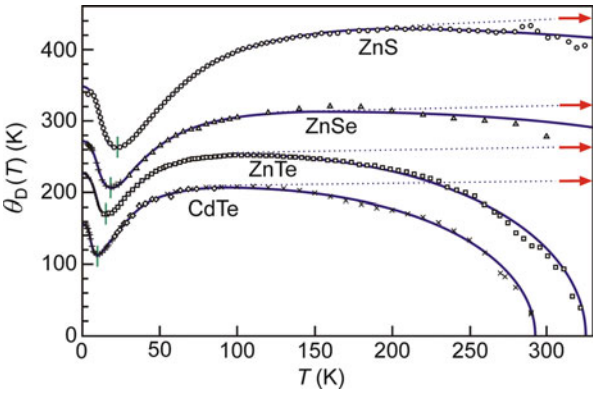


Fig. 3 Effective Debye temperatures $\theta_D(T)$ of II–VI semiconductors according to Eq. 11. Data points are derived from experimental C_P values, *red arrows* indicate high-temperature limits of the theoretical Debye temperatures, and *vertical green bars* denote minima of $\theta_D(T)$ (After Pässler 2011)

shown in Fig. 3. A significant deviation from $\theta = const$ usually appears at low temperatures.

The density-of-state distribution function can be obtained by mode counting from a variety of approximations (Sect. 3.2 in ▶ Chap. 4, "Elasticity and Phonons") and from first-principles computation. The specific heat can then be calculated by integrating the temperature derivative of the product of density distribution and statistical distribution function

$$C_V = \frac{1}{V}\frac{\partial}{\partial T}\int_0^\infty \frac{\hbar\omega\ g(\omega)\ d\omega}{\exp\left(\dfrac{\hbar\omega}{kT}\right) - 1}. \tag{13}$$

At low temperatures ($T \ll \theta$) only acoustic phonons with very long wavelengths ($\lambda \gg a$) participate in the heat content of a solid. As the Debye temperature is approached, optical phonons also become excited; initially, these are the optical phonons with lowest energy. They are near the boundary of the Brillouin zone and have a wavelength close to the lattice constant since $\hbar\omega_0\,(q = \pi/a) < \hbar\omega_0\,(q = 0)$, as shown in Figs. 13 and 15 in ▶ Chap. 4, "Elasticity and Phonons."

1.2.1 Phase Changes

There are major changes in the measured specific heat at temperatures where phase changes occur (*configurational specific heat*). The most obvious change happens at the melting point between solid and liquid state, which will not be discussed further here. Others deal with first- or second-order phase changes (see below) at which new oscillatory modes are able to participate in the phonon spectrum. Examples are molecule crystals or compounds with radicals, such as SO_4, NO_3, or ClO_4, which can become free to rotate within the solid long before melting occurs. The solid reacts to such a change in degrees of freedom by a change in crystal structure, volume, and many other physical properties. Other phase changes below the melting temperature are less drastic and involve only a change in lattice structure (symmetry).

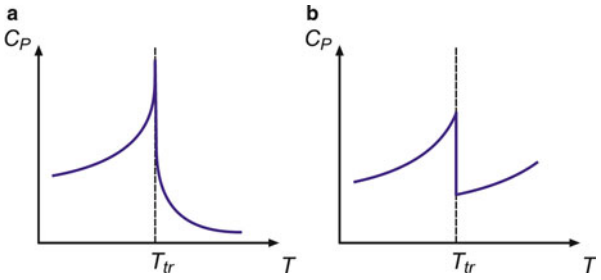


Fig. 4 Specific heat as a function of the temperature for (**a**) first- and (**b**) second-order phase transitions occurring at a temperature T_{tr}

All these changes cause a change in the phonon spectrum and thus also a change in the specific heat. The specific heat develops spikes or jumps at these phase transformations of first or second order, respectively, which, in turn, can be used for the detection of these changes (see Fig. 4). Depending on the sign of the peak, we must distinguish exotherm or endotherm processes responsible for the phase transformation.

The order of the transition is defined by the degree of the lowest derivative of the Gibbs free energy $G(P,T) = U - TS + PV$ that shows a jump. Since $(\partial G/\partial T)_P = -S$ and $(\partial G/\partial P)_T = V$, a first-order transition shows a discontinuity of entropy and volume at the transition temperature. Melting is one such transition. With $(\partial^2 G/\partial T^2)_P = -(\partial S/\partial T)_P = -C_P/T$ and $(\partial^2 G/\partial P^2)_T = (\partial V/\partial P)_T = -\kappa V$, the specific heat or compressibility shows such discontinuity. Dielectric or magnetic ordering at the Curie or Neel temperature and the λ-point transformation of liquid helium are examples for higher-order phase transitions (Landau and Lifshitz 1958). Solids which can rapidly and reversibly be switched between an amorphous and a crystalline phase have recently gained much advertance for optical and electrical memory devices. The phase change in nanostructured materials, e.g., $(GeTe)_m(Sb_2Te_3)_n$, (n, m integer), proceeds on a nanosecond timescale and is accompanied by strong changes of optical and electrical properties; for a review see Lencer et al. 2011.

With sensitive measuring methods (differential thermal analysis or 3ω method), small changes in lattice and defect structures can be detected, such as partial recrystallization, changes in dielectric and magnetic states involving their degree of order, and even in the density of point defects.

1.3 Specific Heat of Amorphous Semiconductors

The specific heat of amorphous solids at decreased temperature is usually dominated by vibrational modes as described by the Debye model, yielding a cubic dependence $C_V \sim T^3$. At low temperature, however, many amorphous materials show an anomalous *linear* increase of the specific heat with temperature. This cannot be explained by a conventional phonon spectrum. However, it can be explained when atoms tunnel within the glass to different *metastable* positions when these different positions of atoms have nearly the same energy (*two-level tunneling system*). If the barrier between the different positions is low and thin enough, such transitions can

be activated at rather low temperatures and result in a specific heat contribution of (Phillips 1987)

$$C_V \propto T. \tag{14}$$

The linear part of C_V is well described by the two-level model. Recent investigations also demonstrate a cubic excess contribution in addition to the part of the Debye model (Molina-Ruiz et al. 2020). This contribution exceeds the Debye part derived from elastic coefficients; it is assigned to local regions of reduced density and resulting localized low-energy vibrational states.

When plotting C_V/T^3 as a function of T, we observe a characteristic maximum near $T = 10$ K which coincides with the plateau of the thermal conductivity in amorphous semiconductors (see Fig. 12) and may point to related causes (Yu and Freeman 1986). Experimental data can be found in Pohl et al. 1974.

2 Thermal Expansion

The anharmonicity of the lattice potential and the resulting increased average amplitude of the lattice oscillations are responsible for the expansion of the semiconductor with increasing temperature. Crystal defect-related additional expansion is neglected here (see Sect. 2.1 in ▶ Chap. 15, "Crystal Defects"). Valuable information can be extracted from a phenomenological continuum approach, which will be discussed before the atomistic model is presented.

2.1 Phenomenological Description

Two thermal expansion coefficients are commonly used, a linear and a volume expansion coefficient:

$$\alpha = \frac{1}{l}\frac{\Delta l}{\Delta T} \quad \text{and} \quad \alpha_V = \frac{1}{V}\frac{\Delta V}{\Delta T}, \tag{15}$$

with $\alpha_V \approx 3\alpha$ for isotropic materials. For anisotropic crystals the linear thermal expansion coefficient is a tensor, which is proportional to the strain tensor $\underline{\boldsymbol{\varepsilon}}$ (see Sect. 1.1 in ▶ Chap. 4, "Elasticity and Phonons") and the temperature difference:

$$\underline{\boldsymbol{\varepsilon}} = \underline{\boldsymbol{\alpha}}\,\Delta T, \tag{16}$$

with the tabulated components in the main axis of the tensor ellipsoid (α_{xx}, α_{yy}, α_{zz}) usually referred to as (α_1, α_2, α_3), or, in uniaxial crystals like wurtzites, as $\alpha_\parallel$ and $\alpha_\perp$ (with respect to the $\boldsymbol{c}$ axis). Because of the varied strengths of the lattice forces in different directions for anisotropic lattices, an anisotropy of α_i appears. Examples of strongly anisotropic (layer) semiconductors are BN, GaS, GaSe, InSe, TlSe, InBi

(White et al. 1975), and some chalcopyrites such as $CuGaSe_2$ (Bodnar and Orlova 1985). Semiconductors with a high degree of anisotropy, e.g., Se, show a minor contraction in the direction of strongest bonding while expanding in the direction of weakest bonding, with the volume expansion coefficient α_V usually positive.[3] This fact provides an opportunity to design composite materials with essentially zero thermal expansion (Roy and Agrawal 1985).

The relation between the strength of the lattice bonding and the thermal expansion yields a useful empirical rule, the *Grüneisen rule*, which connects the melting point T_m of isotropic materials with the thermal expansion and holds well for metals:

$$\left(\frac{\Delta l}{l}\right)_{T_m(\mathrm{K})} \cong 0.07; \quad \text{hence} \quad \alpha \cong \frac{0.07}{T_m}. \tag{17}$$

The maximum thermal expansion at the melting point is about 7% and is independent of the material. The melting point, in turn, is related to the Debye temperature θ and thus to the elastic stiffness constants, i.e., to the strength of bonding via another empirical relation

$$T_m = \gamma \; a^2 \; M \; \theta^2, \tag{18}$$

where M is the atomic mass, γ is the Lindemann parameter (compare Eq. 20), and a is the interatomic distance (in a monatomic lattice). This *Lindemann relation* can be justified with a lattice dynamic model.

Table 2 lists the thermal expansion coefficients and other relevant parameters for several important semiconductors.

2.2 Lattice Dynamic Consideration

The average kinetic energy per atom as a linear harmonic oscillator is

$$E_{\mathrm{osc}} = \frac{1}{2} M \left(\frac{d\overline{u}}{dt}\right)^2 = \frac{1}{2} M \; \omega^2 \overline{u}^2, \tag{19}$$

which is equivalent to $(1/2)kT$ for a linear oscillator, with $\overline{u} = fa$ and f as the fractional displacement from the equilibrium interatomic distance a. The oscillatory energy per atom in a solid is $3kT$; at high temperatures most of the oscillations are in the optical branch. Therefore, we obtain from Eq. 19 with $\hbar\omega = \hbar\omega_0 \cong k\theta$ for the melting point with $E_{\mathrm{osc}} = 3kT_m$

[3]Exceptions like H_2O or In are caused by rather unusual changes in the atomic structure of the solid. (Were it not for these changes in H_2O, lakes would freeze from the bottom up.) Another important anomaly of α_V is observed at very low temperatures (see end of Sect. 2.2 in this chapter).

Table 2 Room temperature thermal expansion coefficients α lattice constant a density ρ and melting point T_m for some semiconductors; data largely from Martienssen and Warlimont 2005

Material	a (Å)	ρ (g/cm^3)	α (10^{-6}K^{-1})	T_m (K)
$C_{Diamond}$	3.5669	3.515	1.0	4100
Si	5.4310	2.329	2.9	1687
Ge	5.6579	5.323	5.9	1210
ZnS	5.4053	4.088		1991
ZnSe	5.6674	5.266[a]	7.4	1799
ZnTe	6.0882	5.636	8.3	1564
CdSe	6.078	5.81	3.8[a]	1537
CdTe	6.46	5.87[b]	5.0	1367
BN[d]	3.615	3.45	1.2	2973
BAs[d]	4.78	5.22	3.85	2075
AlN	3.1111 (a)	3.255	4.4 ($\perp$c)	3025
	4.9788 (c)		3.5 ($\parallel$c)	
AlP	5.451	2.42[a]		2823
AlAs	5.6622	3.81[a]	5.2	2013
AlSb	6.1355	4.218	4.2[a]	1327
GaN	3.190 (a)[a]	6.07[a]	3.2 ($\perp$c)[a]	2791
	5.189 (c)[a]		5.6 ($\parallel$c)[a]	
GaP	5.4506	4.138	4.7	1749
GaAs	5.6536	5.316	5.9	1511
GaSb	6.0959	5.614	7.8	991
InN	3.5446 (a)[a]	6.78[a]	3.6 ($\perp$c)	$\sim$1900
	5.7034 (c)[a]		2.6 ($\parallel$c)	(at 80 kbar)
InP	5.8687	4.81[a]	4.8	1327
InAs	6.0583	5.667	4.5[c]	1221
InSb	6.4794	5.775	5.4	800

[a]Temperature not specified
[b]$T = 4$ K
[c](20–250) K
[d]Zincblende structure

$$T_m = \frac{k}{9\hbar^2} \; M \; \theta^2 f_{\max}^2 a^2, \tag{20}$$

defining $f_{\max}$ as the maximum displacement of atoms at the melting point. This equation is identical to the Lindemann relation (Eq. 18) with $\gamma = k f_{\max}^2/9\hbar^2$. Comparison with the experiment shows that for most solids $0.1 < f_{\max} < 0.15$; that is, the maximum amplitude of the lattice oscillation is roughly 10% of the interatomic spacing at the melting temperature. This is equivalent to the Grüneisen rule since the rms of the linear oscillation amplitude is $f_{\max}/\sqrt{2} \cong 0.07$, and assuming that almost all of these oscillations are totally anharmonic, i.e., atoms are essentially rigid and any oscillation needs additional lattice space.

Any thermal expansion requires a nonharmonic contribution from the interatomic potential (see Fig. 1 in ▶ Chap. 4, "Elasticity and Phonons"). Nonharmonic terms can

be evaluated from higher-order elastic constants (Hiki 1981). Considering the third-order term in the lattice potential (see Eq. 14 in ▶ Chap. 4, "Elasticity and Phonons")

$$E_{\text{pot}} = E_0 + \beta_1 u^2 + \beta_2 u^3, \tag{21}$$

we obtain for the average displacement of a lattice atom

$$\overline{u} = \frac{\int_{-\infty}^{\infty} u \exp\left(-\frac{E_{\text{pot}}}{kT}\right) d\omega}{\int_{-\infty}^{\infty} \exp\left(-\frac{E_{\text{pot}}}{kT}\right) d\omega}, \tag{22}$$

which can be evaluated by introducing Eq. 21 into Eq. 22 and integrating, yielding

$$\overline{u} \cong \frac{3}{4}\frac{\beta_2}{\beta_1^2} kT. \tag{23}$$

Using for the thermal expansion the average relative displacement

$$\alpha = \frac{d}{dT}\left(\frac{\overline{u}}{a}\right) = \frac{3}{4}\frac{\beta_2}{a\beta_1^2} k \tag{24}$$

(a is the interatomic distance for a cubic crystal), we obtain a first-order estimate of the thermal expansion coefficient in terms of the anharmonicity constant (β_2) of the lattice oscillation. In this approximation α is temperature independent. Near the melting point, however, additional lattice expansion is observed because of intrinsic (Schottky or Frenkel) defect generation (see Sect. 2.1 in ▶ Chap. 15, "Crystal Defects").

At lower temperatures, however, α decreases with decreasing T, much like the specific heat. This can be explained in the above theory by replacing the Dulong–Petit value for the lattice energy (used to obtain Eq. 20) with the appropriate function $u(T)$ yielding $C_V(T)$ at lower temperatures (e.g., the Debye function). A direct possibility for introducing C_V is provided through the *Mie–Grüneisen theory* of the equation of state of solids (Grüneisen 1926), which yields

$$\alpha = \frac{\Gamma}{3}\frac{C_V}{BV}, \tag{25}$$

with Γ the *Grüneisen parameter*, which can be expressed as

$$\Gamma = -\frac{d\ln\omega}{d\ln V} = -\frac{V d\omega}{\omega dV}. \tag{26}$$

For sufficiently high temperatures, Γ is nearly constant (see Fig. 5). Equation 25 shows with $\alpha \propto C_V$ the main tendencies of the observed behavior of $\alpha(T)$, neglecting the slight temperature dependence of volume V and bulk inverse compressibility B.

A more sophisticated lattice dynamic theory must consider the temperature dependences of Γ and C_V calculated for the different modes i of lattice oscillations; hence (Mitra and Massa 1982):

$$\alpha(T) = \frac{\sum_i \Gamma_i C_V^{(i)}}{3BV}. \tag{27}$$

This relation is the basis for further lattice dynamic analysis relating to Γ_i and $C_V^{(i)}$ (see Namjoshi et al. 1971).

Recent investigations also calculate thermal properties from first principles. Coefficients of thermal expansion were determined employing density functional perturbation theory with temperature-dependent effective potentials, yielding good agreement with experimental values (Pike and Løvvik 2019).

2.2.1 Negative Thermal Expansion

At low temperature ($T < 0.2\ \theta$), the thermal expansion coefficients of numerous semiconductors (IV, III–V, and II–VI compounds) become negative (observed first by Valentiner and Wallot 1915); at still lower temperatures, a range of positive values of α often reappears (Daniels 1962). This behavior, shown in Fig. 6, can be understood from the residual anisotropy of lattice oscillations that renders the Grüneisen parameters negative for low-frequency transverse acoustic modes in diamond and zinc-blende lattices (Barron 1957 – see also Sect. 2.1). This means a contraction of the crystal perpendicular to a longitudinal oscillation, which can yield a net volume contraction in an intermediate temperature range.

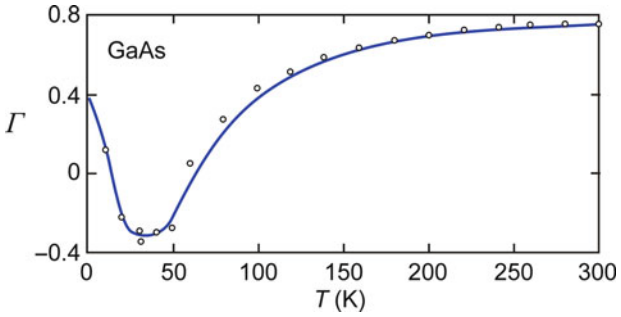


Fig. 5 Grüneisen parameter Γ as a function of the temperature for GaAs. Data points are obtained from linear expansion coefficient and specific heat; the solid line is calculated (After Soma et al. 1982)

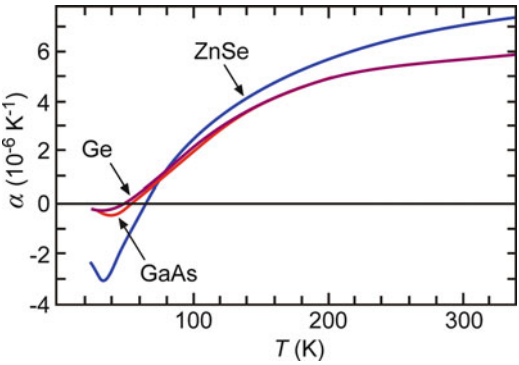


Fig. 6 Linear thermal expansion coefficients of Ge, GaAs, and ZnSe (After Novikova 1966)

2.2.2 Photothermal Expansion

Phonons created as a result of optical excitation cause local heating, thereby expanding the lattice. Such expansion, in turn, can be used to detect the phonon generation. The photothermal expansion method can be used as loss spectroscopy for photoluminescence (nonradiative recombination produces phonons) or in photovoltaic devices. More recent applications also include micro-actuators (Li et al. 2020). A variety of highly sensitive techniques was developed, capable of detecting a change in temperature of 10^{-5} degrees (Amer 1987) and absorption coefficients in the 10^{-2} cm^{-1} range (Luk'yanov et al. 2008). Photoacoustic methods were predominantly applied to study liquids and gases, although the spectral distribution of photo-induced phonon generation contains much information about the electronic and vibronic structure and about deexcitation mechanisms in solids. A review of photoacoustic methods is given by Rosencwaig (1980), Patel and Tam (1981), Tam (1986), and Almond and Tam (1996); a review on various materials is compiled in White (2017). An example for nonradiative relaxation of a transition metal in fluorides measured using photoacoustic spectroscopy is given by Torchia et al. (2002), and the ultrafast dynamics of surface expansion and thermal diffusivity is reported by Pennington and Harris (1992).

3 Thermal Conductivity

3.1 Diffusive Thermal Transport

Heat transport in a solid is caused by phonons and all other mobile quasiparticles which contain excess energy and interact with the lattice, such as electrons (see Sect. 3.3 and Sect. 1.2.1 in ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients"), holes, polarons, excitons, and photons. In semiconductors with low carrier density, and in the absence of light and electric fields, the thermal energy transport is caused mainly by phonons. Such transport would propagate with sound velocity if phonons were not scattered. Such *ballistic*

phonon transport is observed shortly after a heat pulse is applied at low temperatures before thermalization of phonons occurs (Knaak et al. 1986). However, the sample dimension is usually much larger than the mean free path of phonons, being in the range of typically 100 nm at room temperature. Consequently, the thermal energy flux per unit area $J = dQ/(A\ dt)$ is given by a diffusion type of transport, carried by the random motion of phonons and directed by their gradient as the driving force:

$$\mathbf{J} = -\kappa \frac{\partial}{\partial \mathbf{r}} T \tag{28}$$

with κ as the *thermal conductivity* expressed in $\mathrm{Wm^{-1}K^{-1}}$; this is *Fourier's law* of heat conduction. The isotropic form of Eq. 28 applies for cubic crystals as well as polycrystalline and amorphous solids. If additional carriers (such as electrons) are taken into account, their contributions are added; this yields $\kappa = \kappa_1 + \kappa_2 + \ldots$.

The thermal current density can be derived from the phonon flux as a *particle current* of N phonons with an average velocity v_s. After integration over all angles of the phonon flux through an arbitrary surface normal to the flux, we obtain $\frac{1}{6}Nv_s$. With an average thermal energy of $\frac{3}{2}kT$ per phonon, this results in a net thermal energy flux J between two closely spaced parallel planes at T_1 and T_2 of

$$J = \frac{1}{6} N_{\mathrm{av}} v_s \frac{3}{2} k(T_2 - T_1); \tag{29}$$

thereby heat transfer takes place after an inelastic collision, i.e., after a distance of the mean free path for phonons λ is traversed. Replacing the temperature gradient $\partial T/\partial \mathbf{r}$ with $(T_2 - T_1)/2\lambda$, and inserting the classical expression for the specific heat of such a phonon gas $C_V = \frac{3}{2}Nk$, we obtain from Eqs. 28 and 29 the often-cited expression for the thermal conductivity (Debye 1914):

$$\kappa = \frac{1}{3} C_V v_s \lambda. \tag{30}$$

The various modes (i) and branches (j) of the phonon spectrum contributing to κ can be accounted for by appropriate summation while retaining the relation of Eq. 30:

$$\kappa = \frac{1}{3} \sum\nolimits_{i,j} C_{ij} v_{ij} \lambda_{ij}. \tag{31}$$

The contribution of the different components of the *specific heat* introduces a temperature dependence of κ. The *phonon velocities* v_{ij} are group velocities ($\partial\omega/\partial q$) and are less temperature dependent (Sect. 3.1.4 in ▶ Chap. 4, "Elasticity and Phonons").

The critical parameter for the thermal conductivity is the *mean* free *path* λ, which requires a thorough study of the phonon-scattering mechanisms (Sect. 3.2).

For a review of experimental techniques, see Rowe and Bhandari (1986). In anisotropic materials, the thermal conductivity can be anisotropic because of the anisotropy of the elastic constants and therefore of the phonon velocity; see de Goër et al. (1982).

The heat conductivity depends on the phonon velocity and on scattering. The phonon velocity decreases with decreasing lattice bonding. Therefore, the thermal conductivity decreases with an increase of the atomic number when comparing elements within the same column of the periodic system. Examples for III–V semiconductors are listed in Table 3.

3.2 Phonon-Scattering Mechanisms

Phonons can interact with each other because of the anharmonicity of the lattice potential, which permits phonon–phonon scattering. This is the most basic (*intrinsic*) scattering mechanism, limiting the thermal conductivity. Other scattering mechanisms may reduce the mean free path λ_{ij} further according to *Matthiesen's rule*

$$\frac{1}{\lambda_{ij}} = \sum_k \frac{1}{\lambda_{ijk}}, \tag{32}$$

where the index k identifies all scattering mechanisms. The more important ones are:

- Phonon–phonon scattering
- Phonon scattering at point defects
- Phonon scattering at line defects
- Phonon scattering at short- or long-range disorder
- Phonon scattering at interfaces, surfaces, or grain boundaries
- Phonon scattering with carriers

Each one of these scattering categories must be distinguished with respect to the type of phonons involved. A large body of literature describes the different phenomena: e.g., Drabble and Goldsmid (1961), Steigmeier (1969), Touloukian et al. (1970), Childs et al. (1973), Challis et al. (1975), Klemens and Chu (1976), Berman (1976), Slack (1979), Vandersande and Wood (1986), Klemens (1986), and Lindsay et al. (2013). Only a few of these mechanisms will be discussed here, since we will

Table 3 Thermal conductivity of III–V semiconductors at room temperature in W/(cm K)

	Anion	P	As	Sb
Cation	Atomic weight	31	75	122
Al	27	0.9	0.8	0.6
Ga	70	0.8	0.5	0.4
In	115	0.7	0.3	0.2

return to this field when discussing electron scattering with different types of phonons.

3.2.1 Phonon–Phonon Scattering

Conserving total energy and momentum,

$$\hbar\omega_1 + \hbar\omega_2 = \hbar\omega_3 \quad \text{and} \quad \mathbf{q}_1 + \mathbf{q}_2 = \mathbf{q}_3, \tag{33}$$

yields no change per se in heat flow for phonons of low energy and momentum. This scattering process is referred to as an *N-process* and is shown in Fig. 7a. However, when the phonon momentum is large enough so that the summation $\mathbf{q}_1 + \mathbf{q}_2$ leads to a phonon with a momentum outside the first Brillouin zone, Bragg reflection at the zone boundary occurs and the resulting $\mathbf{q}_3$ has its direction essentially reversed. Mathematically, one accounts for the Bragg reflection by adding the reciprocal lattice vector $\mathbf{G}$ with the length of the Brillouin zone and thereby obtains a resulting phonon within this Brillouin zone (see Fig. 7b). Peierls (1929, 1955) defined this as an *Umklapp process* (U-process). Only these processes contribute directly to a thermal resistance.

U-processes require a sufficient $\mathbf{q}$ of the initiating phonons, i.e., sufficient phonon population at higher $\hbar\omega$. This occurs at higher temperatures ($T \gtrsim \theta/4$). More complex lattice structures result in more complicated Brillouin zones with more opportunities for Umklapp processes. Therefore, these crystals usually show a lower heat conductivity.

Leibfried and Schlömann (1963) give an estimate of the thermal conductivity, which increases hyperbolically with decreasing temperature:

$$\kappa \propto \lambda = \frac{12}{5}\sqrt[3]{4}\left(\frac{k}{h}\right)^3 \frac{M\ a}{(\Gamma+\frac{1}{2})^2}\frac{\theta^3}{T} \cong 5.76\frac{M\ a}{(\Gamma+\frac{1}{2})^2}\frac{\theta^3}{T}\ (\text{Å}), \tag{34}$$

where M is the atomic mass, a is the interatomic spacing, and Γ is the Grüneisen anharmonicity parameter (Eq. 26). It follows from Eq. 34 that the thermal conductivity at high temperatures increases rapidly with increasing Debye temperature. The increasing scattering at higher temperature is due to an increase in amplitude and

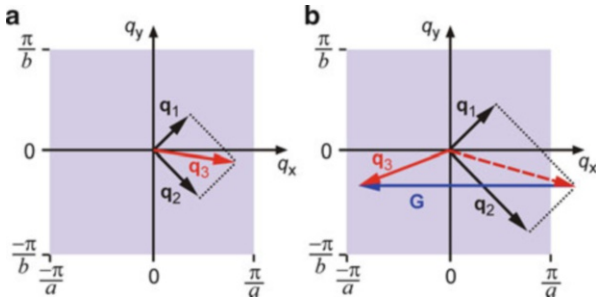


Fig. 7 (**a**) N-processes and (**b**) U-processes for two phonons with momentum $\mathbf{q}_1$ and $\mathbf{q}_2$, producing a third phonon with momentum $\mathbf{q}_3$. The reciprocal lattice vector $\mathbf{G}$ indicates the Bragg reflection at the zone boundary

hence in anharmonicity of the oscillation, causing an increase in phonon–phonon scattering cross-section.

At lower temperatures, U-processes freeze out. This results in a decrease of scattering and thus an increase of the thermal conductivity with decreasing temperature:

$$\kappa_U \propto \lambda = \frac{7}{4}\left(\frac{k}{h}\right)^3 \frac{M\ a}{(\Gamma+\frac{1}{2})^2}\frac{T^3}{\theta}\exp\left(\frac{\theta_U}{T}\right) \tag{35}$$

where $\theta_U < \theta$ (typically $\theta_U \approx 0.5\ \theta$), in agreement with the experiment (Klemens 1958).

3.2.2 Phonon Scattering at Lattice Defects

At lower temperatures, phonon scattering at lattice defects starts to compete with the intrinsic phonon–phonon scattering and limits the thermal conductivity to a degree that depends strongly on the density and distribution of lattice imperfections. Such scattering at lattice defects can be caused by the change in mass and elastic properties of the lattice surrounding the defect, hence changing the wave dispersion-relation. Phonon scattering at *point defects* can be compared to scattering of waves at obstacles. The relaxation time of phonons scattered at foreign atoms of mass M_f can be estimated as

$$\tau = \frac{4\pi {v_s}^3}{N_f {a_f}^3 \omega^4}\left(\frac{M_f - M_0}{M_0}\right)^2 \tag{36}$$

where N_f is the density of foreign atoms, a_f^3 is their atomic volume, and ω is the frequency of the scattered phonons (Klemens 1955). τ is approximately equal to the time between scattering events $\tau \approx \lambda/v_s$. An increased scattering probability is obtained for phonons with a dominant wavelength of a similar dimension as the obstacle. Therefore, point-defect scattering is more pronounced at high and intermediate temperatures, providing more short-wavelength phonons at high q values, i.e., at higher energies.

Scattering at extended defects, such as dislocations (Sproull et al. 1959), stacking faults, colloids (Walton 1967), voids (Vandersande 1980), grain boundaries (Vandersande and Pohl 1982), and surfaces, is also important at low temperatures. Scattering at crystal boundaries is considered in more detail in Sect 4.

When, because of phonon scattering at extended defects, the mean free path levels off at lower temperatures, the temperature dependence of the thermal conductivity is given only by that of the specific heat, i.e., by $C_V \propto (T/\theta)^3$. At higher temperatures where the specific heat levels off, the thermal conductivity is determined by the increase in phonon–phonon scattering with temperature. The combined behavior is sketched in Fig. 8. The influence of additional scattering mechanisms on the thermal conductivity of Si is shown in Fig. 9.

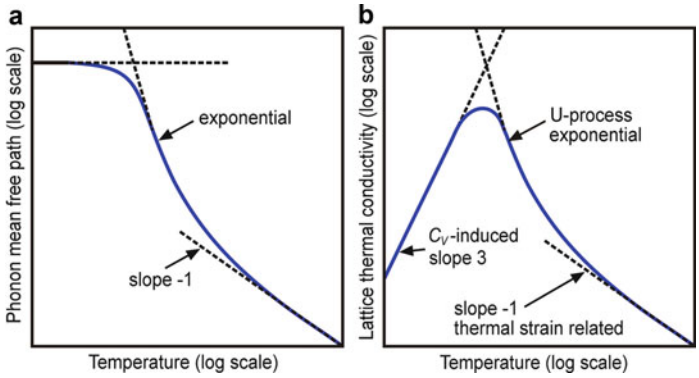


Fig. 8 Typical temperature dependence of (**a**) the mean free path and (**b**) the thermal conductivity in semiconductors

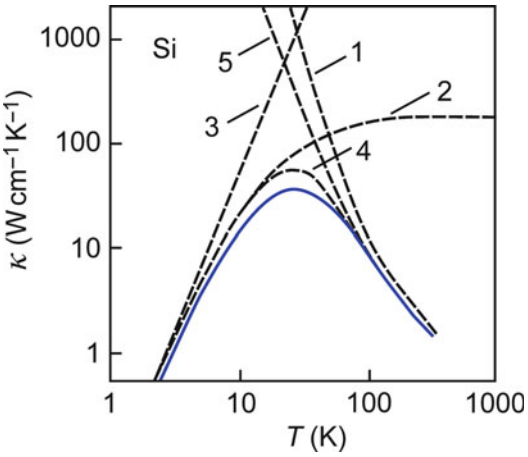


Fig. 9 Typical temperature dependence of the thermal conductivity for Si (*solid curve*) compared with the different scattering contributions. (*1*) Umklapp scattering; (*2*) diffuse boundary and isotope scattering; (*3*) diffuse boundary scattering; (*4*) Umklapp, diffuse boundary, and isotope scattering; and (*5*) Umklapp and isotope scattering (After Glassbrenner and Slack 1964)

Lattice disorder of an alloy type can substantially reduce the thermal conductivity. A similar reduction is observed in the common mixture of different isotopes in any lattice. An example for Ge is shown in Fig. 10a and for SiGe alloys or polycrystals in Fig. 10b. Here, scattering at the surface of thin layers reduces the thermal conductivity at lower temperatures. Phonon mean free paths up to 30 cm are observed at $T < 1$ K in pure Si (Vandersande and Wood 1986). The effect of alloying at a given temperature is illustrated for $Ga_{1-x}Al_xAs$ in Fig. 11a in terms of the inverse quantity of κ, the *thermal resistivity* κ^{-1}. The alloy shows highest thermal resistivity κ^{-1} (i.e., lowest thermal conductivity) for an equal cation sublattice occupation by Ga and Al (maximum disorder).

Finally, one observes strong phonon absorption in the frequency range of local modes (see Sect. 3.2.2 in ▶ Chap. 4, "Elasticity and Phonons") of lattice defects (*resonant scattering*). This can lead to dips in the $\kappa(T)$ curve when, with increasing T, phonon branches above the local mode that could be populated are kept depleted by

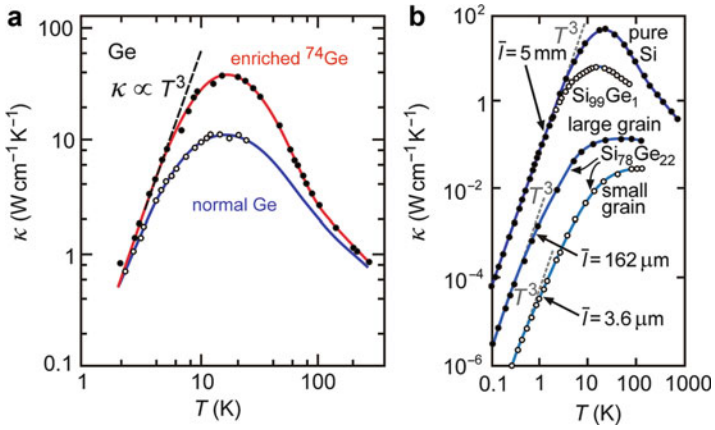


Fig. 10 (**a**) Thermal conductivities of normal Ge and Ge enriched with ^{74}Ge (After Geballe and Hull 1958). (**b**) Thermal conductivity of Si and SiGe alloys as single and polycrystals as a function of the temperature (After Kumar et al. 1985)

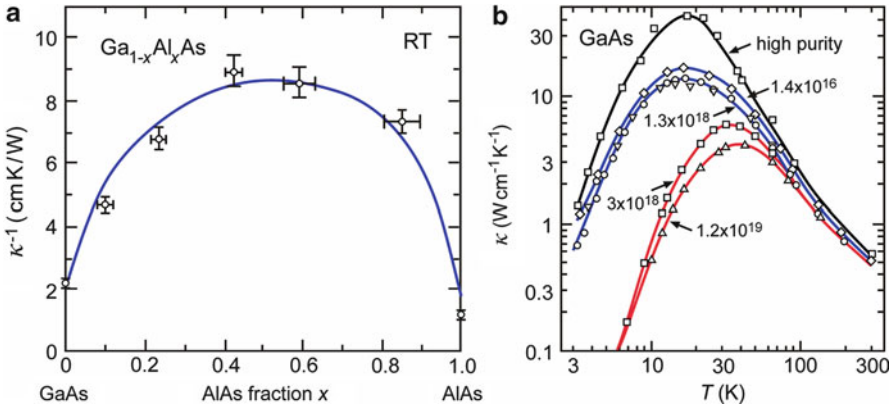


Fig. 11 (**a**) Thermal resistivity κ^{-1} of a $Ga_{1-x}Al_xAs$ alloy for various compositions x at room temperature, the solid line is a theoretical fit to the data; after Afromowitz 1973. (**b**) Influence of doping on the thermal conductivity of GaAs: high-purity (doping $<10^{16}$ cm^{-3}) single crystal (*black*), *n*-type (*blue*) and *p*-type (*red*) samples with carrier concentrations given in cm^{-3} (After Carlson et al. 1965)

this scattering process. Strong interaction is also observed for paramagnetic impurities with resonant scattering involving *spin–lattice relaxation*: see Holland (1964) for Mn doping in GaAs and Adilov et al. (1986) for doping with Ni.

3.3 Phonon Scattering in Crystalline and Amorphous Semiconductors

3.3.1 Phonon Scattering in Crystalline Semiconductors

Doping – the introduction of desired impurities (see Sect. 1.2 in ▶ Chap. 15, "Crystal Defects") – introduces free carriers in addition to the substitutional foreign atoms. Scattering of phonons with electrons and other quasiparticles is discussed later (Sects. 2 and 5 in ▶ Chaps. 23, "Carrier Scattering at Low Electric Fields," and ▶ 24, "Carrier Scattering at High Electric Fields"). Such type of scattering reduces the thermal conductivity especially at low and medium temperatures; the individual contribution of electrons is outlined in Sect. 3.3.2.

The influence of a low density of *p*-type and *n*-type doping on the thermal conductivity is shown for GaAs in Fig. 11b. The thermal conductivity of group IV and III–V semiconductors above 300 K is discussed by Logachev and Vasilev (1973). With high doping densities, other thermal energy transport mechanisms may become important, such as the transport by ambipolar diffusion of electrons and holes or excitons (see Sect. 1 in ▶ Chap. 14, "Excitons"), with consequent scattering or recombination at the cooler end of the sample. For a review of the theory, see Slack (1979); for an experimental review, see Parrott and Stuckes (1975) or Berman (1976).

IR photon emission from the hot end and reabsorption at the cold end, for example, by free carriers (Vandersande and Wood 1986) also contributes to the thermal energy transport at high temperatures in partially transparent semiconductors (Waseda and Ohta 1987). Transparency is needed for the transmission of photons; some absorption is needed for interaction with the lattice (Vandersande and Wood 1986). In Ge this radiative contribution is $\cong 0.01$ W/(cm K); it increases with increasing temperature, but at 200 K it is still small compared to the lattice conductivity of Ge. This *radiative transmission* can contribute a significant fraction of the thermal energy transport at high temperatures with $\kappa \propto T^3$.

In the case of high optical excitation of free electrons via laser pulses, the transmission of energy from the heated electron plasma can become very large, causing a major change in the distribution of phonons (see Sect. 1.2 in ▶ Chap. 32, "Dynamic Processes").

3.3.2 Thermal Conductivity by Free Carriers

Thermal energy can also be carried by electrons and is a major contribution in metals. Both phonon and electronic contributions are additive:

$$\kappa = \kappa_{\mathrm{ph}} + \kappa_{\mathrm{el}}. \tag{37}$$

The electronic contribution can be estimated similarly to the phonon contribution, using the equivalent of Eq. 31:

$$\kappa_{\mathrm{el}} = \frac{1}{3} C_V^{(\mathrm{el})} v_{\mathrm{rms}}^2 \tau_n; \tag{38}$$

with $C_V^{(\mathrm{el})} = 2nk$ for a classical electron gas and $v_{\mathrm{rms}}^2 = 3kT/m_n$, we have

$$\kappa_{\mathrm{el}} = 2nk^2 \left(\frac{\tau_n}{m_n}\right) T \tag{39}$$

where n, τ_n, and m_n are the electron density, their relaxation time, and effective mass, respectively. The electronic contribution is related to the electron conductivity ($\sigma = en\mu = e^2 n\tau_n/m_n$ – see Sect. 2 in ▶ Chap. 22, "Carrier-Transport Equations") by the *Wiedemann–Franz law*

$$\frac{\kappa_{\mathrm{el}}}{\sigma} = 2\left(\frac{k}{e}\right)^2 T \tag{40}$$

and gives a marked contribution only at higher carrier densities (Kittel 1986). The factor 2 in Eq. 40 holds for nondegenerate semiconductors. At higher electron densities, and depending on the scattering mechanism, this factor varies between 2 and 4. The entire proportionality factor at the right side of Eq. 40 is called the *Lorentz number* $L = 2(k/e)^2$. For a strongly degenerate electron gas, $L = (\pi^2/3)(k/e)^2$; for more detail, see Smith (1978).

Bipolar thermal conductivity can have a major contribution in narrow-bandgap semiconductors. Electrons and holes are moving in the same direction without causing a net electric current. The thermal conductivity by bipolar diffusion can be larger by a factor of up to 10^2 than the electronic contribution alone (Vandersande and Wood 1986).

3.3.3 Phonon Scattering in Amorphous Semiconductors

In strongly disordered solids the mean free path is typically on the order of very few interatomic spacings (~10 Å) and is essentially independent of the temperature (Kittel 1949). At low temperatures, the material behaves like an isotropic elastic medium for the long-wavelength acoustic phonons. Below a few degrees K, these phonons freeze out, and all amorphous solids behave remarkably uniformly. Their thermal conductivity has a plateau at ≅10 K and, at lower temperatures, decreases proportionally to T^2. Above the plateau, the thermal conductivity increases linearly with temperature. This behavior is essentially independent of the material and any impurity (see Fig. 12a).

This T^2 dependence can be explained by resonant scattering between *two-level states* (TLS) which may arise from quantum tunneling between atoms (Matsumoto and Anderson 1981). Here the scattering length λ_{ph} caused by spatial fluctuations at a scale of about 10 Å varies inversely with the phonon frequency and hence also inversely with the temperature. Consequently, the thermal conductivity $\kappa = C_V v_s \lambda_{\mathrm{ph}}/3$ (see Eq. 30) varies as $\kappa \propto T^2$ in the range in which such phonons give a T^3 contribution to the specific heat (Anderson and Crowell 1972).

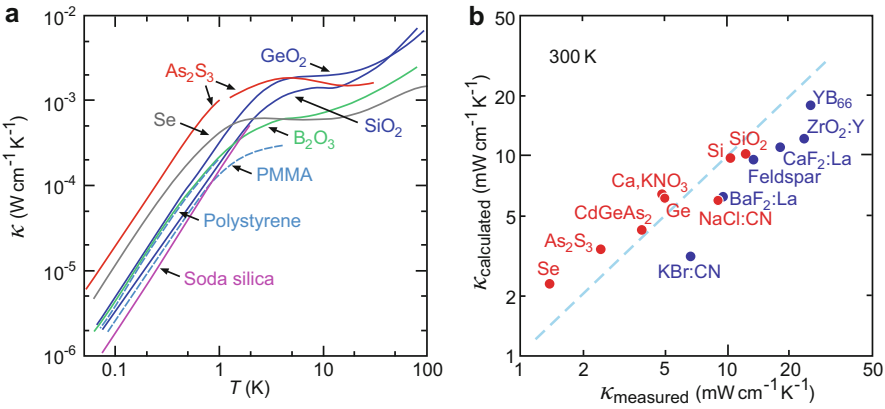


Fig. 12 (**a**) Low temperature thermal conductivities of various amorphous solids; after Stephens 1973. (**b**) Calculated versus measured minimum thermal conductivity of amorphous solids, the *dashed line* indicates perfect agreement; *red and blue circles* denote amorphous solids and disordered crystals, respectively (After Cahill et al. 1992)

Application of Eq. 30 is strictly not appropriate since the mean free path λ of such "phonons" no longer exceeds their wavelength. Assuming λ with a minimum length of half the phonon wavelength, Eq. 30 is still useful to provide a lower bound of κ for amorphous solids (Cahill et al. 1992, Hsieh et al. 2011) as illustrated in Fig. 12b.

The plateau in Fig. 12a is probably caused by Rayleigh scattering (Yu and Freeman 1986) involving localized (nonpropagating) modes (Akkermans and Maynard 1985) or *fractons* (Orbach 1984) with a phonon wavelength $\lambda \cong l_\omega$ as the *fracton localization length*. The source of scattering in this range and the reason for an increase in $\kappa(T) \propto T$ at higher temperatures is assigned to an enhanced scattering in a regime where a peak (*boson peak*) appears in the reduced density of states, see Schirmacher (2006).

A broad range of defect-related phonon-scattering results is reported in the literature; for example, for α-Si doped with Se and Te, see Vakhabov et al. (1985) and for doping with O, see Radhakrishnan et al. (1982); Igamberdiev et al. (1983) reported κ in the range ($6 < T < 300$ K) and various doping. For a representative sample of results, see Anderson and Wolfe (1986).

4 Thermal Transport in Nanostructures

4.1 Scattering at Crystal Boundaries

We now consider small crystals with dimensions in the μm to nm range. Such structures occur in thin layers, superlattices, nanowires, and nanocrystalline solids. If the mean free path exceeds the smallest distance between crystal boundaries, then λ in Eq. 30 is often rescribed by the Casimir length L_C (Casimir 1938)

$$\kappa = \frac{1}{3} C_V v_s L_C; \tag{41}$$

L_C is on the order of the shortest distance (e.g., layer or crystallite thickness, see also Fig. 10b) and depends on the roughness of the surface or the type of interface[4]. Only rough surfaces cause backscattering and thereby increase the thermal resistance. Roughness relates to the wavelength of the scattered elastic wave (low-energy acoustic branch). At smaller dimensions of the roughness, some of the reflection will be specular: a surface appears smoother to longer-wavelength waves. The acoustic phonon wavelength is typically on the order of 100Å at 10K.[5] At such low temperatures, $\kappa(T)$ is then determined by $C_V(T)$ and therefore varies $\propto T^3$ (Fig. 10b). See also de Goër et al. (1965).

4.2 Mean Free Path

Applications based on advanced materials often require managing of heat flow to push limits of device performance. This may be achieved by selecting an appropriate thermal conductivity (either high or low, Qian et al. 2021) and structural design.

High Thermal Conductivity Heat generation in electronic and optoelectronic devices reduces lifetimes, compromises performance, and accelerates electromigration and thermomechanical failures; the same applies for high-power batteries. Downscaling of integrated circuits according to Moore's law leads to increased power density and reduced ability to extract excessive heat (Pop 2010). Localized Joule heating also affects LEDs, semiconductor lasers, concentrator photovoltaics, and supercapacitors. All these devices require solids with *high* thermal conductivity.

Highest thermal conductivity is generally observed in pure crystalline solids with high group velocity and a low anharmonicity for minimizing phonon-phonon scattering. A high group velocity occurs in solids with low atomic mass and stiff bonding. In addition, a simple crystal structure yields a simple phonon spectrum. Best room-temperature κvalues of about 20 $Wcm^{-1}K^{-1}$are hence found in diamond, cubic BN, and graphite. However, heat removal from localized hotspots in semiconductor devices using these solids is seriously compromised by their large mismatch of thermal expansion coefficient with respect to that of common semiconductors such as Si and GaAs (see Fig. 13). Such a mismatch favors a failure of the bonding between the semiconductor and a heat-spreader. Promising recent findings on cubic boron arsenide (BAs) may substantially improve thermal management in modern electronics and photonics (Kang et al. 2019; Chen et al. 2019). This semiconductor with a κ value around 10 $Wcm^{-1}K^{-1}$in zincblende structure is thermally reasonably matched to electronic semiconductors.

[4]In the Casimir model, phonons travel ballistically and scattering occurs only at the boundaries of the nanostructure; these are considered as perfect "phonon black bodies."

[5]The dominant phonon wavelength is in the range of $\Lambda_{dom} = 1.48 v_s \hbar/(kT)$; see Klitsner and Pohl 1987.

Low Thermal Conductivity On the other hand, there are applications which require a *low* thermal conductivity. In thermoelectric applications, the parasitic heat transfer across the device must be minimized; thermoelectric materials simultaneously must be good electrical conductors. The conflicting requirements are expressed by their figure of merit being proportional σ/κ, with σ the electrical conductivity (Hahn et al. 2021). There is also a need for thermal insulators to prevent heat from reaching critical components. Examples are insulation in spacecrafts or thermal barrier coating in gas turbines.

The characteristic length L in nanoscale structures such as superlattices or nanowires is typically between a few nm and 100 nm. The small size leads to enhanced phonon scattering at interfaces and, for the common incoherent processes, to a corresponding strong reduction in the thermal conductivity κ (*coherent* scattering occurs in *ordered* nanostructures as discussed in Sect. 4.4). Applying Matthiesen's rule (Eq. 32), we can separate contributions of bulk and interface to the reduction of the mean free path λ of phonons

$$\lambda^{-1} = \lambda_{\text{bulk}}^{-1} + \lambda_{\text{interface}}^{-1}. \tag{42}$$

λ_{bulk} accounts for all phonon scattering processes in the bulk. $\lambda_{\text{interface}}$ describes scattering at boundaries and becomes significant if $L \leq \lambda_{\text{bulk}}$. The reduction of the thermal conductivity κ for reduced characteristic size L is illustrated for silicon in Fig. 14. We note that κ strongly decreases for smaller size L and is always much smaller than the bulk value of about 150 $\text{Wm}^{-1}\text{K}^{-1}$at room temperature. Such a decrease is also reflected in the contribution of phonons with a short λ to the total thermal conductivity depicted in Fig. 10b.

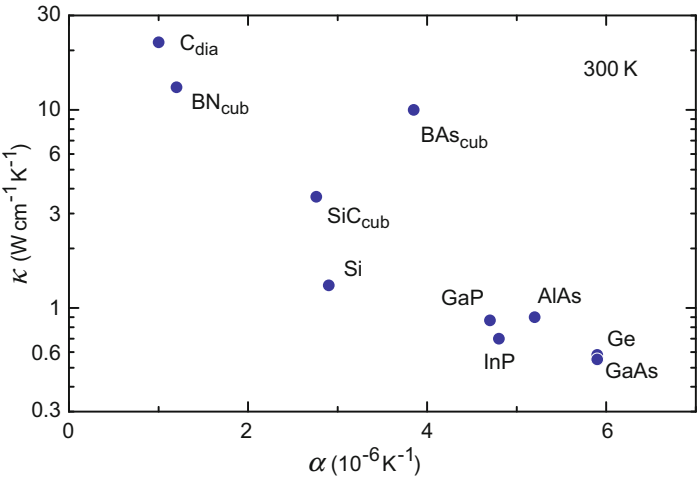


Fig. 13 Thermal conductivity and thermal expansion coefficient for some electronic semiconductors and high-conductivity solids

The mean free path λ_{bulk} comprises a boad spread of values. Since significant boundary scattering only occurs for sizes $L \leq \lambda_{bulk}$, knowledge of this spread is instructive for obtaining an estimate on the size effect (Yang and Dames 2013). Figure 14b shows the thermal conductivity accumulation function depending on the considered mean free path for Si. This quantity denotes the contribution $\kappa_{accumulated}$ of phonons with λ_{bulk} below a considered value to the total thermal conductivity κ_{total} (e.g., about 60% of all phonons in Si have $\lambda_{bulk} < 1$ μm, in the SiGe alloy less than 30%). The distribution indicates that only 10% of the heat conduction in Si at room temperature is carried by phonons with a mean free path below about 90 nm. The calculated distribution given in Fig. 14b is in good agreement with both experiments and first-principles calculations (Minnich et al. 2011). The spread is much broader in the SiGe alloy; the reduction of the fraction of phonons with long λ by alloy scattering leads to an increase of the part with short λ.

4.3 Thermal Transport Parallel to Boundaries

Thermal conductivity in heterostructures is generally anisotropic, and transport parallel or perpendicular to boundaries must be distinguished.

Thermal transport parallel to interfaces is important for nanowires, which may find applications in thermoelectric or (opto-) electronic devices; it also plays a role in in-plane transport of superlattices and thin layers. In-plane phonons can be specularly reflected from interfaces and surfaces (Sect. 4.1). Such reflection depends on the ratio of mean free path λ with respect to interface roughness d_{IF}, and its fraction is estimated to be below $1/e$ for a GaAs surface with $d_{IF} \approx a_{GaAs}$ at $T > 10$ K (Chen and Tien 1993). A high specularity is expected for d_{IF} below one tenth of the phonon wavelength.

The thermal conductivity of Si nanowires along the wire axis is shown at the bottom of Fig. 14a. Here the chosen characteristic length L is the wire diameter d_{NW}. The measurements at room temperature are well described by the size effect of the common diffuse boundary scattering; this Casimir limit (Sect. 4.1) corresponds to $\lambda_{interface} = d_{NW}$ in Eq. 42. Increased specular reflection in Si nanowires indicating $\lambda_{interface} > d_{NW}$ was observed for temperatures below 10 K (Heron et al. 2009). On the other hand, nanowires with very rough surfaces may have very low thermal conductivity also at large diameter as indicated by data at the bottom in Fig. 14a near 100 nm diameter (Hochbaum et al. 2008).

The in-plane conductivity of Si *layers* with different thickness is given at the top of Fig. 14a. The measurements are also described by pure diffuse boundary scattering, despite interfaces of high quality. This holds even for low temperatures (above 20 K), indicating a very small fraction of specular reflection.

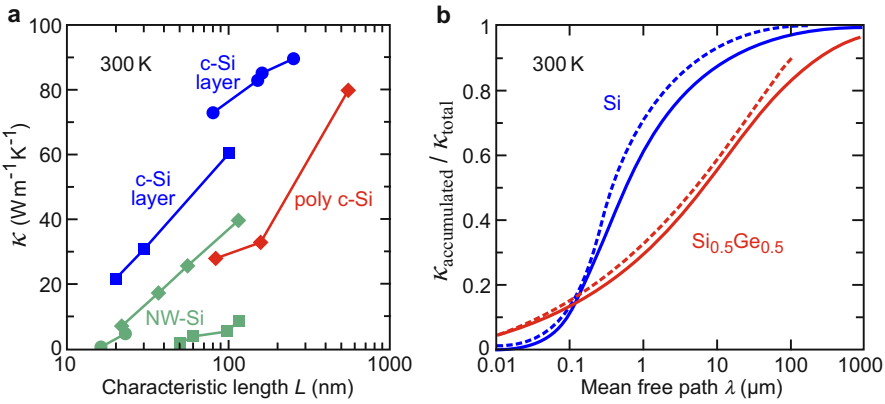


Fig. 14 (**a**) Thermal conductivity κ of nanometer-sized Si for various geometries, depending on the characteristic length L. *c-Si*, *poly c-Si,* and *NW-Si* denote crystalline Si layer ($\kappa_{\text{in-plane}}$), polycrystalline Si of different grain size, and Si nanowires of various diameters, respectively. (**b**) Accumulation function of the thermal conductivity calculated for bulk Si (*blue curves*) and a $Si_{0.5}Ge_{0.5}$ alloy (in *red*). *Solid and dashed curves* indicate different assumptions. (Data from Bera et al. 2010, Dames and Chen 2005, Ju and Goodson 1999, Hochbaum et al. 2008, Li et al. 2003, Liu and Asheghi 2006, and Wang et al. 2011)

4.4 Thermal Transport Normal to Boundaries

Thermal transport across interfaces of heterostructures is important for the thermal management of nanostructured devices made of multiple layers or superlattices. Differences in the acoustic impedance ρv_s of two layers with a common interface lead to reflections and a consequential thermal contact resistance; here ρ is the mass density. This effect leads to an important contribution in structures comprising multiple layers; additional reflection occurs in periodic structures. In the following, we consider superlattices in more detail.

The thermal conductivity of superlattices is anisotropic and usually below that of alloys with a corresponding average composition of the constituent materials. The dimension of the Brillouin zone normal to the layers is reduced (Fig. 18 in ► Chap. 4, "Elasticity and Phonons"), thereby modifying the group velocity of phonons in both, magnitude and direction, and reducing the energy required for Umklapp processes. Phonons are also sensitive to the specularity of the interfaces between the layers in heterostructure; diffuse scattering due to roughness, dislocations, and alloy formation at the interfaces participate significantly in the reduction of thermal conductivity and often make comparison difficult.

Most studies on superlattices focus on heat transport in the cross-plane direction (i.e., perpendicular to the interfaces), since heat flows mainly in this direction (Chen and Tien 1993) and potential benefits of in-plane transport, e.g., waveguiding effects, are easily obscured by interface imperfections. Besides scattering of phonons at rough interfaces, acoustic mismatch of the constituent materials and effects of zone folding which reduce phonon group-velocities affect cross-plane thermal conductivity, since the phonon mean free path is even at room temperature larger than typical layer

thicknesses. Studies on quasi-unstrained GaAs/AlAs superlattices with well-defined interfaces show a decreased conductivity as the interface density increases (Yao 1987; Capinski et al. 1999). The decreased thermal conductivity of GaAs/AlAs and Si/Ge superlattices compared to that of alloys with the same mass ratio (Fig. 15a) is concluded to originate from interface scattering of phonons due to a mismatch in specific heat, group velocity, and density of adjacent layers (Chen 1998).

For various superlattices also an *increase* of thermal conductivity for very short periods was experimentally observed, leading to a minimum for some intermediate period length (Fig. 15b). Such a behavior was reported for Bi_2Te_3/Sb_2Te_3 (Venkatasubramanian 2000) and GaAs/AlAs (Tritt 2004) and also calculated as a general feature by Simkin and Mahan 2000. The trend could be well modeled for Si/Ge by first-principles calculations which include vibrational modes and scattering rates due to anharmonicity and interfacial disorder (Garg and Chen 2013). The analysis of phonon scattering in Si/Ge reveals a dominant heat conduction of *coherent* low-frequency phonons with long mean free path λ at short periods and a transition to mainly incoherent transport of high-frequency phonons at longer periods; similar results were reported for highly perfect oxide superlattices (Ravichandran et al. 2014). While in alloys low thermal conductivity originates from disorder-induced scattering mainly reducing the heat carrying ability of high-frequency phonons, also the heat carrying ability of low-frequency phonons is reduced in superlattices due to a decrease of the cross-plane group velocity; consequently the thermal conductivity drops below the alloy value reached for shortest superlattice periods (11 Å in Fig. 15b). Along the *in-plane* direction, the decrease in group velocity is less pronounced, and the increase of thermal conductivity for longer periods is due to increased phonon lifetimes.

A dominance of coherent heat conduction, concluded from a linear increase with increasing total superlattice thickness, was also observed in GaAs/AlAs samples over a temperature range from 30 to 150 K (Luckyanova et al. 2012). A reduction in thermal conductivity at 300 K due to interface roughness resulted from the loss of coherence of high-frequency phonons, but the lower-frequency phonons that significantly contribute to the thermal conductivity were mostly coherent during their transport through the superlattice, until they were scattered at the sample boundaries.

Theoretical modeling of the thermal conductivity in superlattices is mostly based on a calculation of the actual phonon modes, inserted into the standard formula for the thermal conductivity

$$\kappa = \hbar \sum_{\lambda} \int v_{s\lambda}^2 \omega_\lambda \tau_\lambda \frac{df_{BE}(\omega_\lambda)}{dT} \frac{d^3q}{(2\pi)^3}, \tag{43}$$

where v_s and τ are the **q**-dependent quantities velocity and lifetime of the phonons, and f_{BE} is the Bose–Einstein occupation function. Equation 43 is derived from the Boltzmann transport equation (Sect. 4 in ▶ Chap. 22, "Carrier-Transport Equations"), where the lifetime τ is usually taken from the homogeneous material (Cahill et al. 2003). The Boltzmann approach treats phonons as incoherent particles and neglects

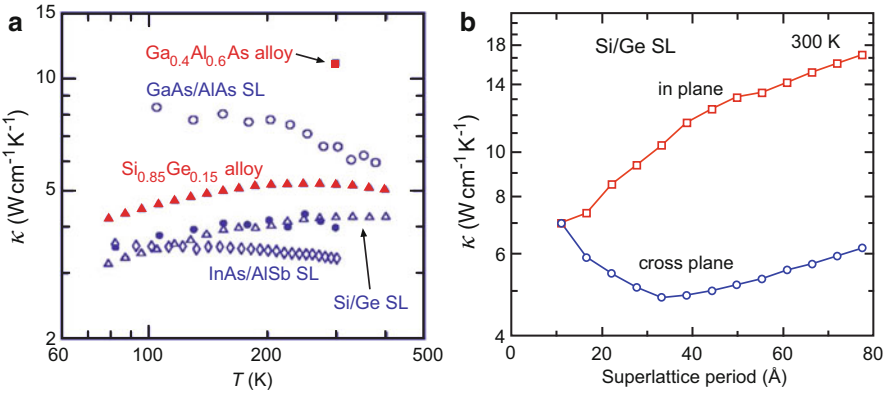


Fig. 15 (**a**) Cross-plane thermal conductivity for various superlattices with bilayer periods of 5 nm: GaAs/AlAs (*open circles*, 5.67 nm), Si/Ge (*open triangles*, 5 nm; *filled circles*, 4.4 nm), InAs/AlSb (*open diamonds*, 6.5 nm); after Cahill et al. 2003. (**b**) Calculated thermal conductivities including anharmonic and interface roughness effects for Si/Ge superlattices as a function of superlattice period (After Garg and Chen 2013)

coherence and the consequential modification of the phonon band-structure. These models describe data in the range of large superlattice periods.

A second group of models assumes the other extreme and treats phonons as coherent waves. Due to interference of traveling toward and away from interfaces, the calculated conductivity typically initially decreases with increasing period length and approaches a constant value beyond about 10 monolayer thickness (Tritt 2004). The models fail to describe the usually observed behavior for longer periods.

A more recently developed partially incoherent approach based on first principles applies force constants derived from density-functional perturbation theory (Debernardi et al. 1995, McGaughey et al. 2019). Additional incorporation of interface-disorder effects appears presently a viable approach to describe actual thermal properties of superlattices.

5 Summary

Phonons represent the thermal energy content of a semiconductor. Only a negligible amount is added by the kinetic energy of free electrons and holes in normal semiconductors since their density is comparatively small. After knowing the phonon density-distribution (a product of density of states and Bose–Einstein distribution), the thermal energy content is obtained by simple summation. The change of the thermal energy content with temperature, the specific heat, is directly accessible experimentally.

Phonons populate at low temperatures almost exclusively the acoustic branches, and only at temperatures approaching the Debye temperature do they extend with a significant fraction to optical branches. The thermal properties due to phonons,

however, change smoothly with temperature because of the additive contribution of all phonons. A thorough knowledge of their distribution permits one to predict details of the thermal properties.

The transport of phonons within the semiconductor is a diffusion-type process, driven by the gradient of the phonon density and described by a random walk of phonons between scattering events. The measured thermal conductivity in a temperature gradient provides experimental access to a variety of such microscopic scattering processes. A simultaneous contribution of both incoherent and coherent scattering with a temperature-dependent fraction describes the thermal transport in superlattice structures.

Thermal expansion is the result of the anharmonicity of interatomic potentials and, in contrast to many other nonharmonic phenomena (e.g., optical), presents itself already at low amplitudes of lattice oscillations (i.e., at relatively low temperatures).

References

Adilov KA, Vakhabov DA, Zakirov AS, Igamberdyev KHT, Mamadalimov AT, Tursunov SHO, Khabibullaev PK (1986) Fiz Tverd Tela 28:1918 (The heat conductivity of nickel-doped n- and p-type silicon, in Russian)

Afromowitz MA (1973) Thermal conductivity of $Ga_{1-x}Al_xAs$ alloys. J Appl Phys 44:1292

Akkermans E, Maynard R (1985) Weak localization and anharmonicity of phonons. Phys Rev B 32:7850

Almond DP, Tam AC (1996) Photothermal science and techniques. Chapman & Hall, London

Amer NM (1987) Characterization of optical and transport properties of semiconductors: a photothermal approach. Proc SPIE 794:11

Anderson CL, Crowell CR (1972) Threshold energies for electron–hole pair production by impact ionization in semiconductors. Phys Rev B 5:2267

Anderson AC, Wolfe JP (eds) (1986) Phonon scattering in condensed matter, vol V. Springer, Berlin

Barron THK (1957) Grüneisen parameters for the equation of state of solids. Ann Phys (NY) 1:77

Bera C, Mingo N, Volz S (2010) Marked effects of alloying on the thermal conductivity of nanoporous materials. Phys Rev Lett 104:115502

Berman R (1976) Thermal conductivity in solids. Claredon Press, Oxford

Bodnar IV, Orlova NS (1985) X-ray study of the thermal expansion anisotropy in $AgGaS_2$ and $AgInS_2$ compounds over the temperature range from 80 to 650 K. Phys Stat Sol A 91:503

Cahill DG, Watson SK, Pohl RO (1992) Lower limit to the thermal-conductivity of disordered crystals. Phys Rev B 46:6131

Cahill DG, Ford WK, Goodson KE, Mahan GD, Majumdar A, Maris HJ, Merlin R, Phillot SR (2003) Nanoscale thermal transport. J Appl Phys 93:793

Capinski WS, Maris HJ, Ruf T, Cardona M, Ploog K, Katzer DS (1999) Thermal-conductivity measurements of GaAs/AlAs superlattices using a picosecond optical pump-and-probe technique. Phys Rev B 59:8105

Carlson RO, Slack GA, Silverman SJ (1965) Thermal conductivity of GaAs and $GaAs_{1-x}P_x$ laser semiconductors. J Appl Phys 36:505

Casimir HBG (1938) Note on the conduction of heat in crystals. Physica 5:495

Challis LJ, Rampton VW, Watt AFG (1975) Phonon scattering in solids. Plenum Press, New York

Chen G (1998) Thermal conductivity and ballistic-phonon transport in the cross-plane direction of superlattices. Phys Rev B 57:14958

Chen X, Li C, Tian F, Gamage GA, Sullivan S, Zhou J, Broido D, Ren Z, Shi L (2019) Thermal expansion coefficient and lattice anharmonicity of cubic boron arsenide. Phys Rev Appl 11:064070

Chen G, Tien CL (1993) Thermal conductivities of quantum well structures. J Thermophys Heat Transf 7:311

Childs GE, Erick LJ, Powell RL (1973) Thermal conductivity of solids at room temperature and below, a review and compilation of literature. National Bureau of Standards Monograph, vol 131. National Bureau of Standards, Washington, DC

Dames C, Chen G (2005) Thermal conductivity of nanostructured thermoelectric materials. In: Rowe DM (ed) Thermoelectrics handbook, macro to nano. CRC Press, Boca Raton, FL

Daniels WB (1962) Low-temperature limit of Grüneisen's gamma of germanium and silicon. Phys Rev Lett 8:3

de Goër AM, Doulat J, Dreyfus B (1965) Conductibilité thermique de l'oxyde de béryllium de 1.5 a 300 K. J Nucl Mater 17:159 (Thermal conductivity of beryllium oxide from 1.5 to 300 K, in French)

de Goër AM, Locatelli M, Nicolau IF (1982) Low-temperature heat transport in α-HgI_2 single crystals. J Phys Chem Solids 43:311

Debernardi A, Baroni S, Molinari E (1995) Anharmonic phonon lifetimes in semiconductors from density-functional perturbation theory. Phys Rev Lett 75:1819

Debye P (1912) Zur Theorie der spezifischen Wärmen. Ann Phys (Lpz) 39:789 (On the theory of the specific heat, in German)

Debye P (1914) Vorträge über kinetische Gastheorie. Teubner Verlag, Berlin (Lectures on the kinetic gas theory, in German)

Drabble JR, Goldsmid HJ (1961) Thermal conduction in semiconductors. Pergamon Press, New York

Dulong PL, Petit AT (1819) Sur quelques points importants de la theorie de la chaleur. Ann Chim Phys 10:399 (On some important points of the theory of heat, in French)

Einstein A (1907) Die Plancksche Theorie der Strahlung und die Theorie der spezifischen Wärme. Ann Phys (Lpz.) 22:180 and 800 ("Berichtigung"). (Planck's theory of radiation and the theory of specific heat, in German)

Garg J, Chen G (2013) Minimum thermal conductivity in superlattices: a first-principles formalism. Phys Rev B 87:140302

Geballe TH, Hull GW (1958) Isotopic and other types of thermal resistance in germanium. Phys Rev 110:773

Glassbrenner CJ, Slack GA (1964) Thermal conductivity of silicon and germanium from 3 K to the melting point. Phys Rev 134:A1058

Grüneisen E (1926) Handbuch der Physik, vol 10. Springer, Berlin, Handbook of Physics, in German

Hahn KR, Melis C, Bernardini F, Colombo L (2021) Intrinsic thermoelectric figure of merit of bulk compositional SiGe alloys: a first-principles study. Phys Rev Mater 5:065403

Heron JS, Fournier T, Mingo N, Bourgeois O (2009) Mesoscopic size effects on the thermal conductance of silicon nanowire. Nano Lett 9:1861

Hiki Y (1981) Higher order elastic constants of solids. Ann Rev Mater Sci 11:51

Hochbaum AI, Chen RK, Delgado RD, Liang WJ, Garnett EC, Najarian M, Majumdar A, Yang P (2008) Enhanced thermoelectric performance of rough silicon nanowires. Nature 451:163

Holland MG (1964) Phonon scattering in doped GaAs from magnetothermal conductivity studies. In: Hulin M (ed) Proceedings of 7th international conference on the physics of semiconductors, Dunod, pp 713–717

Holt M, Wu Z, Hawoong Hong, Zschack P, Jemian P, Tischler J, Haydn Chen, Chiang TC (1999) Determination of phonon dispersions from x-ray transmission scattering: the example of silicon. Phys Rev Lett 83:3317

Hsieh W-P, Losego MD, Braun PV, Shenogin S, Keblinski P, Cahill DG (2011) Testing the minimum thermal conductivity model for amorphous polymers using high pressure. Phys Rev B 83:174205

Igamberdiev KT, Mamadalimov AT, Khabibullaev PK (1983) Izv Akad Nauk Uzb SSR, Ser Fiz Mat 2:39 (in Russian)

Joos G (1945) Lehrbuch der Theoretischen Physik. Akad. Verl. Gesellschaft, Leipzig (Textbook of Theoretical Physics, in German)

Ju YS, Goodson KE (1999) Phonon scattering in silicon films with thickness of order 100 nm. Appl Phys Lett 74:3005

Kang JS, Li M, Wu H, Nguyen H, Hu Y (2019) Basic physical properties of cubic boron arsenide. Appl Phys Lett 115:122103

Kim DS, Smith HL, Niedziela JL, Li CW, Abernathy DL, Fultz B (2015) Phonon anharmonicity in silicon from 100 to 1500 K. Phys Rev B 91:014307

Kittel C (1949) Interpretation of the thermal conductivity of glasses. Phys Rev 75:972

Kittel C (1986) Introduction to solid state physics. Wiley, New York

Klemens PG (1955) The scattering of low-frequency lattice waves by static imperfections. Proc Phys Soc A 68:1113

Klemens PG (1958) Thermal conductivity and lattice vibrational modes. In: Seitz F, Turnbull D (eds) Solid state physics, vol 7. Academic Press, New York

Klemens PG (1986) Thermal expansion of composites. Int J Thermophys 7:197

Klemens PG, Chu TK (1976) Thermal conductivity. Plenum Press, New York, pp 1–98

Klitsner T, Pohl RO (1987) Phonon scattering at silicon crystal surfaces. Phys Rev B 36:6551

Knaak W, Hauß T, Kummrow M, Meißner M (1986) Thermalization of ballistic phonon pulses in dielectric crystals below 1 K using time resolved thermometry. In: Anderson AC, Wolfe JP (eds) Phonon scattering in condensed matter, vol V. Springer, Berlin, p 174

Kumar GS, Vandersande JW, Klistner T, Pohl RO, Slack GA (1985) Low-temperature heat transport by charge carriers in doped semiconductors. Phys Rev B 31:2157

Landau LD, Lifshitz EM (1958) Statistical physics (English translation by Peierls E, Peierls RF). Addison-Wesley, Cambridge, MA

Leibfried G, Schlömann E (1963) Wärmeleitung in elektrisch isolierenden Kristallen. Nachr Akad Wiss Göttingen Kl (II)a 4:71 (1954); English translation AEC-tr-5892, 1 (1963). (Heat conduction in electrically insulating crystals, in German)

Lencer D, Salinga M, Wuttig M (2011) Design rules for phase-change materials in data storage applications. Adv Mater 23:2030

Li J, Zhou X, Liu Z (2020) Recent advances in photoactuators and their applications in intelligent bionic movements. Adv Optical Mater 8:2000886

Li DY, Wu YY, Kim P, Shi L, Yang PD, Majumdar A (2003) Thermal conductivity of individual silicon nanowires. Appl Phys Lett 83:2934

Lindsay L, Broido DA, Reinecke TL (2013) *Ab initio* thermal transport in compound semiconductors. Phys Rev B 87:165201

Liu W, Asheghi M (2006) Thermal conductivity measurements of ultra-thin single crystal silicon layers. J Heat Transf 128:75

Logachev YA, Vasilev LN (1973) Temperature dependence of the phonon thermal conductivity of Ge and Si and $A_{III}B_V$ compounds at high temperatures. Fiz. Tverdogo Tela 15:1612

Luckyanova MN, Garg J, Esfarjani K, Jandl A, Bulsara MT, Schmidt AJ, Minnich AJ, Chen S, Dresselhaus MS, Ren Z, Fitzgerald EA, Chen G (2012) Coherent phonon heat conduction in superlattices. Science 338:936

Luk'yanov AY, Ral'chenko VG, Khomich AV, Serdtsev EV, Volkov PV, Savel'ev AV, Konov VI (2008) Measurement of optical absorption in polycrystalline CVD diamond plates by the phase photothermal method at a wavelength of 10.6 μm. Quantum Electron 38:1171

Martienssen W, Warlimont H (2005) Springer handbook of condensed matter and materials data. Springer, Berlin

Matsumoto DS, Anderson AC (1981) Effect of crosslinking on the low-temperature behavior of polybutadiene. J Non-cryst Solids 44:171
McGaughey AJH, Jain A, Kim HY, Fu B (2019) Phonon properties and thermal conductivity from first principles, lattice dynamics, and the Boltzmann transport equation. J Appl Phys 125:011101
Minnich AJ, Johnson JA, Schmidt AJ, Esfarjani K, Dresselhaus MS, Ka N, Chen G (2011) Thermal conductivity spectroscopy technique to measure phonon mean free paths. Phys Rev Lett 107:095901
Mitra SS, Massa ME (1982) Lattice vibrations in semiconductors. In: Moss TS, Paul W (eds) Handbook on semiconductors, vol 1: Band theory and transport properties. North Holland, Amsterdam, pp 81–192
Molina-Ruiz M, Jacks HC, Queen DR, Wang Q, Crandall RS, Hellman F (2020) Two-level systems and growth-induced metastability in hydrogenated amorphous silicon. Mater Res Express 7:095201
Namjoshi KV, Mitra SS, Vetelino JF (1971) Simple shell-model calculation of lattice dynamics and thermal expansion of alkali halides. Phys Rev B 3:4398
Novikova SI (1966) Thermal expansion. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 2. Academic Press, New York, pp 33–48
Orbach R (1984) Dynamics of fractal structures. J Stat Phys 36:735
Parrott JE, Stuckes AD (1975) Thermal conductivity of solids. Pion, London
Pässler R (2011) Non-Debye behaviours of heat capacities of cubic II–VI materials. J Phys Chem Sol 72:1296
Patel CKN, Tam AC (1981) Pulsed optoacoustic spectroscopy of condensed matter. Rev Mod Phys 53:517
Peierls RE (1929) Zur kinetischen Theorie der Wärmeleitung in Kristallen. Ann Phys (Lpz) 3:1055 (On the kinetic theory of thermal conductivity in crystals, in German)
Peierls RE (1955) Quantum theory of solids. University Press, Oxford
Pennington DM, Harris CB (1992) Dynamics of photothermal surface expansion and diffusivity using laser-induced holographic gratings. IEEE J Quantum Electron 28:2523
Phillips WA (1987) Two-level states in glasses. Rep Prog Phys 50:1657
Pike NA, Løvvik OM (2019) Calculation of the anisotropic coefficients of thermal expansion: A first-principles approach. Computational Mater Sci 167:257
Pohl RO, Love WF, Stevens RB (1974) Lattice vibrations in non-crystalline solids. In: Stuke J, Brenig W (eds) Proceedings of 5th international conference on amorphous and liquid semiconductors. Taylor and Francis, London, p 1121
Pop E (2010) Energy dissipation and transport in nanoscale devices. Nano Res 3:147
Qian X, Zhou J, Chen G (2021) Phonon-engineered extreme thermal conductivity materials. Nat Mater 20:1188
Ravichandran J, Yadav A, Cheaito R, Rossen PB, Soukiassian A, Suresha SJ, Duda JC, Foley BM, Lee C-H, Zhu Y, Lichtenberger AW, Moore JE, Muller DA, Schlom DG, Hopkins PE, Majumdar A, Ramesh R Zurbuchen MA (2014) Crossover from incoherent to coherent phonon scattering in epitaxial oxide superlattices. Nat Mater 13:168
Radhakrishnan V, Roy KP, Sharma PC (1982) Proc Nucl Phys Solid State Phys Symp 25C:314
Rosencwaig A (1980) Photoacoustics and photoacoustic spectroscopy. Wiley, Chichester
Rowe DM, Bhandari CM (1986) Preparation and thermal conductivity of doped semiconductors. Prog Crystal Growth Charact 13:233
Roy R, Agrawal DK (1985) Successful design of new very low thermal expansion ceramics. Mater Res Soc Symp Proc 40:83
Schirmacher W (2006) Thermal conductivity of glassy materials and the “boson peak”. Europhys Lett 73:892
Simkin MV, Mahan GD (2000) Minimum thermal conductivity of superlattices. Phys Rev Lett 84:927
Slack GA (1979) The thermal conductivity of nonmetallic crystals. In: Seitz F, Turnbull D, Ehrenreich H (eds) Solid state physics, vol 34. Academic Press, New York, pp 1–71
Smith RA (1978) Semiconductors. Cambridge University Press, Cambridge

Soma T, Sato J, Matsuo H (1982) Thermal expansion coefficient of GaAs and InP. Sol State Commun 42:889
Sproull RL, Moss M, Weinstock H (1959) Effect of dislocations on the thermal conductivity of lithium fluoride. J Appl Phys 30:334
Steigmeier EF (1969) Thermal conductivity of semi-conducting materials. In: Tye RP (ed) Thermal conductivity. Academic Press, London/New York
Stephens RB (1973) Low-temperature specific heat and thermal conductivity of noncrystalline dielectric solids. Phys Rev B 8:2896
Svenson EC, Brockhouse BN, Rowe JM (1967) Crystal dynamics of copper. Phys Rev 115:619
Tam AC (1986) Applications of photoacoustic sensing techniques. Rev Mod Phys 58:381
Torchia GA, Schinca D, Khaidukov NM, Tocho JO (2002) The luminescent quantum efficiency of Cr_{3+} ions in Cs_2NaAlF_6 single crystals. Opt Mater 20:301
Touloukian YS, Liley PE, Saxena SC (1970) Thermal conductivity; nonmetallic liquids and gases. IFI/Plenum Press, New York
Tritt TM (2004) Thermal conductivity: theory, properties, and applications. Kluwer/Plenum Publishers, New York
Vakhabov DA, Zakirov AS, Igamberdiev KT, Mamadalimov AT, Tursunov SO, Khabibullaev PK (1985) Fiz Tverd Tela 27:3420 (Thermoconductivity and temperature-conductivity of silicon doped with Se and Te. in Russian)
Valentiner S, Wallot J (1915) Über die Abhängigkeit des Ausdehnungskoeffizienten fester Körper von der Temperatur, Ann. Physik 46:837 (On the dependence of the expansion coefficient of solids on the temperature, in German)
Vandersande JW (1980) Phonon scattering by nitrogen aggregates in intermediate type natural diamonds. In: Maris HJ (ed) Phonon scattering in condensed matter. Plenum Press, New York, pp 247–250
Vandersande JW, Pohl RO (1982) Negligible effect of grain boundaries on the thermal conductivity of rocks. Geophys Res Lett 9:820
Vandersande JW, Wood C (1986) The thermal conductivity of insulators and semiconductors. Contemp Phys 27:117
Venkatasubramanian R (2000) Lattice thermal conductivity and phonon localization like behavior in superlattice structures. Phys Rev B 61:3091
Walton D (1967) Scattering of phonons by a square-well potential and the effect of colloids on the thermal conductivity. I. Experimental. Phys Rev 157:720
Wang Z, Alaniz JE, Jang W, Garay JE, Dames C (2011) Thermal conductivity of nanocrystalline silicon: importance of grain size and frequency-dependent mean free paths. Nano Lett 11:2206
Waseda Y, Ohta H (1987) Current views on thermal conductivity and diffusivity measurements of oxide melts at high temperature. Sol State Ionics 22:263
Weißmantel C, Hamann C (1995) Grundlagen der Festkörperphysik. J.A. Barth Verlag, Heidelberg (Fundamentals of Solid State Physics, in German)
White TJ, Davis JH, Walter HU (1975) Thermal expansion and Grüneisen parameters of InBi. J Appl Phys 46:11
White TJ (ed) (2017) Photomechanical materials, composites, and systems. Wiley, Hoboken.
Yao T (1987) Thermal properties of AlAs/GaAs superlattices. Appl Phys Lett 51:1798
Yu CC, Freeman JJ (1986) The thermal conductivity and specific heat of glasses. In: Anderson AC, Wolfe JP (eds) Phonon scattering in condensed matter, vol V. Springer, Berlin, p 20
Yang F, Dames C (2013) Mean free path spectra as a tool to understand thermal conductivity in bulk and nanostructures. Phys Rev B 87:035437

Part III

Energy Bands

The Origin of Band Structure

6

Karl W. Böer and Udo W. Pohl

Contents

1 Approaches for Modeling 194
1.1 The Proximity Approach 194
1.2 The Periodicity Approach 199
1.3 Periodicity Versus Proximity Approach 204
2 The Reduced k Vector 207
2.1 Newtonian Description of a Quasi-Free Electron 208
2.2 The Effective Mass 210
3 The Proximity Approach in Organic Crystals 212
4 Summary 216
References 216

Abstract

Characteristic for much of the electronic behavior in solids is the existence of energy bands, separated by bandgaps. The bands are permitted for occupation with carriers, and their origin can be described by two complementary models. The *proximity approach* considers the effect of the neighborhood in a solid on the energy levels of an isolated atom; this model is particularly suited for organic semiconductors, amorphous semiconductors, and clusters of atoms. The *periodicity approach* emphasizes the long-range periodicity of the potential in a crystal. Electrons near the lower edge of a band in a crystal behave akin to electrons in vacuum; the influence of the crystal potential is expressed by an *effective* electron mass which increases with increasing distance from the band edge. This chapter describes the basic elements of the electronic band structure in solids.

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_6

Keywords

Band structure · Bandgap · Bloch function · Effective mass · HOMO · Kronig-Penney model · LUMO · Organic crystals · Periodicity approach · Proximity approach · Reduced k vector

1 Approaches for Modeling

Electronic transitions in energy and in space are the basic processes of interest in semiconductor physics. The first group is responsible for the large variety of excitation and de-excitation (recombination) processes; the second must be considered for carrier transport. Both are characterized by quantum-mechanical features: the spectrum of electronic energy states (*eigenstates*), the distribution of electrons over these states, and the interactions that cause changes in this distribution.

The general principles that yield the spectrum of energy states typical for the solid semiconductor are discussed first. The main features of this spectrum are obtained using two apparently different models, which both are applied to describe crystals. The first model, referred to as the *proximity approach*, starts from individual atoms and its immediate neighborhood and expands with less and less attention to the atomistic structure the further one extends from the origin. This approach is also successfully used for amorphous semiconductors. The second model, the *periodicity approach*, is at first view rather insensitive to the detailed properties of individual atoms but considers the long-range periodicity of a crystalline lattice. Both yield similar qualitative results: spectra of broad, permitted ranges of energy which, in space, extend as *bands* throughout the entire semiconductor and which are interspersed with forbidden ranges.

In this chapter both the proximity and the periodicity approach are presented, and common features along with some of the differences in the results are pointed out. The discussion starts from a rather heuristic description and introduces sequentially more sophisticated elements.

1.1 The Proximity Approach

In a simple first step, the exchange of electrons between two atoms can be made plausible by considering the splitting of eigenstates of degenerate oscillators, i.e., oscillator states having the same eigenfrequency, when they become coupled with each other. This splitting increases as the coupling gets stronger, corresponding to a closer approach of two atoms.

1.1.1 Formation of Energy Bands

The addition of more atoms of the same kind at increased distances splits the energy levels into more levels which span a *range* of energies. If the levels are spaced

closely enough, *Heisenberg's uncertainty principle* no longer permits distinction between the individual levels.[1] In this case, one obtains an *allowed energy range* in a sufficiently large cluster of atoms, instead of a discrete energy-level spectrum of a single atom or an aggregate of a few atoms – see Fig. 4. Since outer shell electrons can be exchanged more easily, the energy ranges created from valence electrons will be wider than the ranges created from the shielded inner electrons. The latter will more closely resemble the discrete eigenstates of isolated atoms. Since the same atoms behave alike, this allowed energy range extends throughout the crystal. In two dimensions (x,E) one therefore can draw allowed *energy bands* separated by *forbidden gaps* (Fig. 1).

In Fig. 1 the total electron energy is drawn disregarding the potential energy that an electron experiences when separated from an individual atom, which is shown in Fig. 2 for a single atom (upper curve, in green) and for a small one-dimensional cluster (blue curve). The band model emphasizes the *collective behavior*, i.e., the *sharing* of the electron among the atoms of the cluster. The potential distribution picture, on the other hand, emphasizes the *localization* of an electron within each

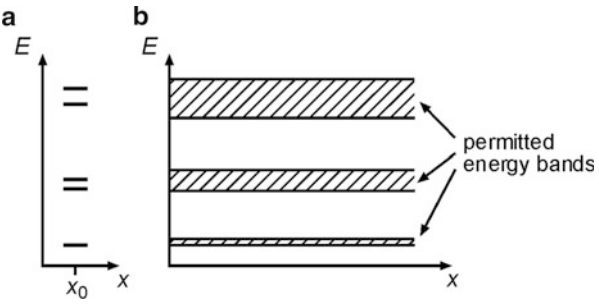


Fig. 1 (**a**) Splitting of eigenstates when two atoms have approached each other. (**b**) Simple band model of a crystal consisting of many atoms

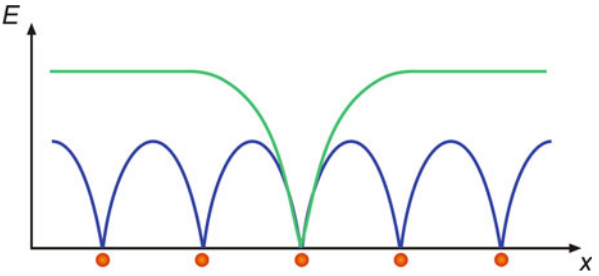


Fig. 2 Potential energy of an electron in an atom (*green curve*) and of an electron in an atomic cluster (*blue curve*); *red dots* below the abscissa indicate positions of atom cores

[1]Applying $\Delta E \Delta t \cong \hbar$ and relating Δt to the time an electron resides at a sufficiently high energy level E_{ik} (later identified as belonging to an upper band), an uncertainty of ΔE results. The time Δt is related to scattering (see Sect. 2 in ▶ Chap. 22, "Carrier-Transport Equations"); the electron is removed from this level after $\lambda/v_{\mathrm{rms}} \approx 10^{-12}$ s, yielding an uncertainty of ~1 meV, which is on the same order as the splitting provided by only 10^4 atoms (assuming a band width of ~1 eV and an equidistant splitting of 1 level per added atom – that is, within a crystallite of <100 Å diameter. With larger crystallites the splitting is even closer and results in a level continuum.

potential funnel. Both pictures are valid: the band picture is more relevant for higher bands, while the potential picture is more relevant for lower (core) levels.

When the band picture is superimposed to the picture of the individual potentials of many adjacent atoms (Fig. 3), we recognize that the semiclassical approach, in which electrons may only move above potential barriers, is inappropriate, since bands indicate that electronic exchange exists well below the crest of the barriers. The origin of the electron transfer through such barriers is *tunneling*, i.e., a *quantum-mechanical* exchange, – see Sect. 2.3 in ▶ Chap. 29, "Carrier Generation."

The formation of energy bands from atomic orbitals is illustrated in Fig. 4 for a successive assembly of atoms X. As an atom X is coupled to another atom X forming a molecule X_2, molecular orbitals are formed from the atomic orbitals as described in Sect. 1.2 of ▶ Chap. 2, "Crystal Bonding"; their eigenenergies are split as shown in Fig. 5a for the lowest states of H_2 formation. Further addition of X atoms introduces further splittings with gradually decreasing spacings until, for an exceedingly large number of atoms X such as ~10^{23} atoms/cm^3 in a crystal, continuous energy bands arise.

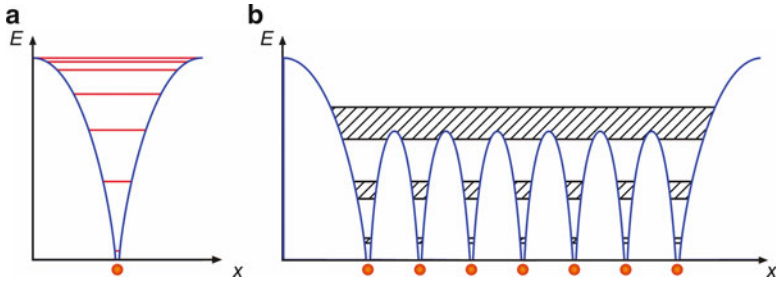


Fig. 3 Potential energy and eigenstates of electrons in (**a**) an atom and (**b**) a small crystal

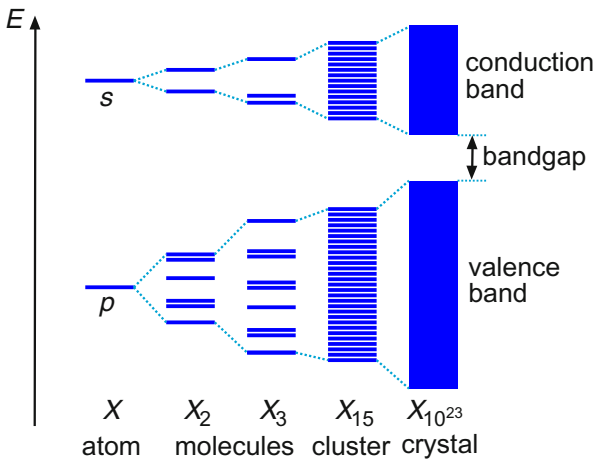


Fig. 4 Schematic illustrating the evolution of energy bands from atomic orbitals of an element X. As more and more X atoms are put together, the energy levels split and their spacing decreases

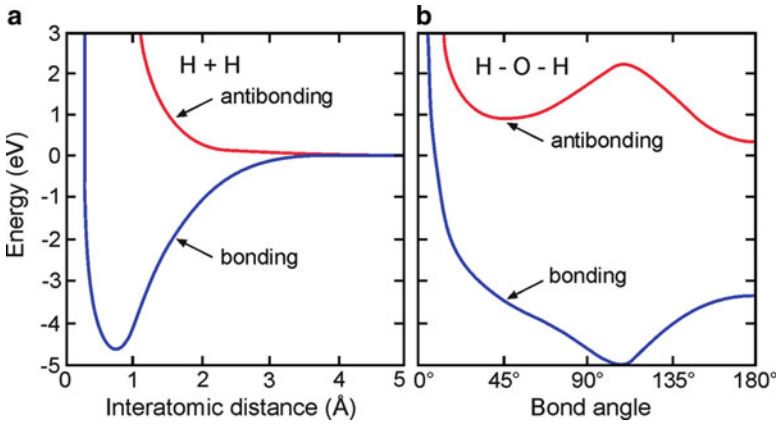


Fig. 5 Electron energy as a function of (**a**) the interatomic distance of a hydrogen molecule and (**b**) the bond angle in H_2O

1.1.2 Electronic Structure of Amorphous Semiconductors

The heuristic approach of the previous section will now be expanded to motivate the formation of bands in a more appropriate quantum-mechanical analysis. The analysis is first applied to a cluster of atoms that forms the building block of an amorphous semiconductor and then to a periodic lattice.

The electron energy depends sensitively on the interatomic distance and bond angle, as shown for a simple H_2 and an H_2O molecule in Fig. 5. The bonding and antibonding curves reflect antiparallel and parallel spin of the electrons in the bonds, respectively.

The electronic structure of an amorphous or crystalline solid can be obtained by starting from an arbitrary atom and including more and more neighbors in an appropriate configuration; the eigenfunctions of such a cluster are determined by solving its Schrödinger equation. This is referred to as a *tight-binding approach*. Solutions can be obtained numerically, using reasonable approximations (Reitz 1955; Heine 1980; Slater and Johnson 1972; Kaplan and Mahanti 1995; Jancu and Voisin 2007 – see also ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"). The analysis can be described as that for a large molecule of, say, 20–50 atoms and delivers a spectrum of *energy eigenvalues* that, when the cluster is large enough, presents a valuable estimate of the energy bands of the solid.

Several Si atoms which form such a cluster produce the typical sp_3 bonding and antibonding states that are generated when the atoms are close enough – see Sect. 1.2 and Fig. 6 in ▶ Chap. 2, "Crystal Bonding." For a small cluster of Si atoms in an "amorphous configuration" (Sect. 3.2 in ▶ Chap. 3, "The Structure of Semiconductors"), we calculate an eigenvalue spectrum (Fig. 6a) and see that the proximity to other Si atoms significantly changes and splits the atomic levels (shown on the left). They are split by a large amount and are distributed unevenly in energy. A much larger number of atoms, however, are required to create a truly band-like level distribution. The spectrum changes significantly when hydrogen is

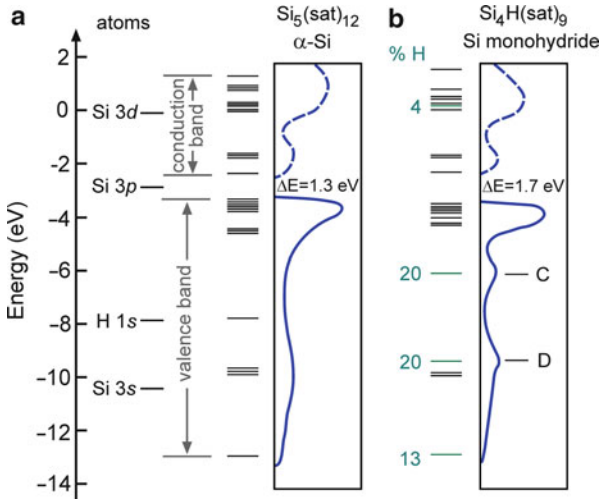


Fig. 6 Electron energy-level distribution for bonding (valence band) and antibonding (conduction band) states of a 17-atom cluster and optical density of states (see Sect. 1 in ▶ Chap. 13, "Band-to-Band Transitions") for (**a**) α-Si and (**b**) α-Si:H. Amorphous silicon α-Si is modeled by a Si-$(SiH_3)_4$ cluster denoted $Si_5(sat)_{12}$ with a central Si bond to 4 Si, each of which is saturated by 3 H. In the hydrogenated cluster 1 H atom is substituted for 1 Si, yielding a SiH-$(SiH_3)_3$ cluster corresponding to a Si:16% H alloy. Principal hydrogen-induced levels are marked in green and labeled according to the H contribution in the molecular orbitals (After Johnson et al. 1980). The agreement with the experimental bandgap of such a small cluster is spurious, however, and should not be over-evaluated. The calculated bandgap depends substantially on the boundary conditions

added to this cluster, which forms a bridging hydrogen structure (Ovshinsky and Adler 1978). It removes states from the gap of α-Si and thereby increases the bandgap from 1.3 eV for α-Si to 1.7 eV for the technically more interesting, hydrogenated α-Si:H (Eberhart et al. 1982; Street 2005).

When more atoms of the same kind are incorporated within such a cluster, more levels appear *within* the two bands, i.e., within the range of bonding and antibonding states. These bands have been labeled in Fig. 6 as the conduction and valence bands; for more detailed definitions, see Sect. 1 in ▶ Chap. 8, "Bands and Bandgaps in Solids." The finite-cluster approach always overestimates the bandgap energy, since, by necessity, this approach omits states present which are far away from the center in **k** space and in direct bandgap materials (Sect. 2 in ▶ Chap. 13, "Band-to-Band Transitions") lie near the edges of the bandgap.

From these examples, we can deduce that some information about the energy width of the upper energy bands and the bandgap can be obtained from clusters containing only ~50 atoms in an appropriate structural configuration. That is, the outer atoms must be kept artificially at positions they would attain when interacting with the surrounding atoms within a much larger amorphous network of atoms. The level distribution *within* a band, however, is poorly represented by such small clusters. The incorporation of many more atoms presents major computational problems for amorphous semiconductors; however, this problem becomes

exceedingly simple in the periodic lattice of a crystalline semiconductor. For more reading, see Adler (1985), Agarwal (1995), Beeby and Hayes (1989), Shinozuka (1999), and Singh and Shimakawa (2003).

1.2 The Periodicity Approach

The behavior of electrons in a semiconductor can be approximated by assuming that they are nearly *free electrons* but interact with the *periodic potential* that simulates the lattice. In order to distinguish the influence of this periodic potential, one should first recall the behavior of a free electron with mass m_0 *in vacuum*. This is determined by the solution of the *Schrödinger equation*

$$\frac{\partial^2}{\partial \mathbf{r}^2}\psi + \frac{2m_0}{\hbar^2}E\psi = 0, \tag{1}$$

which can be described by an *electron wave*

$$\psi(\mathbf{r}) = A\exp(\pm i\,\mathbf{k}\cdot\mathbf{r}), \tag{2}$$

with A as an amplitude factor. The *wavevector* $\mathbf{k}$ relates to electron momentum and energy as

$$\mathbf{k} = \frac{m_0\mathbf{v}}{\hbar} = \frac{\mathbf{p}}{\hbar}, \qquad E = \frac{m_0}{2}v^2 = \frac{p^2}{2m_0} = \frac{\hbar^2k^2}{2m_0}, \tag{3}$$

or, more accurately, to the expectation value of the momentum given by

$$\langle\mathbf{p}\rangle = \int_{-\infty}^{\infty}\psi^*\frac{\hbar}{i}\frac{\partial}{\partial\mathbf{r}}\psi\;\,d\mathbf{r} = \hbar\mathbf{k}\int_{-\infty}^{\infty}\psi^*\psi = \hbar\mathbf{k}. \tag{4}$$

The wavevector $\mathbf{k}$ is the reduced wavevector – see the discussion later in this section and Fig. 14. Hence, $E(\mathbf{p})$ or $E(\mathbf{k})$ is described by a three-dimensional paraboloid (by a parabola in one relevant coordinate) with one electronic parameter, the *electron rest mass* m_0.

Equation 2 represents an electron wave with a wavelength, the *de Broglie wavelength*,[2] of

$$\lambda_{\mathrm{DB}} = \frac{2\pi}{k} = \frac{h}{|p|} = \frac{h}{m_0 v} = \frac{7.27}{v\ (\mathrm{cm/s})}\ \ (\mathrm{cm}) \tag{5}$$

[2]The de Broglie wavelength is on the same order of magnitude as the uncertainty distance obtained from Heisenberg's uncertainty principle $\Delta x \geq \hbar/\Delta p_x$, which has the same form as λ_{DB}. This yields uncertainty distances of 10 Å for thermal (free) electrons at room temperature.

or, when introducing the electron energy from Eq. 3,

$$\lambda_{\mathrm{DB}} = \frac{h}{\sqrt{2m_0 E}} = \frac{12.26}{\sqrt{E(\mathrm{eV})}}(\text{Å}). \tag{6}$$

An electron in the lattice, i.e., when it is exposed to a periodic potential, no longer behaves like a free particle: it experiences interference from the lattice potential when, with increasing electron energy, its de Broglie wavelength becomes comparable to the lattice constant. The ensuing *Bragg reflections* prohibit a further acceleration of the electron, described later in more detail. This simple discussion also indicates the existence of a finite energy range, the energy band in a semiconductor. Near the bottom band edges, the electron behaves to some extent like a free electron, i.e., like a classical particle. The quantum-mechanical nature becomes evident when it gains energy in an electric field or is forced to occupy higher states. At energies of 4 eV, the de Broglie wavelength is 6 Å, i.e., small enough to permit interference effects within the periodic potential of the lattice. This plausibility argument can be substantiated by describing the electron with a wave equation, the Schrödinger equation, and by introducing into the Schrödinger equation a periodic potential $V(\mathbf{r})$,

$$\frac{\partial^2}{\partial \mathbf{r}^2}\psi + \frac{2m_0}{\hbar^2}(E(\mathbf{k}) - V(\mathbf{r})) \quad \psi = 0. \tag{7}$$

The solutions of this Schrödinger equation are so-called Bloch functions which can be expressed as a linear combination of waves

$$\psi_n(\mathbf{k},\mathbf{r}) = u_n(\mathbf{k},\mathbf{r})\exp(i\ \mathbf{k}\cdot\mathbf{r}), \tag{8}$$

with n as the band index specifying a certain band. The waves are plane waves with a space-dependent amplitude factor $u_{\mathrm{n}}(\mathbf{k},\mathbf{r})$, which shows lattice periodicity (Bloch's theorem, 1928; see Sect. 1.2 in ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"). A one-dimensional schematic representation is given in Fig. 7 to indicate the relationship between the lattice potential $V(\mathbf{r})$ and the Bloch function $\psi_{\mathrm{n}}(\mathbf{k},\mathbf{r})$, which contains $u_{\mathrm{n}}(\mathbf{k},\mathbf{r})$ and the plane wavefunction of the electron $\exp(i\mathbf{k}\cdot\mathbf{r})$, to construct the electron wavefunction (Harrison 1980a).

Inserting Eq. 8 into Eq. 7, we obtain the result that solutions exist only for certain ranges of the electron energy $E_n(\mathbf{k})$, which are interspersed with energy ranges in which real solutions do not exist. This confirms the previously obtained results that the energy spectrum in a solid consists of alternating allowed and forbidden energy ranges (energy bands). The periodic-potential approach, however, gives additional information that can be demonstrated readily in a simple one-dimensional model.

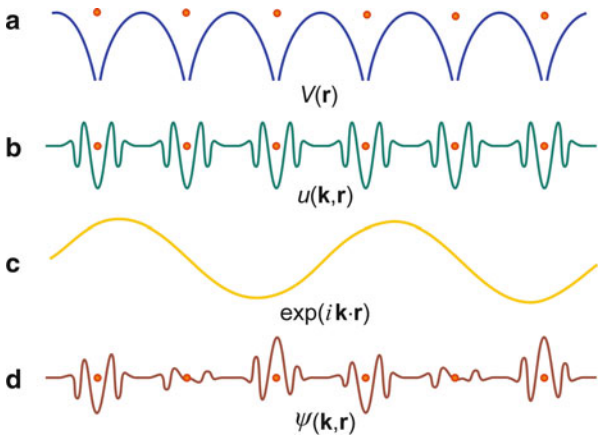


Fig. 7 A schematic representation of electronic eigenstates in a crystal. (**a**) The potential $V(\mathbf{r})$ plotted along a row of atoms; (**b**) $u(\mathbf{k},\mathbf{r})$, which has the periodicity of the lattice; (**c**) a plane electron wave, the real part of which is shown to construct the electron wavefunction; and (**d**) a Bloch function; the state itself is complex, and only the real part is shown. The Bloch function is composed of the product of (**b**) and (**c**)

1.2.1 The Kronig-Penney Model

An enormously simplified periodic potential $V(x)$ is sufficient for introduction into Eq. 9 to show the typical behavior. This is the *Kronig-Penney potential* (Kronig and Penney 1931),[3] which is shown in Fig. 8. Since the discussion of this behavior is rather transparent, it will be used here for an introduction to the basic features of the band model.

Introducing Eq. 8 into Eq. 7 for one relevant dimension, we see that $u(x)$ must satisfy

$$\frac{d^2u}{dx^2} + 2ik\frac{du}{dx} - \left(k^2 - \frac{2m_0[E - V(x)]}{\hbar^2}\right) \quad u = 0. \tag{9}$$

We can split Eq. 9 after the introduction of the Kronig-Penney potential into two differential equations: one for the bottom of the well and one for the top of the barrier with a potential $V = V_0$. The solutions in each part can be expressed as the sum of two waves:

$$\begin{aligned} u_1(x) &= A\exp[i(\alpha - k)x] + B\exp[-i(\alpha + k)x] \quad \text{for} \quad 0 < x < a_1 \\ u_2(x) &= C\exp[i(\beta - k)x] + D\exp[-i(\beta + k)x] \quad \text{for} \quad -a_2 < x < 0, \end{aligned} \tag{10}$$

where α and β are the k values for a free electron in vacuum, for $V = 0$, and for a constant barrier potential V_0, respectively:

[3] In one dimension, there are other periodic potentials for which the Schrödinger equation can be integrated explicitly. $V(x) = -V_0\,\mathrm{sech}^2(\gamma x)$ is one such potential, which yields solutions in terms of hypergeometric functions (see Mills and Montroll 1970). The results are quite similar to the Kronig-Penney potential.

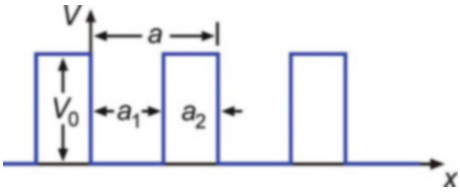


Fig. 8 Kronig-Penney potential with V_0 the barrier height, a_1 and a_2 the well and barrier widths, respectively

$$\alpha = \sqrt{\frac{2m_0E}{\hbar^2}} \quad \text{and} \quad \beta = \sqrt{\frac{2m_0(V_0 - E)}{\hbar^2}}. \tag{11}$$

The integration constants can be determined by the continuity requirements of $u(x)$ and its first derivatives at $x = a_1$ and $x = a_2$, which yields[4]

$$-\frac{\alpha^2 - \beta^2}{2\alpha\beta} \sin(\alpha a_1)\sinh(\beta a_2) + \cos(\alpha a_1)\cosh(\beta a_2) = \cos(ka). \tag{12}$$

Equation 12 provides the *dispersion relation* $E(k)$ (E is contained in α and β).

The dispersion relation is the key to many discussions of electronic properties in solids. Since the wavenumber k is proportional to the electron momentum (Eq. 3), the dispersion equation relates the electron energy to mass and velocity, both of which are essential for understanding the specific behavior of electrons in a semiconductor. This will be explained in detail in several of the following sections.

Equation 12 reveals that a sequence of allowed energy ranges (*bands*) is interspersed with forbidden energy ranges: energy gaps are formed when the left-hand side (LHS) of Eq. 12 exceeds ± 1, which are the limiting values of the equation's right-hand side. In Fig. 9 the hatched ranges show the energy bands; no solution of the Schrödinger equation can be found between the bands for real values of k. This picture describes a situation between a free electron in vacuum, where *all energies* are permitted, and an electron bound to an isolated atom, where the permitted energy ranges shrink to a *set of discrete energy levels*. The height and width of the potential barriers and wells (a_1, a_2, and V_0) determine whether an electron behaves more like an electron bound to a single atom (large a_2/a_1 and V_0) or more like a free electron in vacuum (small a_2 and V_0); see Fig. 10. In the latter example the permitted ranges extend over a wider energy range.

More information can be deduced from the $E(k)$ behavior within each of the permitted energy ranges shown in Fig. 11. At the bottom of the first permitted energy

[4]For $E > V_0$, the square root in β becomes imaginary. Introducing $\gamma = i\sqrt{2m_0(e - V_0)/\hbar^2}$ and with $\sinh(i\gamma) = i\sin\gamma$ and $\cosh(i\gamma) = i\cos\gamma$, we obtain for higher electron energies a similar equation:

$$-\frac{\gamma^2 + \alpha^2}{2\alpha\gamma} \sin(\gamma a_2)\sin(\alpha a_1) + \cos(\gamma a_2)\cos(\alpha a_1) = \cos(ka).$$

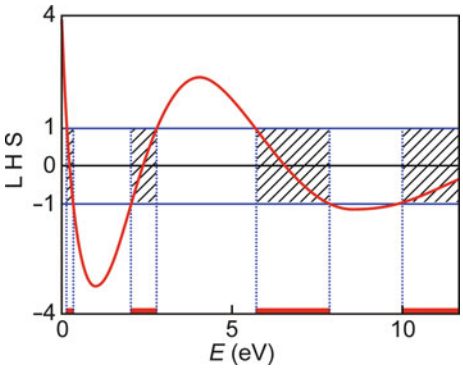


Fig. 9 Left-hand side (LHS) of Eq. 12 as a function of E (contained in α and β), computed for $a_1 = 6$ Å, $a_2 = 1.2$ Å, and $V_0 = 10$ eV

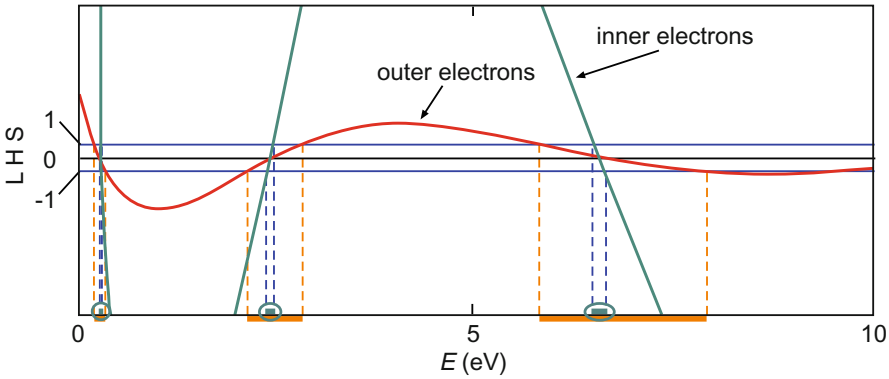


Fig. 10 LHS as in Fig. 9, but for two different values of the parameter V_0 (10 eV for curve labeled "outer electrons" and 40 eV for "inner electrons" – other parameters as in Fig. 9), indicating the reduced width of the permitted bands for higher potential barriers (i.e., for inner electrons that are more tightly bound) represented by encircled *green* bars at the E axis; bands of outer electrons are represented by *orange* bars

range, $E(k)$ is nearly parabolic. Then E moves with increasing k through an inflection point and, at the top of this range, becomes nearly parabolic again but with a negative curvature.

Compared to the parabola of the free electron, the lower part of the $E(k)$ curve is raised. At the upper edge of the first allowed range, i.e., at $k = \pm\pi/a$, the curve coincides again with the free electron parabola. The next permitted band starts after a jump in E from E_1 to E_2 at $k = \pm\pi/a$ and has a similar $E(k)$ behavior as the first energy band, except that the curvatures are larger at the bottom and top of the band. The top is reached at $k = \pm 2\pi/a$, where again a jump of E occurs, from E_2 to E_3, etc. (Fig. 11). This behavior continues for higher bands with broader allowed bands, gradually increased curvature at the bottom and the top, and narrower bandgaps. Figure 11 also contains $E(k)$ for the free electron (Eq. 3), which is parabolic in the entire $E(k)$ range.

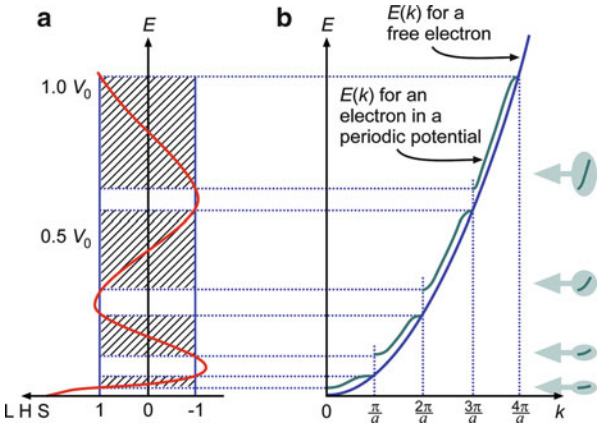


Fig. 11 (**a**) As Fig. 9, however, for a larger a_1/a_2 ratio. (**b**) $E(k)$ for a free electron (parabola) and for a Kronig-Penney potential in an extended wavenumber $E(k)$ representation. *Bottom* segments of the bands shown at the *right* indicate the increased curvature at the edge of higher bands

This general behavior is independent of the actual shape of the periodic potential as long as it has sufficient strength. Although periodicity of $V(x)$ is a necessary – but not sufficient – condition for energy bands with interspersed forbidden gaps, it so happens that in solids, for inner shell electrons, the potential barriers are sufficiently high to cause rather narrow, lower bands. Electrons at sufficiently high energies occupy wider bands and behave more like free electrons: they can move readily through the lattice. They will, however, be subject to interference with the periodic lattice potential (see Sect. 2.2).

When analyzing the effect of a *three-dimensional* periodic potential and using a real lattice potential, the actual $E(k)$ behavior becomes more complex; however, it still maintains the basic features of *energy bands* interspersed with *bandgaps*. This fundamental behavior is the basis for the electronic behavior of semiconductors and is described in more detail in many textbooks of solid state physics, e.g., Anderson (1963), Ashcroft and Mermin (1976), Bube (1992), Callaway (1976), Fletcher (1971), Harrison (1980b), Haug (1972), Kittel (2007), Marder (2010), and Ziman (1972).

1.3 Periodicity Versus Proximity Approach

The proximity of a sufficient number of atoms and the periodic lattice structure of a crystal both lead to allowed energy bands interspaced by bandgaps. We may use one or the other picture to obtain further information about the band structure.

The *periodic lattice-structure approach* is more suited for obtaining the specific $E(k)$ structure of the inner part of the band (near $k = 0$) which cannot be obtained from the proximity approach. It reflects the symmetry of the lattice and allows for obtaining the results in the most economical way. Its results, however, are restricted to periodic lattices, i.e., to crystals. This refers specifically to interference phenomena involving diffraction from further-than-nearest and next-nearest neighbor

distances. These distances, however, can still be discerned in the x-ray diffraction of amorphous semiconductors and therefore may also be expected to influence electron behavior further away from $k = 0$.

The *proximity approach* can be used to obtain some information about the inside of the bands for first orientation. However, the inadvertent inclusion of artificial states at the surface of the cluster and the requirement for an extremely large cluster size to provide band states close to the band edges have been the handicaps of this approach.

A *supershell approach* is sometimes used to avoid some of the shortcomings of the periodic lattice and proximity approaches. This approach takes a cluster of sufficient size and repeats it *periodically* until the entire crystal volume is filled. In this way the mathematical methods developed for studying periodic lattices can be used, while certain elements of an amorphous structure are included in the cluster. The error due to the forced adjustment of each cluster can be minimized by increasing the size of the cluster.

Many of these results are important for understanding the behavior of metals (e.g., overlapping bands), but will not be discussed here. Other results relate to semiconductors, including semiconductor-metal transitions (see Sect. 3 in ▶ Chap. 21, "Equilibrium Statistics of Carriers"). Some heuristic examples of near-band-edge properties are given below.

1.3.1 Band-Edge Fluctuation

The ideal periodicity of a crystal lattice can be modified for a number of reasons, among them lattice oscillations or displaced lattice atoms. An amorphous semiconductor, for example, may be described by having frozen-in large fluctuations of the interatomic distances and bond angles. In some respects such structures are "almost crystalline," but with slightly changing lattice parameters, occurring particularly in the third-neighbor distance and beyond; see Fig. 29 in ▶ Chap. 3, "The Structure of Semiconductors." The good short-range order in an amorphous semiconductor leads to a band structure comparable to that of a perfectly ordered crystal; however, a lattice with a different lattice constant causes a different $E(\mathbf{k})$ with a different width of allowed bands and gaps. Therefore, we expect variations of the *band edges* in time and space.[5] Rather than being perfectly straight, the band edge becomes perturbed. Over a time average, and the band edge appears to be fuzzed out (Fig. 12). For further detail see ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors."

[5]This concept must be used with caution, since $\mathbf{k}$ is a good quantum number only when electrons can move without scattering over at least several lattice distances. That is certainly not the case in most amorphous semiconductors near the "band edge" (see Sect. 4 in ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects"). However, at higher energies further inside the band, there is some evidence that the mean free path (Sect. 2 in ▶ Chap. 22, "Carrier-Transport Equations") is much larger than the interatomic distance even in amorphous semiconductors. In bringing the two approaches together, the argument presented here lacks rigor and has plausibility only in terms of correspondence.

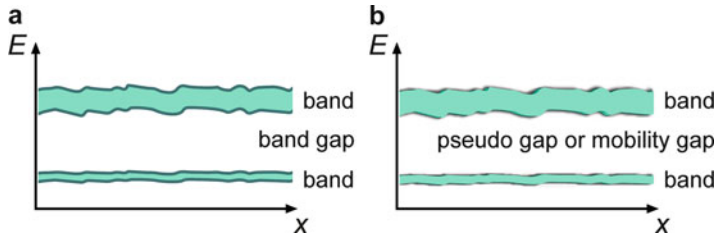


Fig. 12 (**a**) Perturbed and (**b**) fuzzed-out band edges in a crystal with phonons and in an amorphous semiconductor

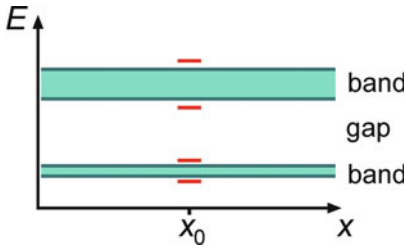


Fig. 13 Simple intrinsic atomic interstitial (i.e., an atom being chemically identical to the atoms of the crystal) in an idealized lattice

1.3.2 Discrete Defect Level in the Bandgap

When a local deviation from the ideal lattice structure is sufficiently large, the eigenstate of a *disordered atom* may lie within the bandgap. A plausibility argument may be obtained from the proximity model.

Assume an extra atom is incorporated in an *interstitial site* of the lattice (later discussed in Sect. 2.1 in ▶ Chap. 15, "Crystal Defects"). This extra atom is much closer to its neighbors; the exchange frequency is substantially larger than that for the nearest neighbors which yield the largest exchange frequency in an ideal lattice (equivalent to the band edges). Thus, the eigenstates of this interstitial atom, here an *intrinsic point defect*, and *its* nearest neighbors lie outside of the allowed bands of the ideal lattice, i.e., within the bandgaps (Fig. 13).

Energy states within the gap are *localized* at the position of this lattice defect (x_0 in Fig. 13) and play an important role in localizing (*trapping*) electrons in real crystals (see Sect. 2 in ▶ Chap. 19, "Deep-Level Centers") and in amorphous semiconductors. It also becomes reasonable to expect an energy distribution of such localized (*trap*) levels in the gap near the band edge, when taking into consideration that in crystalline and amorphous semiconductors a wide variety of lattice imperfections and lattice parameter variations are observed. In Sect. 3 of ▶ Chap. 17, "Optical Properties of Defects" and in ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors" we will return to this level distribution near the band edge.

2 The Reduced k Vector

A general feature of the solutions of the Schrödinger equation is the periodicity of $E(k)$, illustrated in Fig. 14b. The figure shows the periodicity in $\mathbf{k}$ along k_x with a period length of $k_x a = 2\pi$. This means that a shift of the solution $E(k_x)$ by $2\pi/a$ in k_x represents the same behavior. Fig. 14a contains a copy of Fig. 11, indicating the relation to the periodicity of $E(k)$ shown in Fig. 14b. Any full segment of the periodic representation is a *reduced k-vector representation*. One such segment is shown within the first *Brillouin zone* in Fig. 14c, i.e., within $-\pi/a < k_x < \pi/a$. For a three-dimensional lattice, the reduced representation is discussed in Sect. 4 in ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals."

The reduced representation $E(k)$ shows an alternating sign of the curvature at the edge of each band at $k = 0$. It is positive for the first band, negative for the second, etc. This interesting peculiarity occurs in real crystals in a somewhat similar fashion, although it is more complex because of a multiplicity of bands, as will be discussed in Sect. 1.2 in ▶ Chap. 8, "Bands and Bandgaps in Solids."

It is instructive to look at an enlarged detail of Fig. 14 as shown in Fig. 15d. This figure can be constructed from two parabolas of free electrons, shifted by $\frac{2\pi}{a}$ – a situation

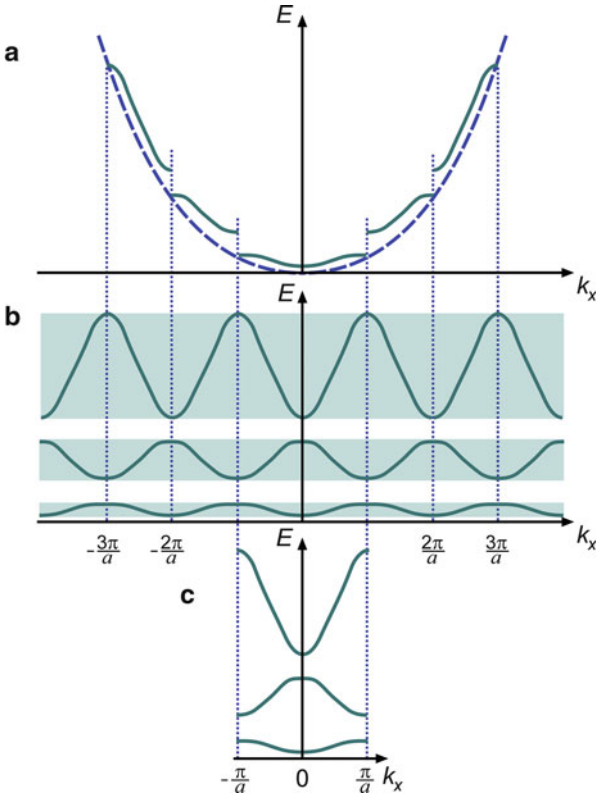


Fig. 14 Comparison between: (**a**) extended wavenumber k, (**b**) periodic, and (**c**) reduced wavenumber representations of $E(k)$

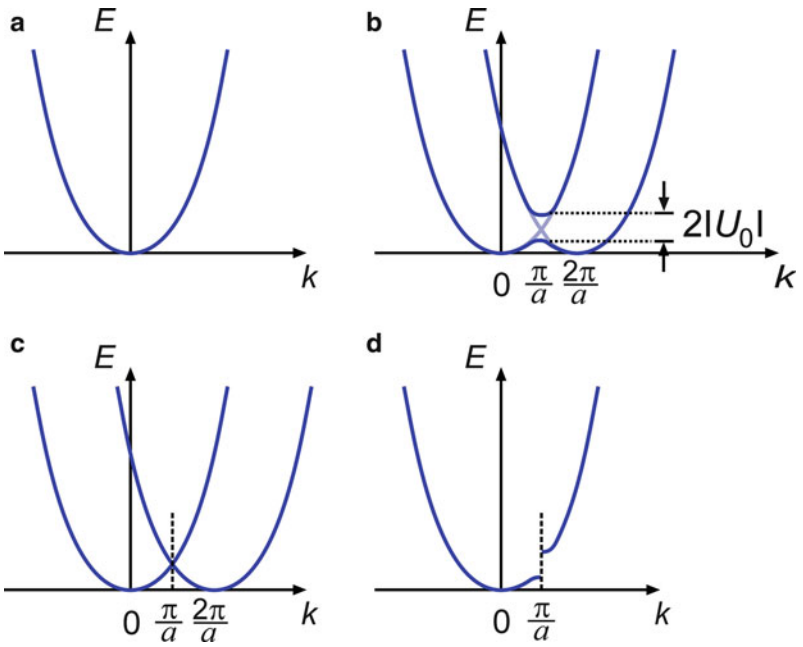


Fig. 15 (**a**) $E(k)$ for a free electron. (**b**) Splitting of $E(k)$ if a small periodic field is introduced. (**c**) Free electron in an empty lattice. (**d**) $E(k)$ of the original parabola, disturbed by the periodic-potential perturbation (compare with Fig. 14a)

that can be thought of by inserting a lattice into the vacuum, although with vanishing lattice potential (an *empty lattice*– see Sect. 4.1 in ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"). The electron in each reference system is described by its corresponding parabola (subfigure 15c). When interacting through a periodic perturbation potential of amplitude U_0, the crossing of both $E(k)$ parabolas is eliminated and a splitting occurs with a gap on the order of $|2\ U_0|$, as shown in subfigures b and d.

2.1 Newtonian Description of a Quasi-Free Electron

In many discussions about electron behavior in solids, a classical particle picture is used rather than the quantum-mechanical one of a wave packet; it is often more intuitive. Electrons behave like little balls, "sliding down" a potential hill and "scattering" upon collision with an atomic lattice defect. The picture is justified by using Bohr's *correspondence principle* near the bottom of the conduction band (the band in which electron conduction takes place). However, since not all of the electron behavior can be described by this model, as explained in the previous section, we can account for the modification by incorporating the information obtained from its dual nature as a wave into one of its classical parameters – its *mass*.

An electron, regarded as a classical (Newtonian) particle, has a momentum

$$p = m_0 v \quad \text{and a kinetic energy} \quad E = \frac{m_0}{2} v^2 = \frac{p^2}{2m_0}. \tag{13}$$

Relativistic effects are excluded here (i.e., $v \ll c$ is assumed): the electron mass is its rest mass m_0. The velocity of such a particle changes with time in response to an acting force $\mathcal{F}$ (*Newton's second law*):

$$\frac{dp}{dt} = m_0 \frac{dv}{dt} = \boldsymbol{\mathcal{F}}. \tag{14}$$

On the other hand, an electron in vacuum, regarded as a wave, has

$$\text{a momentum} \quad p = \hbar k \quad \text{and an energy} \quad E = \frac{p^2}{2m_0} = \frac{\hbar^2 k^2}{2m_0}. \tag{15}$$

When exposed to a force, such as that supplied by an electric field $|\mathbf{E}|$, and with $\mathcal{F} = -e|\mathbf{E}|$, its momentum increases accordingly:

$$\frac{dp}{dt} = \hbar \frac{dk}{dt} = \boldsymbol{\mathcal{F}}. \tag{16}$$

When the electron is described as a *wave packet*, its velocity is the *group velocity*[6]

$$v_g = \frac{\partial \omega}{\partial k} = \frac{1}{\hbar} \frac{\partial E}{\partial k}. \tag{17}$$

Applying Newton's law to such an electron wave packet (in a relation similar to Eq. 14),

$$m_0 \frac{dv_g}{dt} = \boldsymbol{\mathcal{F}}, \tag{18}$$

and with

[6]In an infinite crystal, the electron (when not interacting with a localized defect) is not localized and is described by a simple wavefunction (i.e., having one wavelength and the same amplitude throughout the crystal). The probability of finding it is the same throughout the crystal ($\propto \psi^2$). When localized, the electron is represented by a superposition of several wavefunctions of slightly different wavelengths. The superposition of these wavefunctions is referred to as a *wave packet*. A moving electron is represented by a moving wave packet $\psi = \frac{1}{2\delta k} \int_{k-\delta k}^{k+\delta k} u(x,k) \exp(i(kx - \omega t)) dk$ which quickly spreads out over time. It has its maximum at a position $\overline{x} = \frac{1}{\hbar} \frac{\partial E}{\partial k} t$, yielding for the group velocity, i.e., the velocity of the maximum of the wave packet, $v_g = \frac{\partial \overline{x}}{\partial t} = \frac{1}{\hbar} \frac{\partial E}{\partial k}$. With $E = \hbar\omega$, we obtain $v_g = \frac{\partial \omega}{\partial k}$.

$$\frac{dv_g}{dt} = \frac{1}{\hbar}\frac{d}{dt}\left(\frac{\partial E}{\partial k}\right) = \frac{1}{\hbar}\frac{\partial^2 E}{\partial k^2}\frac{dk}{dt} = \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2}\mathcal{F}, \tag{19}$$

we see by comparison with Eq. 18 that the factor preceding $\mathcal{F}$ has the dimension of an inverse mass. This factor is proportional to the curvature of $E(k)$.

2.2 The Effective Mass

If we want to retain the Newtonian behavior, we have to replace the electron mass m_0 in Eq. 18 with the *effective electron mass* when comparing Eqs. 18 and 19[7]:

$$m^* = \frac{\hbar^2}{\dfrac{\partial^2 E}{\partial k^2}}. \tag{20}$$

This effective mass contains the peculiarities of the interaction of the electron with the lattice (the superscript * distinguishes this mass from the rest mass). However, a possibly important part caused by the adiabatic approximation (see Born-Oppenheimer approximation Sect. 1.1 in ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals") is missing. The influence of this part is discussed in Sect. 1.2 in ▶ Chap. 22, "Carrier-Transport Equations," and can be described by a different effective mass – the polaron mass.

From Fig. 11, we see that the effective electron mass at the lower edge of the third band, here assumed to harbor free electrons (see Sect. 1.1 in ▶ Chap. 8, "Bands and Bandgaps in Solids"), is smaller than the rest mass of a free electron, since the curvature of $E(k)$ is larger here. At higher energies within the band, this curvature decreases, changes sign, and, at the upper edge of the band, becomes negative (as shown in Fig. 16). Consequently, the effective electron mass increases, becomes infinite near the center of an allowed band, and changes sign there. Coming from negative infinity, the effective electron mass returns to a finite but negative value which, at the top of the band, is of the same order of magnitude as at the bottom of the band (Figs. 14 and 16). This behavior is repeated in the next band, except that the sign sequence is exchanged. Here the effective mass is negative at $k = 0$; however, the effective electron mass is always positive at the bottom of any band and negative at the top. For lower bands, i.e., narrower bands, the value of the effective mass becomes larger at the band edge.

When electrons accelerate substantially above the lower edge of the band in sufficiently high fields, the de Broglie wavelength of the electron becomes smaller

[7]For the electron behavior, only *expectation values* can be given. In order to maintain Newton's second law, we continue to use $\hbar k$ (Eq. 15), which is no longer an electron momentum. It is well defined within the crystal and is referred to as *crystal momentum*. We then separate the electron properties from those of the crystal by using $\partial^2 E/\partial k^2$ to define its *effective mass*.

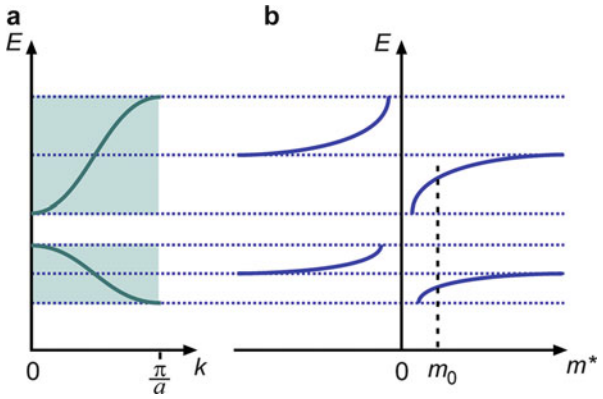


Fig. 16 (**a**) Typical $E(k)$ dependence for two simple bands, and (**b**) derived effective electron masses m^* within these permitted energy bands. Actually, one determines $m^*(k)$; this graph is turned by 90° to show its relation to the band model shown at the *left*

and comparable to the interatomic lattice spacing. Here, *interference effects* of the electron wave with the periodic lattice potential become important: Bragg reflection becomes more prevalent, while more and more frequency components of the wave packet are reflected. Therefore, further acceleration will become more difficult to achieve; in the Newtonian model, the effective mass of the electron increases until, near the center of the band, further acceleration stops. When the energy of the electron is raised above the center of the band, the electron will *decelerate* in the direction of the electric field until it reaches the top of the band, where it will come to a standstill. The electron wave has then reached a perfect diffraction condition.[8] It can be described as a *standing* wave, composed of incoming and refracted waves of exactly the same amplitude. With some caution we may describe the "recoil" of the lattice as being responsible for absorbing an increasing fraction of the electron momentum when it is accelerated. The total momentum is thus still conserved, and Newton's law is fulfilled. When an electron wave impinges on a thin crystal layer in an energy range in which the crystal is partially transparent for the electron, such momentum transfer can be measured directly by changing the electron energy so that diffraction occurs and part of the electron beam is reflected.

This qualitative picture also holds for more realistic periodic potentials, although the quantitative relationship depends on many other factors. Each of the bands usually consists of several branches which often overlap one another and may

[8]In theory, the electron will continue to accelerate in the opposite direction to the field and lose energy, thereby descending in the band, and the above-described process will proceed in the reverse direction until the electron has reached the lower band edge, where the entire process repeats itself. This oscillating behavior is called the *Bloch oscillation*. Long before the oscillation can be completed, however, scattering interrupts the process. Whether in rare cases (e.g., in narrow mini-bands of superlattices or ultrapure semiconductors at low temperatures) such Bloch oscillations are observable, and whether they are theoretically justifiable in more advanced models (Krieger and Iafrate 1986), is controversial. In three-dimensional lattices, other bands overlap and transitions into these bands complicate the picture.

show additional extrema (saddle points) in the first Brillouin zone, making the dependence of the effective mass on the electron energy more complicated. Near the band edge (for electrons), only one – perhaps degenerate – $E(k)$ branch is present in typical semiconductors, so that the above description holds rather well. This branch can be split, for example, by crystal anisotropy or electric or magnetic fields.

In summarizing the much more involved behavior of an electron in such a realistic band, we may wonder if we gained a more intuitive picture using the particle model. If we recognize, however, that the electron will mostly reside close to the bottom of the band, usually within a few kT, the model is quite helpful for an analysis of a number of basic processes. The electron will behave here like a particle with a constant effective mass; the value of this effective mass depends on the actual lattice potential, i.e., on the chemical and crystallographic nature of the material because these determine the shape of $E(k)$. In Sect. 4.5 of ▶ Chap. 22, "Carrier-Transport Equations," we present a more detailed description of the effective mass for the application of this concept to carrier transport in typical semiconductors.

3 The Proximity Approach in Organic Crystals

Organic crystals are composed of molecules as building blocks instead of atoms (Sect. 1.5 in ▶ Chap. 3, "The Structure of Semiconductors"). The electronic band structure of an organic semiconductor may therefore be derived from the proximity approach (Fig. 3) by replacing energy levels of *atomic* orbitals with levels of *molecular* orbitals. A schematic of the level scheme of an isolated single molecule is shown in Fig. 17. If the molecule is electrically neutral and not a radical, it has an even number of electrons. The *highest occupied molecular orbital* of the delocalized π electrons (Sect. 3.3 in ▶ Chap. 2, "Crystal Bonding") in the conjugated molecule, the so-called HOMO, refers to the electronic ground state *of the molecule*. It is occupied by 2 electrons with opposite spin and hence a singlet state with total spin $S = 0$, labeled S_0 in Fig. 17 (valence states below S_0 are not shown).

The first excited singlet state is S_1. There exists also a triplet state labeled T_1 with *parallel* spin of the electron in the HOMO state and the electron in the excited state, yielding a total spin $S = 1$; its energy is larger than that of S_0. Further excited states S_2 and T_2 may exist with an energy separation to S_0 below the ionization energy I_g, but the lifetime of an electron excited to such states is very small compared to that in the states S_1 and T_1. Excited states S_2 and T_2 as well as the vibronic states of the molecule that couple to each of the electronic states are not shown in Fig. 17 for clarity.

The molecule ground state in the gas phase is given by the ionization energy I_g. This quantity represents the energy required to remove the most weakly bound electron from the molecule and is readily accessible in experiment. Vice versa, if an additional electron is bound to the molecule to form a negatively charged molecular ion, an energy A_g, called electron affinity, is released. If the molecule is ionized in the environment of a crystal lattice, the required energy I_c is smaller than I_g, because negative and positive charges are now separated by a polarizable medium

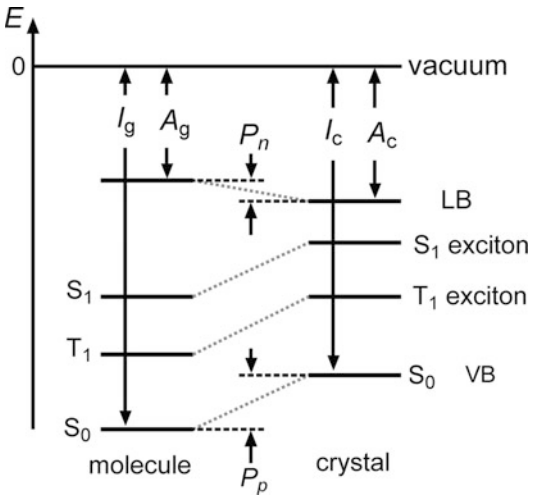


Fig. 17 Schematic of the energy levels of a single molecule (gas phase, *left*) and of a molecule crystal (*right*); I_g, A_g, I_c, and A_c denote, respectively, the ionization energies I and electron affinities A of the molecule in the gas phase and in the crystal; P_n and P_p signify the polarization energies of an electron and a hole in the molecule crystal

with a dielectric constant (typically about 3). I_g is reduced by a polarization energy P_p. Similarly, the electron affinity in a crystal is increased by a polarization energy P_n to a value A_c. The molecular ion states are separated by an energy gap E_g, which for $P_p = P_n \equiv P$ is given by (Karl 1974)

$$E_g = I_c - A_c = I_g - A_g - 2P = 2I_c - I_g - A_g; \tag{21}$$

see also Fig. 18. The energy gap and its position with respect to the vacuum level depend on the spatial extent of the delocalized π electrons in the molecules of the crystal; this dependence is illustrated for acenes in Fig. 18.

In an anthracene crystal the bandgap energy E_g is 4.1 eV; this energy is required to remove an electron from the HOMO level to a quasi-free state of the molecule, leaving a positive charge in the HOMO. The quasi-free state is the *lowest unoccupied molecular orbital*, the LUMO level; in analogy to inorganic semiconductors, the energy of the LUMO in organic semiconductors is referred to as *conduction band*, and that of the HOMO is called *valence band*. The S_1 state indicated in Fig. 17 lies below the LUMO level; in this state the electron is still bound to the positive charge in the HOMO, forming a so-called exciton (see Sect. 1.1.2 of ▶ Chap. 14, "Excitons"); the excitation energy of the S_1 exciton in anthracene is 1 eV lower[9] than E_g.

The polarization energy P in organic crystals has an electronic and a vibronic component. If an electron is added to a molecule, the neighboring molecules are polarized by the negative charge in their vicinity. The characteristic response time for

[9]This energy difference represents the binding energy of the exciton; its value is much larger than values found in inorganic semiconductors. A large binding energy corresponds to a strong spatial localization, a typical feature of excitons in organic crystals.

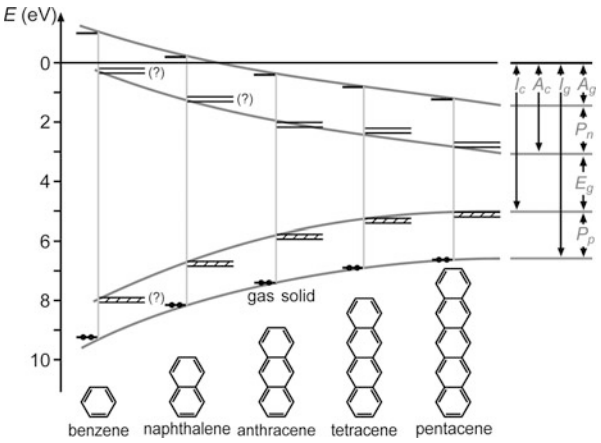


Fig. 18 Ionization energies of the highest occupied level and binding energies of the lowest unoccupied level for various oligoacenes in the gas phase (*left horizontal bars*, referring to I_g and A_g) and in the crystalline state (*right bars*, referring to I_c and A_c); after Karl (1974)

this *electronic polarization* is on the order of the oscillation period in an optical transition. This time is much shorter than that of a vibronic oscillation; the electronic polarization of the molecular neighborhood thus follows the movement of a quasi-free electron, thereby affecting its effective mass. This influence is not related to the effective mass of a quasi-free electron in an inorganic crystal discussed in Sect. 2. The quasi-free electron in the LUMO and its surrounding polarization cloud combined form a mobile quasiparticle referred to as (negative) *polaron*; correspondingly a positive charge in the HOMO builds a positive polaron.[10] We read from Fig. 18 that the polarization energies of the negative and positive polaron are given by

$$\begin{aligned} P_n &= A_c - A_g \\ P_p &= I_g - I_c. \end{aligned} \tag{22}$$

Both quantities are proportional to the polarizability of the organic molecules and are much larger than in conventional semiconductors. Still polarons also occur in inorganic semiconductors and are discussed in Sect. 1.2 of ▶ Chap. 22, "Carrier-Transport Equations," in the framework of transport properties.

The *vibronic component* of the polarization energy is related to a relaxation of the *crystal lattice* in the environment of the charged molecule. The characteristic response time is given by the period of a phonon oscillation and therefore much longer than the electronic response. A consequence of the vibronic relaxation is a decrease of the bandgap energy (in an anthracene crystal both P_n and P_p are increased by 0.15 eV; consequently E_g decreases by 0.3 eV). The energy of E_g before a lattice relaxation is called *optical bandgap* and that after relaxation *adiabatic bandgap*.

[10]The polaron character of mobile carriers in organic crystal is often not explicitly considered; in analogy to the quasiparticles of inorganic semiconductors, the carriers are simply termed electrons and holes.

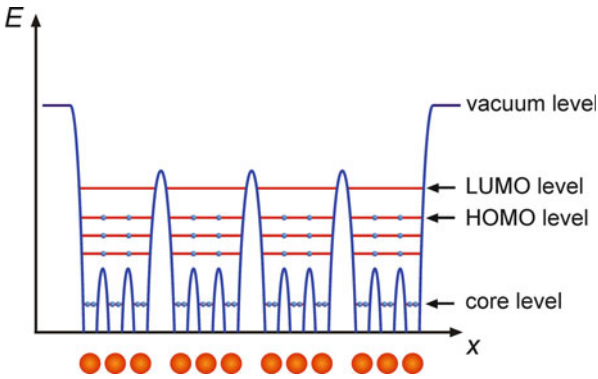


Fig. 19 Schematic of the potential energy and electron levels in a small organic crystal

The conductivity and the carrier mobility in organic semiconductors are generally very low compared to inorganic semiconductors.[11] This holds even if the electron mobility in the LUMO level of a molecule is high. The reason is the weak intermolecular contact in the organic crystal, schematically illustrated in Fig. 19. The figure shows the potential energy and energy levels of electrons in a small crystal comprising a chain of four molecules with three atoms each. Each occupied molecular orbital is filled by two electrons with opposite spin. Above the core levels near the atomic nuclei, the schematic indicates the occupied molecular-orbital levels of electrons which are delocalized within the molecules, with the topmost HOMO state. The filled HOMO and the empty LUMO levels of each molecule are separated from those of neighboring molecules by a potential barrier. The barrier originates from the weak intermolecular van der Waals interactions, leading to a predominant localization of the HOMO and LUMO wavefunctions in each molecule. Height and thickness of this barrier decide whether – in case of small barriers, corresponding to strong intermolecular interaction – conduction and valence bands evolve from these levels or not.

If the intermolecular barrier is low, bands similar to those in inorganic semiconductors are created as illustrated in Fig. 3. Higher barriers may still allow for conductivity by phonon-assisted hopping. The criteria for the occurrence of either band conduction or hopping conduction is derived from the mean time τ between scattering events of a mobile carrier and the width W of the states for carrier transport. For band conduction $\tau \gg \hbar/W$ must apply. Often a decision whether or not the criterion is met is difficult; a typical bandwidth of about 0.5 eV for oligoacene crystals yields $\tau > 10^{-15}$ s, leading to a mean free path length of the carrier that exceeds the crystal lattice constant.

[11]The mobility of electrons is defined in Sect. 2.2 in ▶ Chap. 22, "Carrier-Transport Equations," by $\mu = (q/m^*) \times \tau$, with effective mass m^*, charge q, and a mean time τ between scattering events; in organic crystals μ_{300K} is usually below 1 $cm^2/(Vs)$, often orders of magnitude smaller, compared to values of 10^3 $cm^2/(Vs)$ for inorganic semiconductors.

4 Summary

The electronic band structure of solids is the most significant feature to understand the electronic behavior of semiconductors. General features are described by the proximity approach, which considers the effect of the immediate neighborhood in a solid and, with gradually less attention, that of more distant atoms, on the energy levels of an isolated atom. Similar results are obtained by the periodicity approach, which considers the long-range periodicity of the crystal potential. The typical band structure with alternating bands and bandgaps is characteristic for all solids, in contrast to isolated atoms which show a discrete level spectrum. Details of the band structure depend on the chemistry of the material and its atomic structure (symmetry). Deviation from a periodic structure predominantly influences the energy range near the band edges, while it has little influence near the center of the bands.

Electrons near the lower edge of a band in a periodic lattice behave akin to electrons in vacuum. The influence of the lattice is taken into account by ascribing an *effective* mass to the carriers that, for typical semiconductors, is smaller than the electron rest mass at the band edge and increases with increasing distance from the band edge. In disordered or amorphous semiconductors, the band edge is fuzzy and the electronic states become localized when extended sufficiently beyond the band edge. In organic semiconductors a large polarization affects the effective mass of carriers, accounted for in the polariton model; in the case of large intermolecular barriers hopping conduction instead of band conduction occurs.

References

Adler D (1985) Chemistry and physics of covalent amorphous semiconductors. In: Adler D, Schwartz BB, Steele MC (eds) Physical properties of amorphous materials. Plenum Press, New York, pp 5–103

Agarwal SC (1995) Electronic structure of amorphous semiconductors. Bull Mater Sci 18:669

Anderson PW (1963) Concepts in solids. W. A. Benjamin, New York

Ashcroft NW, Mermin ND (1976) Solid state physics. Holt Reinhart and Winston, New York

Beeby JL, Hayes TM (1989) A calculation of the density of electronic states for amorphous semiconductors. J Non-Cryst Solids 114:253

Bloch F (1928) Über die Quantenmechanik der Elektronen in Kristallgittern. Z Phys 52:555 (On the quantum mechanics of electrons in crystal lattices, in German)

Bube RH (1992) Electrons in solids, an introductory survey, 3rd edn. Academic Press, New York

Callaway J (1976) Quantum theory of solid state. Academic Press, New York

Eberhart ME, Johnson KH, Adler D (1982) Theoretical models for the electronic structures of hydrogenated amorphous silicon. II. Three-center bonds. Phys Rev B 26:3138

Fletcher GC (1971) Electron bond theory of solids. North Holland, Amsterdam

Harrison WA (1980a) Solid state theory. Dover, New York

Harrison WA (1980b) Electronic structure and the properties of solids: the physics of chemical bonds. Freeman, San Francisco

Haug A (1972) Theoretical solid state physics. Pergamon Press, Oxford

Heine V (1980) Electronic structure from the point of view of the local atomic environment. Solid State Phys 35:1

Jancu J–M, Voisin P (2007) Tetragonal and trigonal deformations in zinc-blende semiconductors: A tight-binding point of view. Phys Rev B 76:115202
Johnson KH, Kolari HJ, de Neufville JP, Morel DL (1980) Theoretical models for the electronic structures of hydrogenated amorphous silicon. Phys Rev B 21:643
Kaplan TA, Mahanti SD (1995) Electronic properties of solids using cluster methods. Kluwer/Plenum, New York
Karl N (1974) Organic semiconductors. Festkörperprobleme/Adv Sol State Phys 14:261
Kittel C (2007) Introduction to solid state physics, 7th edn. Wiley, New York
Krieger JB, Iafrate GJ (1986) Time evolution of Bloch electrons in a homogeneous electric field. Phys Rev B 33:5494; and (1987), Quantum transport for Bloch electrons in a spatially homogeneous electric field. Phys Rev B 35:9644
Kronig R de RL, Penney WG (1931) Quantum mechanics of electrons in crystal lattices. Proc R Soc Lond Ser A130:499
Marder MP (2010) Condensed matter physics, 2nd edn. Wiley, Hoboken
Mills RGJ, Montroll EW (1970) Quantum theory on a network. II. A solvable model which may have several bound states per node point. J Math Phys 11:2525
Ovshinsky SR, Adler D (1978) Local structure, bonding, and electric properties of covalent amorphous semiconductors. Contemp Phys 19:109
Reitz JR (1955) Methods of the one-electron theory of solids. Solid State Phys 1:1
Shinozuka Y (1999) Hybridization in electronic states and optical properties of covalent amorphous semiconductors. Mater Res Soc Symp Proc 588:309
Singh J, Shimakawa K (2003) Advances in amorphous semiconductors. CRC Press, Boca Raton
Slater JC, Johnson KH (1972) Self-consistent-field $X\alpha$ cluster method for polyatomic molecules and solids. Phys Rev B 5:844
Street RA (2005) Hydrogenated amorphous silicon. Cambridge University Press, New York
Ziman JM (1972) Principles of the theory of solids. Cambridge University Press, Cambridge

Quantum Mechanics of Electrons in Crystals

7

Karl W. Böer and Udo W. Pohl

Contents

1 The Schrödinger Equation 220
1.1 Born–Oppenheimer Approximation 221
1.2 One-Electron Approximation 222
1.3 Pseudopotentials 224
2 Band-Structure Calculation 226
2.1 Noninteracting Electrons in Crystals 227
2.2 Approaches Explicitly Containing Electron–Electron Interaction 234
3 Relativistic Effects 238
4 Band Structure of Three-Dimensional Lattices 240
4.1 Empty and Nearly Empty Lattices 241
4.2 The Band Structure of Typical Semiconductors 243
4.3 Band Structure of Organic Crystals 247
5 Summary 250
References 251

Abstract

The electronic band structure of semiconductors reveals most of their intrinsic properties. It consists of the dispersion relation $E_n(\mathbf{k})$ for the various bands and is obtained from solving the Schrödinger equation for all electrons and nuclei in the solid. A manageable solution of this many-body problem requires substantial approximations for the interaction potential of all involved particles. Both empirical and ab initio approaches were developed for a one-electron scheme with different ways to approximate the actual interaction potential. Most approaches expand the

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_7

wavefunction in terms of a set of orthogonal trial functions, followed by variation of the expansion coefficients for finding a self-consistent solution. The more recent density-functional method calculates self-consistently the ground-state energy of the many-electron system from the charge-density distribution.

Keywords

Ab initio approaches · APW method · Born-Oppenheimer approximation · Density-functional method · Density of states · Dispersion relation · Electronic band structure · Empty lattice · GW approximation · Hartree-Fock approximation · Hartree approximation · Interaction potential · KKR method · K.p method · LCAO method · LMTO method · Many-body problem · Many-electron system · One-electron scheme · Pseudopotentials · Schrödinger equation · Tight-binding approach · Wavefunction

1 The Schrödinger Equation

In order to understand electronic transitions and the electron transport in a semiconductor, one must analyze its three-dimensional band structure, which is determined by solving the Schrödinger equation with a three-dimensional lattice potential for the actual crystal. To obtain such a potential and to solve the Schrödinger equation amounts to solving a *many-body problem*, because it involves all the lattice atoms and all of their electrons. It requires several sets of approximations for the problem to become manageable. Due to the limited space in this chapter, only some important topics are outlined; further relevant reading includes Ashcroft and Mermin (1976), Bassani and Pastori Parravicini (1975), Harrison (1980), and Phillips and Lucovsky (2009). In this chapter, following conventional descriptions and simplifying Coulomb terms, cgs units are used.

In the previous sections a crude periodic potential was presented for a one-dimensional one-electron model, which yielded some basic results for the band model. In this section we will take the opposite approach. Starting from the most general model, we will simplify it step by step until it can be solved mathematically; this treatment will provide some insight into the present understanding of the band structures. The electron spin, which is coupled to the angular momentum via spin-orbit interaction, is not included here. The spin–orbit interaction is a relativistic effect and considered in Sect. 3.

It is relatively easy to write down in a general form the Schrödinger equation which describes the many-body problem of n electrons in a lattice of N atoms, including atomic motion within the crystal, but excluding relativistic effects:

$$-\sum_{i=1}^{n}\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial \mathbf{r}_i^2}\Phi-\sum_{I=1}^{N}\frac{\hbar^2}{2M_I}\frac{\partial^2}{\partial \mathbf{R}_I^2}\Phi+\sum_{\substack{i,j=1\\ i\neq j}}^{n}\frac{e^2}{2r_{ij}}\Phi$$
$$+V_{\text{ion}-\text{ion}}(\mathbf{R}_1,\mathbf{R}_2,\ldots)\Phi+V_{\text{el}-\text{ion}}(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{R}_1,\mathbf{R}_2,\ldots)\Phi \;=\; E\Phi. \tag{1}$$

Here R_I and M_I are the ion coordinates and masses, and r_i are the electron coordinates; Φ is the wave-function for the system of atoms and electrons. The five terms on the left-hand side describe the kinetic energy of the electrons, the kinetic energy of the atoms, the potential energy due to the electron–electron Coulomb repulsion, the potential energy due to the repulsive interatomic interaction, and, finally, the potential energy due to the attractive electron–atom interaction. A major problem is the extremely large number of terms in the sums of Eq. 1. A first step usually made to reduce the complexity is to separate the chemically inert core and valence electrons and to neglect a deformation of the ion core.[1]

1.1 Born–Oppenheimer Approximation

For a reasonable first approximation, one takes into consideration the large ratio M_I/m_0 of atomic to electron masses. This lets the lattice oscillations appear to stand still for the much faster, more easily accelerated electrons and permits the use of an *adiabatic approximation* (Born and Oppenheimer 1927; Pelzer and Wigner 1932). With the ansatz

$$\Phi(\mathbf{R}_1,\mathbf{R}_2,\ldots,\mathbf{r}_1,\mathbf{r}_2,\ldots)=\varphi(\mathbf{R}_1,\mathbf{R}_2,\ldots)\;\psi(\mathbf{R}_1,\mathbf{R}_2,\ldots,\mathbf{r}_1,\mathbf{r}_2,\ldots), \tag{2}$$

a separation of the atomic eigenfunctions φ and electronic eigenfunctions ψ can be achieved. The resulting *Schrödinger equation for electrons* in a lattice *with atoms at rest* is

$$-\sum_{i=1}^{n}\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial \mathbf{r}_i^2}\psi+\sum_{\substack{i,j=1\\ i\neq j}}^{n}\frac{e^2}{2r_{ij}}\psi+V_{\text{el}-\text{ion}}(\mathbf{R}_1,\mathbf{R}_2,\ldots,\mathbf{r}_1,\mathbf{r}_2,\ldots)\psi$$
$$=E'(\mathbf{R}_1,\mathbf{R}_2,\ldots)\;\psi \tag{3}$$

and the *Schrödinger equation for the oscillating atoms* is

[1]The latter assumption is questionable when explaining certain dielectric properties with strong polarization (Sect. 3 in ▶ Chap. 10, "Interaction of Light with Solids"). If a core–valence electron separation is *not* made, all potentials in Eq. 1 are simple Coulomb potentials. Otherwise, the effective interaction potentials must be obtained.

$$-\sum_{I=1}^{N}\frac{\hbar^2}{2M_I}\frac{\partial^2}{\partial \mathbf{R}_I^2}\varphi+(V_{\text{ion-ion}}(\mathbf{R}_1,\mathbf{R}_2,\ldots)+E'(\mathbf{R}_1,\mathbf{R}_2,\ldots))\varphi = E\varphi. \quad (4)$$

Here the total electron energy E' is added to the potential energy of the lattice. Two terms responsible for electron–lattice interaction

$$-\sum\frac{\hbar^2}{2M_I}\varphi\frac{\partial^2}{\partial \mathbf{R}_I^2}\psi-2\sum\frac{\hbar^2}{2M_I}\frac{\partial}{\partial \mathbf{R}_I}\varphi\frac{\partial}{\partial \mathbf{R}_I}\psi$$

are neglected in Eq. 4 and can be used in a perturbation approach to induce transitions between stationary solutions of Eqs. 3 and 4 (Kubo 1952). A further simplification to the Schrödinger equation for electrons (Eq. 3) can be introduced when the ion motion *induced by the electron configuration* is neglected, yielding

$$-\sum_{i=1}^{n}\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial \mathbf{r}_i^2}\psi+\sum_{\substack{i,j=1\\ i\neq j}}^{n}\frac{e^2}{2r_{ij}}\psi+V_{\text{el-ion}}(\mathbf{r}_1,\mathbf{r}_2,\ldots)\psi = E\psi \quad (5)$$

with $V_{\text{el–ion}}$ as the *periodic potential* of the lattice atoms. The coordinates of electrons are the variables, while ions are assumed to be at their average positions. Lattice oscillations therefore do not enter this analysis.

1.2 One-Electron Approximation

The most drastic step towards a manageable problem is the reduction of the electronic Schrödinger equation to a *one-electron approximation* by separating

$$\psi(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_n) = \psi(\mathbf{r}_1)\ \psi(\mathbf{r}_2)\ \ldots\ \psi(\mathbf{r}_n). \quad (6)$$

The assumption that each electron experiences essentially the same potential $V_{\text{el–el}}(\mathbf{r})$ of an *averaged distribution* of all other electrons, i.e., the electrons are assumed to be *independent* of each other, is called *Hartree approximation*. The principle is illustrated in the simple scheme of Fig. 1. The movement of the electron in atom 1 is generally correlated to that of the electron in atom 2 (and vice versa); in the Hartree approximation electron 1 only senses the *average* charge density of electron 2 illustrated in Fig. 1b. Consequently the correlation gets lost, and computed electron energies are too high. The Hartree approximation yields a single set of terms in the Schrödinger equation:

$$-\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial \mathbf{r}^2}\psi+(V_{\text{el-ion}}(\mathbf{r})+V_{\text{el-el}}(\mathbf{r}))\ \psi = E\psi. \quad (7)$$

The entire problem has now been reduced to the problem of a single electron moving independently of all other electrons in a static potential composed of a

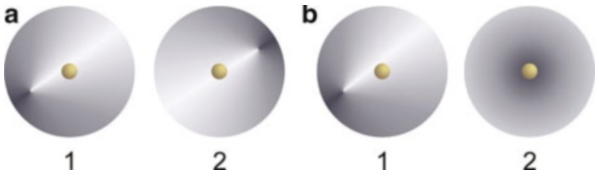


Fig. 1 Scheme of two atoms with one electron each, illustrating the charge density viewed from the position of electron 1, in (**a**) the general case, (**b**) the Hartree approximation

perfect periodic potential of the lattice and an average potential describing its interaction with all the other electrons.[2] The main goal in the analysis of the realistic behavior of electrons in a crystal lattice is to obtain the potential $V_{\text{el–ion}} + V_{\text{el–el}}$; there are many approaches to accomplish this task as outlined in Sects. 2.1 and 2.2. With the potential known, the problem becomes one of numerically integrating Eq. 7. The electron behavior can then be described in one of two fashions, depending on the kind of problem to be discussed.

1. For the electron in higher bands (conduction bands – see Sect. 1 in ▶ Chap. 8, "Bands and Bandgaps in Solids"), a simple periodic function is assumed for the potential, and the wavefunction is best expressed as the *Bloch function*

$$\psi_n(\mathbf{k},\mathbf{r}) = u_n(\mathbf{k},\mathbf{r})\ \exp(i\mathbf{k}\cdot\mathbf{r}) \tag{8}$$

   where n is the band index and $u_n(\mathbf{k},\mathbf{r})$ is periodic with the lattice periodicity (*Bloch theorem*[3]). These Bloch functions are plane waves that are modulated with lattice periodicity; see Fig. 7 in ▶ Chap. 6, "The Origin of Band Structure."
2. For tighter bound states (lower bands), one often uses to better advantage the *Wannier functions*, which are defined as wave packets of the Bloch functions

$$\varphi_n(\mathbf{r}-\mathbf{R}) = \frac{1}{\sqrt{V}}\sum_{\mathbf{k}} u_n(\mathbf{k},\mathbf{r})\ \exp(i\mathbf{k}\cdot\mathbf{r})\ \exp(-i\mathbf{k}\cdot\mathbf{R}) \tag{9}$$

   and are localized near the lattice site $\mathbf{R}$; V is the crystal volume. Such a description assists in finding $V(\mathbf{r})$ in terms of simpler quantities.

[2]For the formal relation between the many-body problem and the one-electron band structure, see Hedin and Lundqvist (1970).

[3]The Bloch theorem states that nondegenerate solutions of the Schrödinger equation in a periodic lattice are also solutions after translation by a lattice vector, with the amplitude function having lattice periodicity $u_{\mathbf{k},\beta}(\mathbf{r}) = u_{\mathbf{k},\beta}(\mathbf{r}+\mathbf{R})$; $\mathbf{R}$ is any translation vector which reproduces the Bravais lattice.

1.3 Pseudopotentials

The initial effort of the band theory is devoted to obtaining the *crystal potential* $V(\mathbf{r})$. The crystal potential has two contributions: the electronic part and the ionic part (see Eq. 7). The *electronic contribution* is in many approaches interwoven with the method for solving the wave equation; therefore, we will return to this subject in Sects. 2.1 and 2.2. The *ionic contribution* is obtained from the charge of the nuclei and that of the core electrons, which are treated as atomic functions; the sum is usually described by Coulomb potentials produced by ions of charge Ze.

The exact ionic potential shows very large amplitudes near the center of each ion and thereby creates substantial and unnecessary computational problems, if only upper bands (i.e., no core states) are of interest. Many methods circumvent these problems by using pseudopotentials which avoid the potential spikes as indicated below. The spiky part of the potential is of importance for inner core electrons but usually not for valence or conduction band electrons. Introduced by Prokofjew (1929) and Fermi (1934) and applied to atoms by Hellmann (1935), the pseudopotential became a major means to provide the most important input to the Hamiltonian relevant for valence electrons. In the late 1950s and 1960s (Phillips and Kleinman 1959; review of Harrison 1966), it was shown that valence electrons are effectively excluded from the ion core of an atom by an almost exact balance between two strong forces: the Coulomb attraction to the core and the quantum-mechanical repulsion from the core electrons (*exclusion*). This is also known as the Phillips cancellation theorem (Phillips and Kleinman 1959). The resulting net force can be described by a rather weak, attractive *pseudopotential* (Ziman 1964; Harrison 1966), which has lattice periodicity

$$V(\mathbf{r}) = \sum_{\mathbf{k}} \sum_{\alpha} V_{\alpha}(\mathbf{k})\, S_{\alpha}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}), \tag{10}$$

where α is the type of atom in the unit cell, V_{α} is the *atomic pseudopotential* of atom α, and S_{α} is a structure factor. A typical plot of an *atomic* pseudopotential is shown in Fig. 2. The repulsive branch starts at ~1/2 of the bond length (tick mark at the r axis in Fig. 2). The strength of the pseudopotential is much reduced compared to the atomic (Coulomb) potential. The use of the weaker periodic lattice potential permits the application of a perturbation formalism. For reviews, see Heine and Weaire (1970), Bassani and Giuliano (1972), Cohen (1984), Vanderbilt (1990), and Laasonen et al. (1993).

A widely applied pseudopotential is the *muffin-tin potential* illustrated in Fig. 3. This potential is composed of a radial-symmetric inner part centered at the location $\mathbf{R}_I$ of the atoms and an outer part at the interstitial positions between the atoms that is often assumed to be constant (e.g., 0), yielding

$$V(\mathbf{r}) = \begin{cases} V(|\mathbf{r} - \mathbf{R}_I|) & \text{for} \quad |\mathbf{r} - \mathbf{R}_I| < r_0 \\ V_0 = V(r_0) = 0 & \text{for} \quad |\mathbf{r} - \mathbf{R}_I| > r_0 \quad \text{for all} \ \mathbf{R}_I. \end{cases} \tag{11}$$

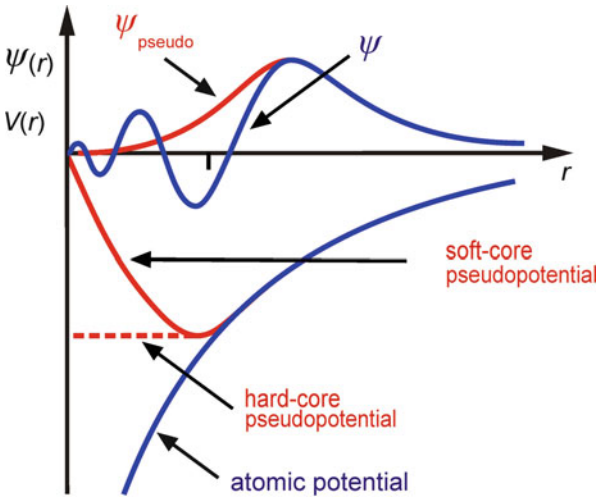


Fig. 2 Radial dependence of typical pseudopotentials compared to the Coulomb-attractive atomic potential (*blue curve*), which dominates further away from the core. The *solid and dashed red curves* show a soft-core and a hard-core pseudopotential, respectively. The *top* part shows the wavefunction resulting from the atomic potential (*blue*) and the pseudo-wavefunction obtained from a pseudopotential

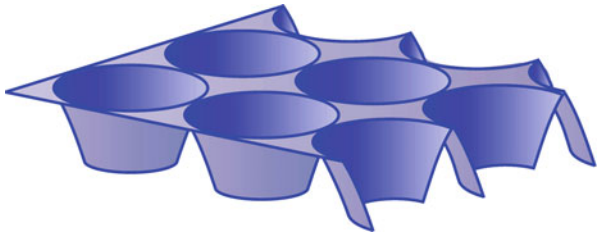


Fig. 3 Muffin-tin potential, showing the typical form of the approximated potential surrounding each atom

The muffin-tin aspect of the potential (Fig. 3) can be modified by assuming an *adjustable* flat potential to fit the experimental results. Also the radius r_0 of the boundary between core and outer region may be chosen, although results are reported to be largely insensitive on the choice.

With a pseudopotential rather than the real lattice potential, we lose information about the bands of core electrons which are of little interest to semiconductor behavior, but gain the simplicity of describing the solutions with *pseudo-wavefunctions*, which are slowly varying as illustrated in Fig. 2 and can be approximated more easily by superposition of fewer terms. The resulting eigenvalues for the upper (valence and conduction) bands, however, are the same as obtained with the exact potential extending into the core region.

The ease of arriving at quantitative results, which can be compared with the experiment, permits the repeated readjustment of the pseudopotential in iterative trials, until agreement with the experiment is satisfactory (*empirical pseudopotential method* (EPM)), as discussed by Cohen and Heine (1970) and Mäder and Zunger (1994). It should be noted that the *empirical* pseudopotential method determines a crystal potential by fitting the *band structure* to the experiment. However, it cannot be made self-consistent, since the screening expressed by electron–electron *and*

exchange interactions discussed later (second and third terms of Eq. 32) are not included. The agreement that can be achieved between theory, when fitted empirically, and experiment is rather good (Chelikowsky and Phillips 1978; Ihm and Cohen 1980). However, when the so-adjusted potential is used to calculate the electron density distribution, there is only fair agreement with the distribution obtained from X-ray diffraction.

First-principles pseudopotentials suitable for computing valence electron total energies and the full band structure show substantially improved agreement with the experiment (Bachelet et al. 1982; Ihm et al. 1979; Froyen 1996). They are derived from microscopic models and are usually applied today (Vanderbilt 1990; Laasonen et al. 1993).

2 Band-Structure Calculation

After a suitable crystal potential $V(\mathbf{r})$ is determined, the Schrödinger equation (7) containing this potential must be solved under the appropriate boundary condition[4] to obtain the dispersion relation $E_n(\mathbf{k})$. A wide variety of methods for efficiently computing such solutions have been developed, all of which are based on variational principles. Usually, the wavefunction is expanded in terms of *trial functions*, followed by a *variation of the expansion coefficients*. This technique proposed by Ritz (see Morse and Feshbach 1953) replaces the problem of solving the wave equation with the simpler one of solving equivalent *secular* equations. A classification of prominent approaches is given in Fig. 4; often, also a mix of different methods is applied.[5] Most of these methods can be used for efficient computation and lead to similar results. With the substantial increase in computer power, nonempirical ab initio calculations have

[4]Namely, $\mathbf{k}$ must be real and $\psi(\mathbf{r})$ periodic with lattice periodicity $\psi(\mathbf{r}) = \psi(\mathbf{r} + n\mathbf{a}_i)$ (*Born–van Karman boundary condition*).

[5]Some of the methods are considered in more detail in this section. The following abbreviations are commonly used:
DFT: Density functional theory
EPM: Empirical pseudopotential method
EEX: Exact-exchange method
(m)GGA: (meta-) Generalized gradient approximation
GWA: Green's function GW approximation
HF(A): Hartree–Fock (approximation)
KKR: Korringa–Kohn–Rostocker (Green's function) method
(L)APW: (Linear) augmented plane wave method
LCAO: Linear combination of atomic orbitals
LDA: Local density approximation
LMTO: Linear muffin-tin orbitals
LSDA: Local-spin-density approximation
OPW: Orthogonal-plane-wave method
PSF: Pseudofunctional method (Kasowski et al. 1986)
PW: Plane wave
TB: Tight binding

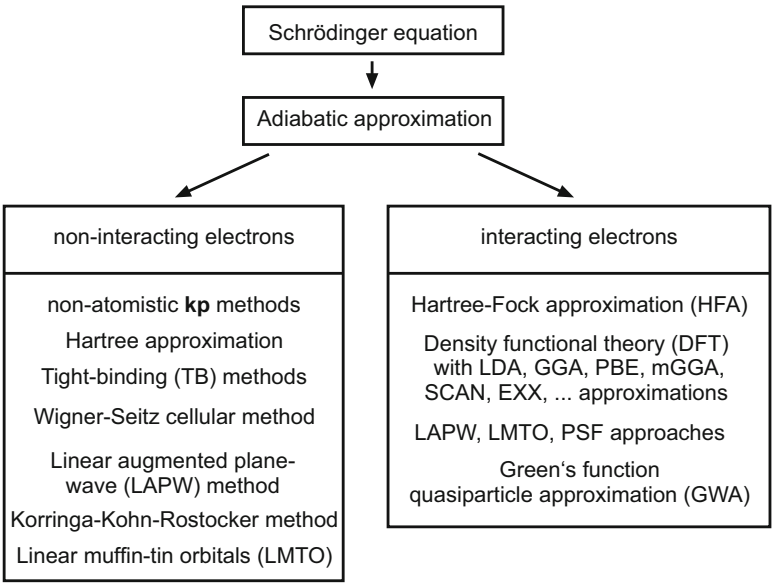


Fig. 4 Classification of methods applied to solve the Schrödinger equation of electrons in the periodic crystal potential

replaced most of the earlier approximate and empirical approaches, where the potential was not treated in a self-consistent manner.

2.1 Noninteracting Electrons in Crystals

The precision of band-structure calculations depends on the ability to correct the drastic simplifications made to obtain the one-electron equation, i.e., to describe electron–electron interaction as realistic as possible. Methods discussed in this section use an *effective* one-electron crystal potential $V(\mathbf{r})$ which includes both the ionic and the electronic part.

2.1.1 The k·p Method

The instructive model of quasi-free electrons outlined in ▶ Chap. 6, "The Origin of Band Structure" (periodicity approach, effective mass), leads to a popular non-atomistic model, which is widely applied to obtain analytic expressions for the band structure near high-symmetry points from experimental optical spectra. The method was introduced by Bardeen (1938) and was used to explore the band structure in the vicinity of $\mathbf{k} = 0$ or other critical points $\mathbf{k}_0$ of the Brillouin zone (Seitz 1940; see Sect. 4). It is based on a symmetry-adapted perturbation approach with the standard $\hbar\,\mathbf{k} \cdot \mathbf{p}/m$ term (with $\mathbf{p} = -i\hbar\partial/\partial\mathbf{r}$) in the Hamiltonian, applied to the lattice-periodic functions in the neighborhood of the critical point $u_n(\mathbf{k},\mathbf{r})$ rather

than to the full Bloch function. Introducing this function into the one-electron Schrödinger equation, one obtains

$$\left(-\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial \mathbf{r}^2}+V(\mathbf{r})+\frac{\hbar}{m_0}\mathbf{k}\cdot\mathbf{p}+\frac{\hbar^2 k^2}{2m_0}\right)\ u_n(\mathbf{k},\mathbf{r}) \ = \ E_n(\mathbf{k})\ u_n(\mathbf{k},\mathbf{r}). \qquad (12)$$

Here the u_n**(k,r)** are a complete, orthogonal set of eigenfunctions in real crystal space referred to as the **k·p** representation. The first two terms in the Hamiltonian of Eq. 12 define the unperturbed Hamiltonian $\mathbf{H}_0$, and the last two terms can be developed as small perturbations in the vicinity of the given $\mathbf{k}_0$. Thus, Eq. 12 may be written as

$$\left(\mathbf{H}_0+\frac{\hbar}{m_0}(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{p}+\frac{\hbar^2}{2m_0}(\mathbf{k}-\mathbf{k}_0)^2\right)\ u_n(\mathbf{k},\mathbf{r}) \ = \ E_n(\mathbf{k})\ u_n(\mathbf{k},\mathbf{r}) \qquad (13)$$

Written in matrix form, the number of **k·p** matrix elements is greatly reduced because of symmetry considerations. These matrix elements

$$\mathbf{M}_{jn}(\mathbf{k}_0) \ = \int \psi_n^*(\mathbf{k}_0,\mathbf{r})\,\mathbf{p}\ \psi_n(\mathbf{k}_0,\mathbf{r})\ d\mathbf{r} \qquad (14)$$

can be evaluated easily (ψ_n is the Bloch function – see Eq. 8). They contain only a small number of parameters which are taken from the experiment or from ab initio calculations. The eigenvalues near a characteristic point $\mathbf{k}_0$ can be expressed as

$$E_n(\mathbf{k}) \ = E_n(\mathbf{k}_0)+\ \frac{\hbar^2(\mathbf{k}-\mathbf{k}_0)^2}{2m_0}+\frac{\hbar^2}{m_0^2}\sum_{\substack{j,n\\ j\neq n}}\frac{(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{M}_{jn}(\mathbf{k}_0)}{E_n(\mathbf{k}_0)-E_j(\mathbf{k}_0)} \qquad (15)$$

Equation 15 can be simplified when interactions between only two bands are of interest (e.g., the valence and conduction bands [see Sect. 1 in ▶ Chap. 8, "Bands and Bandgaps in Solids"]), and the energy difference between these is small compared to the difference with all other bands. Then we can write

$$E_n(\mathbf{k}) \ = E_n(\mathbf{k}_0)+\ \frac{\hbar^2}{2}\sum_{i=1}^{3}\frac{(k_i-k_0)^2}{m_i} \quad \text{with} \ \ i=x,y,z \qquad (16)$$

and can express the effective mass (see Sect. 2.2 in ▶ Chap. 6, "The Origin of Band Structure") for an isotopic system as

$$\frac{1}{m^*}=\frac{1}{m_0}\pm\frac{2\ \left|\mathbf{M}_{nj}(\mathbf{k}_0)\right|^2}{m_0^2\ \left(E_j(\mathbf{k}_0)-E_n(\mathbf{k}_0)\right)} \qquad (17)$$

with + or − for the upper or lower band, respectively.

One can use Eq. 15 for the *deviation* of $E(\mathbf{k})$ from *parabolicity* near a critical point $\mathbf{k}_0$:

$$E(\mathbf{k}-\mathbf{k}_0) = -\frac{E_g}{2} + \frac{\hbar^2(\mathbf{k}-\mathbf{k}_0)^2}{2m_0} \pm \frac{1}{2}\sqrt{E_g^2 + \frac{4\hbar^2(\mathbf{k}-\mathbf{k}_0)^2 \left|\mathbf{M}_{nj}(\mathbf{k}_0)\right|^2}{m_0^2}} \quad (18)$$

which yields with Eq. 17:

$$E(\mathbf{k}-\mathbf{k}_0) = -\frac{E_g}{2} + \frac{\hbar^2(\mathbf{k}-\mathbf{k}_0)^2}{2m_0} \pm \frac{E_g}{2}\sqrt{1 + \frac{2\hbar^2(\mathbf{k}-\mathbf{k}_0)^2}{E_g}\left(\frac{1}{m^*} - \frac{1}{m_0}\right)} \quad (19)$$

with m^* as the effective mass (assumed to be isotropic) in the conduction (+) or valence (−) band (see Sect. 1 in ▶ Chap. 8, "Bands and Bandgaps in Solids"). This illustrates the usefulness of the $\mathbf{k}\cdot\mathbf{p}$ method in a simple example. For a more detailed discussion see Lew Yan Voon and Willatzen (2009), Vurgaftman et al. 2021.

The use of this method for band-structure analysis was initiated by Kane (1956) and has been carried on by many others. It became particularly popular to describe optical transitions in direct-bandgap semiconductors like GaAs using an eight-band $\mathbf{k}\cdot\mathbf{p}$ model, which comprises the three valence-band p orbitals and the one s conduction band (×2 spin states); 8 Bloch functions are sufficient for a reasonable description of ~20% of the Brillouin zone and hence phenomena occurring near the direct bandgap. Examples of $E(\mathbf{k})$ determined by the $\mathbf{k}\cdot\mathbf{p}$ method are given by Cardona and Pollak (1966); see also Burt (1992) and the introduction by Yu and Cardona (2005).

2.1.2 Hartree Approximation

We come back to the approach of Hartree illustrated in Fig. 1 and the related one-electron Schrödinger equation. The *electronic contribution* in Eq. 7 can be approximated by the solution of the corresponding Hartree equation, describing the *Coulomb interaction* only:

$$\left(-\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial\mathbf{r}^2} - \sum_I \frac{Z_I e^2}{|\mathbf{r}-\mathbf{R}_I|} + V_{\text{el-el}}(\mathbf{r})\right)\psi_n(\mathbf{r}) = E_n\ \psi_n(\mathbf{r}) \quad (20)$$

with

$$V_{\text{el-el}}(\mathbf{r}) = \sum_{\substack{m \\ m\neq n}} e^2 \int \frac{\left|\psi_m(\mathbf{r}_j)\right|^2}{|\mathbf{r}-\mathbf{r}_j|} d\mathbf{r}_j \quad (21)$$

Since the *symmetrical*[6] many-electron wavefunction Eq. 6 is incompatible with the Pauli principle, a many-electron wavefunction ψ obeying the Pauli principle is constructed instead by using an *antisymmetric* normalized product of all of the one-electron eigenfunctions $\psi_n(\mathbf{r}_i)$ in a *Slater determinant*:

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(\mathbf{r}_1) & \psi_2(\mathbf{r}_1) & \cdots & \psi_n(\mathbf{r}_1) \\ \psi_1(\mathbf{r}_2) & \psi_2(\mathbf{r}_2) & \cdots & \psi_n(\mathbf{r}_2) \\ \vdots & \vdots & \vdots & \vdots \\ \psi_1(\mathbf{r}_n) & \psi_2(\mathbf{r}_n) & \cdots & \psi_n(\mathbf{r}_n) \end{vmatrix} \tag{22}$$

The normalization factor accounts for the $n!$ possibilities for distributing n indistinguishable electrons over n states. The determinant vanishes if two electrons occupy the same state, since two rows of the determinant are then identical.

The Hamiltonian of the Hartree equation contains its own eigenfunction. Its solution therefore involves first guessing an approximate solution $\psi(\mathbf{r})$ that describes the probability of finding an electron at the position $\mathbf{r}$, and a consequent iteration by introducing Eq. 21 into the Schrödinger equation (Eq. 7). This equation is then solved, using the resulting eigenfunction ψ as the input function in Eq. 21 and continuing until convergence is achieved (*self-consistent field method*). It must be noted that Hartree calculations overestimate substantially the electronic eigenenergies.

2.1.3 Tight-Binding or LCAO Approach

The electronic eigenfunction in a crystal lattice can be approximately derived from the atomic eigenfunctions (*linear combination of atomic orbitals* (LCAO) approach; Heitler and London 1927; Bloch 1928; Slater 1951, 1953a, b) of each lattice atom in the unit cell and consequent periodic repetition. This is the basis for the *tight-binding* (TB) *approximation* in which the crystal wavefunction is composed of a linear combination of the eigenstates of the free atoms. Properly set up, these are the *Bloch tight-binding sums*:

$$\psi_n(\mathbf{k}, \mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_l} \exp(i\,\mathbf{k}\mathbf{R}_l)\, \varphi_n(\mathbf{r} - \mathbf{R}_l) \tag{23}$$

composed from $\varphi_n(\mathbf{r} - \mathbf{R}_l)$, the Wannier functions of the free atoms (Eq. 9) localized at the lattice sites $\mathbf{R}_l$; unlike the atomic eigenfunctions, the Wannier functions already form a complete orthogonal set and therefore can be used directly to describe the electronic eigenfunctions in the crystal. Here N is the number of unit cells in the crystal. The tight-binding sums are used for analyzing low-energy states (deep-lying bands), where the eigenfunctions are rather localized near the nucleus of each atom and are barely disturbed by surrounding atoms. The method can be simplified when

[6] A symmetrical wavefunction does not change sign when the coordinates of two electrons are interchanged, in contrast to an antisymmetrical wavefunction.

the actual atomic eigenfunctions are replaced by a similar but simpler complete set of orthonormalized functions (Kane 1976), a mixed basis set (Bendt and Zunger 1982), or evanescent Bloch solutions (Chang 1982).

Overlap integrals describe the interaction between the atoms that causes broadening of the bands. For higher bands, one must consider the fact that the atomic states extend over much larger than interatomic distances and the overlap integral becomes very sensitive to the tail of the atomic potential. This fundamental difficulty limits the applicability of the tight-binding method to deep states and core bands. Still band structures of semiconductors like Si or GaAs are reasonably well described using tight-binding approaches.

2.1.4 Nearest-Neighbor Tight-Binding Model

A theoretical shortcut to obtain the chemical trend of band structures of tetrahedrally bound semiconductors was introduced by Harrison (1973) and extended by Vogl et al. (1983) and Jancu et al. (1998, 2007). This semiempirical NN-TB model uses a limited set of orbitals, usually one s and three p localized pseudoorbitals plus one unoccupied excited state s^* yielding a sp^3s^* basis, or, for significant improvements, including also the d orbitals for a $sp^3d^5s^*$ basis for also obtaining indirect gap features.[7] These are adjusted to fit optical bandgaps in constructing a pseudo-Hamiltonian, somewhat similar to the older empirical pseudopotential method (EPM). The resulting band structure is easily obtained and is in reasonable agreement with that obtained for such semiconductors from pseudopotentials. When the matrix elements of this model are fixed by the atomic energies of the lattice constituents and by a set of universal constants, certain chemical trends of the electronic structure of zincblende and diamond semiconductors can be predicted.

2.1.5 Cellular Method

The *cellular method* (Wigner and Seitz 1933, 1934) utilizes the fact that the lattice can be divided into Wigner–Seitz cells (Sect. 1.3.1 in ▶ Chap. 3, "The Structure of Semiconductors"); the Schrödinger equation is hence to be solved within one such cell, and the Bloch condition $\psi_n(\mathbf{r}+\mathbf{R}) = \exp(i\mathbf{kR})\psi_n(\mathbf{r})$ is used to obtain solutions for the entire solid. The actual cell potential is approximated by a potential with rotational symmetry, allowing to separate the angular and radial parts of the wavefunction according to

$$\psi(\mathbf{k},\mathbf{r}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} c_{lm}(\mathbf{k})\, Y_{lm}(\theta,\varphi)\, R_l(E,r) \tag{24}$$

with $Y_{lm}(\theta,\varphi)$ as spherical harmonics and $R_l(E,r)$ as the solution of the purely radial wave equation. Solutions can easily be obtained when using appropriate boundary conditions (ψ continuous at r_B and $d\psi/dr|_{r_B} = d\psi/dr|_{r_B+R} = 0$) with nearly spherical

[7] This model is more commonly also referred to as *sps** method or *spds** method.

symmetry (dominant contribution from the spherical core) and the Wigner–Seitz cell (boundary at r_B) containing only one atom in its center (Bell 1953). The difficulty of the cellular method resides in determining appropriate boundary conditions for all **r** from only a few known points on the surface of the Wigner–Seitz cell (Shockley 1938).

2.1.6 Augmented Plane-Wave Method

The difficulties with insufficiently known boundary conditions in the cellular method can be reduced by the *augmented plane-wave* (APW) *method*, in which the potential is prescribed by a muffin-tin type (Fig. 2). The potential is spherically symmetric inside the muffin-tin sphere and constant outside in the so-called interstitial region. For a primitive lattice and spherical symmetry up to one half of the interatomic spacing r_0, we have

$$\begin{aligned}\psi(\mathbf{k},\mathbf{r}) = &\sum_{l=0}^{\infty}\sum_{m=-l}^{l} c_{lm}(\mathbf{k})\,Y_{lm}(\theta,\phi)\,R_l(E,r)\,\eta(\mathbf{r}-\mathbf{r}_0)\\ &+\sum_j b_j(\mathbf{k})\exp\left[i\left(\mathbf{k}+\mathbf{k}_j\right)\cdot\mathbf{r}\right]\left[1-\eta(\mathbf{r}-\mathbf{r}_0)\right]\end{aligned} \tag{25}$$

$$\eta(\mathbf{r}-\mathbf{r}_0) = \begin{cases} 0 & \text{for } r > r_0 \\ 1 & \text{for } r < r_0 \end{cases} \tag{26}$$

Equation 25 has a tight-binding and a plane-wave contribution for the inner and outer parts, respectively. This method was suggested by Slater (1937) and has become very useful for analyzing electronic states for closely packed materials (metals). However, it yields energy gaps of semiconductors which are too small. Some improvement is obtained by assuming an *adjustable* flat potential to fit the experimental results (Loucks 1967; Dimmock 1971).

A difficulty in APW calculations is the search for atomic basis functions at energies that one is trying to obtain. A major advancement was made by using atomic wavefunctions at different energies E_l, usually chosen at the center of the band of interest with angular character l (Andersen 1975; Krakauer et al. 1981; Wei and Krakauer 1985). The convergence of the method was improved by adding a further set of orbitals at a different energy (Singh and Nordstrom 2006). This *linearized augmented plane-wave* (LAPW) method treats all electrons, valence and core, in an equal fashion and contains no shape approximations of the potential. The LAPW method yields very accurate results and is applied in present state-of-the-art band calculations outlined in Sect. 2.2.2.

2.1.7 Green's Function (KKR) Method

The Green's-function method, also referred to as *KKR method*, uses a transformation of the Schrödinger equation into an integral equation (Korringa 1947), which circumvents the difficulty of unknown boundary conditions (Kohn and Rostoker

1954), but also requires a crystal potential in a muffin-tin form (Fig. 2). One obtains $\psi(\mathbf{k}, \mathbf{r})$ from an integral equation

$$\psi(\mathbf{k},\mathbf{r}) = \int_V G[\mathbf{r}-\mathbf{r}', E(\mathbf{k})]\ V(\mathbf{r}')\ \psi(\mathbf{k},\mathbf{r}')\, d\mathbf{r}' \tag{27}$$

where $V(\mathbf{r}')$ is the muffin-tin potential and $G_{\mathbf{k}}(\mathbf{r}\text{-}\mathbf{r}', E(\mathbf{k}))$ is the Green's function. It is determined as the solution of

$$\left(\frac{\hbar^2}{2m}\frac{\partial^2}{\partial \mathbf{r}^2} + E\right) G[\mathbf{r}-\mathbf{r}', E(\mathbf{k})] = \delta(\mathbf{r}-\mathbf{r}') \quad \text{with} \quad G[\mathbf{r}+\mathbf{r}_0, E(\mathbf{k})] = G[\mathbf{r}, E(\mathbf{k})]\exp(i\mathbf{k}\cdot\mathbf{r}_0) \tag{28}$$

as the boundary condition. The Green's function can be expressed as the expansion

$$G[\mathbf{r}-\mathbf{r}', E(\mathbf{k})] = -\frac{1}{V}\frac{\sum_n \exp[i\,(\mathbf{k}+\mathbf{k}_n)\cdot(\mathbf{r}-\mathbf{r}')]}{\left[\hbar^2/2m\right]\ (\mathbf{k}+\mathbf{k}_n)^2 - E(\mathbf{k})} \tag{29}$$

(see Friedman 1956). The Green's function method is closely related to the augmented plane-wave method and can be used to obtain the same results for muffin-tin type of potentials. A substantial reduction of computation is possible when the analysis is restricted to the immediate neighborhood of certain symmetry points in the crystal, with the expansion (Eq. 29).

2.1.8 Linearized Muffin-Tin Orbital Method

This approach is based on ideas of the LAPW and Green's-function methods outlined above (Andersen 1973, 1975; Skriver 1984). It assumes a muffin-tin potential (Eq. 11) and expands the wavefunction inside the muffin-tin sphere as linear combinations of atomic functions as in LAPW. In the interstitial region outside the muffin-tin sphere, however, the wavefunctions are expanded in terms of *spherical* waves as in scattering problems. Such wavefunctions are referred to as muffin-tin orbitals. The Bloch function of the crystal is expanded as muffin-tin orbitals referring to atomic sites. The expansion coefficients and the eigenvalues are obtained from solving the Schrödinger equation with the muffin-tin potential (Martin 2004).

The crystal symmetry and the linearization of the wavefunction included in the LAPW approach lead to a significant simplification, yielding a highly efficient method. The prerequisite of symmetry leads to good results for solids with close packed structures, while the method works less well for solids with open structures like diamond and related lattices or for structures containing surfaces or interfaces.

2.2 Approaches Explicitly Containing Electron–Electron Interaction

A true many-particle Hamilton operator contains an electron–electron interaction term of the type $V(\mathbf{r}_i – \mathbf{r}_j)$ and cannot be represented as a sum of single-particle terms. Methods explicitly including the electron–electron interaction require, however, still approximations to become manageable. Often a system of interacting electrons is mapped onto a system with *non*interacting electrons comprising effective parameters; the interaction is then contained in the parameters, which are usually determined self-consistently.

2.2.1 Hartree–Fock Approximation

Using the Slater determinant for the many-electron wavefunction, Fock obtained an additional contribution to the potential referred to as *electronic exchange interaction* (Fock 1930; Corson 1951), yielding with Eq. 21:

$$\left(-\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial \mathbf{r}^2} - \sum_I \frac{Z_I e^2}{|\mathbf{r} - \mathbf{R}_I|} - V_{\text{el–el}}(\mathbf{r}) - V_{\text{ex}} \right) \psi_n(\mathbf{r}) \;=\; E_n \;\psi_n(\mathbf{r}) \qquad (30)$$

The *exchange interaction* (or *Fock contribution*) is given by

$$V_{\text{ex}} \;\psi_n(\mathbf{r}) = -e^2 \sum_{\substack{m \\ m \neq n}} \int \frac{\psi_m^*(\mathbf{r}_j)\psi_n(\mathbf{r}_j)}{|\mathbf{r} - \mathbf{r}_j|} \psi_m(\mathbf{r})\, d\mathbf{r}_j \qquad (31)$$

This is a Coulomb term arising from the antisymmetry of the wavefunction and the indistinguishability of electrons.

The *Hartree–Fock approximation* (HFA) poses some difficulties, as it renders the potential a nonlocal operator (it is an integral operator). Many attempts deal with a more suitable way of using the Hartree–Fock concept to arrive at an appropriate potential (Löwdin 1956; Pratt 1957; Slater 1953a, b; Wood and Pratt 1957; Pisani et al. 1988). Other methods take care of electron–electron interaction more adequately (Brueckner 1955; Pines 1956; Bohm et al. 1957; Hubbard 1957; Hedin and Lundqvist 1970; Aulbur et al. 2000).

Hartree–Fock-based calculations have successfully been applied to clusters containing not too many atoms. The large number of multicenter integrals imposes limits of the applicability of the method to smaller systems, even if some multicenter integrals and charge self-consistency are neglected. Hartree–Fock calculations can hence usually not be applied directly to mesoscopic or larger systems. It must be noted that Hartree–Fock band-structure calculations similar to Hartree calculations overestimate the bandgap energies of solids.

2.2.2 Density-Functional Theory

Today, methods applying the *density-functional theory* (DFT) are a standard tool for atomistic modeling, due to a good balance between accuracy and computational cost. The approaches are based on the theorem that the energy of the many-electron system is a functional of the electron density $n(\mathbf{r})$ and that this functional $E[n]$ has its minimum at the exact ground-state electron density n_0. The electronic contribution is separated into a Hartree Coulomb potential of the electron–electron interaction according to Eq. 21 and an exchange correlation (xc) term, which both depend on the ground-state charge–density distribution $\rho(\mathbf{r}) = en(\mathbf{r})$ (McWeeny 1957; Chirgwin 1957). The method applies a formalism of Hohenberg and Kohn (1964); see also Kohn and Sham (1965), Levy (1982), and Dreizler and Gross (1990), using a potential

$$V[\rho(\mathbf{r})] = V_{\text{ext}}(\mathbf{r}) + V_{\text{el-el}}[\rho(\mathbf{r})] + V_{\text{xc}}[\rho(\mathbf{r})] \tag{32}$$

where V_{ext} is the external (to the electron) potential imposed by the ions and V_{xc} represents the correction potential including exchange and *correlation*. Using the potential Eq. 15, the intractable N-electron Schrödinger equation is exactly mapped to a manageable set of N self-consistent one-electron equations, with all many-body interactions lumped into the single additive exchange-correlation potential V_{xc}. The correction potential V_{xc} accounts for the exchange interaction and the correlation of all N electrons in the solid. The exchange-correlation potential is formally defined as the functional derivative

$$V_{\text{xc}}(\mathbf{r}) \quad = \frac{\delta E_{\text{xc}}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \tag{33}$$

where $\delta E_{\text{xc}}[n(\mathbf{r})]$ is the functional of the exchange-correlation energy. Since the general form of $\delta E_{\text{xc}}[n(\mathbf{r})]$ is not known[8], several approximations were developed, the most simple being the *local density approximation* (LDA):

$$E_{\text{xc}}^{\text{LDA}}[n(\mathbf{r})] = \int n(\mathbf{r}) \varepsilon_{\text{xc}}^{\text{LDA}}(n(\mathbf{r})) \; d\mathbf{r} \tag{34}$$

$\varepsilon_{\text{xc}}^{\text{LDA}}$ is the exchange-correlation energy *per particle*, representing the interaction of the considered electron with a many-electron system of constant electron density (see Ceperley and Adler 1980), and a parametrization by Perdew and Zunger (1981). The LDA approach proved to describe electron density, atomic geometry, and other quantities of bulk crystals and surfaces with high precision. Larger errors with too large binding or bulk cohesive energies are obtained in the description of

[8]The exchange-correlation potential is a unique functional of the electron density; there hence exists a one-to-one relation from $V_{\text{xc}}(\mathbf{r})$ to $n(\mathbf{r})$ and vice versa. In the so-called *inverse DFT problem* accurate exchange functionals are constructed by transforming $n(\mathbf{r})$ into $V_{\text{xc}}(\mathbf{r})$ (van Leeuwen and Baerends 1994, Jensen and Wassermann 2018). Employing a numerical approach using a finite-element basis and machine learning, an accurate solution to the inverse DFT problem reproducing experimental data was recently presented for polyatomic molecules (Kanungo et al. 2019).

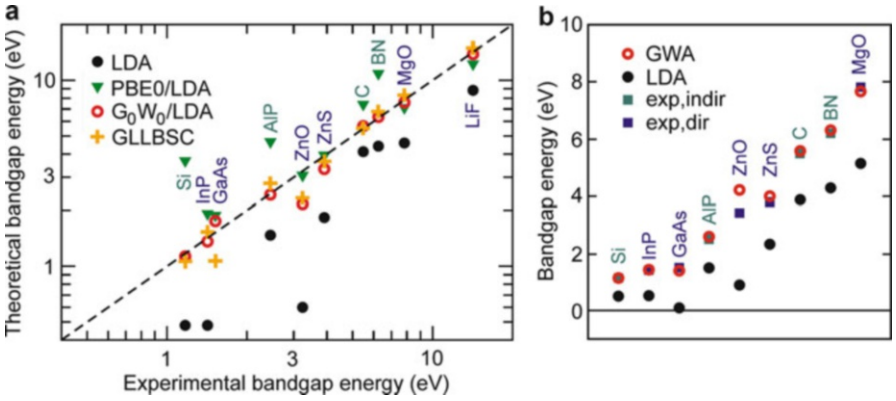


Fig. 5 Calculated and experimental bandgap energies, symbols in a column refer to the labeling solid. (**a**) Method LDA denotes local-density approximation; PBE0/LDA and G_0W_0/LDA refer to the non-self-consistent model-potential method (Gritsenko et al. 1995) and a frequency-dependent non-self-consistent GW method, both based on LDA (After Hüser et al. 2013). (**b**) Comparison of experimental and theoretical bandgap energies obtained from GWA and LDA calculations; *blue* and *green labels* denote direct and indirect bandgaps (Data from Aulbur et al. 2000)

spatially inhomogeneous systems such as atoms or small clusters. Improvements are achieved using also the local *gradient* of the density $[\partial\, n(\mathbf{r})/\partial \mathbf{r}]$; see Perdew et al. (1996). Such *generalized gradient approximations* (GGA) are necessary for describing small systems like molecules, and, addidionally using the PBE parametrization by Perdew et al. (1996), were also successfully applied to describe semiconductors.

In a Bloch-function description, the ground-state charge density

$$\rho_n(\mathbf{r}) = e\, n(\mathbf{r}) = e \sum_n \sum_{\mathbf{k}} |\psi_n(\mathbf{k}, \mathbf{r})|^2 \tag{35}$$

is a sum over occupied bands n and wavevectors $\mathbf{k}$ in the entire Brillouin zone of the crystal. Since $\psi_n(\mathbf{k}, \mathbf{r})$ depends on $V[n(\mathbf{r})]$ and $n(\mathbf{r})$ depends on $\psi_n(\mathbf{k}, \mathbf{r})$, these calculations must be carried out in a self-consistent manner; in the DFT approach the *electron density*, rather than the many-electron wavefunction, is calculated self-consistently from first principles. The density profile can be plotted in a two-dimensional representation as equidensity contour lines, permitting a distinction of electrons from different bands. Examples for electron-density profiles of NaCl and diamond are given in Fig. 10 in ▶ Chap. 2, "Crystal Bonding."

Density-functional theory is widely applied today for electronic-structure calculations ranging from chemistry to condensed-matter physics. Mostly used are LAPW, LMTO, and pseudofunctional methods for all (core + valence)-electron calculations applying the local-density approximation. The fully quantum-mechanical calculation is capable for treating systems with thousands of electrons; DFT based on pseudopotentials and a plane-wave basis set became a standard today for calculating ground-state properties of semiconductors (Martin 2004).

LDA and GGA methods are well suited to predict *ground-state* properties, but usually fail to describe excited-state properties. The bandgap energy of semiconductors is commonly underestimated (bandgap problem, see Fig. 5), because E_g is an excited-state property since it is the unbound limit of the exciton series; the incorrect prediction of E_g may lead to false assignments (Vidal et al. 2011). Associated with this problem, effective masses are often significantly underestimated as well.

The poor agreement of bandgap DFT calculations using the LDA or GGA approaches stimulated studies using the exact exchange within the DFT scheme as proposed by Städele et al. (1999); this *exact-exchange* (EXX) method is also referred to as *optimized effective potential* (OEPx) method. The basic idea is the use of *orbital-dependent* exchange-correlation functionals; the method uses Kohn–Sham orbitals (being functionals of the density) as ingredients in approximate exchange-correlation functionals which are *implicit* density functionals (Engel 2009). A review is given by Kümmel and Kronik (2008).

The EXX approach yields both structural and optical properties in much better agreement with experiment than corresponding LDA or GGA calculations. Improved gaps were found both with all (core + valence)-electron calculations on the basis of the atomic-sphere approximation (Kotani 1995) and with full-potential plane-wave pseudopotential PP calculations (Städele et al. 1999). A disadvantage of the EXX method is the large numerical complexity.

Further developments of the GGA approach termed meta-GGA (mGGA) include also higher order derivatives of the electronic density (similar to a Taylor expansion), or the kinetic energy density (KED). These additions introduce more nonlocal effects. Using the *constrained and appropriately normed* (SCAN) mGGA functional (Sun et al. 2015; Yang et al. 2016) or other mGGA approaches (Patra et al. 2019) the description of atomic interactions was substantially improved; the calculations implemented the optimized effective potential noted above. Also the error in the prediction of E_g was significantly reduced.

2.2.3 Quasiparticle GW Calculations

The GW approach applies the many-body perturbation theory using the *quasiparticle* concept (Hedin 1965). The central idea is based on the electronic self-energy, which is an energy-dependent and spatially nonlocal analogous of the exchange-correlation potential of DFT. In a descriptive picture the Coulomb repulsion between interacting electrons leads to a depletion of the negative charge around a considered electron; the ensemble of this electron and the surrounding positive polarization cloud forms a quasiparticle. This quasiparticle interacts with other quasiparticles via a screened rather than the bare Coulomb potential; the success of one-particle theories such as DFT demonstrates that the quasiparticle concept works well in solids, despite the strong interactions between bare particles (Aulbur et al. 2000).

The energy difference between states of the quasiparticle and those of the bare particle is described by the *self-energy* Σ. This energy equals the energy of the bare particle interacting with itself via the polarization cloud that the particle generates in the many-body system. Σ is nonlocal and energy dependent, and accounts for all exchange and correlation effects. Analogous to the exchange-correlation energy in DFT, the self-energy cannot be determined exactly. The simplest approximation for

the self-energy is the *GW approximation* (GWA), where Σ is expanded to first order in the screened interaction; here G denotes the Green's function and W the dynamically screened Coulomb interaction. Σ should be evaluated self-consistently, but due to the computational complexity rather non-self-consistent G_0W_0 calculations with an initial G_0 and a dielectric constant both obtained from LDA calculations are used.

The GWA method yields quasiparticle energies in excellent agreement with experimentally determined band structures. Figure 5 compares calculated and experimental bandgaps for a number of solids.

Many band-structure calculations apply a combination of various methods. One such approach is an *a posteriori* correction of the LDA approach by the GW method to obtain correct bandgaps. The application of the quasiparticle GW ansatz as a perturbation to the ground state of the fictitious, noninteracting Kohn–Sham electrons of exact-exchange density-functional theory removes the self-interaction by the EXX approach; this yields a stronger localization of cation d electrons and thus a reduced p–d hybridization compared to the local-density approximation (e.g., Rinke et al. 2005).

3 Relativistic Effects

Electrons described as Bloch electrons are free (except for scattering) to move within the periodic potential of the crystal; their velocity in a semiconductor is on the order of the thermal velocity (~10^7 cm/s), and relativistic effects can be neglected. However, this is no longer true for electrons that move in the strong local field near the nuclei, with velocities approaching the velocity of light. Here relativistic terms in the wave equation must be considered. One replaces the Schrödinger equation with the *Dirac relativistic equation* and a four-component spinor wavefunction (Rose 1961), leading to an equation for the upper two-component spinor ψ as

$$\left[\left(-\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial \mathbf{r}^2}+V(\mathbf{r})+\frac{1}{8m_0^3c^2}\frac{\partial^4}{\partial \mathbf{r}^4}-\frac{\hbar^2}{4m_0^2c^2}\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}}\cdot\frac{\partial}{\partial \mathbf{r}}\right)\underline{1}-\frac{i\ \hbar^2}{4m_0^2c^2}\underline{\sigma}\cdot\left(\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}}\times\frac{\partial}{\partial \mathbf{r}}\right)\right]\psi = E\psi \tag{36}$$

with $\underline{1}$ as the unity matrix and $\underline{\sigma}$ as the Pauli operator matrices.[9] There are three relativistic terms; the third and fourth terms represent the relativistic corrections to kinetic and potential energies (the *Darwin and mass–velocity corrections*), and the fifth term represents the spin-orbit coupling; this term describes the interaction of the electron spin with the magnetic moment of the electron in its orbit. It reduces the symmetry, causes removal of some degeneracies of valence-band states, and thereby determines the *spin-orbit splitting* of valence bands.

[9]The components of $\underline{1}$ and $\underline{\sigma}$ are

$$\underline{1}=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \sigma_x=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad \sigma_y=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_z=\begin{pmatrix}1&0\\0&-1\end{pmatrix}.$$

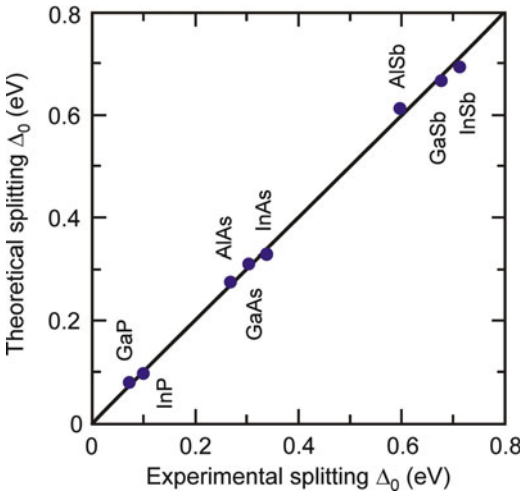


Fig. 6 Calculated and experimental spin-orbit splitting Δ_0 of the valence band in various zincblende semiconductors (After Majewski et al. 2004)

The spin-orbit splitting becomes more pronounced for heavy elements with larger nuclear charges. At the Γ point, the valence band of Si with $Z = 14$ splits by 0.04 eV, while for Ge with $Z = 32$ the valence band splits by 0.30 eV (see Table 5 in ▶ Chap. 8, "Bands and Bandgaps in Solids," for a listing of such splittings). The atomic spin-orbit splitting is nearly the same as that in a crystal since the interaction occurs deep in the atomic core, where the surrounding atoms of the crystal have little influence.

Relativistic corrections to classical (OPW and APW) band-structure calculations have been made by Soven (1965), Loucks (1965), Onodera and Okazaki (1966), and Pay-June Lin-Chung and Teitler (1972).

The density-functional theory is generalized to a spin-polarized form to account for spin-orbit splitting (von Barth and Hedin 1972; Rajagopal and Callaway 1973). For this purpose the two spin orientations $\sigma = \uparrow$ and $\downarrow$ of the carrier density $n_\sigma(\mathbf{r})$ are considered, yielding the total carrier density $n\ (\mathbf{r}) = n_\uparrow + n_\downarrow$. The exchange-correlation potential is then defined by

$$V_{\mathrm{xc},\sigma}\left(\left[n_\uparrow, n_\downarrow\right], \mathbf{r}\right) = \frac{\delta E_{\mathrm{xc}}\left[n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})\right]}{\delta n_\sigma(\mathbf{r})}. \tag{37}$$

The general form of $\delta E_{\mathrm{xc}}[n_\uparrow(\mathbf{r}),\ n_\downarrow(\mathbf{r})]$ is not known. Usually the *local-spin-density approximation* (LSDA) is used; this approach assumes that the exchange-correlation energy per particle at each point in space is given by its value for a homogeneous electron gas, yielding

$$E_{\mathrm{xc}}^{\mathrm{LSDA}}\left[n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})\right] = \int n(\mathbf{r})\varepsilon_{\mathrm{xc}}^{\mathrm{LSDA}}\left(n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})\right)\ d\mathbf{r} \tag{38}$$

analogous to Eq. 34. Merits and shortcomings discussed for the LDA approach also apply here, and respective amendments were developed. For details, see Kümmel

and Kronik (2008); implementation of the relativistic LDA method including the construction of relativistic pseudopotentials is reported by Majewski et al. (2004). The spin-orbit interaction arises from the *p* and *d* electrons in the core and thus enters the electronic Hamiltonian only in the *ionic* pseudopotential explicitly since the gradients of the potential generated by the valence electrons are much weaker than the gradients of the ionic potential.

The valence-band spin-orbit splitting Δ_0 at the Γ point calculated using the relativistic pseudopotential LDA approach is illustrated for various III–V semiconductors in Fig. 6. The results agree well with experimental data and show that the magnitude of Δ_0 is largely determined by the anion, in agreement with the anion *p*-type character of the valence band in the considered zincblende semiconductors.

4 Band Structure of Three-Dimensional Lattices

The dispersion relation $E_n(\mathbf{k})$ for three-dimensional lattices yields additional bands, which may or may not overlap. In any crystallographic direction, however, each band shows single-valued curves. To get a better perception of the topography of this $E_n(\mathbf{k})$ structure, which is a three-dimensional hypersurface in the four-dimensional (E, $\mathbf{k}$)-space, we start our discussion from a two-dimensional display of $E(\mathbf{k})$ in the first Brillouin zone.

As shown in the one-dimensional example in Sect. 2 in ▶ Chap. 6, "The Origin of Band Structure," the translational symmetry of the lattice permits a *reduced representation* of the dispersion relation $E(\mathbf{k})$ (see Fig. 13 in ▶ Chap. 6, "The Origin of Band Structure") with a periodicity ($2\pi/a$ in the k_x direction) related to the *reciprocal lattice periodicity.* This periodicity is maintained in three dimensions. For example, one has within a primitive orthorhombic unit cell

$$-\frac{\pi}{a} < k_x < \frac{\pi}{a}, \quad -\frac{\pi}{b} < k_y < \frac{\pi}{b}, \quad -\frac{\pi}{c} < k_z < \frac{\pi}{c}.$$

For a complete discussion, the momentum vector $\mathbf{k}$ can be restricted to this *reduced* cell, the first Brillouin zone, which contains all relevant information (see Sect. 1.3.1 in ▶ Chap. 3, "The Structure of Semiconductors").

$E(\mathbf{k})$ can be plotted easily in a one-dimensional lattice (see Fig. 14 in ▶ Chap. 6, "The Origin of Band Structure"). The $E(\mathbf{k})$ behavior of the first two bands of a two-dimensional structure is shown in Fig. 7; it represents one curved surface for each of the bands. The center ($k_x = k_y = 0$) is denoted by Γ. Cuts of this surface with a plane parallel to the energy axis show a different $E(\mathbf{k})$ behavior, depending on the orientation of the plane in the k_x and k_y directions. However, because of symmetry of the assumed square lattice, a cut normal to the k_x direction must result in the same $E(\mathbf{k})$ as a cut at the same k value normal to the k_y direction. Another cut at 45° between k_x and k_y shows a different $E(\mathbf{k})$ from the aforementioned, but the same $E(\mathbf{k})$ for all equivalent 45° cuts. All essential elements of the $E(\mathbf{k})$ surfaces are contained in such

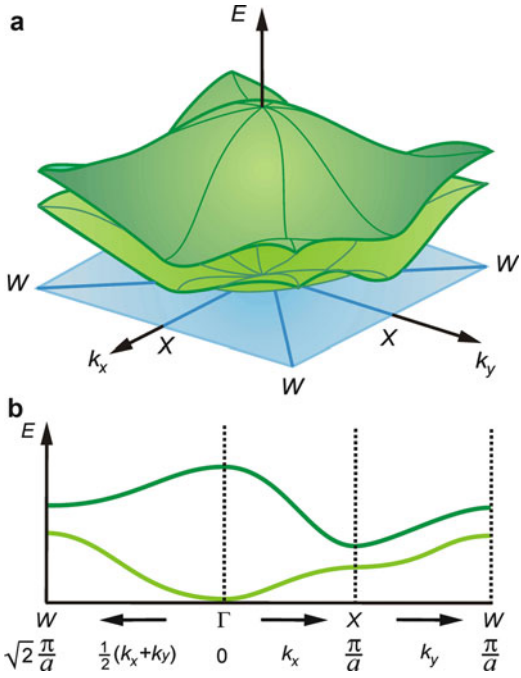


Fig. 7 (**a**) $E(\mathbf{k})$ surfaces for two bands in a hypothetical two-dimensional square lattice, with symmetry points X and W indicated at the boundary of the first Brillouin zone (*bottom plane*). (**b**) $E(\mathbf{k})$ in a conventional one-dimensional $E(k)$ representation. The diagram is broken at the *vertical lines* representing symmetry points W, Γ (center of Brillouin zone) and X

cuts. The intersection of the major crystallographic symmetry axes with the surface of the first Brillouin zone is identified in Fig. 7a with appropriate letters[10] (X for the first and W for the second in the example given above). $E(\mathbf{k})$ in a two-dimensional representation shown in Fig. 7b is plotted along a $\mathbf{k}$ vector connecting these points (compare subfigures a and b of Fig. 7). A similar representation of cuts through the $E(\mathbf{k})$ hypersurface is conventionally made in three-dimensional lattices and will be discussed in Sect. 4.2.

To gain a better perception about the shape of $E(\mathbf{k})$ for a real lattice, we will start from free electrons with well-known paraboloid behavior and progressively introduce a hypothetical lattice with an increasing amplitude of the lattice potential.

4.1 Empty and Nearly Empty Lattices

A crystal lattice with lattice periodicity but vanishing lattice potential is referred to as an *empty lattice* (Shockley 1938). Introducing a very small lattice potential, with just enough of an amplitude to influence slightly the $E(\mathbf{k})$ behavior obtained for the empty lattice, defines this as a *nearly empty lattice*. The development of $E(\mathbf{k})$ with an

[10]See Fig. 8 in ▶ Chap. 3, "The Structure of Semiconductors": Γ for the center at $k = (0,0,0)$; in a face-centered cubic lattice for the point L: $k = \frac{2\pi}{a}(\frac{1}{2},\frac{1}{2},\frac{1}{2})$; for the point X : $k = \frac{2\pi}{a}(1,0,0)$.

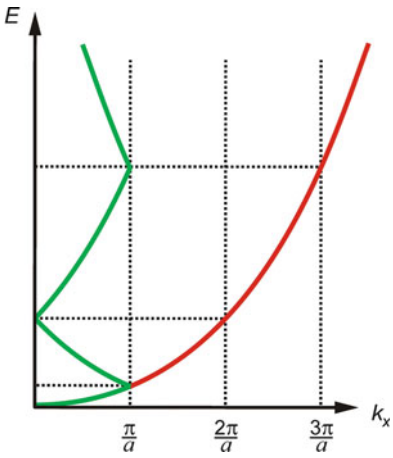


Fig. 8 E(k) for a free electron on an empty one-dimensional lattice (*red curve*) and its reduced representation (*green*) between 0 and π/a

increasing amplitude of the lattice potential is instructive in understanding the origin of the different bands in an actual crystal.

Starting with a free electron in vacuum, one obtains the well-known parabolic $E(\mathbf{k})$ behavior (see Sect. 2 in ▶ Chap. 6, "The Origin of Band Structure"), redrawn in Fig. 8. Adding to the model an empty, one-dimensional lattice with lattice constant a does not change $E(\mathbf{k})$. However, one can now insert multiples of $\frac{\pi}{a}$ on the k_x axis and fold the diagram to a reduced $E(\mathbf{k})$ representation, as indicated by the green curves in Fig 8. The $E_n(\mathbf{k})$ diagram given in Fig. 9a is obtained by replacing the one-dimensional lattice with the empty face-centered cubic lattice of the same geometry as the Ge lattice (Herman 1958). The index n indicates the different branches of $E(\mathbf{k})$. The comparison of Fig. 9a with Fig. 8 shows that several $E(\mathbf{k})$ curves (bands) overlap in the empty Ge lattice while they do not in the empty primitive cubic lattice; this overlap is caused by the additional lattice point in the Ge lattice. The discussion of an empty lattice provides a useful method of determining the sequence of the different bands since such a sequence does not depend on the lattice potential.

The periodic crystal potential, switched on at a very small amplitude, yields $E_n(\mathbf{k})$ of the *nearly empty lattice*, shown in Fig. 9b. A deformation and splitting in the $E_n(\mathbf{k})$ dispersion relation becomes visible. With the knowledge that the unit cell of the diamond lattice contains 2 atoms, and thus 8 valence electrons, one concludes that the lowest four $E_n(\mathbf{k})$ curves, occupied by two electrons each due to spin degeneracy, belong to one set of bands. This set of bands will later be identified as the valence band, disregarding degeneracy, and the curves above the fourth belong to the next higher set of bands, later defined as the conduction band (Sect. 1 in ▶ Chap. 8, "Bands and Bandgaps in Solids"). The bandgap (shaded region in Fig. 9) is expected above the fourth and below the fifth $E_n(\mathbf{k})$ curves. The actual bandgap is the smallest distance between the highest point of the fourth and the lowest point of the fifth curves in Fig. 9b. These points are not necessarily on top of each other; see Fig. 11.

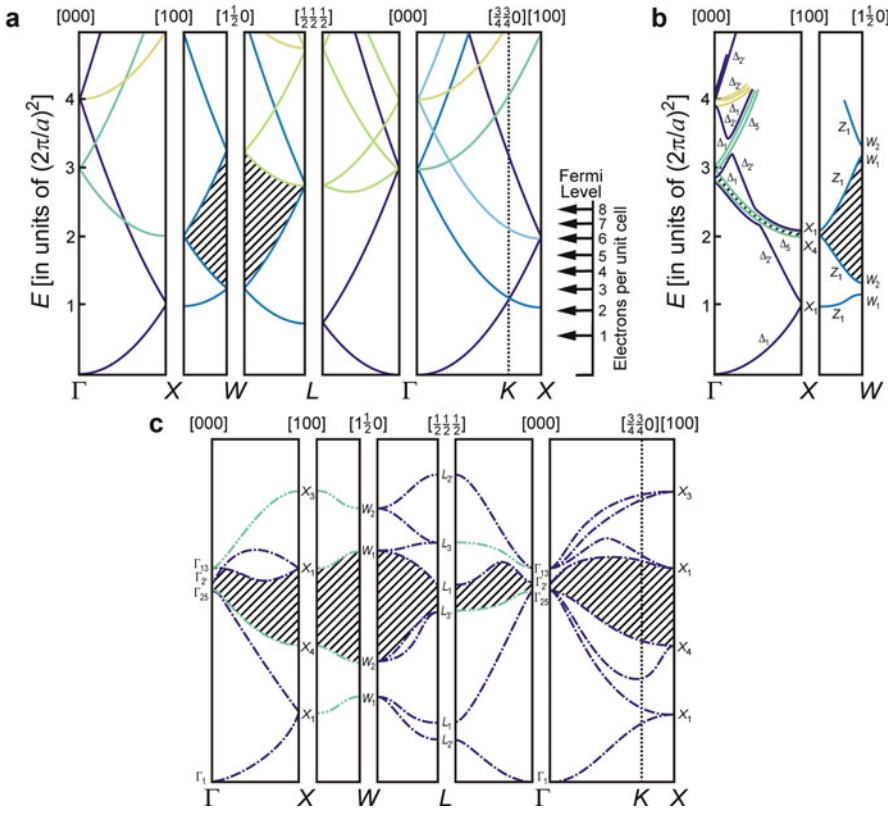


Fig. 9 $E(\mathbf{k})$ of the Ge lattice, disregarding spin-orbit splitting: (**a**) empty lattice, (**b**) segment Γ-*X*-*W* of the nearly empty lattice, and (**c**) actual electronic band structure of Ge. Colors serve solely to distinguish curves. The hatched area represents the region between the uppermost valence band $E_v(\mathbf{k})$ and the lowest conduction band $E_c(\mathbf{k})$ and has no other meaning per se; the respective smallest distance at Γ represents the *direct bandgap*; the distance between the absolute minimum of $E_c(\mathbf{k})$ at *L* and the absolute maximum of $E_v(\mathbf{k})$ at Γ represents the *indirect bandgap* (After Herman 1958)

By comparing Fig. 9a with b, one obtains guidance for the developing band structure. The splitting of the bands and the avoidance of a crossover can be clearly seen in the left diagram between Γ and *X*. The actual band structure is shown in Fig. 9c for the full amplitude of the periodic lattice potential of Ge.

4.2 The Band Structure of Typical Semiconductors

A computation of the band structure of Ge by Herman (1958) is shown in Fig. 9c. This figure also shows the common notation of the bands introduced by Bouckaert et al. (1936) – see Bassani and Pastori Parravicini (1975) for a review. Further modifications of $E_n(\mathbf{k})$, due to the introduction of spin-orbit interaction, can be seen by comparing Figs. 9c and 11b.

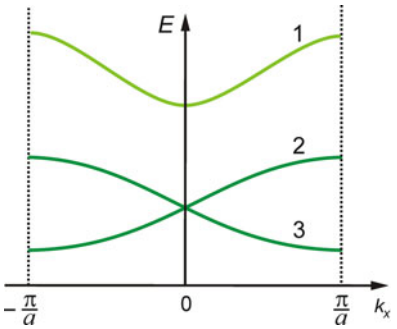


Fig. 10 Symmetry-related information about the shape of $E_n(\mathrm{k})$ at $k = 0$ and at the surface of the Brillouin zone

4.2.1 Symmetry of *E*(k)

The electronic bands $E_n(\mathbf{k})$ have the same point-group symmetry as the crystal to which the Brillouin zone belongs (see Bassani and Pastori Parravicini 1975). In general, one has in a nonreduced (i.e., extended) Brillouin zone representation

$$E_n(\mathbf{k}) = E_n(\mathbf{k} + \mathbf{K}) \tag{39}$$

$$E_n(\mathbf{k}) = E_n(-\mathbf{k}) \tag{40}$$

$$E_n(\mathbf{k}) = E_n(\alpha\mathbf{k}) \tag{41}$$

with **K** as a lattice vector in the reciprocal lattice and α identifying any point-group operation, e.g., a rotation; the relation Eq. 40 is also known as *Kramer's theorem*. Consequently, for nondegenerate $E_n(\mathbf{k})$ at the center (Γ) or at the surface of the Brillouin zone, $E_n(\mathbf{k})$ must have an extremum, as can be seen from fulfilling Eqs. 39 and 40 simultaneously. However, noninteracting branches $E_n(\mathbf{k})$ may cross,[11] thus permitting a finite slope at the symmetry point, as shown for the Γ point for curves 2 and 3 in Fig. 10.

Extrema of $E_n(\mathbf{k})$, however, are not limited to the center or surfaces of the Brillouin zone, as Fig. 11 shows. There is a maximum of $E(\mathbf{k})$ along Σ and along Δ (see Fig. 11 and in ▶ Chap. 3, "The Structure of Semiconductors," and Fig. 8) for one of the Γ_{15} and Γ'_2 branches. In Si, there is, e.g., a secondary minimum at the bottom of the conduction band at ~$0.8\frac{\pi}{a}k_x$. Other minima are at the equivalent points – see Fig. 10 in ▶ Chap. 8, "Bands and Bandgaps in Solids." Such minima are referred to as *satellite minima* and play an important role in the semiconductive properties of the material: they determine the properties of the conduction electrons.

In the neighborhood of extrema or saddle points of $E(\mathbf{k})$ (*critical points*), $E(\mathbf{k})$ can be written in three dimensions as

[11]The crossing $E_n(\mathbf{k})$ must belong to different symmetry states that cannot interact with each other. States of the same symmetry interact and cannot cross (noncrossing rule). Examples of crossing can be seen in Fig. 11 for Si at X_1 and for noncrossing at Δ for GaAs.

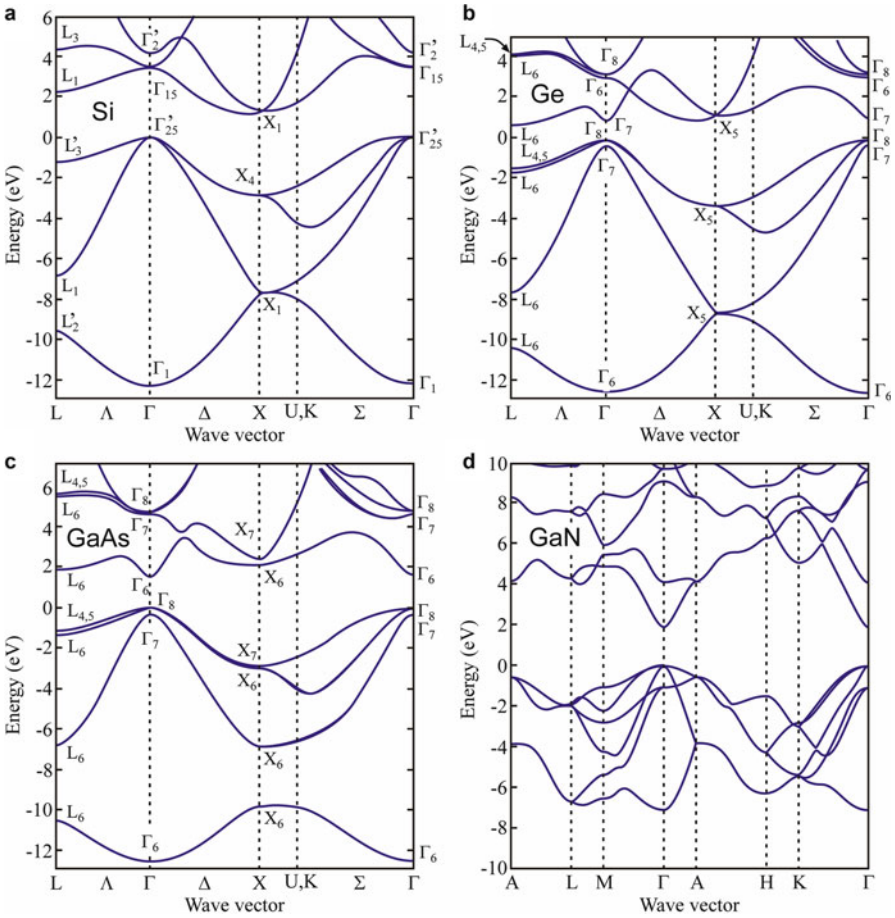


Fig. 11 Energy bands of diamond structure (**a**) Si and (**b**) Ge and of (**c**) zincblende GaAs calculated using the empirical nonlocal pseudopotential method (EPM) (After Chelikowsky and Cohen 1976). (**d**) Energy bands of wurtzite GaN calculated using full-potential linear muffin-tin orbitals (LMTO) (After Lambrecht 1998)

$$E(\mathbf{k}) = E(\mathbf{k}_0) + \sum_{i=1}^{3} a_i (k_i - k_{0i})^2. \tag{42}$$

This permits us to distinguish four types of critical points, which are classified as

M_0	for a_1, a_2, a_3 positive	(minimum)
M_1	for a_1 negative, a_2, a_3 positive	(saddle point)
M_2	for a_1, a_2 negative, a_3 positive	(saddle point)
M_3	for a_1, a_2, a_3 negative	(maximum)

The curvature of $E(\mathbf{k})$ relates to the effective mass. Therefore, M_0 relates to a positive and M_3 to a negative effective mass for electrons. M_1 is characterized by one negative and two positive effective masses, and M_2 is characterized by two negative and one positive mass in the respective directions. In all of these critical points, $\partial E/\partial \mathbf{k} = 0$. The critical points are of importance: here the density of states has a maximum; this will be discussed in the following section.

The parabolic approximation Eq. 42 for $E(\mathbf{k})$ near an extremum is insufficient to describe transport and excitation phenomena at higher energies; higher-order terms must then be taken into consideration. The large variety of bands (Fig. 11), however, necessitates restricting this discussion to the most important ones. These will be identified in Sect. 1 in ▶ Chap. 8, "Bands and Bandgaps in Solids," while the discussion on the shape of these bands will be postponed to Sect. 1.2 in ▶ Chap. 8, "Bands and Bandgaps in Solids."

Any further discussion of the different sets of energy bands shown in Fig. 11 requires an understanding of the symmetry properties of the electronic states, which is not the topic of this book. A systematic introduction can be found, for example, in Bassani and Pastori Parravicini (1975).

4.2.2 Density of States

Bands originate from the splitting of atomic eigenstates. Therefore, each band contains N states, where N is the number of atoms in the crystal times a degeneracy factor v_D (for a simple band $v_D = 2$). The density of states $g(E)$ per energy interval dE is then defined by[12]

$$v_D N = \int_{E_{\min}}^{E_{\max}} g(E)\, dE \tag{43}$$

For computational purposes, it is more convenient to express g as a function of $\mathbf{k}$ rather than of E with

$$g(E)\, dE = \begin{cases} g(k)\, \dfrac{dk}{dE} dE & \text{in one dimension} \\ g(\mathbf{k})\, \dfrac{1}{|\partial E(\mathbf{k})/\partial \mathbf{k}|} dE & \text{in three dimensions.} \end{cases} \tag{44}$$

Within the first Brillouin zone, one can easily follow the filling of the band with electrons. Since electrons are fermions with a spin of 1/2, they can fill each state only to a maximum of two electrons, one with spin up, the other with spin down. The states fill *consecutively* near $T = 0$ K; the highest states filled at $T = 0$ K identifies the Fermi energy, which is also referred to as the *Fermi level* (in ideal semiconductors

[12]Compare with a similar calculation of the density of states for phonons given in Sect. 3.2 in ▶ Chap. 4, "Elasticity and Phonons."

the Fermi level lies near the mid-gap – see Sect. 2 in ▶ Chap. 21, "Equilibrium Statistics of Carriers").

A single band of type 1 in Fig. 10 fills from the center (Γ) to complete first a small sphere (Fig. 7 in ▶ Chap. 8, "Bands and Bandgaps in Solids"), which, with continued filling, will become deformed. The degree of deformation depends on the strength of the lattice forces (see Sect. 1.1 in ▶ Chap. 8, "Bands and Bandgaps in Solids," and on the shape of bands, see Fig. 8 in ▶ Chap. 8, "Bands and Bandgaps in Solids"). With S_E the total area of the enclosing surface to which this filling proceeded, we can then express the volume of the **k** space between E and $E + dE$ by the integral over a closed surface at E:

$$\oint_E \frac{dS_E}{|\partial E(\mathbf{k})/\partial \mathbf{k}|} dE. \tag{45}$$

Since the volume in the entire first Brillouin zone, given here for a cubic crystal, is $(2\pi/a)^3$ and a crystal of volume V has $v_D V/a^3$ electron states per band, there are $8\pi^3/(v_D V)$ electron states in this band; therefore

$$g_n(E)\, dE \quad = \frac{v_D}{(2\pi)^3} \oint \frac{dS_E}{|\partial E(\mathbf{k})/\partial \mathbf{k}|} dE. \tag{46}$$

From Eq. 46 one sees that the increment of the density of states is steepest where the slope of $E(\mathbf{k})$ is the smallest. This is the case at or near the critical points described above.

The band index n is added for distinction, since with more than one band and with overlapping bands the density of states is additive:

$$g(E)\, dE \quad = \sum_n g_n(E)\, dE. \tag{47}$$

The total density of states contains a substantial amount of structure stemming from the critical points of the different bands, as shown in Fig. 12. This figure shows clearly the interrelationship between $E(\mathbf{k})$ for each band (right part of the figure) with the density distribution (left part). The listed symmetry points will assist in finding the related branches.

For most semiconducting properties, only the density of states near the edges of the bandgap is important. This density can be easily estimated in a parabolic $E(\mathbf{k})$ approximation and will be discussed in more detail in Sect. 1 in ▶ Chap. 21, "Equilibrium Statistics of Carriers."

4.3 Band Structure of Organic Crystals

Carriers in organic crystals couple strongly to phonons and are consequently described as polarons (Sect. 1.2 in ▶ Chap. 22, "Carrier-Transport Equations"). In

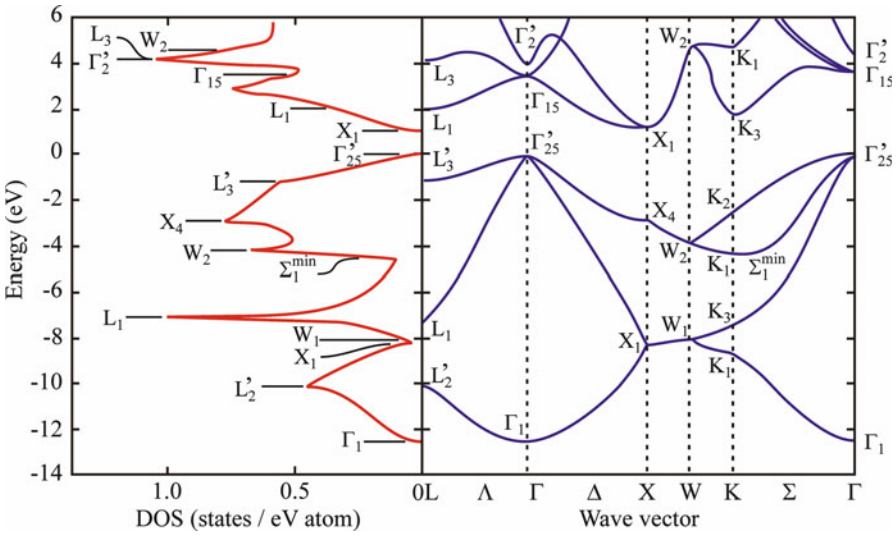


Fig. 12 Band structure $E(\mathbf{k})$ (*blue curves*) and density-of-state (DOS) distribution $g(E)$ (*red*) in valence and conduction bands of Si, with corresponding symmetry points identified. The DOS distribution is turned by 90° from the conventional representation to relate directly to the $E(\mathbf{k})$ representation (After Chelikowsky et al. 1973)

Table 1 Bandwidth W for crystalline organic semiconductors

	Bandwidth W (meV)		
Crystal	HOMO	LUMO	References
Anthracene	509	508	Chen et al. (2003)
Tetracene	625	502	Chen et al. (2003)
Pentacene	738	728	Chen et al. (2003)
Rubrene	420	180	Yanagisawa et al. (2013)
CuPc	96	260	Yang et al. (2008)

the low-temperature regime often band-like conduction similar to that of inorganic semiconductors is observed; carriers are then well characterized in terms of a band structure. Due to the weak intermolecular van der Waals bonds, the electronic bandwidth is generally quite narrow; data of the lowest unoccupied and the highest occupied molecular-orbital (LUMO and HOMO) bands for some organic semiconductors are given in Table 1. The pronounced electron–phonon coupling leads to a significant dependence of the bandwidth W on temperature (see Fig. 41 in ▶ Chap. 8, "Bands and Bandgaps in Solids"): the maximum bandwidth at $T \rightarrow 0$ decreases as the temperature increases (Holstein 1959; Hannewald et al. 2004).

The electronic band structure of an anthracene crystal calculated using the tight-binding approach is given in Fig. 13. The dispersions $E(\mathbf{k})$ refer to directions of the monoclinic Brillouin zone shown in Fig. 8e of ▶ Chap. 3, "The Structure of Semiconductors"; $\mathbf{d}_1 = (½½0)$ and $\mathbf{d}_1 = (-½½0)$ are next-neighbor directions in the xy

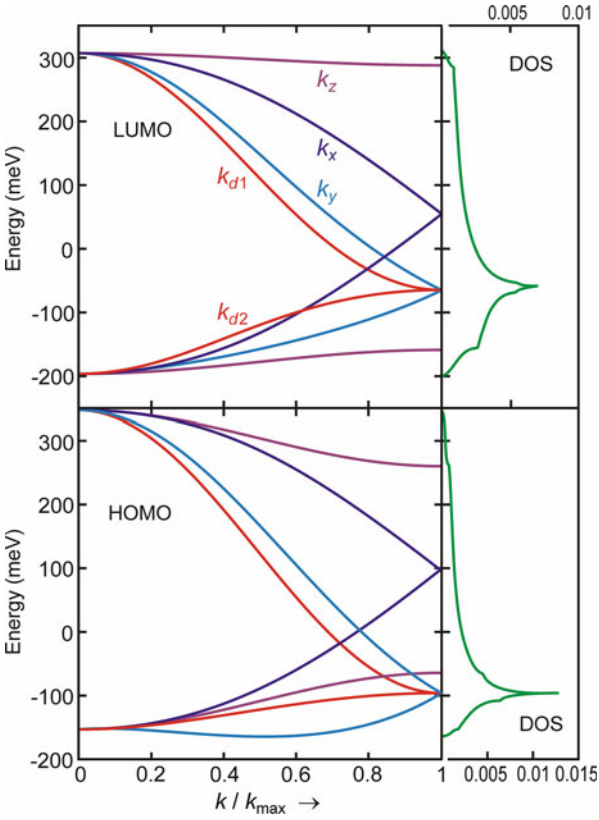


Fig. 13 Dispersion $E(k)$ and density of states (DOS) of the lowest unoccupied and the highest occupied molecular-orbital (LUMO and HOMO) bands for an anthracene crystal. k_x to k_z signify crystallographic directions of the monoclinic unit cell; k_{d1} and k_{d2} refer to nearest-neighbor directions (After Cheng et al. 2003)

plane. Anthracene has two equivalent molecules with different orientations in the unit cell; both the LUMO and HOMO levels are composed of symmetrical and asymmetric linear combinations of molecular orbitals of the two molecules, leading to a twofold splitting. The degeneracy of k_x and k_y at the edge of the Brillouin zone in the xy plane originates from a crystal glide-plane symmetry; no such symmetry exists in triclinic crystals like tetracene and pentacene, where this degeneracy is lifted. The dispersion along k_z is very small, and a large gap occurs between upper and lower band. These features originate in this direction from small interactions of molecules located in adjacent layers and are often found in organic crystals with herringbone packing (Cheng et al. 2003). The density of states of the LUMO and HOMO bands plotted at the right side of Fig. 13 reflects the inverse slope of the contributing dispersion curves.

Rubrene crystals also show band-like conduction; the orthorhombic phase considered here comprises two differently aligned molecule pairs per unit cell in a herringbone packing. The band structure shown in Fig. 14 was calculated applying density-functional theory with two different approaches: a plane-wave

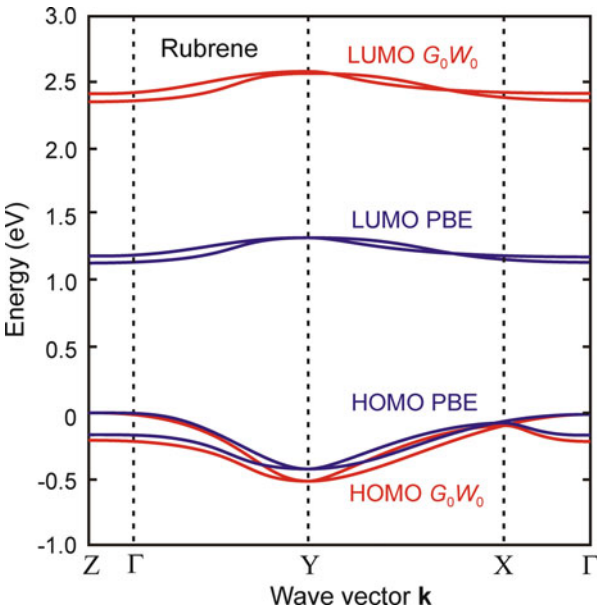


Fig. 14 Electronic dispersion of the LUMO and HOMO bands for a rubrene crystal, calculated using DFT and the PBE exchange-correlation functional (*blue*) or the G_0W_0 approximation (*red*) (After Yanagisawa et al. 2013)

pseudopotential treatment using the Perdew–Burke–Ernzerhof (PBE) exchange-correlation functional and a refined quasiparticle treatment with PBE wavefunctions and eigenvalues plus self-energy corrections within the G_0W_0 approximation (Yanagisawa et al. 2013). The direct bandgap of 2.34 eV of the G_0W_0 calculation agrees reasonably with an experimental optical band gap of 2.2 eV, reflecting the improvement also found for inorganic semiconductors (Fig. 5) by including self-energy corrections. The HOMO bands of rubrene show a large dispersion along Γ-*Y* and only minor dispersion along Γ-*Z*, in accordance with angle-resolved photoelectron-spectroscopy data (Machida et al. 2010) and the experimental large anisotropy of carrier transport.

5 Summary

The analysis of the atomic and electronic behavior of semiconductors rests on the solutions of the appropriate Schrödinger equation. This equation contains the interaction potential of all involved particles, i.e., electrons and atom nuclei. The interaction potential can be approximated in a number of different ways, accounting for electronic and ionic contributions, and can be computed from basic principles. The solution of the Schrödinger equation for a *one-electron* approximation is usually based on variational principles by expanding the wavefunction in terms of a set of orthogonal trial functions, followed by variation of the expansion coefficients.

Methods explicitly containing correction terms for *many-particle* electron–electron interactions account for exchange interaction and correlation of all electrons. Sufficiently accurate boundary conditions reduce the computational effort and yield results in good agreement with the experiment. The band structure and related density of states are computed for most of the common semiconductors and are available for quite a number of new materials with promising semiconducting properties. The development of computational techniques has led to a quantitative description of the band structure and the prediction of stable new compounds before they were synthesized. It permits prediction of their electronic and other properties and has become an essential tool in designing new materials for solid-state devices.

References

Andersen OK (1973) Simple approach to the band-structure problem. Solid State Commun 13:133
Andersen OK (1975) Linear methods in band theory. Phys Rev B 12:3060
Ashcroft NW, Mermin ND (1976) Solid state physics. Holt Reinhart and Winston, New York
Aulbur WG, Jönsson L, Wilkins JW (2000) Quasiparticle calculations in solids. Solid State Phys 54:1
Bachelet GB, Hamann DR, Schlüter M (1982) Pseudopotentials that work: from H to Pu. Phys Rev B 26:4199
Bardeen J (1938) An improved calculation of the energies of metallic Li and Na. J Chem Phys 6:367
Bassani GF, Giuliano ES (1972) Pseudopotential approach to the band structure of alkali halides. Nuovo Cimento 8B:193
Bassani GF, Pastori Parravicini G (1975) Electronic states and optical transitions in solids. Pergamon Press, Oxford/New York
Bell DG (1953) Group theory and crystal lattices. Rev Mod Phys 26:311
Bendt P, Zunger A (1982) New approach for solving the density-functional self-consistent-field problem. Phys Rev B 26:3114
Bloch F (1928) Über die Quantenmechanik der Elektronen in Kristallgittern. Z Phys 52:555 (On the quantum mechanics of electrons in crystal lattices, in German)
Bohm D, Huang K, Pines D (1957) Role of subsidiary conditions in the collective description of electron interactions. Phys Rev 107:71
Born M, Oppenheimer JR (1927) Zur Quantentheorie der Molekeln. Ann Phys 84:457 (On quantum theory of molecules, in German)
Bouckaert LP, Smoluchowski R, Wigner E (1936) Theory of Brillouin zones and symmetry properties of wave functions in crystals. Phys Rev 50:58
Brueckner KA (1955) Many-body problem for strongly interacting particles. II. Linked cluster expansion. Phys Rev 100:36
Burt MG (1992) The justification for applying the effective-mass approximation to microstructures. J Phys Condens Matter 4:6651
Cardona M, Pollak FH (1966) Energy-band structure of Germanium and Silicon: the k·p method. Phys Rev 142:530
Ceperley DM, Alder BJ (1980) Ground state of the electron gas by a stochastic method. Phys Rev Lett 45:566
Chang YC (1982) Complex band structures of zinc-blende materials. Phys Rev B 25:605
Chelikowsky JR, Cohen ML (1976) Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. Phys Rev B 14:556

Chelikowsky JR, Phillips JC (1978) Quantum-defect theory of heats of formation and structural transition energies of liquid and solid simple metal alloys and compounds. Phys Rev B 17:2453
Chelikowsky JR, Chadi DJ, Cohen ML (1973) Calculated valence-band densities of states and photoemission spectra of diamond and zinc-blende semiconductors. Phys Rev B 8:2786
Cheng YC, Silbey RJ, da Silva Filho DA, Calbert JP, Cornil J, Brédas JL (2003) Three-dimensional band structure and bandlike mobility in oligoacene single crystals: a theoretical investigation. J Chem Phys 118:3764
Chirgwin BH (1957) Summation convention and the density matrix in quantum theory. Phys Rev 107:1013
Cohen ML (1984) Application of the pseudopotential model to solids. Ann Rev Mater Sci 14:119
Cohen ML, Heine V (1970) The fitting of pseudopotentials to experimental data and their subsequent application. In: Ehrenreich H, Seitz F, Turnbull D (eds) Solid state physics, vol 24. Academic Press, New York, pp 37–248
Corson EM (1951) Perturbation methods in quantum mechanics. Blackie and Son, London
Dimmock JO (1971) The calculation of electronic energy bands by the augmented plane wave method. In: Ehrenreich H, Seitz F, Turnbull D (eds) Solid state physics, vol 26. Academic Press, New York, pp 103–274
Dreizler RM, Gross EKU (1990) Density functional theory. Springer, Berlin
Engel E (2009) Relevance of core-valence interaction for electronic structure calculations with exact exchange. Phys Rev B 80:161205
Fermi E (1934) Sopra lo spostamento per pressione delle righe elevate delle serie spettrali. Nuovo Cimento 11:157 (On the shift of high-pressure lines of the spectral series, in Italian)
Fock WA (1930) Näherungsmethoden zur Lösung des quantenmechanischen Mehrkörperproblems. Z Phys 61:126 (Approximate methods for the solution of the quantum-mechanical many-body problem, in German)
Friedman B (1956) Principles and techniques in applied mathematics. Wiley, New York
Froyen S (1996) Linear combinations of constituent states: a first-principles Luttinger model. J Phys Condens Matter 8:11029
Gritsenko O, van Leeuwen R, van Lenthe E, Baerends EJ (1995) Self-consistent approximation to the Kohn-Sham exchange potential. Phys Rev A 51:1944
Hannewald K, Stojanović VM, Schellekens JMT, Bobbert PA, Kresse G, Hafner J (2004) The theory of polaron bandwidth narrowing in molecular crystals. Phys Rev B 69:075211
Harrison WA (1966) Pseudopotentials and the theory of metals. W.A. Benjamin, New York
Harrison WA (1973) Bond-orbital model and the properties of tetrahedrally coordinated solids. Phys Rev B 8:4487
Harrison WA (1980) Electronic structure and the properties of solids: the physics of chemical bonds. Freeman, San Francisco
Hedin L (1965) New method for calculating the one-particle Green's function with application to the electron-gas problem. Phys Rev 139:A796
Hedin L, Lundqvist S (1970) Effects of electron–electron and electron–phonon interactions on the one-electron states of solids. In: Ehrenreich H, Seitz F, Turnbull D (eds) Solid state physics, vol 23. Academic Press, New York, pp 1–181
Heine V, Weaire D (1970) Pseudopotential theory of cohesion and structure. In: Ehrenreich H, Seitz F, Turnbull D (eds) Solid state physics, vol 24. Academic Press, New York, pp 249–463
Heitler W, London F (1927) Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik. Z Phys 44:455 (Interaction of neutral atoms and homopolar bond according quantum mechanics, in German)
Hellmann JJ (1935) A new approximation method in the problem of many electrons. J Chem Phys 3:61
Herman F (1958) Theoretical investigation of the electronic energy band structure of solids. Rev Modern Phys 30:102
Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864
Holstein T (1959) Studies of polaron motion Part II: the "small" polaron. Ann Phys (N Y) 8:343

Hubbard J (1957) The description of collective motions in terms of many-body perturbation theory. Proc Roy Soc (Lond) Ser A 240:539
Hüser F, Olsen T, Thygesen KS (2013) Quasiparticle GW calculations for solids, molecules, and two-dimensional materials. Phys Rev B 87:235132
Ihm J, Cohen ML (1980) Calculation of structurally related properties of bulk and surface Si. Phys Rev B 21:1527
Ihm J, Zunger A, Cohen ML (1979) Momentum-space formalism for the total energy of solids. J Phys C: Solid State Phys 12:4409
Jancu J-M, Scholz R, Beltram F, Bassani F (1998) Empirical *spds** tight-binding calculation for cubic semiconductors: General method and material parameters. Phys Rev B 57:6493
Jancu J-M, Voisin P (2007) Tetragonal and trigonal deformations in zinc-blende semiconductors: A tight-binding point of view. Phys Rev B 76:115202
Jensen DS, Wassermann A (2018) Numerical methods for the inverse problem of density functional theory. Int J Quant Chem 118:e25425
Kane EO (1956) Energy band structure in p-type germanium and silicon. J Phys Chem Solids 1:82; and: Kane EO (1957) Band structure of indium antimonide. J Phys Chem Solids 1:249
Kane EO (1976) Gaussian representations of covalent wave functions; silicon. Phys Rev B 13:3478
Kanungo B, Zimmerman PM, Gavini V (2019) Exact exchange-correlation potentials from ground-state electron densities. Nat Commun 10:4497
Kasowski RV, Tsai M-H, Rhodin TN, Chambliss DD (1986) Pseudofunction method: application to a monolayer of CO and to the Si(111) surface. Phys Rev B 34:2656
Kohn W, Rostoker N (1954) Solution of the Schrödinger equation in periodic lattices with an application to metallic Lithium. Phys Rev 94:1111
Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:A1133
Korringa J (1947) On the calculation of the energy of a Bloch wave in a metal. Physica 13:392
Kotani T (1995) Exact exchange potential band-structure calculations by the linear muffin-tin orbital - atomic-sphere approximation method for Si, Ge, C, and MnO. Phys Rev Lett 74:2989
Krakauer H, Posternak M, Freeman AJ, Koelling DD (1981) Initial oxidation of the A1(001) surface: Self-consistent electronic structure of clean A1(001) and Al(001)-p(1×1)O. Phys Rev B 23:3859
Kubo R (1952) Thermal ionization of trapped electrons. Phys Rev 86:929
Kümmel S, Kronik L (2008) Orbital-dependent density functionals: theory and applications. Rev Mod Phys 80:3
Laasonen K, Pasquarello A, Car R, Lee C, Vanderbilt D (1993) Car-Parrinello molecular dynamics with Vanderbilt ultrasoft pseudopotentials. Phys Rev B 47:10142
Lambrecht WRL (1998) Band structure of the group III nitrides. In: Pankove JI, Moustakas TD (eds) Semiconductors and semimetals, vol 50. Academic Press, San Diego, p 369
Leeuwen R, Baerends EJ (1994) Exchange-correlation potential with correct asymptotic behavior. Phys Rev A 49:2421
Levy M (1982) Electron densities in search of Hamiltonians. Phys Rev A 26:1200
Lew Yan Voon LC, Willatzen M (2009) The k.p method: Electronic properties of semiconductors. Springer, Berlin
Lin-Chung P-J, Teitler S (1972) Relativistic pseudopotential method. Phys Rev B 6:1419
Loucks TL (1965) Relativistic electronic structure in crystals. I. Theory. Phys Rev 139:A1333
Loucks TL (1967) Augmented plane wave method. W.A. Benjamin, New York
Löwdin P-O (1956) Quantum theory of cohesive properties of solids. Adv Phys 5:1–171
Machida S-i, Nakayama Y, Duhm S, Xin Q, Funakoshi A, Ogawa N, Kera S, Ueno N, Ishii H (2010) Highest-occupied-molecular-orbital band dispersion of rubrene single crystals as observed by angle-resolved ultraviolet photoelectron spectroscopy. Phys Rev Lett 104:156401
Mäder KA, Zunger A (1994) Empirical atomic pseudopotentials for AlAs/GaAs superlattices, alloys, and nanostructures. Phys Rev B 50:17393

Majewski JA, Birner S, Trellakis A, Sabathil M, Vogl P (2004) Advances in the theory of electronic structure of semiconductors. Phys Stat Sol C 1:2003
Martin RM (2004) Electronic structure: basic theory and practical methods. Cambridge University Press, Cambridge
McWeeny R (1957) The density matrix in self-consistent field theory. III. Generalizations of the theory. Proc Roy Soc (London) Ser A 241:239
Morse PM, Feshbach H (1953) Methods of theoretical physics. Mc Graw-Hill Book, New York
Onodera Y, Okazaki M (1966) An orthogonalized valence orbital approximation in relativistic full-potential linear-combination-of-atomic-orbitals methods. J Phys Soc Jpn 21:1273
Patra B, Jana S, Constantin LA, Samal P (2019) Efficient band gap prediction of semiconductors and insulators from a semilocal exchange-correlation functional. Phys Rev B 100:045147
Pelzer H, Wigner EP (1932) Über die Geschwindigkeitskonstante von Austauschreaktionen. Z Phys Chem B 15:445 (On the rate constant of exchange reactions, in German)
Perdew JP, Zunger A (1981) Self-interaction correction to density-functional approximations for many-electron systems. Phys Rev B 23:5048
Perdew JP, Burke K, Wang Y (1996) Generalized gradient approximation for the exchange-correlation hole of a many-electron system. Phys Rev B 54:16533; and: Perdew JP , Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865
Phillips JC, Kleinman L (1959) New method for calculating wave functions in crystals and molecules. Phys Rev 116:287
Phillips JC, Lucovsky G (2009) Bonds and bands in semiconductors, 2nd edn. Momentum Press, New York
Pines D (1956) Collective energy losses in solids. Rev Mod Phys 28:184
Pisani C, Dovesi R, Roetti C (1988) Hartree-Fock ab initio treatment of crystalline systems, vol 48, Lecture notes in chemistry. Springer, Heidelberg
Pratt GW Jr (1957) Exchange and superexchange coupling between conduction electrons and *d* electrons in magnetic materials. Phys Rev 106:53
Prokofjew W (1929) Berechnung der Zahlen der Dispersionszentren des Natriums. Z Phys 58:255 (Calculation of the number of dispersion centers of sodium, in German)
Rajagopal AK, Callaway J (1973) Inhomogeneous electron gas. Phys Rev B 7:1912
Rinke P, Qteish A, Neugebauer J, Freysoldt C, Scheffler M (2005) Combining GW calculations with exact-exchange density-functional theory: an analysis of valence-band photoemission for compound semiconductors. New J Phys 7:126
Rose ME (1961) Relativistic electron theory. Wiley, New York
Seitz F (1940) The modern theory of solids. Mc Graw-Hill Book, New York
Shockley W (1938) The empty lattice test of the cellular method in solids. Phys Rev 52:866
Singh DJ, Nordstrom L (2006) Planewaves, pseudopotentials, and the LAPW method, 2nd edn. Springer, Berlin
Skriver HL (1984) The LMTO method- Muffin-tin orbitals and electronic structure, Springer series in solid state science, vol 41. Springer, Berlin
Slater JC (1937) Wave functions in a periodic potential. Phys Rev 51:846
Slater JC (1951) A simplification of the Hartree-Fock method. Phys Rev 81:385
Slater JC (1953a) A generalized self-consistent field method. Phys Rev 91:528
Slater JC (1953b) Ferromagnetism and the band theory. Rev Mod Phys 25:199
Soven P (1965) Relativistic band structure and Fermi surface of Thallium. I. Phys Rev 137:A1706
Sun J, Ruzsinszky A, Perdew JP (2015) Strongly constrained and appropriately normed semilocal density functional. Phys Rev Lett 115:036402
Städele M, Moukara M, Majewski J, Vogl P, Görling A (1999) Exact exchange Kohn-Sham formalism applied to semiconductors. Phys Rev B 59:10031
Vanderbilt D (1990) Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. Phys Rev B 41:7892

Vidal J, Zhang X, Yu L, Luo J-W, Zunger A (2011) False-positive and false-negative assignments of topological insulators in density functional theory and hybrids. Phys Rev B 84:041109
Vogl P, Hjalmarson PH, Dow JD (1983) A semi-empirical tight-binding theory of the electronic structure of semiconductors. J Phys Chem Sol 44:365
von Barth U, Hedin L (1972) A local exchange-correlation potential for the spin polarized case: I. J Phys C 5:1629
Vurgaftman I, Lumb M, Meyer J (2021) Bands and photons in III-V semiconductor quantum structures. Oxford University Press, Oxford UK
Wei S-H, Krakauer H (1985) Local-density-functional calculation of the pressure-induced metallization of BaSe and BaTe. Phys Rev Lett 55:1200
Wigner E, Seitz F (1933) On the constitution of metallic sodium. Phys Rev 43:804
Wigner E, Seitz F (1934) On the constitution of metallic sodium II. Phys Rev 46:509
Wood JH, Pratt GW Jr (1957) Wave functions and energy levels for Fe as found by the unrestricted Hartree-Fock method. Phys Rev 107:995
Yanagisawa S, Morikawa Y, Schindlmayr A (2013) HOMO band dispersion of crystalline rubrene: effects of self-energy corrections within the GW approximation. Phys Rev B 88:115438
Yang Y, Yang Y, Wua F, Wei Z (2008) First-principles electronic structure of copper phthalocyanine (CuPc). Solid State Commun 148:559
Yang Z, Peng H, Sun J, Perdew JP (2016) More realistic band gaps from meta-generalized gradient approximations: Only in a generalized Kohn-Sham scheme. Phys Rev B 93:205205
Yu PY, Cardona M (2005) Fundamentals of semiconductors, physics and materials properties, 3rd edn. Springer, Berlin/Heidelberg/New York
Ziman JM (1964) Principles of the theory of solids. Cambridge University Press, London

Bands and Bandgaps in Solids

8

Karl W. Böer and Udo W. Pohl

Contents

1 Valence and Conduction Bands 258
1.1 Insulators, Semiconductors, and Metals 259
1.2 The Effective Mass in Real Bands 268
2 The Bandgap 284
2.1 Bandgap of Alloys 285
2.2 Bandgap Dependence on Temperature and Pressure 288
2.3 Bandgap at High Doping Level 291
3 Electronic States in Low-Dimensional Semiconductors 294
3.1 Quantum Wells and Superlattices 295
3.2 Dimensionality of the Density of States 301
3.3 Quantum Wires 304
3.4 Quantum Dots and Nanocrystals 305
4 Bands in Organic and Amorphous Semiconductors 308
4.1 Bands and Bandgap in Organic Semiconductors 308
4.2 Bands in Amorphous Semiconductors 310
5 Summary 311
References 312

Abstract

Valence and conduction bands and the bandgap in between these bands determine the electronic properties of solids. For *semiconductors*, the band structure of the conduction band and the valence bands near the edge to the fundamental bandgap is of particular interest. Both the band structure and bandgap are influenced by

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_8

external parameters such as temperature and pressure and can also be changed by alloying and heavy doping.

In *low-dimensional semiconductors* like superlattices and quantum wells, quantum wires, and quantum dots anisotropic carrier confinement occurs. The effective gap and energies in conduction and valence bands can be varied by changing spatial dimensions and barrier height of the low-dimensional structure. The bands in *amorphous semiconductors* near the band edge are ill-defined since long-range periodicity is missing. Still the density-of-state distribution shows significant similarities to that of the same material in the crystallite state.

Keywords

Bandgap energy · Bowing · Conduction bands · Confined states · Cyclotron resonance · Density of states dimensionality · Effective mass · Hole effective mass · HOMO · LUMO · Luttinger parameters · Quantum dots · Quantum wells · Quantum wires · Valence bands

1 Valence and Conduction Bands

This chapter outlines the most important bands in semiconductors, the valence and conduction bands, and their dependence on various material and external parameters.

A set of bands (subbands), created by the splitting and hybridization (see below) of the ground state of valence electrons, taken together are referred to as the *valence band*. The number of states contained in this band is given by the multiplicity of its atomic state (typically 4 for sp^3 hybridization), multiplied by the number of atoms creating this band, e.g., the number of atoms in an entire ideal crystal.[1]

The formation of such bands is often shown as it evolves from the spectrum of isolated atoms when they are brought together to form the crystal lattice. Their levels split with decreasing interatomic distance, as shown in Fig. 1 and discussed in Sect. 1 of ▶ Chap. 6, "The Origin of Band Structure." There are several possibilities for the energy and electron distribution over these levels, depending on the crystal bonding and structure. Three relatively simple examples are given in Fig. 1.

Figure 1a shows the splitting and *overlap* of the *s* and *p* bands of a main-group metal. The atomic states remain nearly unchanged when the atoms approach each other (here *s* and *p* bands do not mix) to form a metal.

Figure 1b shows the splitting for a covalent (tetravalent) crystal, e.g., for a *thatogen*, such as C (diamond) or Si. Near the crossover point, a *hybridization* into sp^3 states occurs, providing four sp^3 states for the lower band and four sp^3 states for the upper band. A *bandgap* appears between these two hybrid bands – see Sect. 2.

[1] In a real crystal this number is reduced since electron scattering limits the coherence length of the electron wave (i.e., the length in which quantum-mechanical interaction can take place). With a mean free path λ the number of atoms responsible for the band-level splitting is of the order of $(\lambda/a)^3$ for a primitive cubic lattice with lattice constant *a*. Each of these levels is broadened by collision broadening; therefore, even for λ approaching *a*, the result is still bands rather than discrete levels.

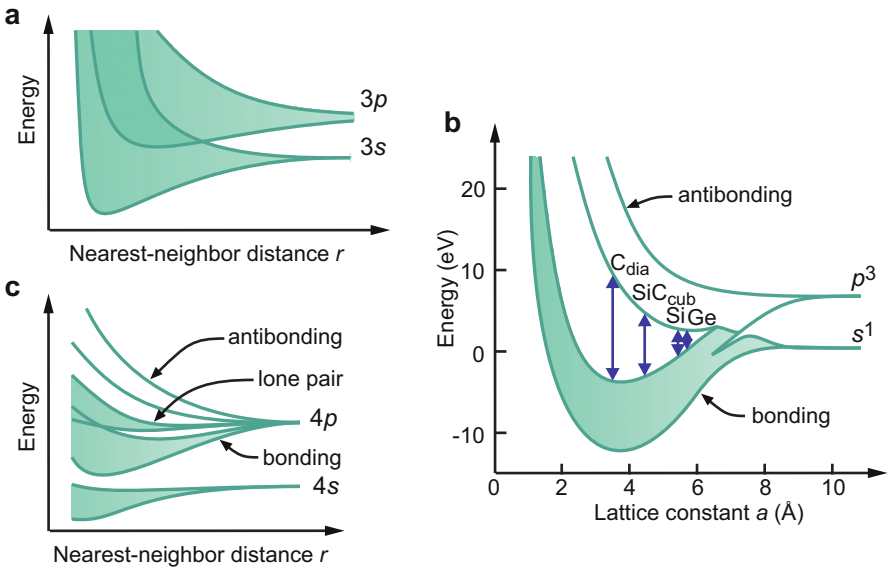


Fig. 1 Electronic energy distribution for a crystal as a function of the interatomic nearest-neighbor distance r (in a and c) or lattice constant a for cubic crystals (in b). (**a**) Overlap of s and p bands in main-group metals. (**b**) sp^3 hybridization in covalent tetravalent crystals; lattice constant a and direct bandgap energy for some crystals are indicated by *blue arrows*. (**c**) Lone-pair semiconductors such as Se

Figure 1c shows the lone-pair configuration typical for higher valency atoms, e.g., a *pnictogen*, such as P or As, or a *chalcogen*,[2] such as Se or Te, where some electrons in the p state do not participate in the bonding states. This is the case when the number of valence electrons is greater than 4, since antibonding states always increase more in energy than bonding states decrease when atoms approach each other (see Fig. 1c). Therefore, a *pair* of electrons when nonbonded (*lone pair*) has lower energy than when split into one bonding state electron and one antibonding state electron; only a total of four electrons can occupy all of the bonding orbitals.

Bands obtained from such splitting, which are completely or partially filled with valence electrons, are called *valence bands*.

1.1 Insulators, Semiconductors, and Metals

1.1.1 Insulators and Semiconductors

In a covalent monatomic crystal, the four valence electrons fill each of the levels in the valence band. In such a totally filled valence band, electronic conduction is impossible, since electrons can only move by an exchange: for every electron moving in one direction, exactly one electron must move in the opposite direction;

[2] *AB* compounds containing these elements are referred to as pnictides or chalcogenides.

there is no free momentum space (*Pauli principle*) for a net electron transport.[3] Diamond is an example of a simple crystal with tetravalent atoms and a totally filled valence band. It is therefore an *insulator*. The band above the valence band has the same number of states but contains no electrons at vanishing excitation. It is an *empty* band. In materials in which this band is relatively close to the valence band (i.e., in materials with a narrow bandgap between these bands, such as germanium with a bandgap of $E_g = 0.64$ eV – see Figs. 1b and 3b), thermal excitation at room temperature will bring a number of electrons into this upper band and partially fill it, although to a very small fraction of the total level density. These electrons can easily gain energy from an external electric field as there are enough free levels available adjacent to each of these electrons; a *net electron transport* (see next section) in the direction of the electric field can take place. Electronic conduction occurs; therefore, this band is referred to as the *conduction band*. Germanium acts as an insulator for vanishing excitation and as a conductor with sufficient thermal excitation: it is called a *semiconductor* and is distinguished from an insulator by a somewhat narrower bandgap.

This description of electronic transport in a partially filled band needs a more precise discussion, which will be given in ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects." However, in order to understand many of the following sections, a heuristic description is supplied below.

Electrons and Holes

When an electron is lifted from the valence band into the conduction band, an empty state is simultaneously created in the valence band. Just as the electron can move in the conduction band, so can the empty state in the valence band. Since it is surrounded by electrons, which can move into the empty state, the empty state moves in the opposite direction of the electron.

The valence-band behavior somewhat resembles the filling of theater seats on the parquet. When all these seats are filled, people still can move but only by exchanging seats. When the balcony (the conduction band) is opened, some people from the front of the parquet may move up to occupy seats in the balcony, leaving their empty seats below. Consequently, people in the parquet may move toward the stage, giving the impression that empty seats move in the opposite direction. However, only people are moving toward the stage, following the force of attraction. Chairs remain fixed to the floor; their *state* of being empty or filled is moving. The situation is similar in the semiconductor: only electrons are moving, not atoms or ions. Ionic conductivity is neglected here. Thus, an empty state deserves its own name: a *hole*. It behaves much like an electron with an effective mass similar in value to the electron mass and not that of an ion.

As shown earlier (Sect. 2.1 of ▶ Chap. 6, "The Origin of Band Structure"), the electron within the conduction band is free to move when considering its mass as an

[3] In a quantum-mechanical picture, any $E(\mathbf{k})$ state represents a certain mass and velocity. In a filled band, all of them add up to zero.

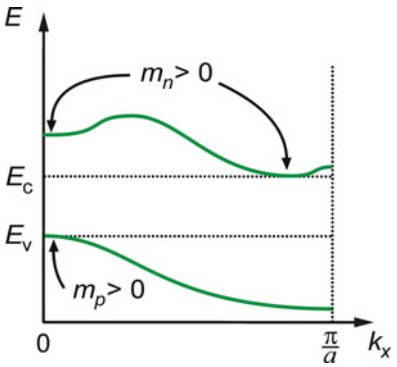


Fig. 2 $E(k)$ for a simple valence and conduction band; signs of the effective masses of holes and electrons are indicated

effective mass. To distinguish it from an electron in vacuum, it is also referred to as a *Bloch electron*. The holes are also free to move. Both contribute to the electrical conductivity and are therefore called *carriers* (of this current).

Electrons tend to occupy the lowest-energy states; that is, they fill a band from the bottom up and occupy states near the bottom of the conduction band designated as E_c. Holes will consequently, like soap bubbles in water, bubble toward its upper surface: they will collect at the upper edge of the valence band designated as E_v.

When an electron is accelerated in the direction of a mechanical force, for instance, by accelerating the semiconductor in the Tolman experiment, the hole is accelerated in the opposite direction. Therefore, its effective mass m_p has the sign opposite to the mass of an electron:

$$m_p = -\frac{\hbar^2}{\left.\frac{\partial^2 E}{\partial k^2}\right|_{E_v}} \quad \text{and} \quad m_n = +\frac{\hbar^2}{\left.\frac{\partial^2 E}{\partial k^2}\right|_{E_c}}; \tag{1}$$

this is illustrated in Fig. 2. At the top of the valence band, $E(k)$ has a *negative curvature* that results in a *negative effective mass for electrons*. However, since the effective mass of the hole has the opposite sign as the electron (Eq. 1), its effective mass is *positive* here, as is the effective mass of an electron at the bottom of the conduction band:

$$m_p(E_v) > 0, \quad m_n(E_c) > 0. \tag{2}$$

Finally, the charge of an electron is negative[4] by convention; therefore, the charge of a hole (i.e., of an atom with a missing electron) is positive. This permits the

[4]A positive charge was arbitrarily related to the charge of a glass rod, rubbed with silk (by Benjamin Franklin); this charge was not caused by an added electron as it became known later, but by a missing electron on the glass rod. This electron was removed by the silk.

Table 1 Typical properties of electrons and holes

	Charge	Near	m^*	$\partial E/\partial k^2$	Mobility	$j_{n,p}(F)$
Electrons	$-e$	E_c	$+m_n$	Positive	$-\mu_n$	Positive
Holes	$+e$	E_v	$+m_p$	Negative	$+\mu_p$	Positive

distinction between a particle flux under the influence of an electric field[5] **F** and the electric current caused by electrons *and* holes. The particle flux proceeds in opposite directions: holes move in the direction of the electric field, and electrons move against it. This is also reflected in the *mobility* of the particles, defined in Sect. 2.2 of ▶ Chap. 22, "Carrier-Transport Equations," by $\mu = (q/m^*) \times \tau$, with particle charge q and a mean time τ (>0) between scattering events. However, both electron and hole *currents* j_n and j_p have the *same* sign and are therefore additive, because $j \propto q\mu F$. Table 1 summarizes the typical properties of electrons and holes in a semiconductor.

1.1.2 Metals

In *monovalent metals*, the valence band is only partially filled; 50% of the *s* states are occupied. Therefore, electronic conduction takes place even with vanishing excitation. *Divalent metals* show allowed eigenvalues in the range of the overlap between *s* and *p* bands[6] – see Figs. 1a and 3a. In this range, therefore, only about 25% of the two *s* and the six *p* states are filled, and thus electronic conduction occurs at vanishing excitation. *Trivalent metals* have about 63% of their states in the overlapping *s* and *p* bands unoccupied. In *transition metals*, inner shell electrons occupy partially filled *d* bands.

All metals are thus distinguished from semiconductors by their substantial (~50%) fraction of free states in the highest partially occupied band with vanishing excitation. This band is therefore both a valence band, containing the valence electrons, and a conduction band where all *metallic conduction* takes place.

This simple model for distinguishing nonmetals from metals by the complete or incomplete filling of bands is due to Wilson (1931). The model is still valid for most cases, except where magnetic properties or strong electron–phonon interactions (polarons; see Sect. 1.2 of ▶ Chap. 22, "Carrier-Transport Equations") interfere.

1.1.3 Semimetals and Narrow-Gap Semiconductors

Metals that show a very small overlap of conduction and valence band exhibit weak electronic conduction. The density of states near the Fermi level (see Sect. 2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers") is very small and only the relatively few electrons there control the conduction. Only these electrons can be scattered. For

[5] In this and all chapters dealing with electrical conductivity, the electric field is identified as *F* and the energy is identified as *E*.

[6] The difference between metals, where the overlap range is allowed, and semiconductors, where the overlap range is forbidden (bandgap), depends on *Wigner's rules* (Wigner 1959), which state that eigenstates belonging to different symmetry groups of the Hamiltonian cannot mix (metals). In semiconductors they do mix, yielding sp^3 hybridization for Si.

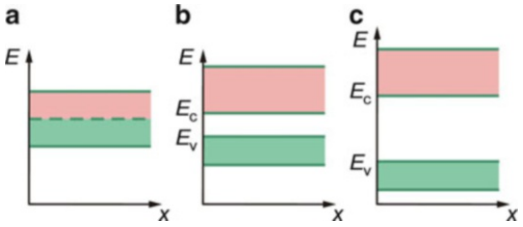


Fig. 3 Empty (*red*) and filled (*green*) states in the upper bands of (**a**) a metal, (**b**) a semiconductor, and (**c**) an insulator. The energies at the top of the valence band and the bottom of the conduction band are identified as E_v and E_c, respectively

Table 2 Band parameters for lead chalcogenides

Material	T (K)	E_g (eV)	$m_{n\parallel}$	$m_{n\perp}$
PbS	290	0.41	0.11	0.11
	77	0.31		
	4	0.29		
PbSe	290	0.27	0.05	0.08
	77	0.17		
	4	0.15		
PbTe	290	0.32	0.025	0.22
	77	0.22		
	4	0.19		

reviews, see Ziman (1969) or Cracknell and Wong (1973). These materials are termed *semimetals*; examples are graphite and bismuth.

On the other hand, semiconductors with very small bandgaps show *relatively* high conductivity compared to other semiconductors. By some alloying, application of pressure, or even at elevated temperature, the bandgap can vanish, turning such semiconductors into metals. Examples include gray tin with a gap of 0–0.08 eV (suggested by magneto-optical experiments (Pidgeon 1969)) and lead chalcogenides and their alloys with other II–VI compounds or with SnTe and mercury chalcogenides. For a review, see Tsidilkovski et al. (1985). The lead chalcogenides are characterized by a small bandgap with a maximum of the valence band and a minimum of the conduction band at the L point rather than the Γ point. The latter is the case for most other direct-gap semiconductors (Sect. 2 of ▶ Chap. 13, "Band-to-Band Transitions") with E_g and m_n as shown in Table 2. In an alloy of PbTe with CdTe ($E_g = 1.5$ eV), the bandgap varies with composition but can dip below zero (to -0.1 eV) because of bowing (see Sect. 2.1) for a few percent Cd (Schmit and Stelzer 1973). Besides optical applications in the infrared range, narrow-gap semiconductors are interesting for thermoelectric applications (Mahan 1997; Graziosi and Neophytou 2020).

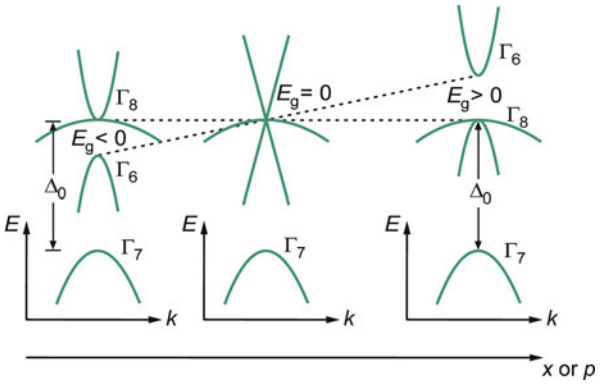


Fig. 4 Conduction and valence bands of $Hg_{1-x}Cd_xTe$ as a function of pressure p or composition x (schematic)

A typical change of the $E(\mathbf{k})$ behavior near the Γ point is shown in Fig. 4 for $Hg_{1-x}Cd_xTe$ as a function of the hydrostatic pressure or the composition (Kane 1979). The bandgap can be expressed by an empirical relation (Hansen et al. 1982)

$$E_g(x,T) = -0.302 + 1.93x + 5.35 \times 10^{-4}(1-2x)T - 0.81x^2 + 0.832x^3(\text{eV}) \quad (3)$$

for $0 < x < 1$ and $4.2 < T < 300$ K.[7] The gap also changes linearly with pressure, with a pressure coefficient of 10^{-2} eV/kbar.

Another interesting property is observed when alloying PbTe and SnTe. The L_6^+ and L_6^- bands are inverted; the L_6^- band is the conduction band in PbTe, while it is the valence band in SnTe, and vice versa for the L_6^+ band. In an alloy of $Pb_{1-x}Sn_xTe$ with $x = 0.62$, both bands touch at 300 K. For $x > 0.62$, the alloy becomes a metal (see Fig. 5). Such materials are also referred to as *gapless semiconductors* when the top of the valence band touches the bottom of the conduction band.

These examples demonstrate the possibility of designing very-narrow-bandgap materials or semimetals by alloying two narrow-bandgap semiconductors with each other (see Fig. 6). The density of states in semimetals at the Fermi level is very small compared to that of a metal. In gapless semiconductors, it is zero (see Fig. 4 at $x = 0.16$ for $T = 0$ K).

A special case of a gapless semiconductor occurs when tailing band states overlap with the conduction band (Fig. 6c). Tailing states are band states which extend into the bandgap due to the disturbed neighborhood of lattice defects as discussed in ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors."

1.1.4 The Shape of Valence and Conduction Bands in Semiconductors

The shape of valence and conduction bands is given by the three-dimensional dispersion relation $E(\mathbf{k})$ and is of interest for semiconducting properties in the energy

[7] For a recent measurement of the conduction- and valence-band dispersion $E(k)$ of a HgTe/HgCdTe layer see Gospodarič et al. (2020).

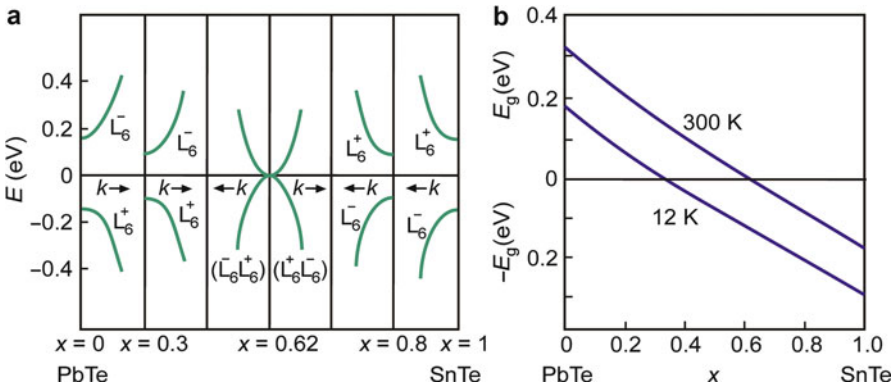


Fig. 5 (**a**) Conduction and valence bands in a schematic $E(k)$ representation of $Pb_{1-x}Sn_xTe$ for different alloy ratios x. For $x = 0.16$ a gapless semiconductor occurs. (**b**) Bandgap as a function of alloy ratio for $T = 12$ and 300 K (After Dimmock et al. 1966)

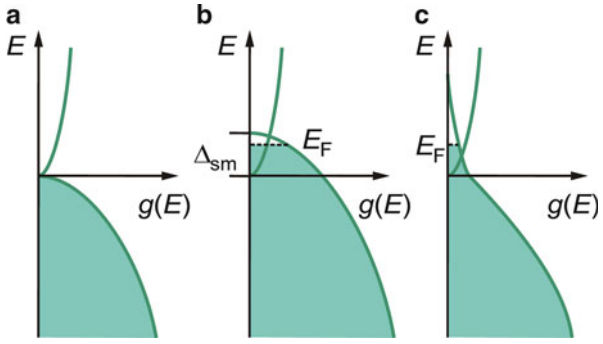


Fig. 6 Density of states $g(E)$ of valence and conduction bands with electron filling indicated for: (**a**) gapless semiconductor, (**b**) semimetal, and (**c**) gapless semiconductor with overlapping tail states (Sect. 2 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors"). E_F denotes the Fermi energy

range near the bottom of the conduction band (near E_c) and near the top of the valence band(s) (near E_v).

The shape of these bands can be visualized from a constant-energy surface near the bottom of the conduction band or the top of the valence band. This surface can be identified by sequentially filling the band[8] as described in Sect. 4.2 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals." Electrons that populate states within $E(\mathbf{k})$ up to a certain energy E_1 [i.e., to the corresponding k_1 (E_1)] are contained in a small sphere in the center of the zone (Fig. 7) that grows parabolically in radius with increasing degree of filling, that is, with an increasing value of k. $k(E)$ is single-valued and monotonic within each band.

When starting from the center, $E(\mathbf{k}_1)$ becomes progressively more deformed with higher energy (higher k). A two-dimensional representation (Fig. 8) demonstrates

[8]Each band contains a large number of energy levels (see Sect. 4.2 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"). Increasing number of electrons first fills the levels at the lowest energy (for $T = 0$) and then successively higher and higher energies. This process is referred to as band filling.

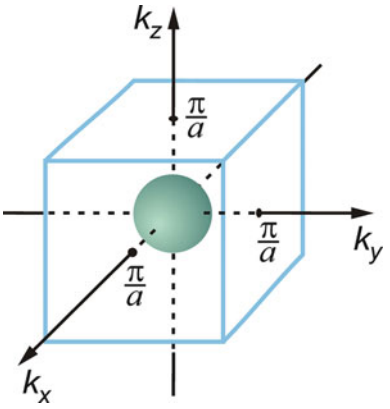


Fig. 7 Equi-energy surface for small electron energies within the first Brillouin zone for a primitive cubic lattice

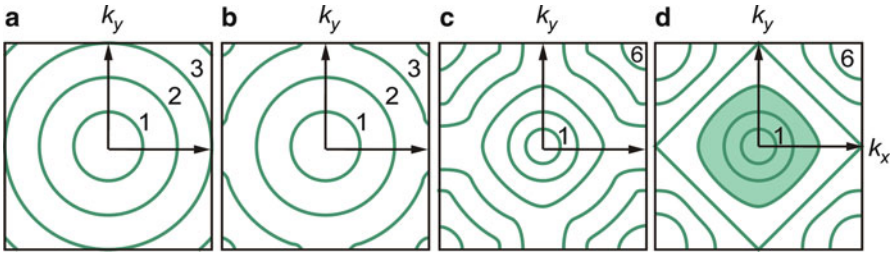


Fig. 8 Two-dimensional cut through a family of equi-energy surfaces for a primitive cubic crystal (corresponding to a cut at E = const in Fig. 7 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"). (**a**–**d**) represent crystals with gradually increasing binding potential. Filling with electrons is indicated by *shading* in (**d**). Spherical surfaces indicate quasi-free behavior of electrons. Deviations from the sphere indicate substantial lattice influence

more clearly some typical shapes of *equi-energy surfaces* in a simple cubic lattice. Figures 8a–d are distinguished by increasing binding forces, showing increasingly more deviation from circles.

A similar development is observed near the top of the valence band: when it becomes filled with holes, the upper part of the Brillouin zone becomes depleted down to a certain energy E_1' with the effect of rounding off the sharp edges of the Brillouin zone.

Depending on the density of electrons within the band, a large fraction of the Brillouin zone is filled. Figure 9 illustrates such a Brillouin zone with a partially filled valence (or conduction) band of a metal. This figure also shows the connection to another neighboring zone. The periodicity of $E(\mathbf{k})$ is indicated at the upper left of Fig. 9. The zones are of the same order, first Brillouin zones. Higher Brillouin zones are not relevant to the discussion of semiconducting properties. In contrast to the one-dimensional $E(k)$ representation of Fig. 13 of ▶ Chap. 6, "The Origin of Band Structure," where several curves in the first Brillouin Chap. 6, are shown, in the **k** space representation only one such equivalent surface can be drawn without confusion. It is the valence band in Figs. 8d, 9, and 10c and the conduction band in Figs. 7 and 10a, b.

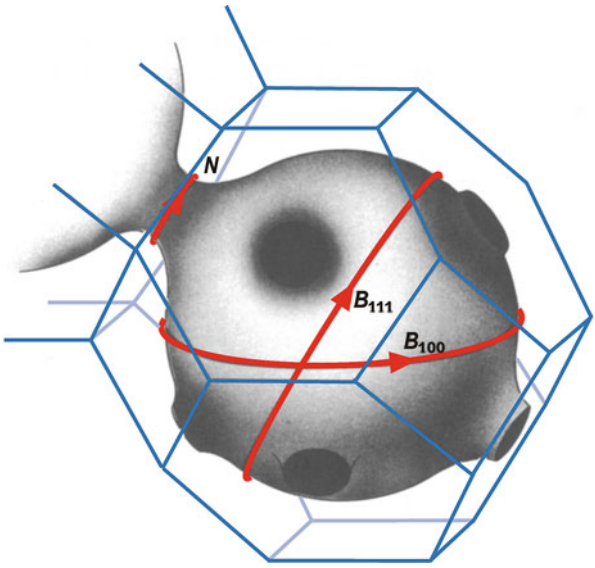


Fig. 9 Equi-energy surface at the Fermi level (Fermi surface) of a face-centered cubic metal, with contacts to the adjacent Fermi surfaces through the center of the hexagonal surfaces of the Brillouin surfaces. Three extremal orbits for electrons along the Fermi surface are identified by B_{111}, B_{100}, and N (After Pippard 1965)

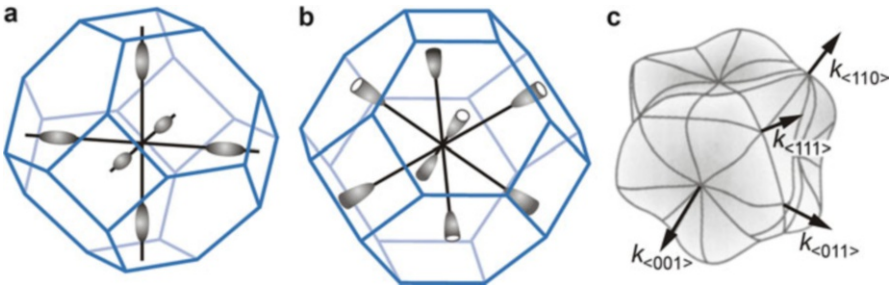


Fig. 10 Brillouin zone of (**a**) Si and (**b**) Ge with equi-energy surfaces near the conduction-band edge and (**c**) equi-energy surface of the valence band for heavy holes in Si or Ge near the upper band edge

In semiconductors, the equi-energy surfaces close to the top of the valence band (nearly full Brillouin zone) and at the bottom of the conduction band (nearly empty Brillouin zone) are of interest for the electron transport. They are shown in Fig. 10 for Ge and Si.

Figure 10a shows for Si six small ellipsoids in the ⟨100⟩ directions which are centered at $\pm 0.8k_x$, $\pm 0.8k_y$, and $\pm 0.8k_z$. Figure 10b indicates for Ge in the ⟨111⟩ directions eight ellipsoids which are centered at the L points (see Fig. 8 of ▶ Chap. 3, "The Structure of Semiconductors"). There are four such L minima per Brillouin zone. Other semiconductors having their lowest minima outside the Γ point are GaP and AlSb. Most other semiconductors have their lowest conduction-band minimum at (000), at the Γ point.

Figure 10c shows the warped equi-energy surface near the upper edge of the valence band for heavy holes (see Sect. 1.2) in Si and Ge. This equi-energy surface fills almost the entire Brillouin zone. Its shape is shown without the surrounding

Brillouin zone in nearly the same scale as in Fig. 10a, b. The filling of the valence band with holes is greatly exaggerated; otherwise, the Brillouin zone with a flat outer surface and only slightly rounded edges would show.

The shape of these equi-energy surfaces is intimately related to the three-dimensional shape of the bands ($E(\mathbf{k})$) – see Sect. 4.2 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals." The anisotropy is evident from Fig. 10. This anisotropic $E_n(\mathbf{k})$ behavior has an influence on the effective mass and therefore on carrier transport. We will first analyze the anisotropy of the effective mass in more detail.

1.2 The Effective Mass in Real Bands

In Sect. 2.2 of ▶ Chap. 6, "The Origin of Band Structure," the effective mass is defined as proportional to the inverse of the second derivative of $E(k)$, i.e., of the curvature of $E(k)$. The anisotropy of $E(k)$ in an actual crystal makes the introduction of a tensor relation for the effective mass necessary:

$$m_n \rightarrow \underline{\mathbf{m}}^* \quad \text{with components} \quad m^*_{ik} = \frac{\hbar^2}{\dfrac{\partial^2 E}{\partial k_i \partial k_j}}. \tag{4}$$

This tensor ellipsoid, when transformed to main axes, contains usually only very few nonvanishing components. We will identify these below for some of the more important semiconductors.

Figure 11 shows the important minima of typical conduction bands and maxima of the valence bands. In most semiconductors, the relevant conduction-band minimum lies at the Γ point (direct gap) or near the L or X point (indirect gap as discussed in Sect. 2 of ▶ Chap. 13, "Band-to-Band Transitions"). The energies of these minima are listed in Table 3. The relevant maxima of the valence band are at the Γ point. Their energies are listed in Table 4.

We will first discuss the shape of the conduction and valence bands near the band edge and thereby provide simple expressions for the respective effective masses.

1.2.1 The Conduction Bands

The relationship of $E(\mathbf{k})$ in the vicinity of the conduction-band minima near the band edge can be described as conduction band, at Γ:

$$E_c(\mathbf{k}) = \frac{\hbar^2}{2m_n} k^2, \tag{5}$$

and at a satellite conduction band:

$$E_c(\mathbf{k}) = \frac{\hbar^2}{2}\left(\frac{(k_x - k_{0x})^2}{m_\parallel} + \frac{(k_y - k_{0y})^2 + (k_z - k_{0z})^2}{m_\perp}\right), \tag{6}$$

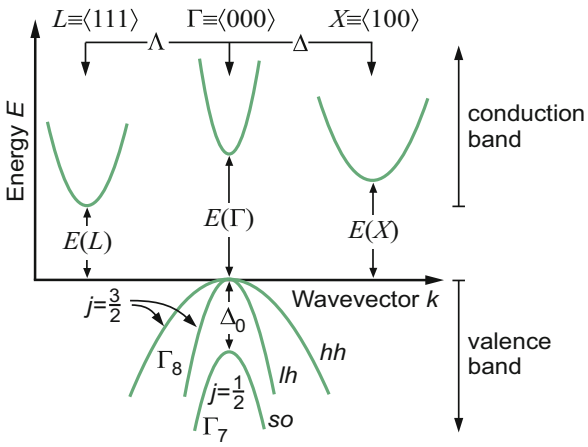


Fig. 11 Simplified band structure of typical semiconductors near the most important extrema of valence and conduction bands

relating $E(\mathbf{k})$ directly to the corresponding effective masses. Satellite minima of $E(\mathbf{k})$ at $\mathbf{k}_0 \neq 0$ show ellipticity, such as those of the indirect gap semiconductors Ge, Si, and GaP; so do the higher satellite minima of others III–V compounds, such as GaAs. For symmetry reasons, these ellipsoids are ellipsoids of revolution about the main axes; hence, we distinguish only two effective masses, parallel and orthogonal to these axes, $m_\parallel$ and $m_\perp$, respectively (Eq. 6). The main axes of these ellipsoids of revolution lie in ⟨100⟩ direction for Si and in ⟨111⟩ direction for Ge. Another example is the camel's back conduction band of some III–V compound semiconductors (see Fig. 12); it can be described by

$$E_c(\mathbf{k}) = \frac{\hbar^2 k_\parallel^2}{2m_\parallel} + \frac{\hbar^2 k_\perp^2}{2m_\perp} - \sqrt{\left(\frac{\Delta^c}{2}\right)^2 + \Delta_0^c \frac{\hbar^2 k_\parallel^2}{2m_\parallel}}, \tag{7}$$

where $k_\parallel$, $m_\parallel$, $k_\perp$, and $m_\perp$ are the components parallel and perpendicular to [100] and Δ_0^c describes the nonparabolicity; ΔE and Δ^c are identified in Fig. 12.

The energies of the conduction-band minima, as measured from the valence-band maximum, and the corresponding effective masses are listed in Table 3.

1.2.2 The Valence Band of Cubic Semiconductors

The valence band develops from the twofold degenerate s states and the sixfold degenerate p states of an atom forming sp^3 hybrids. Without spin–orbit interaction it breaks into two bands at $\mathbf{k} = 0$ (Γ_1 and $\Gamma_{25'}$, as shown in Fig. 9c of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"). With spin–orbit interaction, the upper edge $\Gamma_{25'}$ breaks into two sets of bands: one set with a total momentum of 3/2 is fourfold degenerate at $\mathbf{k} = 0$; this degeneracy can be removed in anisotropic crystals with crystal-field coupling, for example, in hexagonal GaN.

Table 3 Electron effective masses in units of m_0, conduction bands at 0 K in eV with characteristic points and band numbers identified, and spin–orbit splitting Δ_0 and camel's back parameters Δ^c and Δ_0^c in eV. D, W, and ZB in column "Structure" denote diamond, wurtzite, and zincblende, respectively. The two values of m_n for wurtzite crystals refer to data parallel and perpendicular to the c axis. Lowest gap values defining the bandgap are printed bold

Crystal	Struct.	1. Indir. gap	$m_{n\|\|}$	$m_{n\perp}$	2. Dir. gap	m_n	3. Gap	$m_{n\|\|}$	$m_{n\perp}$	Δ^c	Δ_0^c or ΔE
Si	D	(X_1) **1.16**	0.92	0.19	(Γ_2) 4.19	0.188					
Ge	D	(L_6) **0.76**	1.57	0.081	(Γ_7) 0.898	0.038					
AlN	W				**(Γ) 6.25**	0.32 0.30					
AlP	ZB	(X_6) **2.52**	2.68	0.155	(Γ_6) 3.63	0.22	(L_6) 3.57				
AlAs	ZB	(X_6) **2.24**	0.97	0.22	(Γ_6) 3.099	0.15	(L_6) 2.46	1.32	0.15		
GaN	W				**(Γ) 3.51**	0.20 0.20					
GaP	ZB	(X_6) **2.35**	0.91	0.25	(Γ_1) 2.72	0.13	(L_1) 2.72	1.2	0.15	(Δ) 0.355	(Δ) 0.433 Δ_0^c
GaAs	ZB	(L_6) 1.815	1.9	0.075	(Γ_6) **1.519**	0.067	(X_6) 1.981	1.3	0.23	(X) 0.304	(X) 0.009 ΔE
GaSb	ZB	(L_6) 0.875	1.3	0.10	(Γ_6) **0.812**	0.039	(X_6) 1.141	1.51	0.22	(X) 0.178	(X) 0.025 ΔE
InN	W				**(Γ) 0.78**	0.07 0.07					
InP	ZB	(L_6) 2.014			(Γ_6) **1.424**	0.080	(X_6) 2.384				
InAs	ZB	(L_6) 1.133	0.64	0.05	(Γ_6) **0.417**	0.026	(X_6) 1.433	1.13	0.16		
ZnSe	ZB	(L_6) 3.96			(Γ_6) **2.82**	0.16	(X_6) 4.57				
CdS	W				**(Γ) 2.579**	0.15 0.17					
CdTe	ZB	(L_6) 2.82			(Γ_6) **1.475**	0.096	(X_6) 3.48				

The other set is twofold degenerate with $j = 1/2$ and shifted by the spin–orbit splitting energy Δ_0 (Figs. 11b of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals," and 11); it is referred to as the *spin–orbit split-off band* "so" with an effective mass m_{so}. For $\mathbf{k} \neq 0$ the degeneracy of the upper band is removed. Here, the band splits into two bands, each of which is twofold degenerate. The one with lower curvature is the *heavy-hole band* "hh" with an effective mass m_{hh}; the other with larger curvature is the *light-hole band* "lh" with an effective mass m_{lh}. The valence-

Table 4 Upper valence bands (except Γ_6) in eV for some semiconductors

Crystal	Γ-related maxima			*L*-related	*X*-related
Si	(Γ_1) –12.5	$\Gamma_{25} = (\Gamma_8)$ 0	(Γ_7) –0.044	(L_3) –2.82	(X_4) –6.27
Ge	(Γ_6) –12.7	(Γ_8) 0	(Γ_7) –0.29	$(L_{6,5})$ –1,43	(X_5) –3.29
GaP	(Γ_1) –13.0	$\Gamma_{15} = (\Gamma_8)$ 0	(Γ_7) –0.08	(L_3) –1.1	(X_5) –2.7
GaAs	(Γ_7) –12.6	(Γ_8) 0	(Γ_7) –0.35	$(L_{4,5})$ –1.2	(X_7) –2.87
GaSb	(Γ_6) –12.0	(Γ_8) 0	(Γ_7) –0.76	$(L_{4,5})$ –1.1	(X_7) –2.37
InP	(Γ_6) –11.4	(Γ_8) 0	(Γ_7) –0.21	$(L_{4,5})$ –0.94	(X_7) –2.06
InAs	(Γ_6) –12.7	(Γ_8) 0	(Γ_7) –0.43	$(L_{4,5})$ –0.9	(X_7) –2.37
ZnSe	(Γ_6) –12.5	$\Gamma_{15} = (\Gamma_8)$ 0	(Γ_7) –0.45	$(L_{4,5})$ –0.76	(X_7) –1.96
CdTe	(Γ_6) –11.1	(Γ_8) 0	(Γ_7) –0.89	$(L_{4,5})$ –0.65	(X_7) –1.60

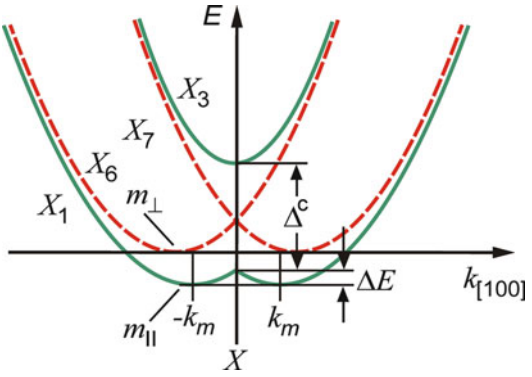


Fig. 12 Camel's back conduction band in GaP along the Δ axes near the Brillouin zone boundary at *X*. *Dashed* and *solid curves* refer to diamond and zincblende structures, respectively

band energies at a number of symmetry points are given in Table 4 for some typical semiconductors. Near the Γ point ($\mathbf{k} = 0$) the Hamiltonian describing the three sets of upper valence bands can be written as (D'yakonov and Perel' 1971)

$$\mathbf{H} = -(A + 2B)k^2 + 3B\left(\mathbf{k} \cdot \widehat{\mathbf{L}}\right)^2 - \frac{1}{3}\Delta_0\left(\underline{\sigma} \cdot \widehat{\mathbf{L}}\right) + \frac{1}{3}\Delta_0, \tag{8}$$

with

$$A = -\frac{\hbar^2}{4}\left(\frac{1}{m_{\mathrm{lh}}} + \frac{1}{m_{\mathrm{hh}}}\right), \quad B = \frac{\hbar^2}{4}\left(\frac{1}{m_{\mathrm{lh}}} - \frac{1}{m_{\mathrm{hh}}}\right), \tag{9}$$

where $\widehat{\mathbf{L}}$ is the angular-momentum matrix, $\underline{\sigma}$ is the Pauli operator spin matrix, and Δ_0 is the spin–orbit splitting.

Only the light- and heavy-hole bands need to be considered when Δ_0 is larger than certain energies, e.g., kT or the ionization energy of shallow acceptors (see Sect. 1 of ▶ Chap. 18, "Shallow-Level Centers"). For these two bands, the Hamiltonian reduces considerably. Luttinger (1956) gives it in a form that does not require spherical symmetry

$$\mathbf{H} = \frac{\hbar^2}{2m_0}\left[\left(\gamma_1 + \frac{5}{2}\gamma_2\right)\mathbf{k}^2 - 2\gamma_3(\mathbf{k}\cdot\mathbf{J})^2 + 2(\gamma_3 - \gamma_2)\left(k_x^2 J_x^2 + k_y^2 J_y^2 + k_z^2 J_z^2\right)\right], \quad (10)$$

where **J** is a pseudovector representing the spin–momentum operator (Bir and Pikus 1974) and γ_i are the *Luttinger parameters*. Along the principal directions of a cubic semiconductor, the eigenvalues of this Hamiltonian yield two respective parabolic bands

$$E_{\mathrm{lh}} = \frac{\hbar^2 k^2}{2m^*_{\mathrm{lh}}} \text{ and } E_{\mathrm{hh}} = \frac{\hbar^2 k^2}{2m^*_{\mathrm{hh}}} \quad (11)$$

for light and heavy holes with effective masses

$$\begin{aligned} m^*_{\mathrm{lh},[001]} &= \frac{m_0}{\gamma_1 + 2\gamma_2} \text{ and } m^*_{\mathrm{hh},[001]} = \frac{m_0}{\gamma_1 - 2\gamma_2}, \\ m^*_{\mathrm{lh},[110]} &= \frac{m_0}{2\gamma_1 + \gamma_2 + 3\gamma_3} \text{ and } m^*_{\mathrm{hh},[110]} = \frac{m_0}{2\gamma_1 - \gamma_2 - 3\gamma_3}, \\ m^*_{\mathrm{lh},[111]} &= \frac{m_0}{\gamma_1 + 2\gamma_3} \text{and } m^*_{\mathrm{hh},[111]} = \frac{m_0}{\gamma_1 - 2\gamma_3}. \end{aligned} \quad (12)$$

With γ_3 substantially different from γ_2, the two sets of valence bands become more warped and are determined from the dispersion equation:

$$E_{\mathrm{lh,hh}} = \frac{\hbar^2}{2m_0}\left[\gamma_1 k^2 \pm \sqrt{4\gamma_2^2 k^2 + 12\left(\gamma_3^2 - \gamma_2^2\right)\left(k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2\right)}\right] \quad (13)$$

with the + and − signs for light- and heavy-hole bands, respectively. The degree of warping can be judged from the tabulated values of γ_i (Table 5). For example, it is small for Ge and GaAs ($\gamma_2 \approx \gamma_3$) and much larger for Si and GaP (see also Fig. 13).

Often an equivalent description of these two bands is given[9]:

$$E_{\pm} = Ak^2 \pm \sqrt{\left(Bk^2\right)^2 + C\left(k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2\right)}, \quad (15)$$

with the + and − signs for the light- and heavy-hole bands, respectively. The band parameters can be interpreted as A giving the average curvature, B giving the splitting between heavy- and light-hole bands, and C describing the warping. These parameters are related to the *Luttinger band parameters* by

[9]There are also Dresselhaus parameters L, M, and N (Dresselhaus et al. 1955) to describe the valence band. They are related to the Luttinger parameters by

$$\gamma_1 = -2m_0(L + 2M)/3, \quad \gamma_2 = -2m_0(L - M)/6, \quad \gamma_3 = -2m_0 N/6. \quad (14)$$

Table 5 Hole effective masses in units of m_0,* valence-band splitting Δ_0 in eV, and Luttinger valence-band parameters of some cubic semiconductors

Crystal	γ_1	γ_2	γ_3	Δ_0	$m_{hh}[100]$	$m_{hh}[110]$	$m_{hh}[111]$	$m_{lh}[100]$	m_{so}
Si	4.29	0.34	1.45	0.043	0.28		0.72	0.202	0.29
Ge	13.38	4.24	5.69	0.295	0.20		0.50	0.046	0.10
AlP	3.35	0.71	1.23	0.07	0.52	0.87	1.12	0.205	0.28
AlAs	3.76	0.82	1.42	0.28	0.47	0.82	1.09	0.180	0.28
GaP	4.05	0.49	1.25	0.08	0.33	0.52	0.65	0.189	0.25
GaAs	6.98	2.06	2.93	0.341	0.35	0.64	0.89	0.084	0.17
GaSb	13.4	4.7	6.0	0.76	0.25	0.49	0.71	0.04	0.12
InP	5.08	1.6	2.1	0.108	0.53	0.89	1.14	0.114	0.14
InAs	19.7	8.4	9.3	0.367	0.35	0.65	0.91	0.024	0.09
ZnSe	3.94	1.00	1.52	0.424	0.52		1.11	0.168	0.24
CdTe	4.14	1.09	1.62	0.90	0.51		1.11	0.158	0.24

*For the more isotropic light hole only data for the [100] direction is given

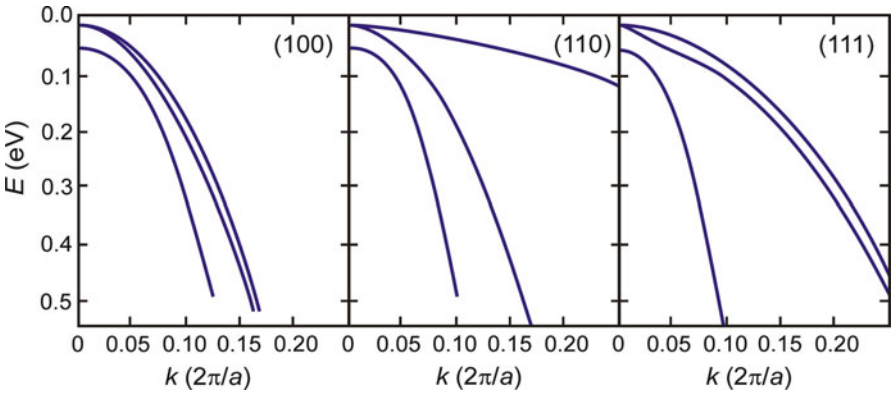


Fig. 13 $E(k)$ diagram for the valence bands of Si near the Γ point in three symmetry directions

$$\frac{\gamma_1}{m_0} = -\frac{2}{\hbar^2}A, \quad \frac{\gamma_2}{m_0} = -\frac{1}{\hbar^2}B, \quad \frac{\gamma_3}{m_0} = -\frac{1}{\sqrt{3}\hbar^2}\sqrt{C^2 + 3B^2}. \tag{16}$$

The different shapes of the surfaces of constant energy for the different conduction and valence bands are summarized in Fig. 16. An impression about the degree of warping for Si can be obtained from Fig. 13, which shows cuts through the $E(\mathbf{k})$ surface in different crystallographic directions (Pantelides 1978).

1.2.3 The Valence Band of Wurtzite Semiconductors

The valence band of wurtzite crystals and related hexagonal polytypes develops from sp^3 hybrids similar to the cubic case. However, the behavior of the bands along [0001], i.e., along the c axis, is distinctly different from that perpendicular. The valence-band states transform p-like at the center of the Brillouin zone as X, Y, and Z representations

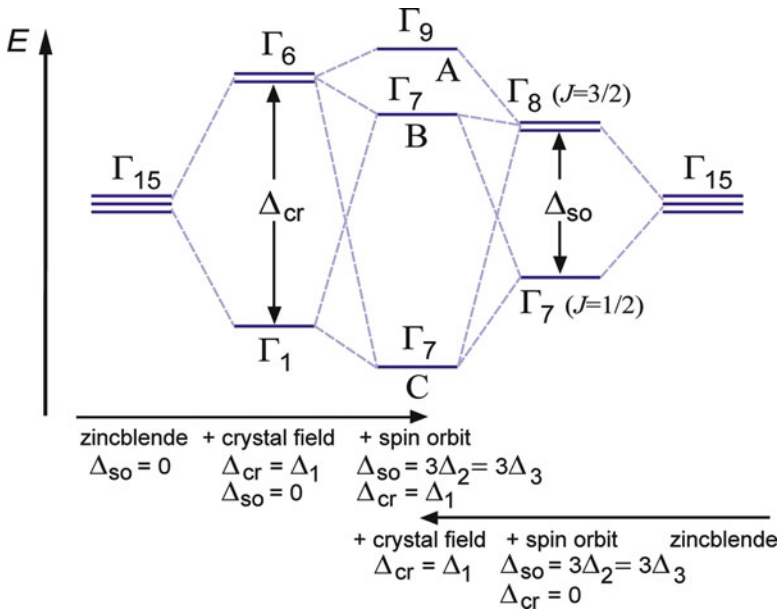


Fig. 14 Splitting of the top of the valence band at $\mathbf{k} = 0$ for wurtzite crystals by the effect of the crystal-field and the spin–orbit interaction. The three resulting heavy-hole, light-hole, and crystal-hole bands are referred to as *A*, *B*, and *C*, respectively

of the wurtzite C_{6v}^4 space-group symmetry. The hexagonal symmetry causes a crystal-field splitting Δ_{cr}. The additional effect of the spin–orbit interaction Δ_{so} creates a splitting of the valence-band maximum at $\mathbf{k} = 0$ into three valence bands labeled *A*, *B*, and *C* as illustrated in Fig. 14. The spin–orbit splitting Δ_{so} is expressed in terms of two parameters Δ_2 and Δ_3, which commonly are assumed to be equal, yielding $\Delta_2 = \Delta_3 = \Delta_{so}/3$. The crystal-field splitting $\Delta_{cr} = \Delta_1$ is generally not related to Δ_{so}. The energy splitting of the Δ_{cr} and Δ_{so} parameters is given by

$$\langle X|H_{cr}|X\rangle = \langle Y|H_{cr}|Y\rangle = E_v + \Delta_1, \tag{17a}$$

$$\langle Z|H_{cr}|Z\rangle = E_v, \tag{17b}$$

$$\langle X|H_{so,z}|Y\rangle = -i\,\Delta_2, \tag{17c}$$

$$\langle Y|H_{so,x}|Z\rangle = \langle Z|H_{so,y}|X\rangle = -i\,\Delta_3. \tag{17d}$$

This yields at the Γ point the energies

$$\begin{aligned} E_1 &= \Delta_1 + \Delta_2 \\ E_{2,3} &= \tfrac{1}{2}\left(\Delta_1 - \Delta_2 \pm \sqrt{(\Delta_1 - \Delta_2)^2 + 8\Delta_3^2}\right), \end{aligned} \tag{18}$$

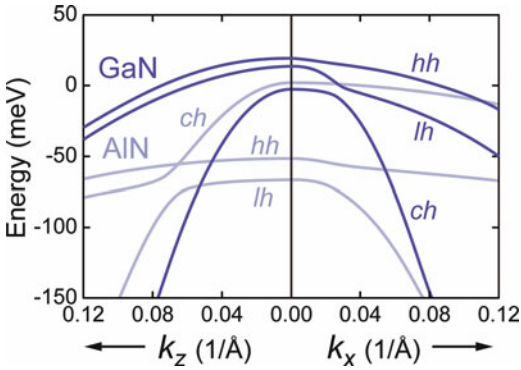


Fig. 15 Valence-band structure of GaN and AlN near the center of the Brillouin zone. The k_z axis points along the [0001] direction and k_x axis perpendicular along an in-plane direction (After Chuang and Chang 1996)

with E_1 corresponding to the Γ_9 level in Fig. 14 and $E_{2,3}$ to the two Γ_7 levels. The order of the levels depends on the values of the parameters, which are affected by the kind of material, and the hexagonal c/a and u lattice parameters. In GaN the splitting by Δ_{cr} leads to an upper Γ_6 level and a lower Γ_1 level, while the order is reversed for AlN as shown in Fig. 15.

Unlike in zincblende crystals, the valence band of wurtzite crystals deviates significantly from parabolic behavior. The valence-band structure is commonly expressed in terms of seven (sometimes only 6) parameters A_i and the three splitting parameters Δ_i mentioned above, using the Hamiltonian reported from Bir and Pikus (1974):

$$\begin{aligned} H = {} & \Delta_1 L_z^2 + \Delta_2 L_z \sigma_z + \sqrt{2}\Delta_3 (L_+ \sigma_- + L_- \sigma_+) \\ & + (A_1 + A_3 L_z^2) k_z^2 \; + \; (A_2 + A_4 L_x^2)\left(k_x^2 + k_y^2\right) \\ & - A_5 \left(L_+^2 k_-^2 + L_-^2 k_+^2\right) \; - \; 2iA_6 k_z ([L_z L_+] k_- - \; [L_z L_-] k_+) \\ & + A_7 (L_+ k_- + L_- k_+) \end{aligned} \tag{19}$$

with the abbreviations given in atomic units (i.e., $\hbar = 1$, $m_0 = ½$, $e^2 = 2$)

$$k_\pm = k_x \pm i k_y, \;\; L_\pm = \pm\left(i/\sqrt{2}\right)(L_x \pm i L_y), \;\; \sigma_\pm = \pm(i/2)(\sigma_x \pm i\sigma_y),$$

and $2[L_z L_\pm] = L_z L_\pm + L_\pm L_z$. The L and σ symbols represent the angular-momentum matrices and Pauli spin-matrices, and the A_i and Δ_i are fitting parameters listed in Table 6.

The parameters A_i of wurtzite crystals correspond to the Luttinger parameters γ_i of zincblende crystals. The six hole masses are accordingly expressed by (Suzuki et al. 1995):

Table 6 Parameters for the valence band of some hexagonal semiconductors; SiC(2H) is the wurtzite modification. Quantities of A_1 to A_6 are dimensionless

Crystal	Δ_{cr} (meV)	Δ_{so} (meV)	A_1	A_2	A_3	A_4	A_5	A_6	A_7 (meV Å)	References
GaN	10	17	−7.21	−0.44	6.68	−3.46	−3.40	−4.90	93.7	Vurgaftman and Meyer (2003)
AlN	−169	19	−3.86	−0.25	3.58	−1.32	−1.47	−1.64	0	Vurgaftman and Meyer (2003)
InN	40	5	−8.21	−0.68	7.57	−5.23	−5.11	−5.96	0	Vurgaftman and Meyer (2003)
ZnO	38	−13.6 (‖) −9.2 (⊥)	−3.78	−0.44	3.45	−1.63	1.68	−2.23	25	Lambrecht et al. (2002)
CdS	27	65	−4.53	−0.39	4.02	−1.92	1.92	−2.59		Jeon et al. (1996)
CdSe	39	416	−5.06	−0.43	4.50	−1.29	1.29	−0.47		Jeon et al. (1996)
SiC (2H)	144	−4.83	−4.83	−0.576	4.18	−1.24	−1.41	−1.49	23.2	Wellenhofer and Rössler (1997)
SiC (4H)	74	−4.76	−4.76	−0.628	4.14	−1.11	−1.45	−1.49	12.7	Wellenhofer and Rössler (1997)
SiC (6H)	44	−4.76	−4.76	−0.590	4.14	−1.15	−1.50	−1.34	11.9	Wellenhofer and Rössler (1997)

$$\begin{aligned} m_0/m_{\mathrm{hh}}^{\parallel} &= -(A_1 + A_3), & m_0/m_{\mathrm{hh}}^{\perp} &= -(A_2 + A_4 - A_5) \\ m_0/m_{\mathrm{lh}}^{\parallel} &= -(A_1 + A_3), & m_0/m_{\mathrm{lh}}^{\perp} &= -(A_2 + A_4 - A_5) - 2A_7^2/|\Delta_1| \\ m_0/m_{\mathrm{ch}}^{\parallel} &= -A_1, & m_0/m_{\mathrm{ch}}^{\perp} &= -A_2 + 2A_7^2/|\Delta_1|, \end{aligned} \tag{20}$$

where $\parallel$ and $\perp$ refer to the orientation with respect to the c axis and the indices hh, lh, and ch refer to light, heavy, and crystal hole. We note that the inversion parameter A_7 breaks the spin degeneracies of light hole and crystal hole in the in-plane direction denoted by the symbol $\perp$.

The A parameters can be related to the Luttinger parameters in the *cubic approximation*: the z axis of the cubic zincblende system is aligned along [111] and the corresponding x and y axes are aligned along the $[11\bar{2}]$ and $[\bar{1}10]$ directions, respectively; the hexagonal crystal field is assumed zero, i.e., $\Delta_{\mathrm{cr}} = \Delta_1 = 0$ and $\Delta_2 = \Delta_3 = \Delta_{\mathrm{so}}/3$. This leads to the dependences $A_1 = A_2 + 2A_4$, $A_3 = -2A_4 = \sqrt{2}A_6 - 4A_5$, $A_7 = 0$. For these conditions the A parameters read:

$$\begin{aligned} A_1 &= -(\gamma_1 + 4\gamma_3) \\ A_2 &= -(\gamma_1 - 2\gamma_3) \\ A_3 &= 6\gamma_3 \\ A_4 &= -2\gamma_3 \\ A_5 &= -(\gamma_2 + 2\gamma_3) \\ A_6 &= -\sqrt{2}(2\gamma_2 + \gamma_3). \end{aligned} \tag{21}$$

For GaN and AlN the cubic approximation was reported to yield a good description of the valence bands at $\mathbf{k} = 0$ when compared to the first-principles LAPW calculations of the wurtzite structure (Suzuki et al. 1995).

1.2.4 Probing Bands with Cyclotron Resonance

By applying a sufficiently strong magnetic field, the band shape can be probed by forcing the electrons into circles perpendicular to the direction of the magnetic induction (for vanishing electric field), following the Lorentz force $\mathcal{F}_{\mathrm{L}}$[10]:

$$\mathcal{F}_{\mathrm{L}} = \hbar \frac{d\mathbf{k}}{dt} = e\,(\mathbf{F} + \mathbf{v} \times \mathbf{B}). \tag{22}$$

These circling electrons are accelerated by the magnetic induction with a centripetal acceleration

$$\mathbf{a} = \frac{\hbar}{m^*}\frac{d\mathbf{k}}{dt} = \frac{e}{m^*}(\mathbf{v} \times \mathbf{B}) = \frac{e}{\hbar m^*}\left(\frac{\partial E(\mathbf{k})}{\partial \mathbf{k}} \times \mathbf{B}\right); \tag{23}$$

this acceleration is a measure of the effective mass, m^*, hence of the band shape at

[10]Note that **F** represents the electric field and E an energy.

the specific electron energy (see below). The accelerated electrons can interact with a high-frequency, small-amplitude probing electromagnetic field and show a resonance absorption – the *cyclotron resonance* (in the classical limit)[11] – when the *ac* frequency of the probing field coincides with the *cyclotron frequency* of the circling electrons:

$$\omega_c = \frac{eB_0}{m^*} = 178.4\frac{m_0}{m^*}B_0 \quad (\mathrm{GHz/T}), \tag{24}$$

where B_0 is the stationary magnetic induction. This resonance absorption is quite distinct when the electrons are permitted to complete many circles[12] before being interrupted by scattering (see ▶ Chap. 23, "Carrier Scattering at Low Electric Fields"). Since the path of an electron to complete a cycle is smaller with the decreasing radius of the circle and scattering is reduced with the decreasing density of defects and phonons (see ▶ Chap. 23, "Carrier Scattering at Low Electric Fields"), cyclotron-resonance measurements are usually performed at high magnetic fields in materials of high purity and at low temperature. Excessively high magnetic fields, however, cause inconveniently large high-resonance frequencies (for B_0 of 1 Tesla, one obtains $\nu = \omega_c/2\pi \cong 100$ GHz). It should be noted that in *low-dimensional* systems with a resonator structure the frequency of *ac* absorption may substantially be redshifted, see Zabolotnykh and Volkov (2021).

Although electrons along all cross sections of the Fermi surface perpendicular to the direction of the magnetic field cause resonance absorption, absorption *maxima* are observed at the extrema of the cross section (belly or neck, see Fig. 9) because of the higher electron density here. In addition, one observes the so-called dog-bone cross sections, in which electrons circle between four adjacent Fermi surfaces near the neck of each surface, as well as many other cross-section shapes depending on the direction of the magnetic field and the shape of the Fermi surfaces.

[11]Here cyclotron resonance is discussed within the same band, and quantum effects are neglected. This can be justified when, neglecting scattering, each electron describes full circles which have to be integers of its de Broglie wavelength Eq. 6 of ▶ Chap. 6, "The Origin of Band Structure." This integer represents the quantum number n_q of the circle, and for the magnetic induction discussed here, it is a large number. Resonance means absorption (or emission) of one quantum $\hbar\omega_c$, hence changing n_q by $\Delta n_q = \pm 1$, which is the selection rule for cyclotron transitions. Since $\Delta n_q \ll n_q$, a change in circle diameter is negligible; hence, the classical approach is justified. At higher fields the circles become smaller, and when approaching atomic size, the quantum levels (Landau levels – see Sect. 2 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients") become wider spaced and a quantum-mechanical approach is required. For reviews, see Lax (1963), Mavroides (1972), and McCombe and Wagner (1975).

[12]The circle diameter ($\pi v_n/\omega_c$) is typically on the order of 10^{-3} cm for a magnetic induction of 1 T; here v_n is the thermal velocity of an electron. In metals, however, one also has to consider the skin penetration of the probing electromagnetic field. The skin depth of a metal is usually a very small fraction of the circle diameter, so that the probing *ac* field can interact only at the very top part of each electron cycle close to the surface. This enhances information about near-surface behavior in metals, while in semiconductors, probing extends throughout the bulk.

In semiconductors, the cyclotron resonance can be used to probe the shape of the bands near the edge of the conduction or valence band. Since the cyclotron-resonance frequency depends only on the effective mass, its measurement yields the most direct information about its behavior (see the following section and reviews by (Smith 1967; Pidgeon 1980).

1.2.5 Measurement of Effective Masses with Cyclotron Resonance

The effective mass is obtained directly from the cyclotron frequency (Eq. 24). By changing the relative alignment of the magnetic induction and the crystal axis, one can probe the anisotropy of the effective mass, such as that caused by the elliptical satellite bands – see Fig. 16b.

The resonance frequency in such elliptical bands is

$$\omega_c = e\sqrt{\frac{B_{0x}^2}{m_y m_z} + \frac{B_{0y}^2}{m_z m_x} + \frac{B_{0z}^2}{m_x m_y}}, \tag{25}$$

with the components of a matrix tensor (m_x, m_y, m_z) in diagonal form. With $m_{\parallel} = m_x$ and $m_{\perp} = m_y = m_z$, the longitudinal and transverse effective masses, one obtains from Eq. 25 for the resonance frequency

$$\omega_c = eB_0\sqrt{\frac{\cos^2\theta}{m_{\perp}^2} + \frac{\sin^2\theta}{m_{\parallel} m_{\perp}}}, \tag{26}$$

where θ is the angle between $\mathbf{B}_0$ and the principal axis of the $E(\mathbf{k})$ ellipsoid (Fig. 17).

Figure 18 shows the measured resonance spectrum in Ge at a constant probing frequency of 23 GHz (cm wave), with varying magnetic induction applied 10° out of the (110) plane and 30° from the [100] direction. Resonances shown in Fig. 18 are caused by electrons and holes in the different valence bands (see Fig. 10c). Hole resonances can be expressed (Dresselhaus et al. 1955) with $\omega_c = eB_0/m_p$ by introducing an anisotropic hole effective mass:

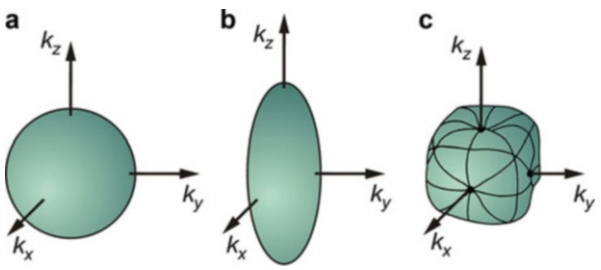


Fig. 16 Typical shapes of surfaces of constant energy: (**a**) spherical, (**b**) ellipsoidal, and (**c**) warped

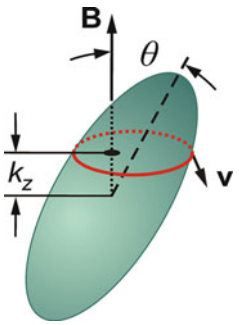


Fig. 17 Magnetic orbit in **k** space near the bottom of the conduction band in a satellite valley of Ge, with arbitrary orientation of the magnetic induction **B**

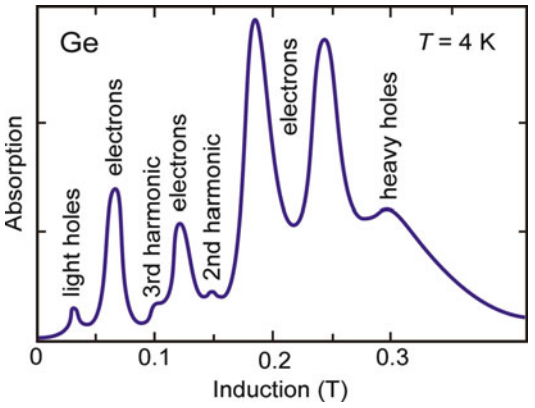


Fig. 18 Cyclotron resonance in Ge with magnetic induction applied in a direction showing the eight resonances observed in this crystal (After Dexter et al. 1956)

$$\frac{m_p}{m_0} = \frac{1}{A \pm \sqrt{B^2 + C^2/4}} \left(1 \pm \frac{C^2 (1 - 3\cos^2\theta)^2}{64\sqrt{B^2 + C^2/4}\left(A \pm \sqrt{B^2 + C^2/4}\right)} \right). \quad (27)$$

The parameters A, B, and C are the empirical valence-band parameters explained in the section on conduction bands above. Because of the warping of the valence band, higher harmonics of the heavy-hole resonance are also found in cyclotron resonance (Fig. 18).

1.2.6 The Conduction Band at Higher Energies

The parabolic approximation is no longer sufficient when a significant fraction of the electrons is at higher energies in the band, for example, at elevated temperatures, high electric fields, after optical excitation, or when pushed up by high doping – see Sect. 1 of ▶ Chap. 24, "Carrier Scattering at High Electric Fields," and Sect. 3 of ▶ Chap. 17, "Optical Properties of Defects." With increasing E, the band curvature decreases; hence, usually m^* increases, introducing an *energy-dependent effective mass*. When expressed by $m_n(T)$, the changes are relatively small between 0 and

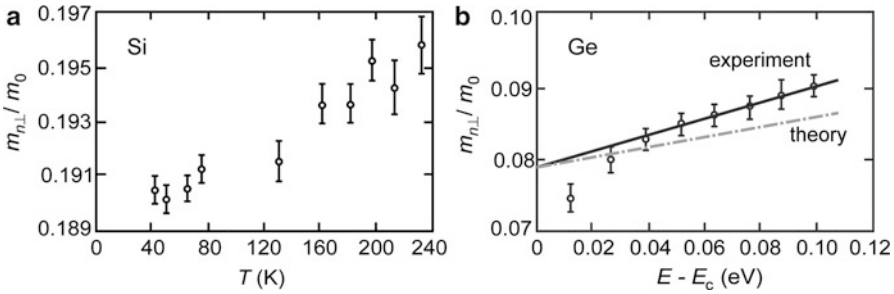


Fig. 19 Effective transverse electron mass (**a**) as a function of the temperature for Si (After Ousset et al. 1976) and (**b**) as a function of the energy for Ge (After Aggarwal et al. 1969)

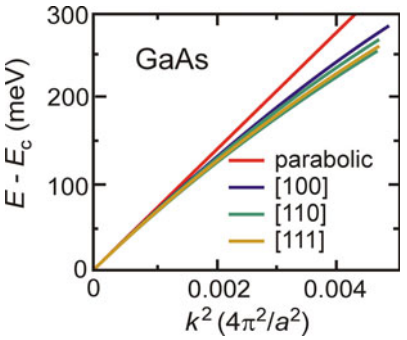


Fig. 20 $E(\mathbf{k})$ in the lowest conduction band of GaAs as a function of the square of the wavevector for various directions from $k = 0$ to $k = 0.07 \times (2\pi/a)$. The two curves for the [110] direction show the spin splitting of the conduction band (After Rössler 1984)

300 K (1–5%). This dependence is shown in Fig. 19 as a function of the temperature for Si and as a function of the energy above the band edge for Ge.

When a more accurate description of the band shape is needed beyond $E_c(k = 0)$, we can use an expression obtained from the $\mathbf{k} \cdot \mathbf{p}$ theory. For the shape of the conduction band of GaAs, for example, we have

$$E \ \ (k \text{ near } \Gamma_6) = \frac{\hbar^2 k^2}{2m_n} + (\alpha + \beta s)k^4 \pm \gamma\sqrt{s - 9t}\,k^3 \tag{28}$$

with

$$s = \frac{k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2}{k^4} \quad \text{and} \quad t = \frac{k_x^2 k_y^2 k_z^2}{k^6}. \tag{29}$$

The second term in Eq. 28 describes the deviation from parabolicity, and the third term describes a slight band warping of the conduction band. In GaAs there is also a slight spin splitting of the conduction band (Rössler 1984). See Fig. 20.

1.2.7 The Momentum Effective Mass

The effective mass of a carrier is conventionally described in relation to Newton's second law, yielding the well-known relation involving the curvature of $E(\mathbf{k})$ – see

Sect. 2.2 of ▶ Chap. 6, "The Origin of Band Structure." Another way to introduce this mass in a semiconductor with a spherical band is given by the relationship between the carrier velocity and the pseudo-momentum:

$$v_i = \sum_j \frac{\hbar}{m_{ij}} k_j \quad \text{with} \quad v_i = \frac{1}{\hbar}\frac{\partial E}{\partial k}. \tag{30}$$

In a spherical band, therefore, the *momentum effective mass* is a scalar:

$$\frac{1}{m^*} = \frac{1}{\hbar^2}\frac{1}{k}\frac{\partial E}{\partial k}. \tag{31}$$

In a parabolic dispersion $E(k)$, this definition is identical to the conventional one, relating the effective mass to the second derivative of $E(k)$. The common description is satisfactory as long as the discussion is restricted to the energy range near the band edge, where parabolicity is a reasonable approximation. At higher energies the momentum effective mass is more appropriate (Zawadzki 1982), and it will therefore be used in the following section.

1.2.8 The Effective Mass at Higher Energies

A large fraction of the electrons can reach substantially higher energies within a band when the Fermi level is shifted into the conduction band by higher doping. Since the curvature decreases with increasing distance from the lower band edge, the effective mass increases. In semiconductors with a very low effective mass (here one has a low density of states near the band edge), one can reach this condition at moderate doping levels, for example, with a donor density in excess of 10^{17} cm^{-3} for InSb (see Sect. 3 of ▶ Chap. 17, "Optical Properties of Defects") as can be seen by comparing upper and lower abscissae of Fig. 21. Kane (1957) estimates the shape of the conduction band as a function of the wavevector for a three-band model (Γ_6, Γ_7, and Γ_8) near $k = 0$ for InSb. In GaAs the corrections are somewhat smaller

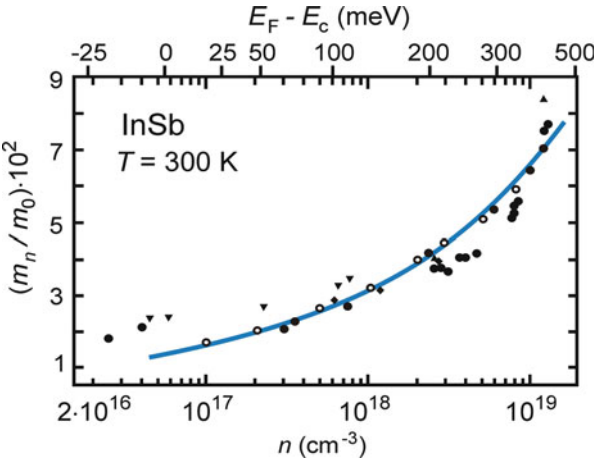


Fig. 21 Effective masses of electrons in InSb as calculated from the three-band Kane model (*solid curve*) and measured by various authors as a function of the position of the Fermi level below and inside the conduction band (After Zawadzki 1974)

(Vrehen 1968) because of a larger bandgap. Thus, one obtains for the nonparabolic conduction band

$$E(k) = \frac{\hbar^2 k^2}{2m_n^0} - \left(1 - \frac{m_n^0}{m_0}\right)\left(\frac{\hbar^2 k^2}{2m_n^0}\right)^2 \left(\frac{3E_g^2 + 4\Delta_0 E_g + 2\Delta_0^2}{E_g\left(E_g + \Delta_0\right)\left(3E_g + 2\Delta_0\right)}\right), \tag{32}$$

where the energy is normalized to $E = 0$ at E_c; Δ_0 is the spin–orbit splitting, and m_n^0 is the effective mass at the bottom of the conduction band, which can be expressed (Kane 1957) as

$$\frac{1}{m_n^0} = \frac{1}{m_0} + \frac{4P^2}{3\hbar^2 E_g} \frac{\Delta_0 + \frac{3}{2}E_g}{\Delta_0 + E_g}. \tag{33}$$

Here P is the matrix element connecting the conduction band with the three valence bands.

The effective mass slightly above the bottom of the conduction band can be approximated as

$$m_n(E) = m_n^0\left(1 + \frac{2E}{E_g}\right) \cong m_n^0\sqrt{1 - \left(\frac{v_g}{v}\right)^2}, \tag{34}$$

with $v = \sqrt{E_g/\left(2m_n^0\right)}$ and v_g as the group velocity (Eq. 17 of ▶ Chap. 6, "The Origin of Band Structure"; Landsberg 1987). Equation 34 is in fair agreement with the experiment for InSb (Fig. 21).

With increasing temperature, one must also consider the lattice expansion and consequent relative shift of the different bands if they are near enough to E_c. This influences the effective mass as a function of temperature and, in the given example of GaAs, causes a reduction in m_n with increasing temperature rather than an increase with increasing band filling – see Fig. 22.

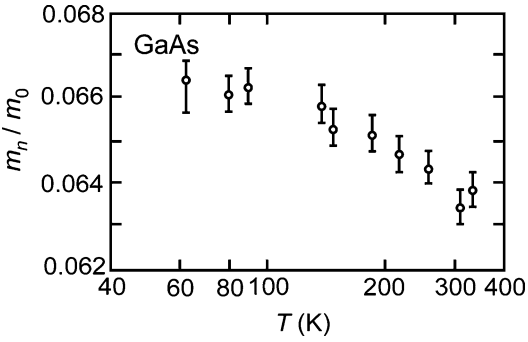


Fig. 22 Electron effective mass as a function of temperature for GaAs (After Blakemore 1982)

2 The Bandgap

In *semiconductors*, the valence band is separated from the conduction band by a relatively narrow bandgap (for further distinction, see Sect. 1 of ▶ Chap. 13, "Band-to-Band Transitions"); in insulators, the gap is much wider (Fig. 3). The distinction between semiconductors and insulators is arbitrary at a bandgap of ∼3 eV. A wide variety of materials provides a continuous transition of behavior from that of *insulators* and *wide-*, *narrow-*, *and zero-bandgap semiconductors* to metals.

The bandgap shows a distinctive trend that is seen for various *AB* compounds in Fig. 23 (see also Sect. 1, Tables 2, 3, and 8): it decreases with decreasing ionicity of the lattice binding forces and decreases steeply in the same class of compounds for homologous components with increasing atomic number, e.g., increasing ionic radius (interatomic spacing) or decreasing binding energy (see Phillips 1970). We must be careful, however, when using this trend for predicting the bandgap of unknown compounds, since substantial deviations occur between materials with the gap at the Γ, *L*, or *X* point in $E(\mathbf{k})$.

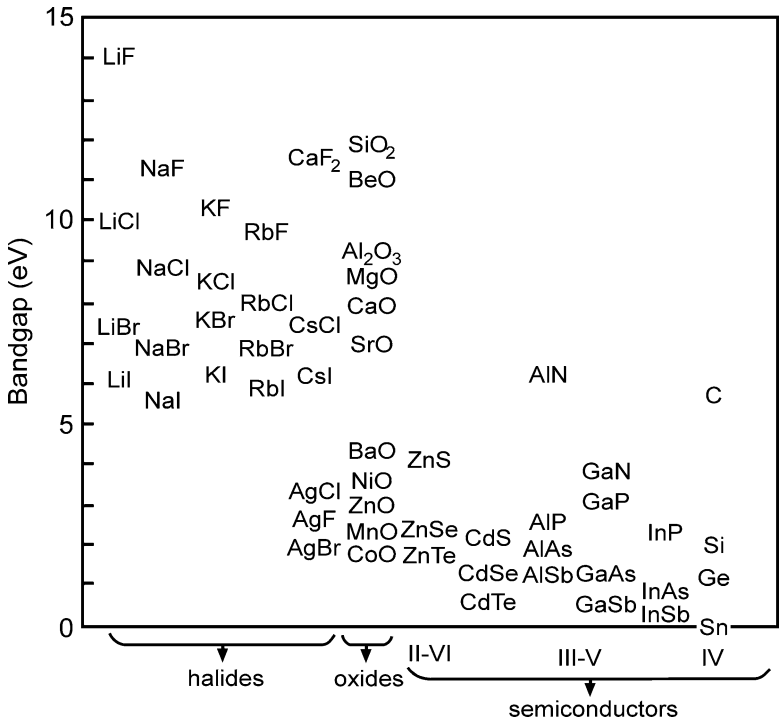


Fig. 23 Bandgap energies for various *AB* compounds and element semiconductors; *C* refers to the diamond modification (After Hayes and Stoneham 1984)

2.1 Bandgap of Alloys

A continuous variation of the width of the bandgap can be achieved by *alloy* formation or by application of hydrostatic pressure, which thereby changes the lattice composition or the lattice constant and thus the bandgap (see Sect. 2.2). Alloys[13] can be formed between similar metals or by mixing similar, homologous elements within a compound. An example is $Ga_xAl_{1-x}As$, in which the metal-atom sublattice is a homogeneous and statistical mixture (substitutional random alloy) of Ga and Al, with a resulting bandgap between that of GaAs and AlAs. Another example is $GaAs_{1-x}P_x$, in which the nonmetal sublattice is alloyed. There is an infinite number of such pseudo-compounds, which provides the possibility of creating any desired bandgap energy. For quaternary alloys such as $Ga_xAl_{1-x}P_yAs_{1-y}$, see Pearsall (1982).

Alloy formation between components AC and BC produces a linear interpolation of the bandgap

$$\overline{E}_g(x) = E_g(AC) + x\ \left[E_g(BC) - E_g(AC)\right] \tag{35}$$

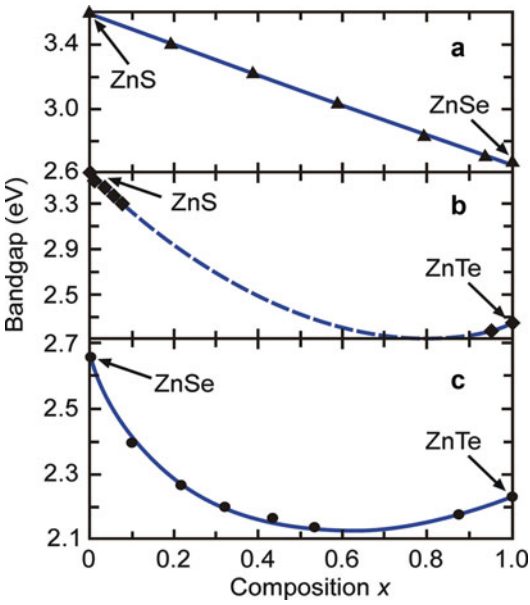


Fig. 24 Bandgap energy as a function of the composition parameter x of mixed crystals between ZnS, ZnSe, and ZnTe (After Larach et al. 1957)

[13]The conventional term alloy of metals also encompasses crystallite mixtures of nonintersoluble metals, such as lead and tin (solder). Here, however, only materials within their solubility ranges are discussed. The Hume-Rothery rule identifies these metals as having similar binding character, similar valency, and similar atomic radii (Hume-Rothery 1936). Corresponding guidelines apply to the intersolubility of cations or anions in compounds.

only if the alloying atoms (A and B) have nearly identical binding forces to atom C and have nearly the same atomic radii. The bar over E_g indicates the averaging over various local configurations with different numbers of nearest neighbors of a certain atom. An example is ZnS_xSe_{1-x} (Fig. 24a). If the radii are substantially *different*, a strong *bowing* of the bandgap is observed, as shown for ZnS_xTe_{1-x} and $ZnSe_xTe_{1-x}$ in Fig. 24b, c (Bernard and Zunger 1987). For a review, see Jaros (1985).

The bowing is described with an empirical bowing parameter b according to

$$E_g(x) = \overline{E_g}(x) - b\,x\ (1-x). \tag{36}$$

This bowing parameter is listed in Table 7 for a selection of II–VI and III–V compounds. It must be noted that reasonable values for b in many cases apply only for a limited composition range; sometimes also composition-dependent parameters $b(x)$ are reported (see, e.g., Vurgaftman et al. 2001).

Since the lattice parameters of semiconductor alloys are well described by the linear dependence expressed by Vegard's rule (Eq. 11 of ▶ Chap. 2, "Crystal Bonding"), the composition parameter x used as abscissa in Fig. 24 is usually replaced by a lattice parameter (generally a in both zincblende and wurtzite semiconductors). The resulting presentation of alloy bandgaps is shown in Fig. 25 for various semiconductors. Kinks in the curves indicate crossings of conduction bands.

The bowing is caused by a change of the lattice energy of the alloy, which is due to chemical and structural differences when a lattice atom is replaced by an alloying atom. The change in bandgap energy is given by

$$\Delta E_g = \Delta E_g^{\text{chem}} + \Delta E_g^{\text{struc}}. \tag{37}$$

The varying chemical nature of the alloying atom may be expressed by its different *electronegativity* "en" and *hybridization* "pd" when forming the lattice bond:

Table 7 Bowing parameter b in eV for various compound semiconductors (computed from Richardson 1973)

	GaAs	AlSb	InP	GaSb	InAs	InSb	ZnS	ZnSe	ZnTe	CdTe
GaP	0.38	2.31	1.36	3.51	2.67	5.68	0.87	1.41	3.51	2.96
GaAs		0.44	0.22	1.44	0.84	2.86	1.20	0.76	1.55	0.76
AlSb			0.01	0.005	0.05	0.92	3.64	1.88	1.17	1.55
InP				0.49	0.22	1.55	2.12	0.92	0.84	0.46
GaSb					0.008	0.24	4.08	1.85	0.44	−0.24
InAs						0.52	3.13	1.25	0.35	0.05
InSb							6.20	3.21	0.90	0.44
ZnS								0.60	3.02	2.07
ZnSe									0.90	0.02
ZnTe										−0.57

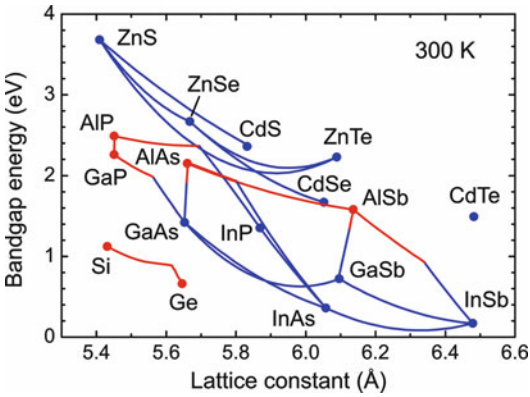


Fig. 25 Bandgap energy as a function of lattice constant a for pure (*dots*) and alloyed zincblende and diamond semiconductors at room temperature. *Blue and red drawing* denotes direct and indirect bandgap, respectively

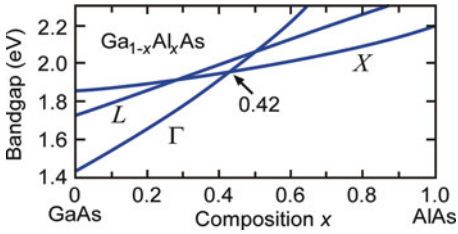


Fig. 26 Changes of the conduction-band minima of $Ga_{1-x}Al_xAs$ as a function of the composition parameter x at room temperature (From data of Saxena 1981)

$$\Delta E_g^{\text{chem}} = \Delta E_g^{\text{en}} + \Delta E_g^{\text{pd}}. \tag{38}$$

The changes in structure are induced by different *bond lengths* (u) and *tetragonal* (*bond angle*) *distortion* (η) (see below):

$$\Delta E_g^{\text{struc}} = \Delta E_g^{\text{u}} + \Delta E_g^{\eta}. \tag{39}$$

The shifting of different conduction-band minima by alloying can be substantially dissimilar from each other and consequently may change their relative position. An example is shown in Fig. 26 for the Γ, X, and L points of $Ga_{1-x}Al_xAs$.

The structural changes are more easily discussed when starting from a 50% alloy. When ordered, this alloy can be described as a chalcopyrite. For example, an ordered $Zn_{0.5}Cd_{0.5}S$ can be described as $ZnCdS_2$, which in structure is similar to $CuGaS_2$ (see Fig. 13b of ▶ Chap. 3, "The Structure of Semiconductors"). The bond alternation is given by an anion displacement parameter $u = \left(\overline{AC}^2 - \overline{BC}^2\right)/a^2 + 1/4$ with (A, B, C) standing for the components (Zn, Cd, S in the given example) and $\overline{AC}$ as the average distance between atoms A and C, etc. The tetragonal distortion is described by the ratio $\eta = c/2a$. The more the measured u in the actual alloy deviates from the ideal ratio 1/4 and η from 1 for an ideal chalcopyrite, the larger is the contribution of $\Delta E_g^{\text{struc}}$ to the bowing.

In ternary compounds, e.g., chalcopyrites, little bowing is observed when the anion sublattice is alloyed (e.g., $CuInS_xSe_{1-x}$, $CuInSe_xTe_{1-x}$, and $CuInS_xTe_{1-x}$), while there is substantial bowing when the cation sublattice is alloyed (e.g., $CuGa_xIn_{1-x}Se_2$ or $Cu_xAg_{1-x}GaSe_2$). For reviews, see Martins and Zunger (1986) and also Jaffe and Zunger (1984).

Strong bowing indicates a large change in lattice energy, which acts as a driving force for ordering, that is, for compound formation when a stoichiometric atomic ratio is reached rather than the formation of a statistical alloy. Examples are SiGe, $GaInP_2$, or Ga_2AsP (Jen et al. 1986; Srivastava et al. 1986; Ourmazd and Bean 1985). For $In_xGa_{1-x}N$ and $In_xAl_{1-x}N$ alloys, see Ferhat and Bechstedt (2002).

2.2 Bandgap Dependence on Temperature and Pressure

The bandgap (and other band features – see Sect. 1 of ▶ Chap. 13, "Band-to-Band Transitions") changes with temperature and pressure:

$$\Delta E_g = \left(\frac{\partial E_g}{\partial T}\right)_P \Delta T + \left(\frac{\partial E_g}{\partial P}\right)_T \Delta P. \tag{40}$$

Temperature-induced changes are due to changes in:

- Lattice constants
- Electron–phonon interaction

The first term of Eq. 40, however, is closely related to the second term, which gives the changes of the bandgap under pressure. The latter can be divided into several contributions, solely related to a change in the lattice constant. These changes influence the optical behavior due to changes in:

- Matrix element, which depends on $1/a$ (lattice constant in a cubic crystal)
- Density of states inducing changes in the effective mass
- Energy of electronic levels
- Plasma frequency containing a changed density of dipoles and a changed effective charge
- Phonon frequency as the lattice stiffens with increasing pressure (anharmonicity of oscillations)

The changes in *electron–phonon interaction* dominating the first term of Eq. 40 are more involved. They have attracted substantial interest and can be divided into three different approaches:

1. The approach suggested by Fan (1951), involving an electron self-energy term that arises from spontaneous emission and reabsorption of a phonon; this approach was expanded by Cohen (1962) to include intervalley scattering.

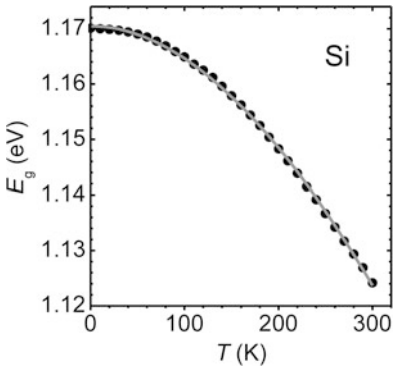


Fig. 27 Decrease of the bandgap energy with increasing temperature for Si. The curve is a fit to Eq. 44 (Data are from Bludau et al. 1974)

2. The approach suggested by Antončik (1955), who introduced a temperature-dependent structure factor (*Debye–Waller factor*),[14] that is experimentally accessible from the temperature dependence of the Bragg reflections.
3. The approach suggested by Brooks (1955) and refined by Heine and van Vechten (1976), relating to a change in lattice vibrations from ω to ω' when an electron is excited from the valence into the conduction band:

$$E_g(T) = E_g(0) - kT\sum_i \ln\left(\frac{\omega_i'}{\omega_i}\right). \tag{42}$$

See Lautenschlager et al. (1985) for later work.

Finally, the influence of *uniaxial stress* lowers the crystal symmetry and thus removes the degeneracy of the valence-band maximum in cubic crystals. Uniaxial stress experiments can be designed to yield information on level and effective-mass symmetries and on deformation potentials. These changes will be discussed in Sect. 3.2 of ▶ Chap. 18, "Shallow-Level Centers."

Since the influences of temperature and pressure are manyfold and involve all bands, it is not possible to describe features which apply universally to all semiconductors. A few examples will be presented here. More comprehensive literature includes a large number of odd cases that behave substantially differently from the given examples, e.g., referenced by Martinez (1980).

The bandgap usually decreases with increasing temperature – see Fig. 27. Exceptions are the lead chalcogenides, which show an increase of the gap with temperature; this bandgap is determined at the L point – see Fig. 5.

In a wide temperature range, the bandgap changes linearly with temperature

[14]This Debye–Waller factor (W) is related to the probability of phonon emission during electron or x-ray diffraction and is given in the Debye approximation by

$$p = \exp(-2W) = \exp\left(-\frac{6E_R}{k\Theta}\left[\frac{1}{4} + \frac{T}{\Theta}\int_0^{\Theta/T}\frac{x\ dx}{\exp(x)-1}\right]\right), \tag{41}$$

where E_R is the recoil energy $Mv^2/2$.

Table 8 Direct (d) and indirect i) bandgap energies for various semiconductors and their temperature dependence expressed in terms of parameters α and β according to Eq. 44 and their linear temperature and pressure coefficients

Crystal	E_g (0 K) (eV)	E_g (300 K) (eV)	α (meV/K)	β (K)	$\frac{dE_g}{dT}\left(\frac{meV}{K}\right)\Big\|_{300\,K}$	$\frac{dE_g}{dP}\left(\frac{meV}{kbar}\right)$
Si	1.1695 (i)	1.110 (i)	0.473	636	−0.28	−1.41
Ge	0.744 (i)	0.664 (i)	0.477	235	−0.37	5.1
AlN	6.25 (d)	6.13 (d)	1.799	1462	−0.6	4.9
AlP	2.52 (i)	2.45 (i)	0.318	588	−0.2	
AlAs	2.24 (i)	2.15 (i)	0.70	530	−0.4	10.5
GaN	3.510 (d)	3.42 (d)	0.909	830	−0.4	3.9
GaP	2.350 (i)	2.272 (i)	0.5571	372	−0.37	10.5
GaAs	1.519 (d)	1.411 (d)	0.540	204	−0.39	11.3
GaSb	0.812 (d)	0.69 (d)	0.417	140	−0.37	14.5
InN	0.78 (d)	∼0.75 (d)	0.245	624	−0.1	3.0
InP	1.424 (d)	1.34 (d)	0.363	162	−0.29	9.1
InAs	0.417 (d)	0.356 (d)	0.276	93	−0.34	10.0
ZnSe	2.820 (d)	2.713 (d)	0.490	190	−0.45	7
CdS	2.579 (d)	2.501 (d)	0.47	230	−0.38	4.4
CdSe	1.841 (d)	1.751 (d)	1.7	1150	−0.36	5.0
CdTe	1.606 (d)	1.43 (d)	0.310	108	−0.54	7.9

$$E_g(T) = E_{g0} + \beta_E \; T, \tag{43}$$

where β_E is typically in the -10^{-4} eV/deg range – see dE_g/dT at 300 K in Table 8. For lower temperature, however, $|\beta_E|$ decreases and vanishes for $T \to 0$ according to the third law of thermodynamics.

The temperature dependence of the gap over a larger temperature range can be approximated by (Varshni 1967)

$$E_g(T) = E_{g0} - \frac{\alpha \; T^2}{\beta + T}. \tag{44}$$

The empirical parameters α and β are typically in the range (0.4–0.6) meV/K and (100–600) K as listed in Table 8 for some semiconductors.

The change of the bandgap energy due to pressure is expressed as

$$\Delta E_g(P) = \left(\frac{\partial E_g}{\partial P}\right)_T \Delta P. \tag{45}$$

However, it cannot be linearized in a wider pressure range, since it often involves several bands which are influenced by different pressure coefficients that may even have different signs. As an example, the pressure dependence of the gap in GaAs is given in Fig. 28a. The gap first increases, shows a maximum at $\sim 6 \times 10^4$ bar, and then decreases again with increasing pressure. The lower valley in the conduction

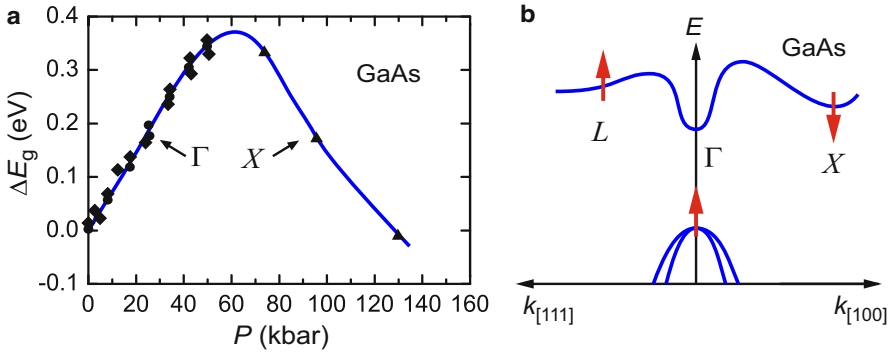


Fig. 28 (**a**) Pressure dependence of the bandgap of GaAs, and (**b**) trends (*arrows*) of $E(k)$ with increasing pressure (After Edwards et al. 1959)

band in GaAs (Γ point) increases, while the (100) valley decreases with pressure. At pressures above 8×10^4 bar, GaAs becomes an indirect bandgap material. The pressure coefficient can approximately be described by the relation E_g (eV) = 1.45 + 0.0126 P – 3.77 × $10^{-5}P^2$ (P in kbar).

Pressure coefficients near 1 bar are typically on the order of 10 meV/kbar. They are usually positive for Γ ⟨000⟩ and L ⟨111⟩ valleys and are negative for valleys along ⟨l00⟩ – see Fig. 28b.

2.3 Bandgap at High Doping Level

The bandgap in heavily doped semiconductors is substantially reduced. There are many reasons for the reduction of the bandgap, summarized in an extensive review by Abram et al. (1978); see also Mahan (1980), van Overstraeten (1982), and Kalt (1996). The more important phenomena which influence the bandgap at high doping densities ($N_d > 10^{17}$ cm^{-3}) are given below:

(a) The exchange energy of electrons due to their fermion nature. It tends to keep the electrons with parallel spin away from each other but is attractive to electrons with opposite spin, resulting in a net attractive term. As a consequence, one obtains a lowering of the conduction-band edge δE_{cex} relative to the Fermi level in equilibrium and at $T = 0$:

$$\delta E_{\text{cex}} = -\frac{e^2}{\pi\varepsilon\varepsilon_0}\Lambda_\delta k_F \cong B_{\delta e}\left(\frac{N_d\ (\text{cm}^{-3})}{10^{18}}\right)^{\frac{1}{3}}\quad (\text{meV}); \tag{46}$$

the wavenumber at the Fermi surface is given by

$$k_F = \left(\frac{3\pi^2 N_d}{\nu_{\text{D}}}\right)^{\frac{1}{3}}, \tag{47}$$

assuming total ionization of uncompensated donors of density N_d and ν_D the

Table 9 Empirical parameters for bandgap narrowing (Mahan 1980)

Crystal	ν_D	$A_{\delta F}$	λ_δ	$B_{\delta e}$	$C_{\delta e}$	$D_{\delta p}$	$E_{\delta p}$
Si	6	3.3	0.95	−6.5	−12.1	−13.1	6.1
Ge	4	6.6	0.84	−4.9	−6.0	−8.2	4.4

degeneracy factor of the band; Λ_δ and $B_{\delta e}$ are numerical factors listed in Table 9 (Mahan 1980); for further many-body corrections, see Kalt (1996).

(b) The attractive interaction between free electrons and charged donors, causing another reduction in the conduction-band edge:

$$\delta E_{\text{ced}} = -\frac{e^2}{\pi\varepsilon\varepsilon_0\lambda_{TF}} \cong C_{\delta e}\left(\frac{N_d\ (\text{cm}^{-3})}{10^{18}}\right)^{\frac{1}{6}} \quad (\text{meV}), \tag{48}$$

with the Thomas–Fermi screening length

$$\lambda_{TF} = \pi^{\frac{2}{3}}\sqrt{\frac{\varepsilon_{\text{st}}\varepsilon_0\hbar^2}{m_n e^2(3N_d)^{\frac{1}{3}}}} \tag{49}$$

which is responsible for screening at sufficient doping density – see (Landsberg et al. 1985).

(c) The exchange energy for holes, causing a relative increase of the valence-band edge by

$$\delta E_{\text{vhx}} = -\frac{m_p\ e^4\sqrt{\omega_p}}{\pi(\varepsilon\varepsilon_0)^2\hbar}J \cong D_{\delta p}\left(\frac{N_d\ (\text{cm}^{-3})}{10^{18}}\right)^{\frac{1}{4}} \quad (\text{meV}), \tag{50}$$

with ω_p as the *plasma frequency* (Sect. 1.1 of ▶ Chap. 12, "Photon–Free-Electron Interaction") and J as an integral (Mahan 1980) of nearly constant value ($J \cong 0.8$).

(d) The interaction between holes and donors is repulsive and thus causes a relative lowering of the valence-band edge by

$$\delta E_{\text{vhd}} = -0.48\frac{e^2}{\varepsilon\varepsilon_0}N_d^{\frac{1}{3}} \cong E_{\delta p}\left(\frac{N_d\ (\text{cm}^{-3})}{10^{18}}\right)^{\frac{1}{3}} \quad (\text{meV}). \tag{51}$$

The total change in the bandgap energy is obtained as the sum of all of these contributions,

$$\Delta E_{\text{g}} = \delta E_{\text{cex}} + \delta E_{\text{ced}} + \delta E_{\text{vhx}} + \delta E_{\text{vhd}} = \Delta E_{\text{c}} + \Delta E_{\text{v}}. \tag{52}$$

The influence of the first two contributions on the conduction band is shown in Fig. 29a for Si. The total reduction of the bandgap energy as a function of the doping

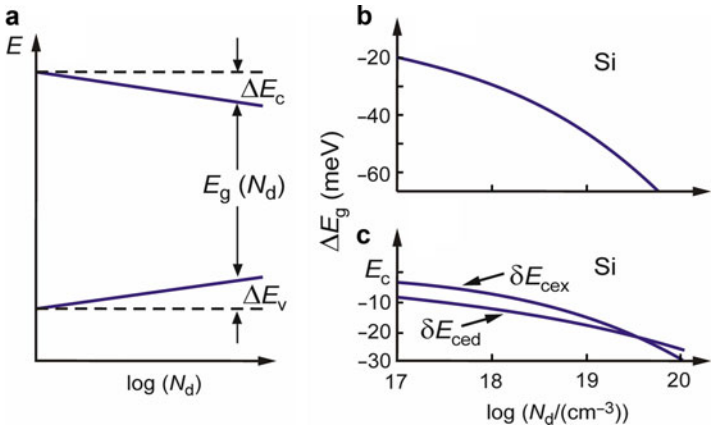


Fig. 29 (**a**) Schematic illustrating the change of the bandgap energy with increasing donor density. (**b**) Bandgap narrowing as a function of the density of shallow donors in Si; (**c**) contributions of the electron exchange energy δE_{cex} and the interaction of electrons and charged donors δE_{ced} to the relative lowering of the conduction band in Si (After Mahan 1980)

Table 10 Doping-dependent bandgap energy

Crystal	Bandgap energy
Si	$E_g\,(\text{eV}) = 1.206 - 0.0404\,\ln(N_a\,(\text{cm}^{-3})/7.5 \times 10^{16})$
GaAs	$E_g\,(\text{eV}) = 1.45 - 1.6 \times 10^{-8}\sqrt[3]{n\ (\text{cm}^{-3})}$
InP	$E_g\,(\text{eV}) = 1.344 - 2.25 \times 10^{-8}\sqrt[3]{n\ (\text{cm}^{-3})}$
InN	$E_g\,(\text{eV}) = E_g(n=0) - 2.0 \times 10^{-8}\sqrt[3]{n\ (\text{cm}^{-3})}$

density is shown for Si in Fig. 29b, and contributions of δE_{cex} and δE_{ced} are given in Fig. 29c.

Pantelides et al. (1985) estimated that multivalley interaction and density fluctuation cause a further reduction in the electrically obtained bandgap, which is on the order of 50 meV at $N_d \sim 10^{20}$ cm^{-3} in GaAs. This agrees with the optically determined bandgap reduction via absorption, luminescence, or photoluminescence spectroscopy (Wagner 1984). See also Berggren and Sernelius (1984).

Simple empirical relations to approximate the change in bandgap energies with doping are given for some semiconductors in Table 10; see also Jain and Roulston (1991).

2.3.1 Fermi Level at High Doping Densities

When the doping density N_d exceeds the effective level density N_c at the lower edge of the conduction band, the Fermi level moves from the bandgap into the conduction band. This results in an "effective widening" of an optical gap, since, for example, optical excitation from the valence band can only proceed to empty states above the

Fermi level. The shift is given for parabolic bands, using the density of states in these bands (Sect. 1 of ▶ Chap. 21, "Equilibrium Statistics of Carriers") and $n \cong N_d$, by

$$\delta E_{\mathrm{cF}} = \frac{\hbar^2}{2m_{\mathrm{dsn}}}\left(\frac{3\pi^2 N_d}{\nu_{\mathrm{D}}}\right)^{\frac{2}{3}} \cong A_{\delta F}\left(\frac{N_d\ (\mathrm{cm}^{-3})}{10^{18}}\right)^{\frac{2}{3}}\ (\mathrm{meV}), \tag{53}$$

where ν_{D} is the degeneracy factor of the conduction band and m_{dsn} is the density-of-state mass for electrons. This equation is valid when all donors are ionized and uncompensated. A similar expression can be obtained for high doping with acceptors, replacing m_{dsn} with the density-of-state mass for holes m_{dsp} and ν_{D} with the degeneracy factor for the valence bands. However, the larger effective mass of holes in the heavy-hole band makes this effect less favorable. Table 9 lists $A_{\delta F}$ and ν_{D} for Si and Ge. Such a shift is substantial even at moderate electron densities for semiconductors with low effective mass (e.g., InP). See Sect. 3.2 of ▶ Chap. 17, "Optical Properties of Defects," for the optical manifestation of this shift (*Burstein–Moss effect*).

3 Electronic States in Low-Dimensional Semiconductors

When a semiconductor with a smaller bandgap is embedded in another one of a larger gap, the latter acts as a barrier for carriers. To find out whether electrons or holes or both are confined in the small-bandgap semiconductor and to obtain the actual barrier height, it is important to know the offset of the valence-band edge (Kroemer 1983; Yu et al. 1992) and the actual bandgap of the (generally strained) confined solid. Three cases referred to as type I to III may be distinguished in the alignment of the band edges of the two semiconductors as illustrated in Fig. 30. The *straddled alignment* of a *type I* double heterostructure provides a confinement for both electrons *and* holes, while the *staggered alignment* of *type II* confines *either* electrons (as depicted in Fig. 30b) *or* holes (if $E_{\mathrm{v}B} > E_{\mathrm{v}A}$). The *misaligned band lineup* of *type III*, also referred to as broken-gap configuration, occurs if the bandgaps of the two semiconductors do not overlap, i.e., if $E_{\mathrm{v}A}$ lies above $E_{\mathrm{c}B}$ or $E_{\mathrm{c}A}$ lies below $E_{\mathrm{v}B}$.

A *quantum well* is formed if the confining potential (illustrated for electrons in Fig. 30) acts solely in one spatial direction and extends over a small length in the range or below the de Broglie wavelength of confined carriers[15]; the motion of carriers is restricted in this direction and unrestricted in the two perpendicular directions, creating a two-dimensional (2D) well. If an additional confinement acts along one of these two perpendicular directions, a *quantum wire* (1D) is formed with an unrestricted carrier motion only along one direction: the wire axis. If a further

[15] At room temperature this is $\lambda = h/p = h/\sqrt{2m^*E} \cong 17$ nm for 300 K thermal energies and a typical effective mass of $0.2 \times m_0$, i.e., in the 10 nm range. For excitons the relevant Bohr radius is also in this range, e.g., 11.5 nm for GaAs and 3.2 nm for GaN.

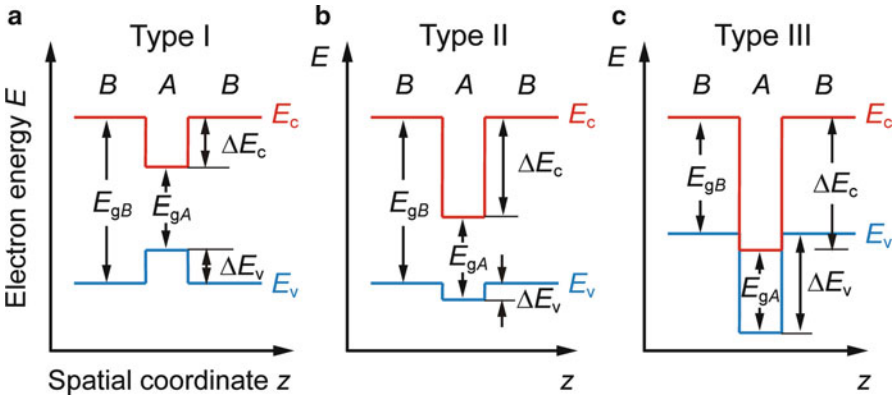


Fig. 30 Alignment of band edges in (**a**) type I, (**b**) type II, and (**c**) type III double heterostructures made from small-bandgap semiconductor *A* with a small extension along the spatial coordinate *z* clad by a wide-bandgap semiconductor *B*. E_v denotes the upper edge of the valence band, E_c the lower edge of the conduction band, the difference $E_c - E_v$ is the energy of the bandgap E_g

confinement potential acts also along this axis, a *quantum dot* is created with a three-dimensional confinement, leading to 0D states. 2D through 0D quantum structures are discussed in the following.

3.1 Quantum Wells and Superlattices

3.1.1 Quantum Wells

A quantum well is made from a thin semiconductor layer sandwiched between semiconductors with a larger bandgap forming barriers. Usually the same material is used for lower and upper barriers leading to a symmetrical square potential as illustrated for E_c in Fig. 30. For an *infinite* confinement potential, the wavefunction of a confined carrier cannot penetrate into the barriers. The eigenvalues along this direction[16] for a well of thickness L_z are given by

$$E_{z,n} = \frac{\hbar^2}{2m_z^*}\left(\frac{n\pi}{L_z}\right)^2, \quad n = 1, 2, 3, \ldots . \tag{54}$$

The quantum number gives the number of the carrier's half-wavelengths in the confined state. We note that the energy of the ground state with the quantum number $n = 1$ is increased by the *quantization energy* $\Delta E = E_{z,1}$. The eigenvalues of the wavefunction are given by the sum of those for the *x* and *y* directions and (Eq. 54),

[16]Quantum wells are fabricated from a layer sequence *BAB* grown using epitaxy (Sect. 3.3 of ▶ Chap. 1, "Properties and Growth of Semiconductors"); the growth direction is usually referred to as *z* direction, which we designate here as the direction where the confinement occurs.

yielding $E = E_x + E_y + E_{z,n}$. For unrestricted carrier motion in the xy plane according to the parabolic approximation, this sum reads

$$E = E_n = \frac{\hbar^2 k_x^2}{2m_x^*} + \frac{\hbar^2 k_y^2}{2m_y^*} + \frac{\hbar^2}{2m_z^*}\left(\frac{n\pi}{L_z}\right)^2. \tag{55}$$

The dispersion $E(\mathbf{k})$ along k_x and k_y therefore consists of a series of parabola each labeled by a particular value of n. The parabola are also referred to as *subbands*.

The confinement of a semiconductor quantum-well is given by the *band offsets* illustrated in Fig. 30, which are typically some tenth of eV. For such *finite* barrier height $W(z) = \Delta E \equiv W_0$, the wavefunctions of a confined charge carrier penetrate into the barriers. The eigenvalues for a *square potential* are obtained from the transcendental equation

$$\tan\left(\sqrt{\frac{m_w E_n L_z^2}{2\hbar^2}}\right) = \sqrt{\frac{m_w}{m_b}\frac{W_0 - E_n}{E_n}} \tag{56a}$$

for even wavefunctions, i.e., even quantum numbers n, and

$$\cot\left(\sqrt{\frac{m_w E_n L_z^2}{2\hbar^2}}\right) = \sqrt{\frac{m_w}{m_b}\frac{W_0 - E_n}{E_n}} \tag{56b}$$

for odd wavefunctions (Bastard and Brum 1986). m_w and m_b are the effective masses of the carriers in well and barriers, respectively, and L_z is the well width. Numerically obtained solutions are given in Fig. 31 (Dingle et al. 1974). Energies in the figure are

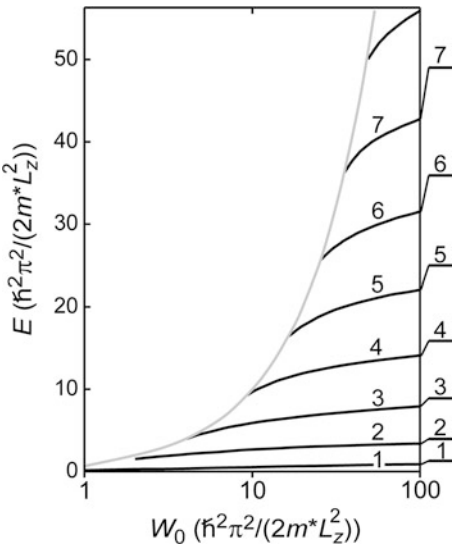


Fig. 31 Calculated bound-state energies of a carrier in a symmetrical rectangular potential well of finite depth W_0 indicated by the *gray line*. Bars at the right-hand side mark energy levels with quantum numbers n for infinite high barriers (After Dingle 1975)

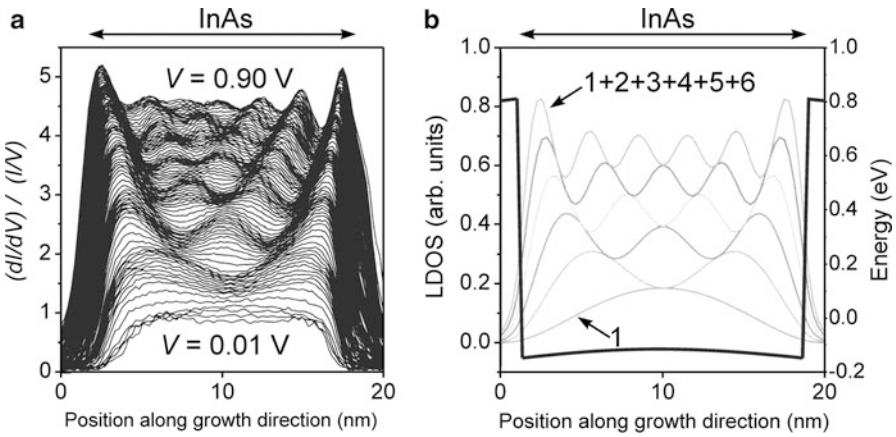


Fig. 32 (**a**) Scanning tunneling $(dI/dV)/(I/V)$ spectra locally probed across a 17 nm wide InAs/GaSb quantum well for bias varied from 0.01 to 0.9 V in steps of 0.01 V. (**b**) Confining potential (*black lines*) and calculated local density of states given as the sum of the squared subband wavefunctions (*gray curves*) (After Suzuki et al. 2007)

scaled in units of the ground-state energy E_1 of a well with infinite barriers (Eq. 54). The gray line signifies the top of the well at $E = W_0$. Discrete bound states are found for $E < W_0$, while continuum states exist for $E \geq W_0$. We note that the number of bound states of a confined carrier decreases as W_0 decreases. Furthermore, the level spacing and consequently the energy of the levels decrease. The topmost bound level approaches the top of the well as W_0 is gradually reduced. It should be noted that at least one bound level exists in any quantum well.

A direct demonstration of subband formation in InAs/GaSb quantum wells by imaging the local density of states is shown in Fig. 32 (Suzuki et al. 2007). The two semiconductors form a broken-gap band alignment with the conduction-band edge of the InAs quantum well lying below the valence-band edge of the cladding GaSb (Fig. 30c), thereby providing a large confinement potential for electrons in the well. Using the tip of a low-temperature scanning-tunneling microscope, the local density of states was imaged across the InAs well by the spatially resolved normalized differential conductance dI/dV: the tunneling tip locally probes the probability amplitude of electrons (i.e., the electron density) across the well (Rontani et al. 2007).

The $E(\mathbf{k})$ behavior of the *valence band* in an actual strained semiconductor quantum well is a bit more complicated, since we must distinguish between light and heavy holes in cubic (zincblende, diamond) or light, heavy, and crystal holes in wurtzite semiconductors, which result in two or three sets of mini-bands. Considering excited states in the valence band of a cubic crystal, we observe a crossing of states between heavy- and light-hole bands, as shown in Fig. 33. This results in a mixing between these states (Collins et al. 1987). As shown on an enlarged scale in Fig. 33c, this indicates that the dispersion curves cannot cross and that the interaction (mixing) that takes place near the points of *intended* crossing results in a splitting,

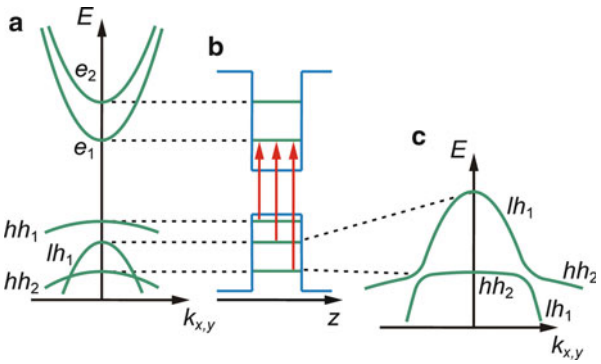


Fig. 33 (**a**) Dispersion relations for electron and hole subbands in a quantum well confining two electron states, one light-hole and two heavy-hole states. (**b**) Optical excitation transitions into the ground state of the conduction band. (**c**) Light- and heavy-hole dispersions showing the crossing at enlarged scale

making the top of the upper band light-hole-like and its bottom part heavy-hole-like and vice versa for the lower band; compare also to Fig. 15.

3.1.2 Superlattices

Permitted bands separated by bandgaps occur in a bulk crystal as a consequence of the quantum-mechanical properties of electrons in a periodic potential. This remains true in superstructures. The additional periodicity in the lattice potential can be provided by alternating layers of semiconductors with different bandgaps, i.e., in *superlattices* – see Esaki and Tsu (1970) and Ando et al. (1982). Some of the effects observed in a superlattice are already present in a single layer that provides a two-dimensional quantum-well confinement of the electronic eigenfunctions as discussed above.

The periodic alternation of barrier and well layers produces a potential (Fig. 34a) of the same form as the previously discussed Kronig–Penney potential[17] (Sect. 1.2 of ▶ Chap. 6, "The Origin of Band Structure"). The resulting eigenvalue spectrum is similar to the spectrum of free electrons exposed to the periodic potential of a crystal, except that now the periodic potential is imposed on *Bloch electrons* with an effective mass m_n, and the potential has a lower amplitude and a larger period length than the periodic potential in a bulk lattice. Consequently, within the conduction band, one observes a subband structure of *mini-bands* located within the valleys of this band; the higher mini-bands extend beyond the height of the potential barriers. The lower mini-bands are separated by *mini-bandgaps* (Fig. 34b) in the direction of the superlattice periodicity l_z. Within the plane of the superlattice layers (x, y), however, the electron eigenfunction experiences only the regular lattice periodicity. Therefore, the dispersion relations $E(k_x)$ and $E(k_y)$ are much like those for the unperturbed lattice except for the mixing with the states in the z direction; this results in lifting the lowest energy

[17]The alternating potential shown in Fig. 30a is of type I, i.e., a minimum of $E_c(z)$ coincides with a maximum of $E_v(z)$. Both minima and maxima coincide in a type-II superlattice (Fig. 30b). An example for type II is the $GaAs_xSb_{1-x}$ superlattice. For x below 0.25, the valence band of GaAs extends above the conduction band of GaSb, resulting in quasimetallic behavior. For a review of type-II superlattices, see Voos and Esaki (1981).

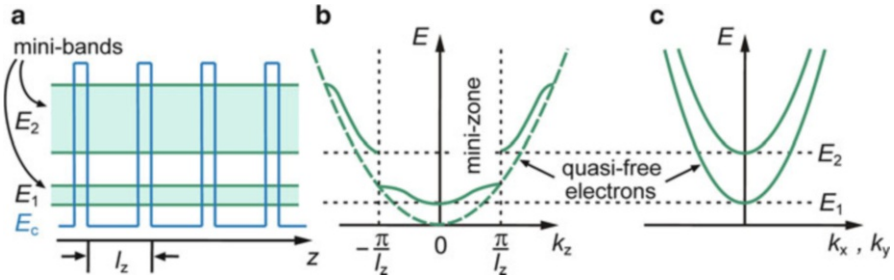


Fig. 34 (**a**) Mini-bands and (**b**) mini-zones in the conduction band in the k_z direction for a superlattice. Carriers are confined in the z direction in the lower mini-bands. There is no confinement in the x and y directions; (**c**) the ordinary band scheme applies, but with the band minimum lifted to the respective mini-band minimum

(at $k = 0$) of the $E(k)$ parabola above E_c of the bulk well material (Fig. 34b, c). The second mini-band results in a second, shifted parabola, etc.

The mini-band structure is a direct-bandgap structure, independent of whether the host (well) material has a direct or indirect bandgap. It permits optical transitions from the valence band, which has a similar mini-band structure, to the lowest conduction mini-band state at $k = 0$ (Fig. 33b).

Variations of the period length, barrier width, and barrier height change the width of the allowed mini-bands and the interfacing mini-gaps. Using the Kronig–Penney model, these are determined by Eq. 12 of ▶ Chap. 6, "The Origin of Band Structure"; however, m_0, contained in α and β (see Eq. 11 of ▶ Chap. 6, "The Origin of Band Structure"), is replaced by m_n and the lattice parameters a_1 and a_2 (Fig. 7 of ▶ Chap. 6, "The Origin of Band Structure") are replaced by the superlattice well and barrier widths l_1 and l_2 (Fig. 35). This yields an implicit equation for the band edges (i.e., for $\cos(ka) = 1$ in Eq. 12 of ▶ Chap. 6, "The Origin of Band Structure"):

$$\frac{\alpha^2 - \beta^2}{2\alpha\beta} = \mathrm{ctg}\ (\alpha l_1)\ \ \mathrm{ctgh}\,(\beta l_2). \tag{57}$$

The resulting mini-band structure and $E(k)$ dispersion relation shown in Fig. 34b is very similar to the band structure shown in Figs. 3b and 10 of ▶ Chap. 6, "The Origin of Band Structure," for the periodic crystal potential, except that the edge of the Brillouin zone in the k_z direction (*mini-zone*) lies at $\pm\pi/l$ with $l = l_1 + l_2$. Typically, l is on the order of 5. . .50 lattice constants; thus the mini-zone is only a small fraction (1/5 . . . 1/50) of the Brillouin zone of the host lattice. In contrast, the Brillouin zone in the k_x and k_y directions extends to the full width $\pm\pi/a$ and $\pm\pi/b$. There are no mini-gaps in the x and y directions.

Figure 35 shows the computed widths of mini-bands and intermittent gaps as a function of the period length for a symmetrical well/barrier structure with a barrier

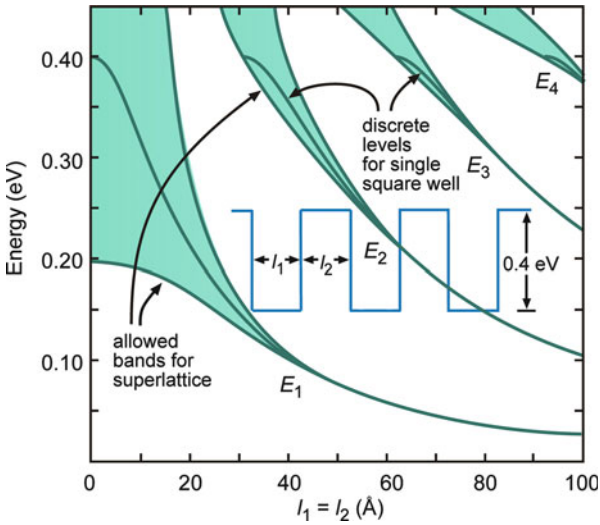


Fig. 35 Computed mini-bands for a symmetrical superlattice (After Esaki 1985)

height of 0.4 eV.[18] For $l_1 = l_2 = 40$ Å, the lowest band is rather narrow and lies at 100 meV above the well bottom. The second band extends from 320 to 380 meV. Higher bands (E_3, E_4 ...) overlap above the top of the barrier. Increasing the thickness or the height of the barrier layer reduces the tunneling through the barriers. The electronic eigenvalues within each separated well can be estimated from Eq. 54 putting $L_z = l_1$ and considering the effect of finite barriers shown in Fig. 31. The dependence of the energy of these states on the quantum-well thickness ($E_n \propto 1/l_1^2$) was verified from optical absorption, providing a nice confirmation of the quantum-mechanical model. The width of these levels is lifetime broadened and can be estimated from the uncertainty relation

$$\Delta E_n \ \Delta t \cong \frac{\Delta E_n \ l_1}{v_{\text{rms}}} T_e \cong \hbar, \tag{58}$$

with the tunneling probability T_e through such barriers (Sect. 2.3 of ▶ Chap. 29, "Carrier Generation," and Sect. 2.3 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors") given by

$$T_e \propto \exp\left(-\sqrt{\frac{8m_n}{\hbar^2}(\Delta E_c \ - E_n)} \ l_2\right), \tag{59}$$

where ΔE_c is the barrier height and E_n is the energy of the level from which tunneling takes place.

[18]This is a typical band offset at heterointerfaces; for the important GaAs/$Ga_xAl_{1-x}As$ interface, 62% of the x-dependent bandgap discontinuity ΔE_g form the discontinuity at the conduction-band edge ΔE_c and 38% at the valence-band edge ΔE_v (Watanabe et al. 1985).

A more sophisticated approach in dealing with superlattices of various dimensions resorts to a quantum-mechanical description of the periodic superlattice as given, for example, by Schulman and McGill (Schulman and McGill 1981); see also the review by Bastard and Brum (1986).

Quantum wells and superlattices that show a beautiful illustration of quantum-mechanical behavior have been fabricated from a number of semiconductor pairs that have little lattice mismatch. The best studied is the GaAs/$Ga_xAl_{1-x}As$ couple. The height of the barrier can be changed by varying the concentration of Al in $Ga_xAl_{1-x}As$; typically, it is a few tenths of an eV (Cho 1971; Woodall 1972; Chang et al. 1973).

3.1.3 Ultrathin Superlattices

As the width of the layers in superlattices becomes thinner and thinner, the superlattice structure finally disappears and is replaced by the electronic structure of a single compound. A distinction between a true superlattice and a bulk semiconductor can be made when all bandgaps between mini-bands disappear and the density of states increases monotonically from the band edge into the band.

For instance, stacking single layers of GaAs and AlAs in the [100] direction results in a $(GaAs)_1$–$(AlAs)_1$ structure identical to bulk $GaAlAs_2$. Consequently, the band structure must be the same. Extending this discussion to the symmetrical $(GaAs)_n$–$(AlAs)_n$ superlattices, Batra et al. (1987) have shown that for $n \geq 3$ the band alignment of the valence band becomes staggered with hole confinement in GaAs. However, electrons are confined in AlAs for $3 < n < 10$; only for $n > 10$ are electrons also confined in GaAs, as is expected for thicker superlattices with near-bulk gap properties for each layer – see also (Kamimura and Nakayama 1987). For Si_nGe_m ultrathin superlattices, see Pearsall et al. (1987) and Froyen et al. (1987).

3.2 Dimensionality of the Density of States

The electronic density of states $g(E)$ is obtained from the number of electron states per volume and per energy interval, $g = 2 \times (1/V)\,(dN/dE)$, the factor 2 accounting for the spin degeneracy (see also Sect. 4.2 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"). dN is obtained from the volume in $\mathbf{k}$ space (where states are equally spaced) between two planes of constant energy at E and at $E + dE$, respectively. The volume apparently depends on the dimension of the considered electronic system and so does $g(E)$. We consider an isotropic parabolic energy band where

$$E(k) - E_0 \propto k^2 \tag{60}$$

applies; see Fig. 36a. E_0 represents the energy of the band edge E_c (or E_v for holes as carriers) in the three-dimensional case and that of a confined state in lower dimensions.

For a *one-dimensional band*, we obtain from Eq. 60

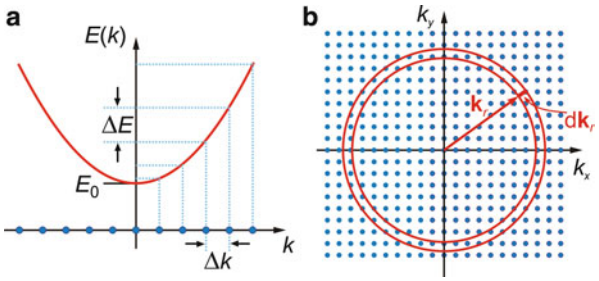


Fig. 36 (**a**) Energy of carriers in an isotropic parabolic energy band. States are equidistant along the k axis. (**b**) Two-dimensional **k** space with equally spaced states along the k_x and k_y axes. The *red ring* with radius k_r and thickness $d\mathbf{k}_r$ comprises states of equal energy $E(k_r)$

$$k \propto \sqrt{E(k) - E_0}, \text{ and } dk \propto \frac{dE}{\sqrt{E(k) - E_0}},$$

yielding

$$g_{1D}(E) \propto \frac{1}{\sqrt{E(k) - E_0}}. \tag{61}$$

The proportionality factor is $\sqrt{m^*/\left(2\pi^2\hbar^2\right)}$; hence, $g_{1D}(E)$ is expressed in units of $m^{-1} \times J^{-1}$ or $cm^{-1} \times eV^{-1}$.

In a *two-dimensional band*, all states of equal energy between E and $E + dE$ lie in a ring of radius $k_r \propto \sqrt{E(\mathbf{k}) - E_0}$ and thickness dk as illustrated in Fig. 36b; for their number $dN \propto k_r\, dk$ applies. Since $E(\mathbf{k}) - E_0 \propto k_r^2$ we obtain a constant density of states

$$g_{2D}(E) \propto \theta(E - E_0), \tag{62}$$

$\theta(E - E_0)$ being the unit step function (i.e., 0 for $E < E_0$ and 1 for $E \geq E_0$). The factor is $m^*/(\pi\hbar^2)$, yielding $g_{2D}(E)$ in units of $cm^{-2} \times eV^{-1}$.

In a *three-dimensional band*, the states of equal energy between E and $E + dE$ are contained in a shell of radius k_r and thickness dk with a number $dN \propto k_r^2 dk$, yielding

$$g_{3D}(E) \propto \theta(E - E_0) \times \sqrt{E - E_0}. \tag{63}$$

The proportionality factor is $(2\pi^2)^{-1}(2m^*/\hbar^2)^{3/2}$; consequently $g_{3D}(E)$ is expressed in units of $cm^{-3} \times eV^{-1}$.

Equations 61 through 63 can be combined in the relation

$$g(E) \propto (E - E_0)^{d/2-1}, \tag{64}$$

with the dimensionality $d = 1$, 2, or 3. The dependence of g on the energy is illustrated in Fig. 37 for various dimensionalities $d \leq 3$ and three confined levels.

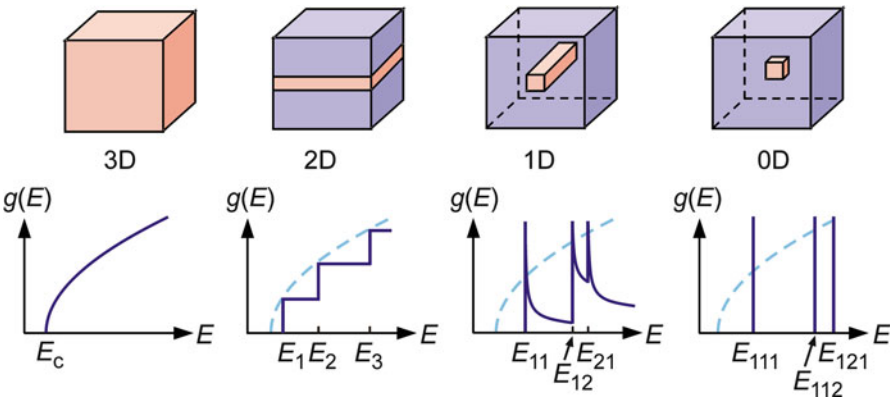


Fig. 37 Electronic density of states $g(E)$ in isotropic semiconductors (*red*), clad by barriers (*blue*) which reduce the dimensionality: 3D bulk semiconductor, 2D quantum well, 1D quantum wire, and 0D quantum dot. The dashed curve represents the 3D case drawn for reference (Note: its unit differs)

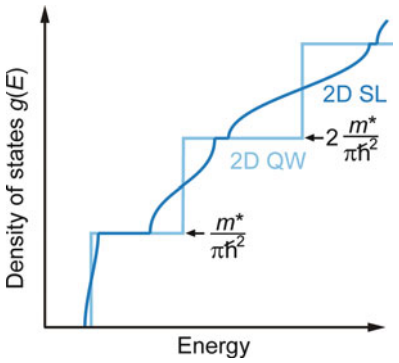


Fig. 38 Two-dimensional density of states for electrons in a quantum well (2D QW) and in mini-bands of a superlattice (2D SL). The mini-bands extend in the energy ranges of the S-shaped segments around the steps of the QW-related density of states

We note that the dimensionality has a strong effect on the density of states. In particular, $g(E)$ increases strongly at the lowest occupiable energy (band edge) as the dimensionality decreases. This feature makes low-dimensional structures particularly interesting for applications with carriers in or near the ground state (see, e.g., Asada et al. 1986).

The 2D density of states of a quantum well and of a superlattice differ. When proceeding from the level structure of an isolated quantum well to mini-bands in a superlattice with sufficiently permeable barriers, we can follow the broadening of the energy level in a given state. $g(E)$ has a staircase character for the isolated well with equidistant steps of the height $\Delta g = m^*/(\pi\ \hbar^2)$ as shown in Fig. 38. Each level can be occupied by the number of electrons given by its degeneracy multiplied by the number of atoms in the wells. When significant tunneling becomes possible, each level splits into bands, and the staircase behavior (light-blue steps) becomes

somewhat softened (Fig. 38). The effective density of states near the bottom of the first mini-band (for thin enough barriers, so that the mini-band width is less than several kT) is given by $N_{\text{mini}} = m^* kT/(\pi \hbar^2)$, measured in cm^{-2}.

3.3 Quantum Wires

While a quantum well follows naturally from the epitaxy of a double heterostructure, a further reduction of dimensionality toward a one-dimensional quantum wire or a zero-dimensional quantum dot requires some patterning to define an additional *lateral* confinement. The interface-to-volume ratio of 1D and 0D structures increases as compared to 2D quantum wells; the electronic properties of such structures are therefore sensitive to interface defects, usually precluding fabrication by patterning a quantum well using lithography techniques. Techniques employed instead are outlined in Sect. 2.2 of ▶ Chap. 3, "The Structure of Semiconductors." Most epitaxial techniques lead to complicate and often only weak lateral confinement potentials, and early studies suffered from thickness fluctuations on a length scale of the exciton Bohr radius, leading to 0D behavior; peculiarities of bound states in quantum wires are reviewed by Akiyama (1998) and Hurt (2000).

The basic changes of eigenvalues and density of states occurring for a reduction of the dimensionality from 3D bulk to a 2D quantum well were outlined above (Sects. 3.1 and 3.2) and apply also for a further reduction to a 1D quantum wire. The additional confinement $W(y)$ leads to a quasi-free mobility only along the x axis. We again assume infinite barriers and the parabolic approximation. Analogous to Eq. 55 the eigenvalues of the wavefunction are then given by the sum of that for the unrestricted carrier motion along the x direction and those for the confined y and z directions, yielding

$$E = E_{l,n} = \frac{\hbar^2 k_x^2}{2m_x^*} + l^2 \frac{\hbar^2}{2m_y^*} \left(\frac{\pi}{L_y}\right)^2 + n^2 \frac{\hbar^2}{2m_z^*} \left(\frac{\pi}{L_z}\right)^2, \tag{65}$$

where L_y and L_z are the (independent) extensions of the wire along the respective axes, m_x^* and m_y^* are the corresponding effective masses, and l, n are the integer quantum numbers defining the subbands. The actual geometry and finite depth of the potential lead to some modifications of this dependence, but the general trend is preserved (for the effect of a finite potential on a quantum well, see Fig. 31). The quantum-mechanical behavior was nicely demonstrated for V-shaped (Vouilloz et al. 1997, 1998) and T-shaped quantum wires (Wegscheider et al. 1996) fabricated from GaAs/$Ga_xAl_{1-x}As$ heterostructures. See also Sect. 3 of ▶ Chap. 13, "Band-to-Band Transitions," and Sect. 2 of ▶ Chap. 14, "Excitons."

3.4 Quantum Dots and Nanocrystals

A quantum dot represents the ultimate limit of carrier confinement; the three-dimensional confinement restricts the carrier motion in all three spatial dimensions and leads to fully quantized electron and hole states, similar to the discrete states in an atom. Assuming again infinite barriers, we obtain the eigenvalues of the wavefunction analogous to Eqs. 55 and 65,

$$E = E_{j,l,n} = \frac{\hbar^2 \pi^2}{2} \left(\frac{j^2}{m_x^* L_x} + \frac{l^2}{m_y^* L_y} + \frac{n^2}{m_z^* L_z} \right). \tag{66}$$

The zero-dimensional density of states is a sum of δ functions given by $g(E) = \Sigma\, 2\, \delta(E\text{-}E_{j,l,n})$, with the factor 2 accounting for the spin degeneracy. The function is shown in Fig. 37 for the three energy levels of lowest energy. Similar to the two-dimensional case, the peaks are not equidistant and not degenerate if the products $m^* \times L$ differ along the spatial directions.

With finite barriers a minimum size of a quantum dot is required to allow for confining a carrier, in contrast to structures of higher dimensionality. For a dot with spherical shape, the minimum diameter $D_{\min}$ required to confine at least one bound state is given by (Bimberg et al. 1999)

$$D_{\min} = \frac{\pi \ \hbar}{\sqrt{2 m^* W_0}}, \tag{67}$$

where W_0 is the confining potential and m^* the effective mass (assumed to be isotropic and identical in dot and barrier). For a rough estimate of the minimum size to confine a single electron in a spherical InAs/GaAs dot, we use a conduction-band offset of ~0.9 eV for W_0 and an effective electron mass in InAs of 0.03 m_0, yielding $D_{\min} \cong 6$ Å. The finite barrier height provided by the matrix environment of the quantum dot usually allows for only few confined electronic states.

Electrons *and* holes are confined to quantum dots with type I band alignment; therefore, the binding energy of this electron–hole pair (exciton) is increased compared to that of the bulk material (see Sect. 3 of ▶ Chap. 14, "Excitons").

The wavefunctions of the confined states can be calculated by considering realistic size, shape, and composition of the dot, in addition to material properties like dielectric constants, strain tensors, and piezoelectric tensors (Schliwa et al. 2007). Results for one electron and one hole confined in an epitaxially grown InAs dot with pyramidal shape of 11.3 nm base length and {101} side facets in a GaAs matrix (Sect. 2.2.3 of ▶ Chap. 3, "The Structure of Semiconductors") are shown in Fig. 39. The iso-surfaces encase 65% probability and resemble atomic *s*-like ground states (left column) and *p*- and *d*-like excited states. Corresponding wavefunctions of single-electron states in uncovered InAs/GaAs dots were experimentally imaged using a low-temperature scanning-tunneling microscope. The images were obtained from spatially resolved differential voltage–current curves

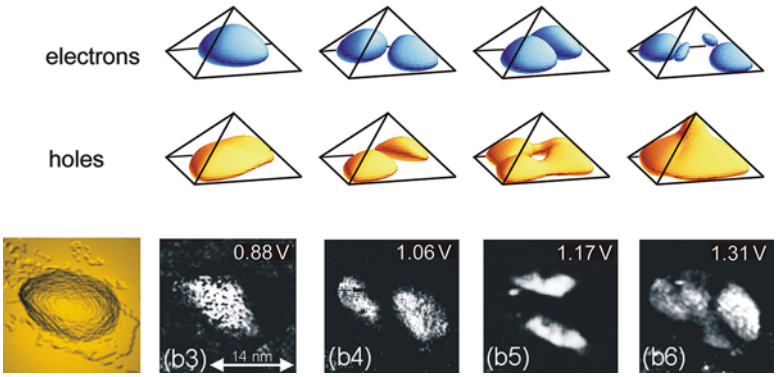


Fig. 39 *Top*: Calculated probability densities of electron and hole wavefunctions confined in a pyramid-shaped InAs quantum dot in a GaAs matrix (After Stier et al. 1999). *Bottom*: Low-temperature STM images of an uncovered InAs quantum dot on GaAs. *Left image*: Constant-current image showing the dot shape. *Right four images*: Single-electron densities of different excited states sampled at different bias voltage (After Maltezopoulos et al. 2003)

dI/*dV* taken at different sample voltage (Maltezopoulos et al. 2003). In the same way states of *holes* confined in InAs quantum dots embedded in GaAs matrix were imaged from cleaved samples (Urbieta et al. 2008).

The electronic structure of very small isolated crystallites (nanocrystals), e.g., as a suspension in a liquid dielectric or embedded in glass or polymer matrices (Fig. 26 of ▶ Chap. 3, "The Structure of Semiconductors"), may be compared with a zero-dimensional quantum dot with infinite barriers; the barrier height is generally in the eV range. The bandgap increases compared to that of the bulk when the crystallite size decreases below a few hundred Å. For reviews see, e.g., (Woggon 1997; Scholes 2008).

Often such nanocrystals have a spherical shape. In this case three regimes of quantization are usually distinguished, depending on the ratio of the nanocrystal radius R with respect to the Bohr radius of the confined exciton or with respect to the orbit radii of electron and hole in the exciton (around the common center of mass); the regimes are also referred to as weak, intermediate, and strong confinement (Klingshirn 2012).

For an estimate of the quantization energy, we assume again isotropic parabolic bands. *Weak confinement* is characterized by $R \geq a_B$, yielding a quantization energy

$$E_{\text{weak}} \cong \frac{\hbar^2 \pi^2}{2MR^2}, \tag{68}$$

where $M = m_e + m_h$ is the mass of the exciton. The relation expresses a quantization of the center-of-mass motion of the exciton, while relative motion of electron and hole is hardly affected because the nanocrystal is larger than the exciton. In the *intermediate regime* the nanocrystal radius lies between the Bohr radii of electron and hole, $a_{B,e} \geq R \geq a_{B,h}$, yielding

$$E_{\text{intermediate}} \cong \text{const} \times \text{Ry}^* \left(\frac{a_B \pi}{R}\right)^2. \tag{69}$$

Ry* is the effective Rydberg constant, and the constant is of the order unity. At this medium confinement the confinement energy is of the same order as the Rydberg energy, while it is smaller in the weak and larger in the strong confinement regime. *Strong confinement* occurs if the nanocrystal radius is significantly smaller than the electron and hole Bohr radii, $R \ll a_{B,e}, a_{B,h}$. The quantization energy can be roughly approximated as

$$E_{\text{strong}} \cong \frac{\hbar^2 \pi^2}{2\mu R^2}, \tag{70}$$

μ being the reduced mass of electron and hole.

The estimates above neglect the exciton binding energy (which depends on the nanocrystal size) and the contribution of the Coulomb energy; the latter scales with $1/R$, while the quantization energy scales as $1/R^2$. A minor correction term $-\text{const} \times e^2/(4\pi\varepsilon\varepsilon_0 R)$ may be included for a better description in the intermediate and strong confinement regimes; still other factors like effects of the confinement potential or limited validity of the effective-mass approach may be more important.

The increase of the bandgap energy by quantization is illustrated for CdSe nanocrystals in Fig. 40. The graph gives the measured energy of the lowest confined state, compared to an effective-mass calculation which applied a fit to the Luttinger parameters (for VB nonparabolicity) and the potential energy for the electron (Norris and Bawendi 1996). The trend – implying also excited states – is reasonably well reproduced by the simple approach, though some discrepancy in the curvature and at large sizes occurs.

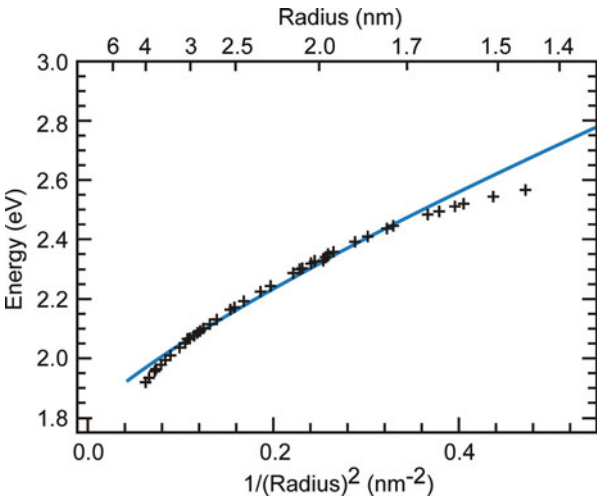


Fig. 40 Measured and calculated energy of the lowest state of electron–hole pairs in CdSe nanocrystal quantum-dots as a function of the radius (After Norris and and Bawendi 1996)

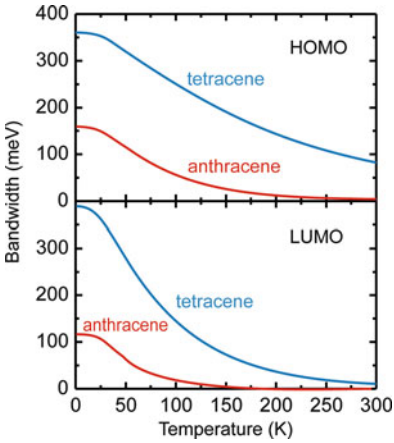


Fig. 41 Temperature-dependent narrowing of the LUMO and HOMO bands in anthracene and tetracene crystals (After Hannewald et al. 2004)

4 Bands in Organic and Amorphous Semiconductors

4.1 Bands and Bandgap in Organic Semiconductors

The charge transport in organic semiconductor crystals and highly organized thin films can often be described in a band-like regime similar to that in inorganic solids. The lowest unoccupied and the highest occupied molecular-orbital (LUMO and HOMO) bands are then also termed valence and conduction bands, and the mobile polaron[19] carriers in these bands are referred to as electrons and holes.

The *width of the conduction and valence bands* is generally small; it is determined by the weak intermolecular interaction.[20] The bandwidth given for some organic semiconductors in Table 1 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals," is maximum at $T \rightarrow 0$. At increasing temperature the bandwidth is progressively reduced (Fig. 41) by an increased electron–phonon coupling and therefore an enhancement of the polaron mass. The carriers get more and more localized over single molecules (or molecule chains), leaving only transport by a thermally activated hopping mechanism. The transport thus gradually changes from coherent band-like motion at low temperatures to phonon-assisted hopping transport at high temperature. This transition occurs in oligoacenes at about room temperature (Cheng et al. 2003).

[19]For the concept of polaron quasiparticles, see Sect. 1.2 of ▶ Chap. 22, "Carrier-Transport Equations."

[20]The intermolecular interaction is expressed in terms of a *transfer integral* (see the chapters on defects and on transport controlled by defects later in the book). In the tight-binding approximation of a one-dimensional molecule chain, the total bandwidth equals four times the transfer integral between neighboring molecules. The bandwidth for any molecular packing can be expressed from the amplitude of the transfer integrals between the various interacting units (Brédas et al. 2002).

The width of the bands illustrated in Fig. 41 is intimately connected with the temperature-dependent effective mass of electrons and holes. The minimum limit at low temperature is obtained from the curvature of the HOMO and LUMO bands according to Eq. 1 resulting from band-structure calculations. Instead of effective masses generally the *mobility* is reported for organic semiconductors (Sect. 4 of ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects"). This quantity is readily accessible in experiment and usually much larger for holes than for electrons. Furthermore, it is strongly anisotropic due to the low symmetry of molecules and crystal lattices and hence the oriental dependence of the transfer integrals.[20]

The *bandgap energy* of organic solids is given by the difference between the LUMO and the HOMO energy levels. It is the relevant quantity for electrical conductivity and also referred to as transport gap $E_g^{\text{transport}}$. The value differs from the optical bandgap E_g^{opt} determined from the long-wavelength edge of absorption spectra. Organic solids are easily polarized and have large binding energies for excitons; the absorption edge reflects the formation of (Frenkel) excitons with the electron and hole on the same molecule. The charge separation energy, $E_g^{\text{transport}} - E_g^{\text{opt}}$, is the binding energy of the exciton, which is far larger than in inorganic semiconductors (Hill et al. 2000). Figure 42 shows the transport and optical bandgaps for many organic semiconductors; their difference is given by the separation of the straight and dotted lines, illustrating an increase of the exciton binding energy for larger bandgap energies.

The bandgap of organic semiconductors is generally large (typically well above 2 eV). The thermal activation of carriers is therefore negligible: pure organic solids are insulating. Since reliable and robust doping proved difficult up to now, charge carriers (both electrons and holes) are injected from the contacts in all organic semiconductors used for devices.

The energy of the transport gap $E_g^{\text{transport}}$ can be directly obtained from electron spectroscopy: ultraviolet photoelectron spectroscopy (UPS) measures the HOMO energy, and inverse photoelectron spectroscopy (IPES) determines the LUMO energy. Data for some organic crystals are given in Table 11.

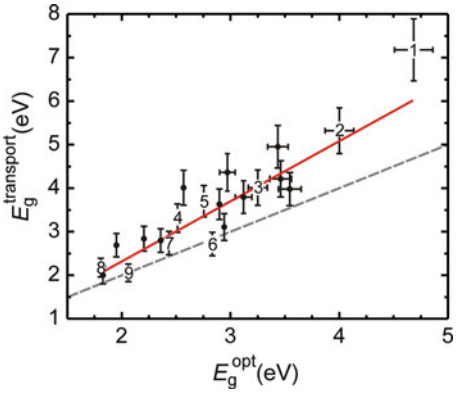


Fig. 42 Transport-bandgap energy $E_g^{\text{transport}}$ versus optical bandgap energy E_g^{opt} for 24 organic semiconductors. *Numbers at data points* refer to the compounds listed in Table 11; the *solid line* signifies a linear fit to the data, and the *dashed straight line* represents $E_g^{\text{transport}} = E_g^{\text{opt}}$ (After Djurovich et al. 2009)

Table 11 LUMO and HOMO energies for organic semiconductor compounds measured by photoelectron spectroscopy (IPES and UPS, respectively), transport bandgap $E_g^{transport} = E^{LUMO} - E^{HOMO}$, and measured optical bandgap E_g^{opt}. The numbers in the first column refer to those in Fig. 42 (Data from Djurovich et al. 2009)

No	Compound	E^{LUMO} (eV)	E^{HOMO} (eV)	$E_g^{transport}$ (eV)	E_g^{opt} (eV)
1	Benzene	−0.4	−7.58	7.2	4.68
2	Naphthalene	−1.1	−6.4	5.3	4.00
3	Anthracene	−1.7	−5.70	4.0	3.25
4	Tetracene	−1.8	−5.10	3.3	2.51
5	Alq3	−1.96	−5.65	3.69	2.75
6	Perylene	−2.5	−5.2	2.7	2.83
7	α-Hexathiophene	−2.57	−5.3	2.7	2.43
8	CuPc	2.65	−4.82	2.17	1.80
9	Pentacene	−2.8	−4.85	2.1	2.06

4.2 Bands in Amorphous Semiconductors

The band structure in amorphous semiconductors cannot be determined in the same fashion as for the crystalline state (described in ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals") since long-range periodicity is missing and **k** is no longer a good quantum number. Therefore, an $E(\mathbf{k})$ diagram cannot be drawn for amorphous materials. Consequently, the effective-mass picture, which depends on an analysis of $E(\mathbf{k})$, cannot be used in its classical form.

Nevertheless, there is a strong experimental evidence from optical absorption, reflection, and photoemission spectroscopy that similarities exist between the band structure in amorphous and crystalline states of the same material. This can be understood by recognizing that the major features of the electronic properties of a solid are determined by short-range order, as proposed by Ioffe and Regel (1960) and shown more convincingly by Weaire and Thorpe (1971) for amorphous Si and Ge, using a tight-binding model.

Density-of-state distributions for amorphous and crystalline Ge are shown in Fig. 43 – as obtained experimentally (a) and compared with the theoretical distribution (b). The distribution of the amorphous state is much smoother. The sharp van Hove singularities due to long-range order are absent, while several of the major features remain in both states.

The more refined tight-binding calculation of Bullett and Kelly (1975) shows a substantially improved agreement with the experiment. More recent calculations employ density functional methods and molecular dynamics simulations (Prasai et al. 2016). For general reviews on the subject see Connell and Street (1980), Mott and Davis (1979), and Robertson (1983); see also Sect. 4 of ▶ Chap. 13, "Band-to-Band Transitions," and Sect. 2 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors").

The properties near the band edges are more sensitive to the actual amorphous structure. Therefore, it is no longer possible to describe the dispersion relation and the level distribution here in general terms. There seems to be a rather smooth

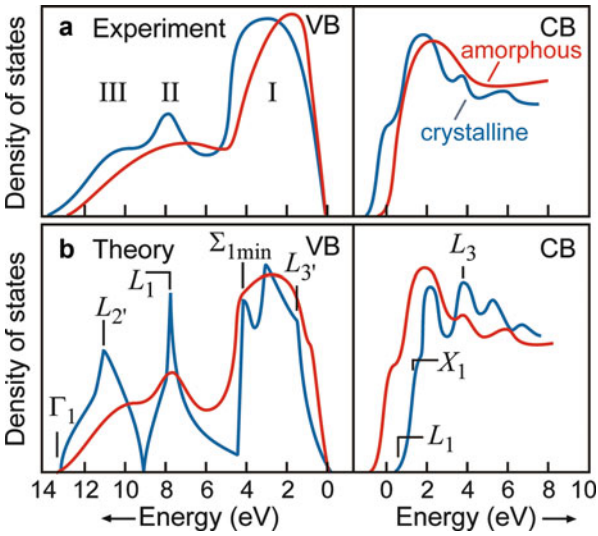


Fig. 43 Electron density-of-state distributions of amorphous and crystalline Ge. (**a**) Experimental data derived from photoemission, (**b**) theory (After Economou et al. 1974)

transition between extended (band) states and localized states due to the lack of long-range order. This does not permit a cohesive discussion of the band-edge behavior similar to the discussion for crystalline semiconductors. On the other hand, the measurement of transport properties suggests the existence of an edge. This will be analyzed later in Sect. 4 of ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects."

5 Summary

For semiconducting purposes, only the band structure of conduction and valence bands near the edge to the bandgap is of direct interest. This structure is well investigated for important semiconductors. It has nearly spherical equi-energy surfaces for the conduction band at Γ and ellipsoids of rotation at X and L. The lowest minimum of $E(\mathbf{k})$ of the conduction bands and the highest maximum of $E(\mathbf{k})$ of the valence bands determine the bandgap. In the valence band of cubic semiconductors, two subbands need to be considered; the light- and heavy-hole bands are degenerated at Γ in isotropic semiconductors and split by the crystal field in anisotropic ones; in hexagonal semiconductors three nondegenerate subbands occur at Γ, originating from light, heavy, and crystal holes. The spin–orbit split-off band may additionally become important at higher hole energy, especially for low atomic number elements where such a splitting is relatively small. The $E(\mathbf{k})$ behavior of the bands near the band edges is described by the effective mass of electrons or holes and is directly accessible to cyclotron-resonance measurements.

The band structure and the bandgap are influenced by external parameters. The bandgap usually shrinks with increasing temperature and expands with increasing pressure. It can be changed by alloying, where, for alloying with similar elements,

the gap interpolates linearly between the value of the pure compounds. With dissimilar element alloying, a substantial bowing is observed. The bandgap is also changed (decreased) by heavy doping.

In low-dimensional semiconductors anisotropic carrier confinement occurs and energy levels get quantized to discrete values along the axes of restricted carrier motion. The quantization energy is inverse to the effective carrier mass and approximately inverse to the square of the confinement length for a square potential. Bands along unrestricted carrier motion remain essentially unaffected; different quantized states along the confined directions create subbands along the unrestricted axes. In superlattices, the coupling of adjacent quantum wells leads to mini-bands. The effective gap and bandwidth of mini-bands can be varied easily by changing layer width and barrier height of the superlattice. With reduced layer width, the electronic behavior of such superlattices approaches that of simple chemical compounds, assuming bulk properties. The electronic density of states changes strongly as the dimensionality is reduced; it gradually increases at lowest energy when going from a three-dimensional bulk crystal to a two-dimensional quantum well, a one-dimensional quantum wire, and eventually to a zero-dimensional quantum dot.

In organic semiconductors conduction and valence bands arise from the lowest unoccupied and the highest occupied molecular orbitals LUMO and HOMO. The bandwidth is typically small and narrows as the temperature increases, and the bandgap is large. At increased temperature band-like conduction gradually changes to hopping transport.

The bands in amorphous semiconductors near the band edge are ill-defined, since **k** is no longer a good quantum number due to missing long-range periodicity. Nevertheless, the density-of-state distribution shows significant similarities to that of the same material in the crystallite state. Near the band edge, a smooth transition between extended and localized states occurs.

References

Abram RA, Res GJ, Wilson BLH (1978) Heavily doped semiconductors and devices. Adv Phys 27:799

Aggarwal RL, Zuteck MD, Lax B (1969) Magnetopiezotransmission studies of the indirect transition in germanium. Phys Rev 180:800

Akiyama H (1998) One-dimensional excitons in GaAs quantum wires. J Phys Condens Matter 10:3095

Ando T, Fowler AB, Stern F (1982) Electronic properties of two-dimensional systems. Rev Mod Phys 54:437

Antončik E (1955) On the theory of temperature shift of the absorption curve in non-polar crystals. Czech J Phys 5:449

Asada M, Miyamoto Y, Suematsu Y (1986) Gain and threshold of three-dimensional quantum-box lasers. IEEE J Quantum Electron QE 22:1915

Bastard G, Brum JA (1986) Electronic states in semiconductor heterostructures. IEEE J Quantum Electron QE 22:1625

Batra IP, Ciraci S, Nelson JS (1987) Confined states and stability of GaAs–AlAs superlattices. J Vac Sci Technol B 5:1300

Berggren K-F, Sernelius BE (1984) Intervalley mixing versus disorder in heavily doped *n*-type silicon. Phys Rev B 29:5575
Bernard JE, Zunger A (1987) Electronic structure of ZnS, ZnSe, ZnTe, and their pseudobinary alloys. Phys Rev B 36:3199
Bimberg D, Grundmann M, Ledentsov NN (1999) Quantum dot heterostructures. J. Wiley, Chichester
Bir GL, Pikus GE (1974) Symmetry and strain-induced effects in semiconductors. Wiley, New York
Blakemore JS (1982) Semiconducting and other major properties of gallium arsenide. J Appl Phys 53:R123
Bludau W, Onton A, Heinke W (1974) Temperature dependence of the band gap of silicon. J Appl Phys 45:1846
Brédas JL, Calbert JP, da Silva Filho DA, Cornil J (2002) Organic semiconductors: a theoretical characterization of the basic parameters governing charge transport. Proc Natl Acad Sci U S A 99:5804
Brooks H (1955) Theory of the electrical properties of germanium and silicon. Adv Electron Electron Phys 7:85
Bullett DW, Kelly MJ (1975) The valence bands of model amorphous semiconductor structures. Solid State Commun 16:1379
Chang LL, Esaki L, Howard WE, Ludeke R (1973a) The growth of a GaAs–GaAlAs Superlattice. J Vac Sci Technol 10:11
Chang LL, Esaki L, Howard WE, Ludeke R, Schul G (1973b) Structures grown by molecular beam epitaxy. J Vac Sci Technol 10:655
Cheng YC, Silbey RJ, da Silva Filho DA, Calbert JP, Cornil J, Brédas JL (2003) Three-dimensional band structure and bandlike mobility in oligoacene single crystals: a theoretical investigation. J Chem Phys 118:3764
Cho AY (1971) Growth of periodic structures by the molecular-beam method. Appl Phys Lett 19:467
Chuang SL, Chang CS (1996) Effective-mass Hamiltonian for strained wurtzite GaN and analytical solutions. Appl Phys Lett 68:1657
Cohen ML (1962) Electron–phonon self-energies in many-valley semiconductors. Phys Rev 128:131
Collins RT, Vina L, Wang WI, Mailhiot C, Smith DL (1987) Electronic properties of quantum wells in perturbing fields. Proc SPIE 792:2
Connell GAN, Street RN (1980) Amorphous semiconductors. In: Moss TS, Keller SP (eds) Handbook on semiconductors, vol 3, Materials properties and preparation. North Holland, Amsterdam, pp 689–801
Cracknell AP, Wong KC (1973) The fermi surface: its concepts, determination, and use in the physics of metals. Oxford University Press, Oxford
Dexter RN, Zeiger HJ, Lax B (1956) Cyclotron resonance experiments in silicon and germanium. Phys Rev 104:637
Dimmock JO, Melngailis I, Strauss AJ (1966) Band structure and laser action in $Pb_xSn_{1-x}Te$. Phys Rev Lett 16:1193
Dingle R (1975) Confined carrier quantum states in ultrathin semiconductor heterostructures. In: Queisser HJ (ed) Festkörperprobleme, advances in solid state physics, vol 15. Pergamon/ Vieweg, Braunschweig, p 21
Dingle R, Wiegmann W, Henry CH (1974) Quantum states of confined carriers in very thin $Al_xGa_{1-x}As$–GaAs–$Al_xGa_{1-x}As$ heterostructures. Phys Rev Lett 33:827
Djurovich PI, Mayo EI, Forrest SR, Thompson ME (2009) Measurement of the lowest unoccupied molecular orbital energies of molecular organic semiconductors. Org Electron 10:515
Dresselhaus G, Kip AF, Kittel C (1955) Cyclotron resonance of electrons and holes in silicon and germanium crystals. Phys Rev 98:368
D'yakonov MI, Perel' VI (1971) On spin orientation of electrons in interband absorption of light in semiconductors. Sov Phys JETP 33:1053
Economou N, Cohen MH, Freed KF, Kirkpatrick ES (1974) Electronic structure of disordered materials. In: Tauc J (ed) Amorphous and liquid semiconductors. Plenum Press, New York, p 101

Edwards AL, Slykhouse TE, Drickamer HG (1959) The effect of pressure on zinc blende and wurtzite structures. J Phys Chem Solids 11:140
Esaki L (1985) Compositional superlattices. In: Parker EMC (ed) The technology and physics of molecular beam epitaxy. Plenum Press, New York, p 143
Esaki L, Tsu R (1970) Superlattice and negative differential conductivity in semiconductors. IBM J Res Dev 14:61
Fan HY (1951) Temperature dependence of the energy gap in semiconductors. Phys Rev 82:900
Ferhat M, Bechstedt F (2002) First-principles calculations of gap bowing in $In_xGa_{1-x}N$ and $In_xAl_{1-x}N$ alloys: relation to structural and thermodynamic properties. Phys Rev B 65:075213
Froyen S, Wood DM, Zunger A (1987) Structural and electronic properties of epitaxial thin-layer Si_nGe_n superlattices. Phys Rev B 37:6893
Gospodarič J, Dziom V, Shuvaev A, Dobretsova AA, Mikhailov NN, Kvon ZD, Novik EG, Pimenov A (2020) Band structure of a HgTe-based three-dimensional topological insulator. Phys Rev B 102:115113
Graziosi P, Neophytou N (2020) Ultra-high thermoelectric power factors in narrow gap materials with asymmetric bands. J Phys Chem C 124:18462
Hannewald K, Stojanovi VM, Schellekens JMT, Bobbert PA, Kresse G, Hafner J (2004) Theory of polaron bandwidth narrowing in organic molecular crystals. Phys Rev B 69:075211
Hansen GL, Schmit JL, Casselman TN (1982) Energy gap versus alloy composition and temperature in $Hg_{1-x}Cd_xTe$. J Appl Phys 53:7099
Hayes W, Stoneham AM (1984) Defects and defect processes in nonmetallic solids. Wiley, New York
Heine V, van Vechten JA (1976) Effect of electron–hole pairs on phonon frequencies in Si related to temperature dependence of band gaps. Phys Rev B 13:1622
Hill IG, Kahn A, Soos ZG, Pascal RA Jr (2000) Charge-separation energy in films of π-conjugated organic molecules. Chem Phys Lett 327:181
Hurt NE (2000) Mathematical physics of quantum wires and devices – from spectral resonances to Anderson localization. Kluwer, Dordrecht
Hume-Rothery W (1936) The structure of metals and alloys. The Institute of Metals, London
Ioffe AF, Regel AR (1960) Non-crystalline, amorphous, and liquid electronic semiconductors. Progr Semicond 4:237
Jaffe JE, Zunger A (1984) Electronic structure of the ternary pnictide semiconductors $ZnSiP_2$, $ZnGeP_2$, $ZnSnP_2$, $ZnSiAs_2$, and $MgSiP_2$. Phys Rev B 30:741
Jain SC, Roulston DJ (1991) A simple expression for band gap narrowing (BGN) in heavily doped Si, Ge, GaAs and Ge_xSi_{1-x} strained layers. Sol State Electron 34:453
Jaros M (1985) Electronic properties of semiconductor alloy systems. Rep Prog Phys 48:1091
Jen HR, Cherng MJ, Stringfellow GB (1986) Ordered structures in $GaAs_{0.5}Sb_{0.5}$ alloys grown by organometallic vapor phase epitaxy. Appl Phys Lett 48:1603
Jeon J-B, Sirenko YM, Kim KW, Littlejohn MA, Stroscio MA (1996) Valence band parameters of wurtzite materials. Sol State Commun 99:423
Kalt H (1996) Optical properties of III–V semiconductors – the influence of multi-valley band structures, vol 120, Springer series in solid-state sciences. Springer, Berlin
Kamimura H, Nakayama T (1987) Comments Cond Mat Phys 13:143
Kane EO (1957) Band structure of indium antimonide. J Phys Chem Sol 1:249
Kane EO (1979) Band structure of narrow gap semiconductors. In: Zawadzki W (ed) Proceedings of the international summer school physics narrow gap semiconductor, Nimes. Springer, Berlin, p 13
Klingshirn CF (2012) Semiconductor optics, 4th edn. Springer, Heidelberg/New York
Kroemer H (1983) Heterostructure devices: a device physicist looks at interfaces. Surf Sci 132:543
Lambrecht WRL, Rodina AV, Limpijumnong S, Segall B, Meyer BK (2002) Valence-band ordering and magneto-optic exciton fine structure in ZnO. Phys Rev B 65:075207

Landsberg PT (1987) The band-to-band Auger effect in semiconductors. Solid State Electron 30:1107
Landsberg PT, Neugroschel A, Lindholm FA, Sah CT (1985) A model for band-gap shrinkage in semiconductors with application to silicon. Phys Stat Sol (b) 130:255
Larach S, Shrader RE, Stocker CF (1957) Anomalous variation of band gap with composition in zinc sulfo- and seleno-tellurides. Phys Rev 108:587
Lautenschlager P, Allen PB, Cardona M (1985) Temperature dependence of band gaps in Si and Ge. Phys Rev B 31:2163
Lax B (1963) In: Smith RA (ed) Enrico Fermi school of physics semiconductors. Academic Press, New York, p 240
Luttinger JM (1956) Quantum theory of cyclotron resonance in semiconductors: general theory. Phys Rev 102:1030
Mahan GD (1997) Good thermoelectrics. Solid State Phys 51:81
Mahan GD (1980) Energy gap in Si and Ge: impurity dependence. J Appl Phys 51:2634
Maltezopoulos T, Bolz A, Meyer C, Heyn C, Hansen W, Morgenstern M, Wiesendanger R (2003) Wave-function mapping of InAs quantum dots by scanning tunneling spectroscopy. Phys Rev Lett 91:196804
Martinez G (1980) Optical properties of semiconductors under pressure. In: Moss TS, Balkanski M (eds) Handbook on semiconductors, vol 2, Optical properties of solids. North Holland, Amsterdam, pp 181–222
Martins JL, Zunger A (1986) Ordering and decomposition in semiconductor alloys. J Mater Res 1:523
Mavroides JG (1972) Magneto-optical properties. In: Abeles F (ed) Optical properties of solids. North Holland, Amsterdam
McCombe BD, Wagner RJ (1975) Intraband magneto-optical studies of semiconductors in the far infrared. Adv Electron Electron Phys 37:1
Mott NF, Davis EA (1979) Electronic processes in non-crystalline materials. Claredon Press, Oxford
Norris DJ, Bawendi MG (1996) Measurement and assignment of the size-dependent optical spectrum in CdSe quantum dots. Phys Rev B 53:16338
Ourmazd A, Bean JC (1985) Observation of order–disorder transitions in strained-semiconductor systems. Phys Rev Lett 55:765
Ousset JC, Leotin J, Askenazy S, Skolnik MS, Stradling RA (1976) Cyclotron resonance measurements of the non-parabolicity of the conduction bands in silicon and germanium. J Phys C (London) 9:2803
Pantelides ST, Selloni A, Car R (1985) Energy-gap reduction in heavily-doped silicon: causes and consequences. Solid State Electron 28:17
Pantelides ST (1978) The electronic structure of impurities and other point defects in semiconductors. Rev. Mod. Phys 50:797
Pearsall TP (ed) (1982) GaInAsP-alloy semiconductors. Wiley, New York
Pearsall TP, Bevk J, Feldman LC, Bonar JM, Mannaerts JP, Ourmazd A (1987) Structurally induced optical transitions in Ge-Si superlattices. Phys Rev Lett 58:729; Phys Rev Lett 58:1053
Phillips J (1970) Ionicity of the chemical bond in crystals. Rev Mod Phys 42:317
Pidgeon CR (1969) In: Haidemanakis ED (ed) Electron structure in solids. Plenum Press, New York, p 47
Pidgeon CR (1980) Free carrier optical properties of semiconductors. In: Balkanski M (ed) Handbook on semiconductors, vol 2, optical properties of solids. North Holland, Amsterdam, pp 223–328
Pippard AB (1965) Dynamics of conduction electrons. Gordon and Breach, London
Prasai K, Biswas P, Drabold DA (2016) Electrons and phonons in amorphous semiconductors. Semicond Sci Technol 31:073002

Richardson D (1973) An investigation of their film alloys of compound semiconductors. PhD thesis, University of Newcastle. See also: Hill R (1974) Energy-gap variations in semiconductor alloys. J Phys C 7:521

Robertson J (1983) Electronic structure of amorphous semiconductors. Adv Phys 32:361

Rontani M, Molinari E, Maruccio G, Janson M, Schramm A, Meyer C, Matsui T, Heyn C, Hansen W, Wiesendanger R (2007) Imaging correlated wave functions of few-electron quantum dots: theory and scanning tunneling spectroscopy experiments. J Appl Phys 101:081714

Rössler U (1984) Nonparabolicity and warping in the conduction band of GaAs. Solid State Commun 49:943

Saxena AK (1981) Non-Γ deep levels and the conduction band structure of $Ga_{1-x}Al_xAs$ alloys. Phys Stat Sol (b) 105:777

Schliwa A, Winkelnkemper M, Bimberg D (2007) Impact of size, shape, and composition on piezoelectric effects and electronic properties of In(Ga)As/GaAs quantum dots. Phys Rev B 76:205324

Schmit JL, Stelzer EL (1973) Semiconductor opto-electronics. Butterworths, Stoneham, p 398

Scholes GD (2008) Controlling the optical properties of inorganic nanoparticles. Adv Funct Mater 18:1157

Schulman JN, McGill TC (1981) Complex band structure and superlattice electronic states. Phys Rev B 23:4149

Smith SD (1967) Magneto-Optics in crystals. In: S Flügge (ed) Handbuch der Physik vol XXV, 2a, Springer, Berlin/Heidelberg/New York, p 234

Srivastava GP, Martins JL, Zunger A (1986) Atomic structure and ordering in semiconductor alloys. Phys Rev B 31:2561

Stier O, Grundmann M, Bimberg D (1999) Electronic and optical properties of strained quantum dots modeled by 8-band kp theory. Phys Rev B 59:5688

Suzuki M, Uenoyama T, Yanase A (1995) First-principles calculations of effective-mass parameters of AlN and GaN. Phys Rev B 52:8132

Suzuki K, Kanisawa K, Janer C, Perraud S, Takashina K, Fujisawa T, Hirayama Y (2007) Spatial imaging of two-dimensional electronic states in semiconductor quantum wells. Phys Rev Lett 98:136802

Thorpe MF, Weaire D (1971) Electronic properties of an amorphous solid. II. Further aspects of the theory. Phys Rev B 4:3518

Tsidilkovski IM, Harus GI, Shelushinina NG (1985) Impurity states and electron transport in gapless semiconductors. Adv Phys 34:43

Urbieta A, Grandidier B, Nys JP, Deresmes D, Stiévenard D, Lemaître A, Patriarche G, Niquet YM (2008) Scanning tunneling spectroscopy of cleaved InAs/GaAs quantum dots at low temperatures. Phys Rev B 77:155313

van Overstraeten RJ (1982) Proceedings in physics semiconductor devices. Wiley, New Delhi, p 69

Varshni YP (1967) Temperature dependence of the energy gap in semiconductors. Physica 34:149

Voos M, Esaki L (1981) InAs–GaSb superlattices in high magnetic fields. In: Chikazumi S, Miura N (eds) Physics in high magnetic fields. Springer Series in Solid-State Sciences, vol 24. Springer, Berlin, Heidelberg, p 292

Vouilloz F, Oberli DY, Dupertuis MA, Gustafsson A, Reinhardt F, Kapon E (1997) Polarization anisotropy and valence band mixing in quantum wires. Phys Rev Lett 78:1580

Vouilloz F, Oberli DY, Dupertuis MA, Gustafsson A, Reinhardt F, Kapon E (1998) Effect of lateral confinement on valence-band mixing and polarization anisotropy in quantum wires. Phys Rev B 57:12378

Vrehen QHF (1968) Interband magneto-optical absorption in gallium arsenide. J Phys Chem Sol 29:129

Vurgaftman I, Meyer JR (2003) Band parameters for nitrogen-containing semiconductors. J Appl Phys 94:3675

Vurgaftman I, Meyer JR, Ram-Mohan LR (2001) Band parameters for III-V compound semiconductors and their alloys. J Appl Phys 89:5815

Wagner J (1984) Photoluminescence and excitation spectroscopy in heavily doped *n*- and *p*-type silicon. Phys Rev B 29:2002
Watanabe MO, Yoshida J, Mashita M, Nakanisi T, Hojo A (1985) Band discontinuity for GaAs/AlGaAs heterojunction determined by *C-V* profiling technique. J Appl Phys 57:5340
Weaire D, Thorpe MF (1971) Electronic properties of an amorphous solid. I. A simple tight-binding theory. Phys Rev B 4:2508
Wegscheider W, Pfeiffer LN, West KW (1996) GaAs/AlGaAs quantum wire lasers and other low-dimensional structures fabricated by cleaved edge overgrowth. Festkörperprobleme/Adv Solid State Phys 35:155
Wellenhofer GA, Rössler U (1997) Global band structure and near-band-edge states. Phys Status Solidi B 202:107
Wigner EP (1959) Group theory and its application to quantum mechanics of atomic spectra. Academic Press, New York
Wilson AH (1931) The theory of electronic semiconductors. Proc R Soc London A133:458
Woggon U (1997) Optical properties of semiconductor quantum dots. Springer, Berlin
Woodall JM (1972) Solution grown $Ga_{1-x}Al_xAs$ superlattice structures. J Crystal Growth 12:32
Yu ET, McCaldin JO, McGill TC (1992) Band offsets in semiconductor heterojunctions, vol 46, Solid State Physics. Academic Press, New York, pp 1–146
Zawadzki W (1974) Electron transport phenomena in small-gap semiconductors. Adv. Phys 23:435
Zawadzki W (1982) Mechanisms of electron scattering in semiconductors. In: Moss TS, Paul W (eds) Handbook on semiconductors, vol 1, Band theory and transport properties. North Holland, Amsterdam, pp 713–803
Ziman JM (1969) The physics of metals, vol 1, Electrons. Cambridge University Press, London
Zabolotnykh AA, Volkov VA (2021) Electrically controllable cyclotron resonance. Phys Rev B 103:125301

Magnetic Semiconductors

9

Karl W. Böer and Udo W. Pohl

Contents

1 Magnetic Interaction in Solids 320
1.1 Paramagnetic Ions 321
1.2 Magnetic Ordering in Semiconductors 324
2 Diluted Magnetic Semiconductors 330
2.1 II–VI Diluted Magnetic Semiconductors 331
2.2 III–V Diluted Magnetic Semiconductors 335
2.3 Exchange Mechanisms in Diluted Magnetic Semiconductors 339
3 Spintronics 345
3.1 Generation of a Spin Polarization 347
3.2 Spin-Polarized Transport 351
3.3 Optical Spin Detection 361
4 Summary 362
References 362

Abstract

Magnetic properties are introduced into solids by paramagnetic ions. These are transition-metal ions of the iron series with a partially filled electronic 3*d* shell or rare-earth ions of the lanthanide series with an incomplete 4*f* shell. In *magnetic semiconductors*, they represent a cation *component* of the crystal, while in *diluted magnetic semiconductors*, they are a substitutional *alloy* component on the cation sublattice. The magnetic moments of the paramagnetic ions

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_9

are coupled by different kinds of exchange interactions. Superexchange mediated by *p* states of anion ligands favors antiferromagnetism with antiparallel alignment of the magnetic moments, while double exchange and *p–d* exchange favor ferromagnetism with parallel alignment. Magnetic ordering is disturbed if the thermal energy exceeds the exchange energy; critical Curie and Néel temperatures exist for the transition from the paramagnetic high-temperature range to magnetically ordered respective ferromagnetic and antiferromagnetic regimes at lower temperature.

The spin of carriers is utilized in spintronics for current control by creating a nonequilibrium spin polarization. When phase effects can be neglected, the transport of spin-polarized carriers is described by a two-current model assuming two independent channels of different spin projections. Spintronic structures typically comprise two ferromagnetic layers, which are separated by a thin nonmagnetic spacer layer. For insulating spacers and antiparallel aligned ferromagnetic moments a large difference in resistance of the two spin-polarized channels is found, referred to as tunneling magnetoresistance. Similarly a giant magnetoresistance is observed for a thin *conductive* spacer. The orientation of the ferromagnetic magnetization can be altered by spin-transfer torque exerted by a spin-polarized current. Diluted magnetic semiconductors serve as a valuable test bench for exploratory physics of spintronic devices.

Keywords

Antiferromagnetism · Curie temperature · Diluted magnetic semiconductors · DMS · Double exchange · Exchange interactions · Ferromagnetism · Giant magnetoresistance · GMR · Magnetic moments · Magnetic properties · Magnetic semiconductors · Magnetic tunnel junction · Magnetoresistance · MRAM · Néel temperature · Optical spin orientation · Paramagnetic ions · *p–d* exchange · Rare-earth ions · Superexchange · TMR · Transition-metal ions · Tunneling magnetoresistance · Spintronics · Spin FET · Spin injection · Spin polarization · Spin-polarized transport · Spin-transfer torque · Spin valve

1 Magnetic Interaction in Solids

Carriers in solids have a spin in addition to their charge, giving rise to a magnetic moment. Control of the carrier spin in semiconductors is interesting for novel devices such as, e.g., low-power nonvolatile memories. *Magnetic semiconductors* are compounds with paramagnetic ions on a sublattice. These ions may either represent a *component* of the compound semiconductor (e.g., Eu in EuO) or a magnetic *impurity* introduced into a nonmagnetic host semiconductor (e.g., Mn in $Cd_{1-x}Mn_xTe$). The latter are *alloys* with magnetic ions and are referred to *diluted magnetic semiconductors* (Sect. 2). We will first provide a general description of magnetic properties in solids and then focus on this kind of semiconductors.

1.1 Paramagnetic Ions

Magnetic ions introduced into a nonmagnetic semiconductor at low concentration (doping regime, composition parameter x below 10^{-3}) behave similar to isolated ions. We therefore first refer to the magnetic properties of a free ion (or, atom). The magnetic moment of an ion may originate from the spin of the electrons and from their orbital angular momentum.[1]

The magnetic moment of an ion tends to align in the presence of an external magnetic field H. Any system containing electrons in a volume V experiences a *magnetization density* M given by

$$M = \frac{1}{\mu_0 V} \frac{\partial E}{\partial H}, \tag{1}$$

where $\mu_0 = 1.25664 \times 10^{-6}$ Vs/(Am) is the vacuum permeability.[2] Magnetic properties of matter are described by the *magnetic susceptibility* per unit volume, which is defined as

$$\chi = \frac{\partial M}{\partial H}. \tag{2}$$

The dimensionless quantity χ describes how strong the externally applied field H magnetizes the matter; it is positive for paramagnetic and negative for diamagnetic materials.[3] Often M is linear in H and Eq. 2 simply reduces to $\chi = M/H$.

The magnetic moment $\boldsymbol{\mu}$ of a free ion is composed of the contributions arising from the spin $\mathbf{S}$ and the angular momentum $\mathbf{L}$, which are coupled by Russell–Saunders coupling to a total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$, yielding

$$\boldsymbol{\mu} = -g\mu_B \mathbf{J}/\hbar, \tag{3}$$

where g is the Landé factor. Taking the g-factor g_0 of a free electron to be exactly 2 (instead of actually 2.0023...), the Landé g-factor of the free ion in Eq. 3 reads

$$g = 1 + \frac{1}{2} \frac{J(J+1) + S(S+1) - L(L+1)}{J(J+1)}; \tag{4}$$

[1] There is also a small diamagnetic contribution arising from filled electronic shells with zero spin and orbital angular momentum known as Larmor (or Langevin) diamagnetic susceptibility; this contribution does not depend on temperature and is generally counteracting paramagnetism in solids.

[2] The vacuum permeability μ_0 was redefined in 2019 by the update of SI units; it is no longer a constant (by the former definition of the SI ampere), but needs to be determined experimentally.

[3] The susceptibility χ is generally a tensor. For simplicity we assume a magnetization vector $\mathbf{M}$ parallel to $\mathbf{H}$ in Eq. 1, yielding a scalar χ. In a vector relation a *set* of equations according to Eq. 1 applies for the vector components.

here, J, L, and S are the respective quantum numbers of total angular momentum, angular momentum, and spin of the coupled electrons of the considered ion. The *Bohr magneton* μ_B in Eq. 3 is the absolute value of the magnetic moment of a free electron with mass m_0 and spin $|\mathbf{s}| = \frac{1}{2}\hbar$ and is given by

$$\mu_B = \frac{e\hbar}{2m_0}. \tag{5}$$

To obtain the magnetic susceptibility of N identical ions in a volume V, we consider the effect of the magnetic field H on the populated ground state of the ions. The degenerate ground state of an ion with total angular momentum **J** splits into $(2J + 1)$ levels with different quantum numbers $m_J = J, J-1, \ldots, -J$ with respective energies

$$E_{m_J} = m_J\, g\ \mu_B\ B, \tag{6}$$

where $B = \mu_0\,(H + M)$ is the magnetic flux density (often *also* termed magnetic field). If there are N identical ions of angular momentum **J** in a volume V, their magnetization according to Eq. 1 is obtained from the added contributions of each thermally excited level for all ions. If only the $(2J + 1)$ levels of the split ground state are occupied (and not an additional excited state of the ion), then, using Eqs. 1 and 2, the magnetic susceptibility of N such ions is

$$\chi = (N/V)\ \ g\ \ J\ \ \mu_B\ \ B_J(x); \tag{7}$$

here $x \equiv g\ J\ \mu_B\ B/(kT)$, and the Brillouin function $B_J(x)$ accounts for the thermal occupation of the states, which are split by the magnetic field.[4] For small splitting compared to thermal energies ($x \ll 1$), Eq. 7 reads

$$\chi \cong \frac{N}{V}\frac{p^2 {\mu_B}^2}{3kT}, \tag{8}$$

where p, the effective number of Bohr magnetons μ_B, is defined by

$$p \equiv g\sqrt{J(J+1)}. \tag{9}$$

Magnetic ions in semiconductors are generally rare-earth ions of the lanthanide series with a partially filled electronic $4f$ shell or, particularly in diluted magnetic semiconductors, transition-metal ions of the iron series with a partially filled $3d$ shell. The effective numbers p of Bohr magnetons for these ions are listed in Table 1; the values were measured from the $1/T$ dependence of the susceptibility according to

[4]The Brillouin function is defined by $B_J(x) = \frac{2J+1}{2J}\coth\left(\frac{2J+1}{2J}x\right) - \frac{1}{2J}\coth\left(\frac{x}{2J}\right)$ and varies from 0 to 1 for $x = 0$ to ∞. For high temperature or small splitting, $x \ll 1$ applies, and the hyperbolic cotangent can be approximated by $\coth y = 1/y + y/3 - y^3/45 + \ldots$.

Table 1 Measured and calculated effective magneton numbers p for rare-earth ions of the lanthanide series and transition-metal ions of the iron series incorporated in insulating crystals; for transition metals an electron g-factor $g_0 = 2$ is used

Electron configuration	Ground state $^{2S+1}L_J$	Ion	p measured	$p = g[J(J+1)]^{1/2}$	$p = g_0[S(S+1)]^{1/2}$
$4f^1$	$^2F_{5/2}$	Ce^{3+}	2.4	2.54	
$4f^2$	$^3H_4$	Pr^{3+}	3.5	3.58	
$4f^3$	$^4I_{9/2}$	Nd^{3+}	3.5	3.62	
$4f^4$	$^5I_4$	Pm^{3+}	–	2.68	
$4f^5$	$^6H_{5/2}$	Sm^{3+}	1.5	0.84	
$4f^6$	$^7F_0$	Eu^{3+}	3.4	0.00	
$4f^7$	$^8S_{7/2}$	Gd^{3+}	8.0	7.94	
$4f^8$	$^7F_6$	Tb^{3+}	9.5	9.72	
$4f^9$	$^6H_{15/2}$	Dy^{3+}	10.6	10.63	
$4f^{10}$	$^5I_8$	Ho^{3+}	10.4	10.60	
$4f^{11}$	$^4I_{15/2}$	Er^{3+}	9.5	9.59	
$4f^{12}$	$^3H_6$	Tm^{3+}	7.3	7.57	
$4f^{13}$	$^2F_{7/2}$	Yb^{3+}	4.5	4.54	
$3d^1$	$^2D_{3/2}$	Ti^{3+}	–	1.55	1.73
$3d^1$	$^2D_{3/2}$	V^{4+}	1.8	1.55	1.73
$3d^2$	$^3F_2$	V^{3+}	2.8	1.63	2.83
$3d^3$	$^4F_{3/2}$	V^{2+}	3.8	0.77	3.87
$3d^3$	$^4F_{3/2}$	Cr^{3+}	3.7	0.77	3.87
$3d^3$	$^4F_{3/2}$	Mn^{4+}	4.0	0.77	3.87
$3d^4$	$^5D_0$	Cr^{2+}	4.8	0	4.90
$3d^4$	$^5D_0$	Mn^{3+}	5.0	0	4.90
$3d^5$	$^6S_{5/2}$	Mn^{2+}	5.9	5.92	5.92
$3d^5$	$^6S_{5/2}$	Fe^{3+}	5.9	5.92	5.92
$3d^6$	$^5D_4$	Fe^{2+}	5.4	6.70	4.90
$3d^7$	$^4F_{9/2}$	Co^{2+}	4.8	6.54	3.87
$3d^8$	$^3F_4$	Ni^{2+}	3.2	5.59	2.83
$3d^9$	$^2D_{5/2}$	Cu^{2+}	1.9	3.55	1.73

Eq. 8. The given ground-state terms obey Hund's rules, which state that the occupation of the $2(2l + 1)$ states of a partially filled shell maximizes (1) the total spin S and (2) the orbital momentum L (consistent with maximum S) and (3) takes either maximum J (for more than half-filled shells) or otherwise minimum J. We note a very good agreement of the measured values and those calculated from Eq. 9 for the $4f$ ions.[5] By contrast, $J = S$, i.e., $L = 0$ must be assumed for the $3d$ ions to obtain an agreement.

[5]The poor agreement for Sm and Eu originates from excited states lying closely above the ground state; such conditions were excluded for the validity of Eqs. 7, 8, and 9.

The reason for the deviating behavior of the 3*d* ions is due to the effect of the crystalline environment. The 3*d* shell is the outermost shell of transition metals; electrons in this shell experience the electric field of neighboring ions. The interaction with this crystal field is much stronger than the spin-orbit coupling and leads to a crystal-field splitting of orbital *L* multiplets (Sects. 1.4 and 3 in ▶ Chap. 19, "Deep-Level Centers"). As a consequence, the *mean value* of every **L** component averages to zero by a precession of the angular orbital momentum in the crystal field; **L** is said to be quenched. The crystal field does not affect electrons in the 4*f* shell of rare-earth ions, because this shell lies beneath the filled 5*s* and 5*p* shells deep in the ions and is thereby well screened. Since the electric crystal field does not split the spin degeneracy, the ground-state multiplet is determined by *S*.

Equation 8 is known as *Curie's law* and usually written in the form

$$\chi \cong \frac{C}{T}, \tag{10}$$

where *C* is the Curie constant related to the quantities appearing in Eq. 8. This temperature-dependent paramagnetism of magnetic ions in semiconductors is complemented by a paramagnetic component of carriers (electrons and holes).[6] In metals the *Pauli paramagnetic susceptibility* of conduction electrons is essentially independent on temperature; such conditions apply similarly for highly doped, degenerate semiconductors. In nondegenerate semiconductors the carrier concentration depends strongly on temperature, and likewise their paramagnetic contribution; furthermore, band effects and diamagnetic contributions appear in the susceptibility of semiconductors (Sect. 2 in ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients").

1.2 Magnetic Ordering in Semiconductors

1.2.1 Exchange Interactions

In a solid the magnetic ions are located on the sites of a Bravais lattice. Their magnetic moments are affected by an interaction with those of neighboring magnetic ions, in addition to the interaction with an externally applied field. The coupling to neighboring magnetic ions can be described in the simplified spin-Hamiltonian formalism by the sum of two-spin contributions $\mathbf{S}_i \cdot \mathbf{S}_j$ and corresponding interaction parameters J_{ij}, yielding

$$E_{\text{exc}} = -\sum\nolimits_{i \neq j} J_{ij} \ \mathbf{S}_i \mathbf{S}_j. \tag{11}$$

The energy E_{exc} gets minimum if the scalar product of the vector spin operators $\mathbf{S}_i$ and $\mathbf{S}_j$ tends to parallel spins for positive J_{ij} and to antiparallel spins for negative J_{ij}.

[6]The additional paramagnetism of nuclear spins is negligible compared to electronic contributions (a fraction below 10^{-3}).

Equation 11 is known as Heisenberg model. The Heisenberg exchange parameter J_{ij} is an exchange integral of overlapping charge distributions of neighboring magnetic ions. The sign of the exchange integral depends on the ratio of the interatomic distance a_{ij} of the interacting atoms i and j with respect to the radius r of the shell of the magnetic electrons. A small interatomic distance, i.e., strong overlap of the electrons in the paramagnetic shell, leads to antiparallel alignment of the spins due to Pauli principle; at larger a_{ij} this condition is relaxed and parallel spin alignment is favored. As a rule of thumb, a distance a_{ij} exceeding three times the radius of the magnetic shell is required for a parallel alignment. A representation of the exchange parameter J for transition metals, i.e., metallic solids, as a function of the ratio of a_{ij} to the radius r_{3d} of the 3*d* electron shell (4*f* shell for Gd) is given in Fig. 1. A parallel spin alignment is referred to as ferromagnetism and the antiparallel alignment to antiferromagnetism.

The Heisenberg model in Eq. 11 was introduced above for the case of a *direct exchange*, arising from the direct Coulomb interaction among electrons of two neighboring magnetic ions (Fig. 2a). There are also other types of exchange mechanisms illustrated below and discussed in more detail in Sect. 2.3. In transition metals characterized in Fig. 1, conduction-band electrons also contribute to the interaction.

In a *semiconductor* like europium monoxide, the magnetic ions occupy a sublattice and are surrounded by nonmagnetic nearest neighbors on another sublattice (e.g., Eu^{2+} and O^{2-} ions on the cation and anion sublattices of the rock-salt structure of EuO, respectively). The concept of an exchange mechanism can still be applied; the magnetic interaction of the Eu ions is mediated by their common nonmagnetic O neighbors. This type of interaction is termed *superexchange* and illustrated in Fig. 2b.

The exchange interaction may also be mediated by electrons in the conduction band or holes in the valence band. Such kind of interaction is of particular importance for a low concentration of magnetic ions in semiconductors with mediating holes (*p–d* exchange, Sect. 2.3). Conduction-band electrons may also have a significant effect which acts simultaneously to the direct exchange: the partially filled 4*f* shells of rare-earth metals, e.g., lie deep beneath filled 5*s* and 5*p* shells and do hence overlap very little. Consequently the direct exchange is weak among next neighbors, and the contribution of free electrons, referred to in this case as *indirect exchange*, may even prevail.

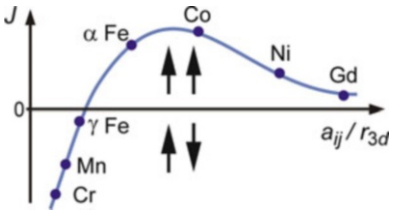


Fig. 1 Exchange parameter J for transition metals as a function of the ratio a_{ij}/r_{3d} of the interatomic distance a_{ij} to the radius r_{3d} of the 3*d* (Gd: 4*f*) electrons. Parallel and antiparallel *arrows* indicate ferromagnetic and antiferromagnetic coupling for $E_{exc} > 0$ and $E_{exc} < 0$, respectively

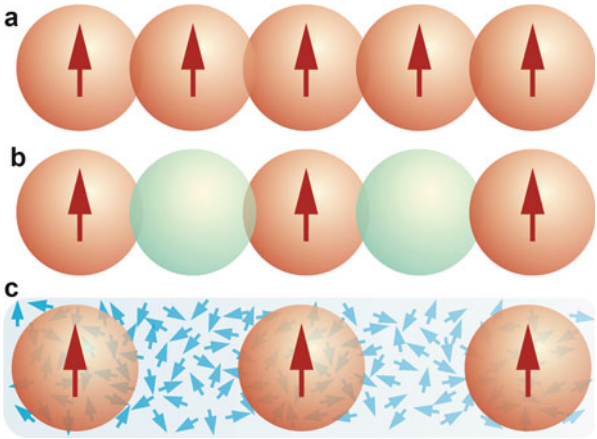


Fig. 2 Schematic of exchange mechanisms. (**a**) Direct exchange with overlapping charge distributions of magnetic ions, (**b**) superexchange with overlapping charge distributions of magnetic ions and nonmagnetic mediating ions, and (**c**) indirect exchange or itinerant exchange with mediating conduction carriers

The exchange mechanisms outlined above are illustrated in Fig. 2.

The magnetic properties of the europium chalcogenides EuO (E_g = 1.12 eV), EuS (1.65 eV), EuSe (1.8 eV), and EuTe (2.00 eV) can be described by the Heisenberg model with dominating nearest-neighbor Eu–Eu exchange and next-nearest-neighbor Eu–anion superexchange interactions J_1 and J_2, respectively. These compound semiconductors crystallize in rock-salt structure with increasing Eu–anion distance for larger atomic number of the anion. The resulting variation of the exchange parameters shown in Fig. 3 leads to a decreasing strength of ferromagnetic order from EuO to EuS, a tendency to antiferromagnetic alignment in EuSe, and to pure antiferromagnetic order in EuTe. The nearest-neighbor interaction J_1 is related to the excitation of an Eu 4*f* electron to a more extended 5*d* band, where it experiences an exchange interaction with a 4*f* electron of a neighboring Eu ion before returning to its initial state. This mechanism produces a ferromagnetic exchange ($J_1 > 0$). In the weaker next-nearest-neighbor interaction J_2 an anion *p* electron is excited to an Eu cation *d* state, yielding a ferromagnetic coupling for small and an antiferromagnetic coupling for larger distances.

1.2.2 Ferromagnetic Domains

The exchange interaction acts solely between *two* respective neighboring magnetic ions and leads in a ferromagnet to a parallel alignment of the adjacent magnetic moments. Successive coupling of all neighboring pairs would thus lead to a spontaneously occurring (i.e., without the action of an external magnetic field) homogeneous alignment of all magnetic moments in the solid. There is, however, an additionally acting very weak magnetic *dipolar* interaction among the magnetic moments of the ions. Though being weak, its interaction range only falls inversely to the cube of the distance, in contrast to the exponentially falling short-ranged exchange interaction. The dipolar interaction thus couples *many* magnetic moments and has hence a sizeable contribution to the magnetic energy.

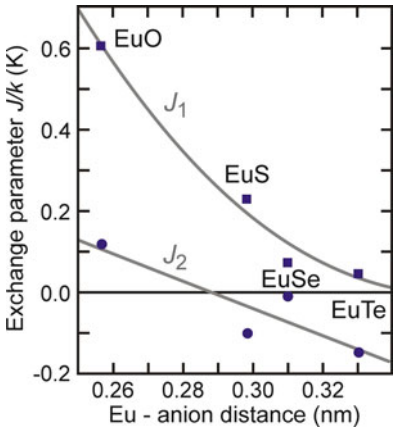


Fig. 3 Heisenberg exchange parameters for nearest (J_1, *squares*) and next-nearest neighbors (J_2, *circles*) for europium chalcogenides with different Eu–anion distances. *Gray lines* are guides to the eye (Data from Wachter 1979)

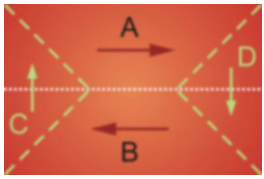


Fig. 4 A ferromagnet with two magnetic domains *A* and *B* with a reduced dipolar energy compared to a homogeneous single domain *A* extending throughout the solid. *Green dashed lines* indicate the introduction of two further domains *C* and *D* leading to a still lower dipolar energy

Due to the dipolar contribution, a homogeneous magnetization throughout a sample has a larger energy than the segmentation into two antiparallel *domains* as illustrated in Fig. 4. This is intuitively clear if the two halves *A* and *B* were two bar magnets with opposite poles near one another: a parallel alignment would bring repelling like poles near one another. Adding two more domains *C* and *D* further reduces the dipolar energy, thereby also reducing the unfavorable magnetic stray field outside the solid. The energy cost of forming domain walls (with magnetic orientations being unfavorable for the short-ranged exchange interaction) is overbalanced by the dipolar energy gain.

A ferromagnetic solid contains many uniformly magnetized domains with a random mutual orientation similar to grains in a polycrystalline solid, leaving an only minor external magnetization. Application of a *weak* external magnetic field leads to the growth of domains with a favorable orientation on expense of the others; at *stronger* fields the orientation of all domains gets aligned to the external field, until a *saturation magnetization* is attained (such saturation existed already within the randomly oriented domains). Removing the external field leaves a preferred orientation of domains, leading to a remanence. Since the same behavior occurs at

reversed orientation of the external field, a hysteresis curve appears in the field-dependent magnetization (Figs. 9b and 12).

1.2.3 Ferromagnetic Susceptibility

The magnetic ordering described above exists only at sufficiently low temperature. The alignment of magnetic moments is disturbed if the thermal energy is of the same order as the exchange energy. Well above a critical temperature, which is referred to as Curie temperature T_C, a ferromagnet behaves like a paramagnet. In this region the susceptibility is described by a relation termed Curie–Weiss law, which can be motivated by a *mean-field approach*. The approximation assumes that all interactions expressed by Eq. 11 can be replaced by a mean external field

$$\mathbf{H}_{\text{eff}} = \mathbf{H} + \lambda \mathbf{M} \tag{12}$$

acting on each spin (also assumed to have the same average value), where the molecular-field constant λ is given by

$$\lambda = \frac{V}{N} \frac{J_0}{(g\mu_B)^2}, \quad J_0 = \sum\nolimits_i J_i. \tag{13}$$

J_0 is an effective single-ion term of magnetic ions replacing the two-ion exchange parameters appearing in Eq. 11. The field H acting on a magnetic moment of a paramagnet is replaced by H_{eff}, yielding with Eq. 10 in a linear approximation

$$\frac{M}{H + \lambda M} \cong \frac{C}{T}. \tag{14}$$

Since, according to Eq. 2, the susceptibility is defined with respect to the *external* field H, we obtain from Eq. 14, after rearranging, the *Curie–Weiss law*

$$\chi = \frac{M}{H} \cong \left(\frac{T}{C} - \lambda\right)^{-1} = \frac{C}{T - T_C} \text{ for } T > T_C, \tag{15}$$

where $T_C = \lambda\, C$ is often taken as the Curie temperature and C is the Curie constant. The critical temperature in Eq. 15 actually does not coincide with the Curie temperature T_C, which is defined for the transition from the ferromagnetic low-temperature region to the paramagnetic region at high temperature. Near T_C the actual susceptibility is rather described by a law $\chi \propto (T - T_C)^{-\gamma}$, with an exponent γ between 1.3 and 1.4. Keeping Eq. 15, i.e., $\gamma = 1$, yields a reasonable description well above T_C but requires a critical temperature T_{crit} somewhat above the transition temperature T_C. Curie temperatures for ferromagnets are given in Table 2.

Table 2 Curie temperature T_C for ferromagnetic solids. The rare-earth metals terbium (Tb) to thulium (Tm) become antiferromagnetic above T_C; see Table 3

Crystal	Structure	T_C (K)
Fe	bcc	1044
Co	hcp	1388
Ni	fcc	624
Gd	hcp	292
Tb	hcp	219
Dy	hcp	85
Ho	hcp	19
Er	hcp	19
Tm	hcp	32
Fe_2O_3	Corundum Al_2O_3	948
CrO_2	Rutile TiO_2	386
MnAs	Hexagonal MnAs	318
MnBi	Hexagonal NiAs	630
MnSb	Hexagonal NiAs	587
EuO	Cubic NaCl	69
EuS	Cubic NaCl	17

1.2.4 Antiferromagnetic Susceptibility

The transition of an ordered alignment of magnetic moments to a paramagnetic phase at higher temperature applies also for the *antiferromagnetic coupling*.[7] The temperature for such transition is called Néel temperature T_N. Antiferromagnetism does not lead to a macroscopic magnetization outside the solid; instead, it leads to an anomalous dependence of susceptibility and specific heat on temperature below T_N and can also be measured in neutron scattering. Below T_N the susceptibility depends on the orientation of the external magnetic field with respect to the spin axis; see Fig. 5c. A perpendicular orientation leads to an approximately constant value $\chi_\perp$; on the other hand, an orientation of antiparallel and parallel spins along the axis of the magnetic field leads to a vanishing $\chi_\parallel$ at $T = 0$ and a steady increase up to T_N.

The susceptibility of an antiferromagnet in the paramagnetic high-temperature range is described by the Curie–Weiss law

$$\chi = \frac{C}{T + \Theta} \text{ for } T > T_N, \tag{16}$$

with a characteristic Curie–Weiss temperature Θ (<0) in the range $-\Theta/T_N = 0.4 \ldots 0.53$; C is referred to as *Curie–Weiss constant*. Comparison with Eq. 15 indicates that the Curie–Weiss temperature Θ of an antiferromagnet represents a negative Curie temperature; this is a consequence of the negative exchange interaction J and

[7] Besides ferro- and antiferromagnetic coupling, the magnetic moments can be ordered in a *ferri*magnetic coupling with a not canceling antiparallel alignment for the moments of neighboring (not identical) magnetic ions.

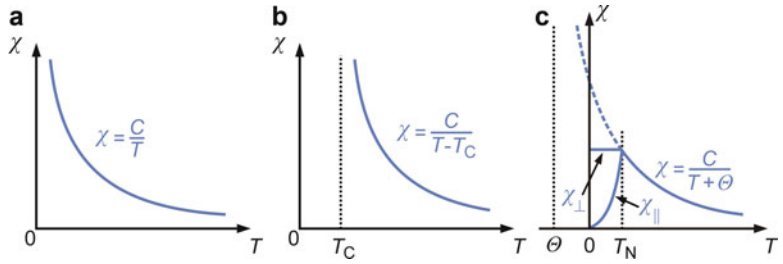


Fig. 5 Qualitative temperature dependence of the magnetic susceptibility in (**a**) paramagnetic, (**b**) ferromagnetic, and (**c**) antiferromagnetic solids. Susceptibilities $\chi_\perp$ and $\chi_\parallel$ in (*c*) denote values perpendicular and parallel to an external magnetic field, respectively

Table 3 Néel temperature T_N of antiferromagnetic solids; for compounds also the Curie–Weiss temperature Θ is listed

Crystal	Structure	T_N (K)	Crystal	Structure	T_N (K)	$-\Theta$ (K)
Cr	bcc	308	MnO	NaCl	116	610
Mn	Cubic α-Mn	103	MnS	NaCl	160	528
Tb (>219 K)	hcp	230	MnSe	NaCl	247	
Dy (>85 K)	hcp	179	MnTe	NiAs	307	690
Ho (>19 K)	hcp	132	Cr_2O_3	Al_2O_3	320	
Er (>19 K)	hcp	85	FeO	NaCl	198	570
Tm (>32 K)	hcp	57	CoO	NaCl	291	330
			NiO	NaCl	525	~2,000
			EuSe	NaCl	5	
			EuTe	NaCl	10	

indicated by the vertical dotted line in Fig. 5c. The Néel and characteristic temperatures of antiferromagnetic solids are listed in Table 3.

2 Diluted Magnetic Semiconductors

Diluted magnetic semiconductors (DMS) are compound semiconductors alloyed with paramagnetic ions. The magnetic ions are predominantly 3*d* transition metals like Mn, usually located on a cation site[8] where it substitutes a nonmagnetic ion (Zn, Cd, Hg) with completed *d* shell (Hg also *f* shell) and no net spin. Most work was devoted to II–VI and III–V semiconductors with zincblende or wurtzite structure. Carrier-induced ferromagnetism found in these compounds fueled investigations for creating novel devices based on spin manipulation of carriers

[8]Occasionally also incorporation on an interstitial site of the host crystal is found. Such incorporation in parallel to substitutional alloying may be detrimental for the intended magnetic properties as, e.g., pointed out for $Ga_{1-x}Mn_xAs$ in Sect. 2.2.

(see Sect. 3). The goal to achieve an ordering temperature well above room temperature proved challenging, and so did as well the understanding of the magnetic interactions. An early review on II–VI DMS was given by Furdyna and Kossut (1988), and more recent aspects emphasizing III–V DMS were reviewed by Jungwirth et al. (2005) and Dietl et al. (2008, 2014); an excellent review on first-principles theory was provided by Sato et al. (2010).

2.1 II–VI Diluted Magnetic Semiconductors

In II–VI semiconductors a large concentration of 3*d* elements can be substituted for the nonmagnetic group-II cation in the lattice. This holds particularly for manganese, which is about an order of magnitude more miscible than other magnetic transition metals (Fe, Co, Cr) and is the most studied element due to its large spin (Table 1). The composition range for ternary $A^{II}{}_{1-x}Mn_xB^{VI}$ alloys along with the thermodynamically stable phases is depicted in Fig. 6. We note a large miscibility, usually well above $x = 0.35$ in the crystal structure of the host. At large composition parameters x, alloying with Mn not only leads to the usual effects of altering the bond length (Sect. 3.2 in ▶ Chap. 2, "Crystal Bonding"; for $Cd_{1-x}Mn_xTe$, see Balzarotti et al. 1984) and bandgap (Sect. 2.1 in ▶ Chap. 8, "Bands and Bandgaps in Solids") but is also accompanied by a phase change. MnS and MnSe crystallize in the cubic NaCl (rock-salt) structure and MnTe in the hexagonal NiAs structure; there are also small alloy phases with NaCl structure near $x = 1$ for $Cd_{1-x}Mn_xSe$ and $Zn_{1-x}Mn_xS$. The alloys shown in Fig. 6 all show similar magnetic properties, outlined in the following for various ranges of the composition parameter x.

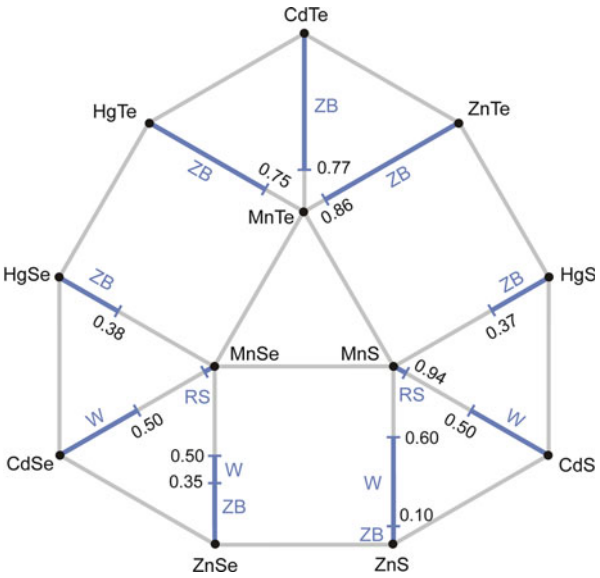


Fig. 6 $A^{II}{}_{1-x}Mn_xB^{VI}$ ternary alloy semiconductors and their crystal structures. *ZB, W, and RS* signify, respectively, zincblende, wurtzite, and rock-salt modifications; the *numbers* denote the Mn composition parameters x (molar fraction) on the cation sublattice, for which homogeneous crystal phases form (After Giriat and Furdyna 1988)

2.1.1 Doping Regime

The magnetic properties of the $A^{II}{}_{1-x}Mn_xB^{VI}$ DMS depend on the concentration of the substitutional Mn^{2+} ions. For dilute samples ($x < 0.001$) only isolated paramagnetic Mn^{2+} ions with $^6S_{5/2}$ ground state need to be considered. The susceptibility for $T > 1$K is described in this *doping regime* by a Curie law[9] according to Eq. 8:[10]

$$\chi \cong \frac{N_{Mn}}{V} \frac{g_0^2 \mu_B^2 S(S+1)}{3kT}. \tag{17}$$

2.1.2 Low-Alloying Regime

At increased concentration of Mn^{2+} ions ($0.002 < x < 0.10$), there is an increased probability for the occurrence of interacting Mn clusters. Pairs of Mn^{2+} ions on neighboring fcc sublattice sites have antiparallel spin alignment, i.e., they have an antiferromagnetic interaction ($J_{nn} < 0$), coupled by the superexchange mechanism. The susceptibility can be approximated at high temperatures by a Curie–Weiss law according to Eq. 16 with a composition-dependent characteristic temperature; see Fig. 7a. A diamagnetic contribution χ_{dia} accounting for the magnetic properties of the host crystal is added, yielding in this *low-alloying regime*

$$\chi = \frac{C}{T + \boldsymbol{\Theta}} + \chi_{dia}. \tag{18}$$

$\chi_{dia} \cong -3.5 \times 10^{-7}$ emu/g applies for the alloys of Fig. 6.[11] The Mn^{2+} ions were found to be randomly distributed in the cation sublattice.[12] For such random distributions the probability of small clusters was calculated by Pohl and Busse (1989) for cubic lattices and by Kreitman and Barnet (1965) for hexagonal lattices. Using these probabilities the susceptibility can be described in terms of a spin-cluster model, yielding

$$\chi \cong \sum_i \frac{N_i}{V} \chi_i. \tag{19}$$

Here i is the type of cluster (e.g., pair or triple; various configurations must be distinguished; see also Pohl and Busse 1989) and N_i/V is the respective number per

[9] A slight departure from the Curie law found for $Cd_{1-x}Mn_xTe$ and $Cd_{1-x}Mn_xSe$ at $T < 0.1$K was attributed to the small crystal-field splitting of the ^6S ground state enabled by minute contributions of excited states and the additional effect of the nuclear spin.

[10] Also a Curie–Weiss law according to Eq. 16 with a characteristic temperature $|\boldsymbol{\Theta}| \ll 1$ K was widely applied, indicating some minor residual coupling effects among the magnetic ions.

[11] The mass magnetization is given here in the conventionally used units of emu/g; the conversion factor to the SI unit is 1 emu/g = 1 A × m^2 / kg = 1 J / (T × kg).

[12] The sublattice of the cations is fcc in (cubic) zincblende and hcp in (hexagonal) wurtzite lattices.

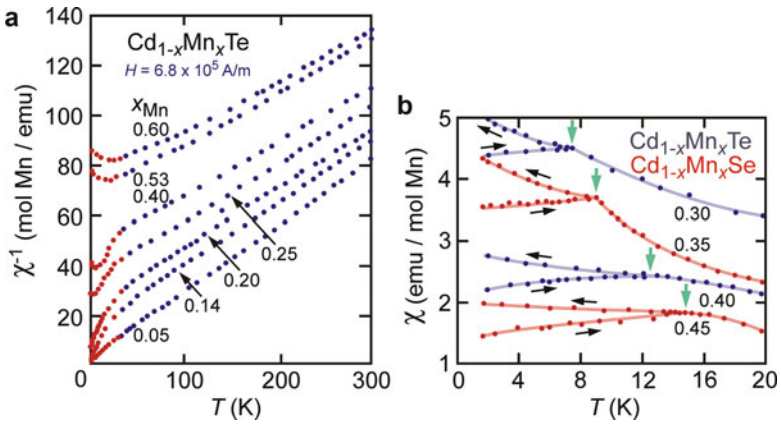


Fig. 7 (**a**) Inverse susceptibility of $Cd_{1-x}Mn_xTe$ for various composition parameters *x*. *Red* and *blue* data points were measured after zero-field cooling for increasing *T* and after field cooling, respectively. (**b**) Susceptibility of $Cd_{1-x}Mn_xTe$ (*blue* symbols) and $Cd_{1-x}Mn_xSe$ (*red* symbols) for zero-field cooling (symbol →) and field cooling (symbol ←); *vertical arrows* indicate cusps in the data obtained for zero-field cooling (After Oseroff 1982)

unit volume; χ is thus composed from the contributions of all cluster types. An explicit evaluation of the contributions χ_i leads to the expression (Nagata et al. 1980)

$$\sum_i \frac{N_i}{V}\chi_i = \sum_i \frac{N_i}{V}\frac{g_0^2\mu_B{}^2}{3\ kT}\frac{\sum_S\sum_m S(S+1)(2S+1)\exp(-(E_S - g\mu_B mH)/kT)}{\sum_S\sum_m (2S+1)\exp(-(E_S - g\mu_B mH)/kT)};$$

E_S is the energy at zero magnetic field *H*, *S* is the total spin, and *m* is the magnetic quantum number; the second summation is performed over $-S \leq m \leq S$. In the exchange interaction, only nearest neighboring Mn^{2+} ions were taken into account, with $|J_{nn}/k| \cong 10$ K (7…8 K for CdMnTe and CdMnSe and 9…15 K for ZnMnS, ZnMnSe, and ZnMnTe).

2.1.3 High-Alloying Regime

As the concentration of Mn^{2+} ions is further increased ($x \geq 0.17$), the clusters get more extended; at a critical concentration x_c given by the *site-percolation threshold*, the cluster extends over the entire sample. The value of x_c depends on the lattice structure and interaction range; for *fcc* lattice $x_c \cong 0.20$ and 0.14 for nearest and second-nearest neighbors, respectively. In this *high-alloying regime*, a *spin-glass* behavior occurs for antiferromagnetic coupling of nearest-neighbor ions, resulting from the *frustration* mechanism: if three neighboring spins are considered that form an equilateral triangle, *two* of them may be aligned antiparallel, but the third cannot also be aligned antiparallel simultaneously to both of the two others. Since both fcc and hcp sublattices in zincblende and wurtzite structures are composed of tetrahedrons (which comprise such triangles), frustration occurs in both structures.

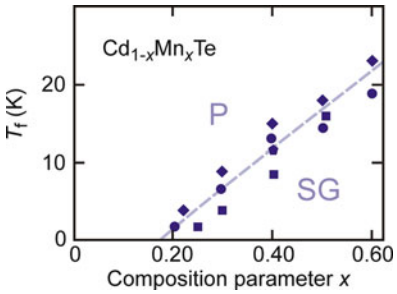


Fig. 8 Magnetic phase diagram showing the maximum temperature (cusp) of χ measured under zero-field cooling versus manganese composition x; T_f is the transition temperature separating the paramagnetic phase P from the spin-glass phase SG (After Oseroff and Keesom 1988)

In the spin-glass regime, the low-temperature susceptibility shows a different temperature dependence upon sample cooling either with or without an externally applied field H. When the sample is cooled in zero field and H is applied at lowest temperature (zero-field cooling, ZFC), the susceptibility increases and passes a maximum as the temperature is raised; by contrast, cooling in presence of an applied field yields a larger low-T susceptibility, which steadily decreases as T is increased (Fig. 7b). The temperature T_f of the ZFC cusp indicated by vertical arrows in the figure broadens, gets weaker, and shifts to lower temperature as the magnitude of H is increased (Oseroff 1982). Such a cusp is also observed in classical metallic spin-glass alloys CuMn and AuFe.

The value of T_f depends also on the composition parameter x as illustrated in Fig. 8. All diluted magnetic semiconductors of Cd and Zn given in Fig. 6 show a composition $x \cong 0.17$ for $T_f \to 0$ (intersection of dashed line with abscissa in the figure) which is close to the percolation threshold x_c for next-nearest-neighbor interaction; a spin-glass is generally expected for antiferromagnetically coupled next-nearest neighbors which are subjected to the frustration mechanism (de Seze 1977).

For applications employing the magnetic properties of semiconductors, a *ferromagnetic* coupling of magnetic moments is important. Such a coupling is difficult to achieve in II–VI semiconductors, particularly for high operation temperature.[13] Reliable room-temperature ferromagnetism was reported for $Zn_{1-x}Cr_xTe$, where the d^4 electrons of Cr^{2+} experience a p–d exchange interaction (Sect. 2.3). Figure 9 shows respective calculated and experimental Curie temperatures. Based on a mean-field approach, ferromagnetic coupling resulting in a Curie temperature exceeding 300 K was also predicted for ZnO alloyed with 5% Mn (Dietl et al. 2000). This theory requires mediation by free holes, which can hardly be obtained in the notorious n-type ZnO. On the other hand, such a mechanism operates well in III–V semiconductors alloyed with Mn considered in the following section.

[13]In some cases an observed ferromagnetism did solely originate from *precipitates* of magnetic ions alloyed into nonmagnetic II–VI host crystals (Saito et al. 2003).

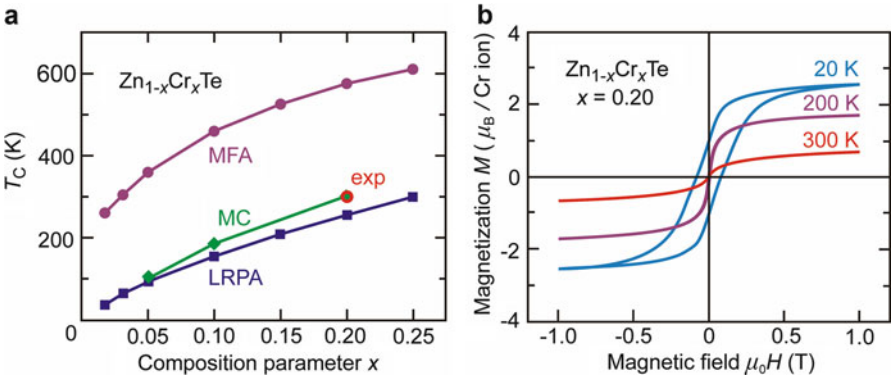


Fig. 9 (**a**) Curie temperature of ferromagnetic $Zn_{1-x}Cr_xTe$ for varied Cr composition. *MFA*, *LRPA*, and *MC* denote computations by a mean-field approximation, the local random-phase approximation, and the Monte Carlo method, respectively (After Sato et al. 2010). The *red* data point labeled exp was experimentally obtained from the sample studied in panel (*b*). (**b**) Magnetization of a $Zn_{1-x}Cr_xTe$ layer grown using molecular-beam epitaxy (After Saito et al. 2003)

2.2 III–V Diluted Magnetic Semiconductors

The solubility of paramagnetic 3*d* ions in III–V semiconductors is much lower than in II–VI compounds ($\sim 10^{-3}$ in equilibrium, ~8% at nonequilibrium); usually nonequilibrium low-temperature growth employing epitaxy (Sect. 3.3 in ▶ Chap. 1, "Properties and Growth of Semiconductors") is applied to achieve a sufficient alloying composition with magnetic ions. Substitutional incorporation of – usually divalent – transition-metal ions for a trivalent cation in a III–V semiconductor leads to localized acceptor or donor states in the bandgap of the host crystal. The location of these states with respect to the valence band of the host crystal and the position of the Fermi level determine the kind of magnetic interaction; see Sect. 2.3. For some III–V semiconductors, the position of the localized 3*d* states is shown in Fig. 10. $Ga_{1-x}Mn_xAs$, $In_{1-x}Mn_xAs$, and $Ga_{1-x}Mn_xN$ are the most studied semiconductors. In GaAs alloying with Mn induces a fairly shallow acceptor level (denoted $Mn^{3+/2+}$ in Fig. 10), leading to one hole carrier per substitutional Mn ion.

2.2.1 Mediation of Ferromagnetism by Holes

Control of defects is important as demonstrated for the Curie temperature measured for $Ga_{1-x}Mn_xAs$: the early finding of 110 K (Ohno et al. 1996) was improved later to a value of 170 K by annealing the samples; the difference was attributed to the removal of interstitial Mn_i and As_{Ga} antisite point defects, which were unintentionally introduced during the low-temperature epitaxy.[14] The ferromagnetic

[14]Molecular-beam epitaxy is performed at low deposition temperatures down to 180 °C to allow for high nonequilibrium alloying levels; due to the limited kinetics at such low temperature, point defects are easily created.

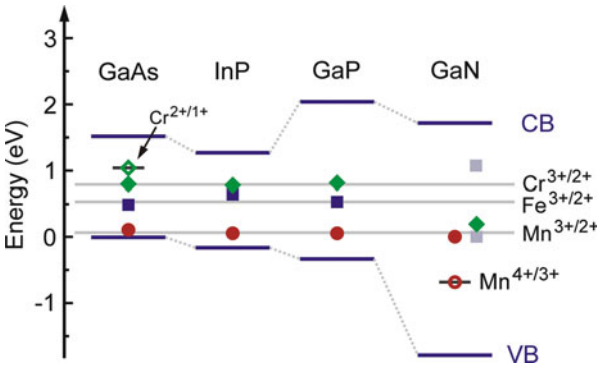


Fig. 10 Measured locations of 3*d*-impurity acceptor or donor levels in III–V semiconductors: Mn (*circles*), Fe (*squares*; different values for GaN are from different references), and Cr (*diamonds*). *Gray lines* are guides to the eye; (After Wolos and Kaminska 2008)

interaction of Mn^{2+} ions in GaAs is mediated by holes. Mn_i and As_{Ga} defects are both donors: they introduce free electrons and thereby compensate the hole carrier-density, which is created via the intended substitution of trivalent Ga^{3+} by divalent Mn^{2+} on a cation lattice site. The Curie temperature of ferromagnetic $Ga_{1-x}Mn_xAs$ given in Fig. 11a shows an increase for Mn composition. Calculated values agree well with those of samples, which were annealed to remove compensating donor defects, while as-grown samples with defects have significantly lower T_C values.[15]

Calculated T_C of Fig. 11a assumed one hole per Mn ion. Evidence for the mediation of ferromagnetic coupling by holes is provided by calculations, which assume a *reduced* number of holes per Mn ion (Fig. 11b). Below unity the Curie temperature decreases strongly and gets negative for a very low hole density: in this range the ferromagnetic state becomes unstable and the short-range antiferromagnetic coupling mediated by superexchange (illustrated in Fig. 2a) dominates.

Mediation of the ferromagnetism by holes is experimentally demonstrated in the isothermal magnetization of (InMn)As by an electric field (Ohno et al. 2000); the example also demonstrates electric control of magnetic properties. A thin layer of $In_{0.03}Mn_{0.97}As$ is used as magnetic channel material of an insulating-gate field-effect transistor; the magnetization M of the layer in a magnetic field B can be measured in terms of the sheet resistance R_{Hall} in a Hall geometry (Sect. 1.2.2 in ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients"), which is given by

$$R_{Hall} = \frac{R_0}{d}B + \frac{R_S}{d}M;$$

d is the thickness of the layer, and R_0 and R_S are, respectively, the ordinary and anomalous Hall coefficients. The measurement of R_{Hall} at $B = 0$ slightly below the Curie temperature of the (InMn)As channel is shown in Fig. 12 for various gate bias voltages V_G. At zero bias a small hysteresis loop indicates some spontaneous magnetization; at positive bias holes are depleted from the channel, and the loop

[15] At highest Mn composition, a deviation from expected values even for annealed samples indicates some onset of compensation.

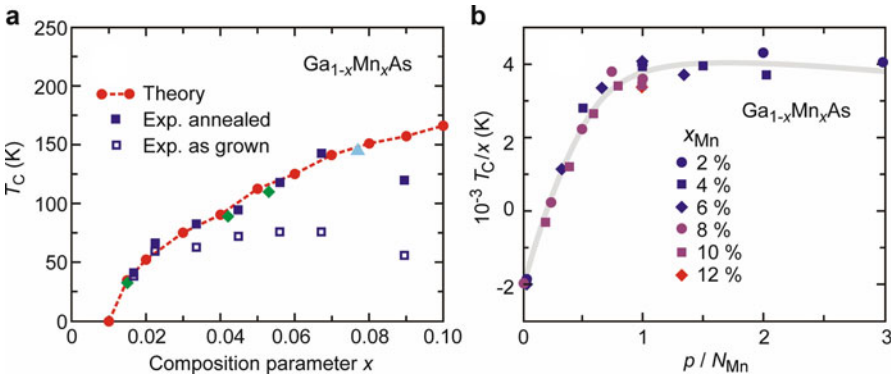


Fig. 11 (**a**) Curie temperature T_C of $Ga_{1-x}Mn_xAs$ for varied Mn composition. *Open* and *filled squares* denote values of as-grown and of subsequently annealed samples, respectively. The *red circles* represent theoretical values obtained from the local random-phase approximation, assuming one free hole per Mn ion (After Sato et al. 2010). (**b**) Normalized Curie temperature T_C per Mn composition x, calculated using a tight-binding coherent-potential approximation for various numbers p of holes per Mn ion and Mn composition parameters x (After Jungwirth et al. 2005)

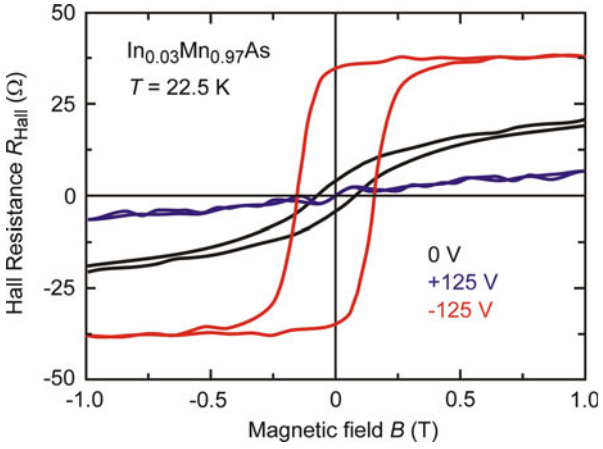


Fig. 12 Magnetization of $In_{0.03}Mn_{0.97}As$, measured via the proportional Hall resistance R_{Hall} in a gated field-effect transistor, measured at different gate biases (After Ohno et al. 2000)

develops to a paramagnetic linear response without hysteresis. By contrast, at negative bias holes are accumulated in the channel, and a pronounced hysteresis loop appears.

2.2.2 Stability of the Ferromagnetic State

The stability of the ferromagnetic (FM) state with respect to antiferromagnetic (AFM) coupling can be expressed by a stabilization energy $\delta = E_{FM} - E_{AFM}$. Table 4 lists stabilization energies for substitutional Mn ion pairs in various III–V

Table 4 Calculated ferromagnetic stabilization energies $\delta = E_{FM} - E_{AFM}$ given in (meV/Mn ion) for substitutional first (1nn) to fourth (4nn) next-neighbor Mn ion pairs in various diluted magnetic semiconductors; ZB denotes the cubic zincblende (β) modification of GaN

Semiconductor	1nn	2nn	3nn	4nn
GaN ZB	−188	−11	−63	−161
GaP	139	131	−89	−132
GaAs	−124	−30	−76	−114
GaSb	−56	−7	−28	−54
$CuGaS_2$	−81			−83
$CuGaSe_2$	−71			−61
$CuGaTe_2$	−52			−59

Data from Zhao et al. (2004)

and chalcopyrite I–III–VI_2 DMS, calculated using density-functional theory; in both kinds of semiconductors, Mn^{2+} ions substitute Ga^{3+} ions. The locations of the first (1nn) to the fourth (4nn) next neighbor on the cation *fcc* sublattice noted in the table refer to $(\frac{1}{2}, 0, \frac{1}{2})a$, $(1, 0, 0)a$, $(\frac{1}{2}, \frac{1}{2}, 1)a$, and $(1, 1, 0)a$, respectively, for the zincblende semiconductors.[16] Large negative values denote a stable ferromagnetic state; the calculations yield for all cases ferromagnetic coupling, with largest stability for the 1nn pair $(\frac{1}{2}, \frac{1}{2}, 0)a - (0, 0, 0)a$ and the 4nn pair $(1, 1, 0)\ a - (0, 0, 0)\ a$, which both are oriented along the $\langle 110 \rangle$ direction. The reason for this orientation dependence is the coupling of the Mn–As bonds to similarly directed states of the dangling-bond holes introduced by the replacement of trivalent Ga^{3+} ions by divalent Mn^{2+} ions (Zhao et al. 2004).

A clear trend of the ferromagnetic stabilization energy with respect to paramagnetism was found within the series of transition-metal ions from calculations based on density-functional theory with the local-density approximation and the coherent-potential approximation (Sato and Katayama-Yoshida 2002). The results given in Fig. 13 show the total energy difference per unit cell ΔE between the ferromagnetic and the paramagnetic state; here, positive values denote a stable ferromagnetic state. The first half of the 3*d* transition-metal series (V and Cr, in II–V compounds also Mn) tends to ferromagnetism (positive ΔE), while – except for ZnO – the second half (Fe and Co) tends to paramagnetism. In the II–VI semiconductors, the trend is shifted by one to lower atomic number; this originates from the divalent charge state and a corresponding occupation of 3*d* states increased by one compared to a trivalent charge state in III–V semiconductors. The universal trend does not depend on the concentration of the magnetic ions and indicates some general rule in the exchange mechanism discussed in the next section.

[16] In the chalcopyrite structure (Fig. 13 in ▶ Chap. 3, "The Structure of Semiconductors"), the 1nn and 3nn neighbors are located at $(\frac{a}{2}, 0, \frac{c}{4})$ and $(\frac{a}{2}, \frac{a}{2}, \frac{c}{2})$, respectively.

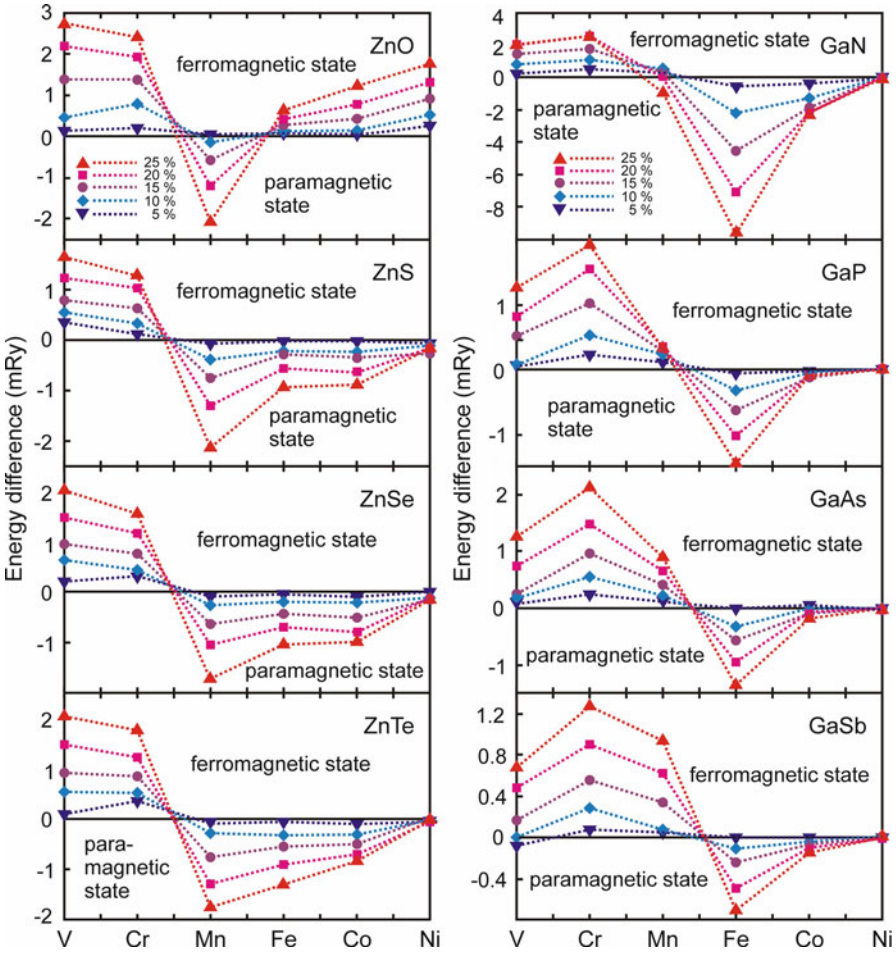


Fig. 13 Chemical trend of the energy difference ΔE per unit cell between ferromagnetic and paramagnetic states in diluted magnetic semiconductors (After Sato and Katayama-Yoshida 2002)

2.3 Exchange Mechanisms in Diluted Magnetic Semiconductors

In diluted magnetic semiconductors, several simultaneously acting exchange mechanisms exist, which either induce ferromagnetism or antiferromagnetism; a physical understanding of mechanisms leading to ferromagnetism or magnetic disorder is still incomplete. In the following the most relevant and usually dominating interactions of double exchange, *p–d* exchange, and superexchange are outlined.

Transition-metal ions alloyed into nonmagnetic semiconductors introduce electronic impurity bands with low dispersion. The energy of these bands with respect to those of the host crystal and to the Fermi level determines the prevailing kind of

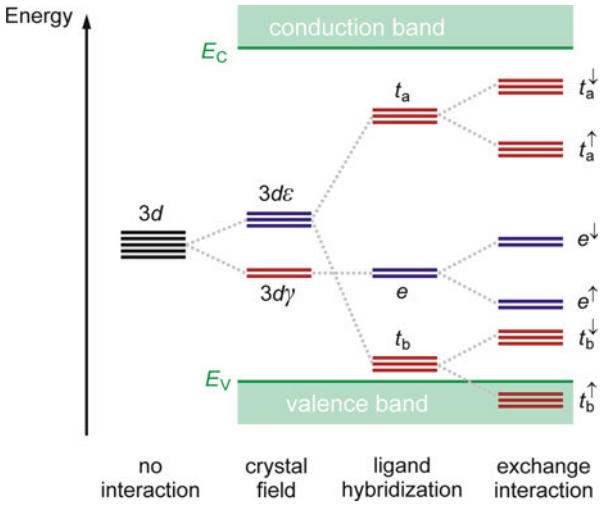


Fig. 14 Electronic structure of 3*d* electrons of a transition-metal ion on a substitutional tetrahedral site, subjected to successive interactions

interaction. We therefore first consider the electronic structure of the 3*d* states of a transition-metal ion in the crystal environment of the host semiconductor.

In semiconductors with wurtzite or zincblende structure, substitutional 3*d* ions occupy lattice sites with a tetrahedral environment of anion ligands (T_d symmetry). As schematically illustrated in Fig. 14, the tetrahedral crystal field of the ligands leads to a spitting of the states of the five *d* orbitals, separating states of the three $3d\varepsilon$ orbitals with *e* symmetry (d_{xy}, d_{yz}, d_{zx}) from those of the two $3d\gamma$ orbitals ($d_{x^2-y^2}$, $d_{3z^2-r^2}$) with t_2 symmetry (Sects. 1.4 and 3 in ▶ Chap. 19, "Deep-Level Centers"). The $3d\varepsilon$ wavefunctions are extended to the neighboring host-crystal anions and therefore hybridize with their *p* states (which form the host valence band) and to some extent also with their *s* orbitals (*p*–*d* hybridization), forming bonding states t_b in (or, near) the valence band; the corresponding antibonding states t_a are raised high into the bandgap of the host. The $3d\gamma$ orbitals are strongly localized and extend more to the interstitial region; they hence show only weak *p*–*d* hybridization, leading to non-bonding *e* states (Zunger 1986; Sato and Katayama-Yoshida 2002). The degree of hybridization depends on the mutual location of the impurity *d* states with respect to the host *p* and *s* states. The charge–charge Coulomb interaction is complemented by spin–spin interaction, which splits each of the *e* and t_2 states into spin-up and spin-down states. The different exchange mechanisms are illustrated below in a scheme representing the density of spin-polarized states, with up-spin and down-spin DOS drawn separately; in the host crystal half of the states refer to the spin-up DOS and the other part to the spin-down DOS.

2.3.1 Superexchange

Coupling of the magnetic moments of magnetic ions by superexchange does not require a finite density of states at the Fermi energy; it hence also occurs in insulating magnetic solids (Goodenough 1955; Kanamori 1959). In Fig. 2b the mechanism is

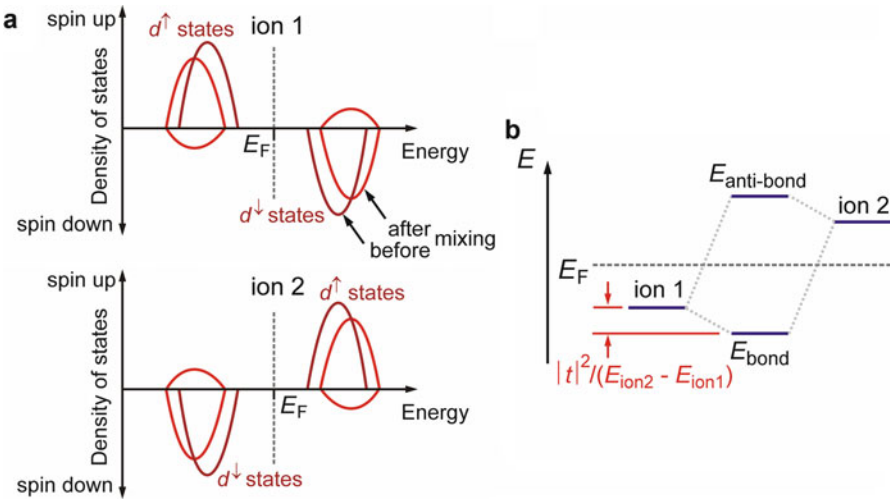


Fig. 15 (**a**) Density of spin-polarized states of two alike magnetic ions with their antiparallel moments coupled by superexchange. *Dark-red* and *red lines* indicate *d* states before and after hybridization. (**b**) Energies of two unlike interacting ions in a molecular model

illustrated by a coupling of two magnetic ions mediated by the *p* orbitals of the nonmagnetic anions, i.e., by an overlap of *d* states of the magnetic impurity and *p* states of a ligand. Hybridization of two *different* states generally leads to the level scheme drawn in Fig. 15b. The simple molecular model yields an energy gain for the low-energy state of ion 1 by approximately $|t|^2/(E_{\text{ion2}} - E_{\text{ion1}})$, where *t* is the hopping integral. We assume for simplicity the state of ion 1 as a gap state of a magnetic impurity and that of ion 2 as host states, albeit strictly speaking the gap states originating from an impurity are already hybrids. The electronic structure of two coupled magnetic ions is shown in Fig. 15a. Due to their antiparallel alignment which is usually induced by superexchange, their respective density of spin-polarized states is mirrored at the energy axis. The Fermi energy is assumed to lie between the energy of the *d* states; hybridization of the *d* states with equal spin will then lower the energy of the occupied low-energy $d^{\uparrow}$ states and raise the energy of the unoccupied $d^{\downarrow}$ states in ion 1 and vice versa in ion 2. This yields the energy gain for stabilizing the antiferromagnetic coupling.

In a diluted magnetic semiconductor, the energy gain considered for two coupled ions in Fig. 15b is proportional to the impurity concentration *x*. The energy gain for superexchange therefore reads

$$\Delta E_{\text{SExc}} \cong x\,|t|^2 / \left(E_{t_2^{\uparrow}} - E_{t_2^{\downarrow}}\right). \tag{20}$$

Equation 20 shows that the superexchange depends strongly on the exchange splitting appearing in the denominator. The interaction is short-ranged but quite strong, because the overlap of t_2 states can be rather large. Superexchange does not

depend on the location of the Fermi level, provided it lies *between* the spin-polarized *d* states. It gets weaker if E_F lies in one of these bands and disappears if E_F lies outside.

Superexchange may also induce ferromagnetic alignment. This occurs if hybridization occurs between t_2 and e impurity states with E_F lying in between. The energy gain is similar to Eq. 20 with an additional factor of 2 and the denominator replaced by $\left(E_{t_2^\uparrow} - E_{e^\uparrow}\right)$. Due to the stronger localization of e states, the hopping integral is, however, much smaller. Usually double exchange and p–d exchange dominate ferromagnetic coupling, with contributions from both mechanisms.

2.3.2 Double Exchange

Double exchange is the prevailing ferromagnetic coupling mechanism in diluted magnetic II–VI semiconductors. The interaction introduced by Zener (1951a) is a strong, short-range ferromagnetic coupling mechanism; it occurs if relevant *partially filled* 3*d* bands of magnetic impurities lie in the bandgap of the host semiconductor. The electronic structure of 3*d* impurities which experiences double exchange is illustrated in Fig. 16a. The split 3*d* states correspond, e.g., to $t_b^\uparrow$ and $t_b^\downarrow$ of Fig. 14, and the Fermi level E_F is assumed to lie in the middle of the lower impurity band labeled $d^\uparrow$. A coupling of two ions with two states of *equal* energy leads generally to a symmetric bonding–antibonding splitting indicated in the scheme in Fig. 16b; here, both of these states lie in the broad impurity band. If the impurity band is only half filled, solely the bonding states E_{bond} are occupied, yielding an energy gain $\Delta E_{DExc} \propto |t|$ for a coupled impurity pair, where t is the hopping matrix element. We notice from Fig. 16 that the energy gain gets maximum for a half-filled impurity band and zero for a completely filled or empty band; since E_F lies within a partially filled band, a stabilization of ferromagnetism is also accompanied by a considerable itinerant character of the electrons; the mobility is, however, lower than that of conduction-

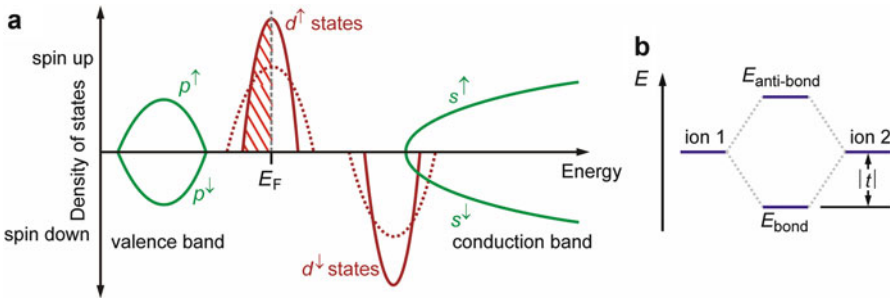


Fig. 16 (**a**) Schematic of the density of spin-polarized states for double exchange. *The hatched area* designates filled spin-up *d* states for low-alloying concentration; at high concentration the *d* band broadens as indicated by the *dotted line*. (**b**) Energies of two like interacting ions in a molecular model yielding bonding and antibonding states

band electrons, which mediate itinerant exchange in transition-metal ferromagnets: the deeper lying states of *d* electrons are more atomic-like (also expressed by a low dispersion; compare Sect. 1.1 and Fig. 3 in ▶ Chap. 6, "The Origin of Band Structure").

The energy gain ΔE_{DExc} increases for larger impurity concentration x. At higher doping the impurity band broadens (dotted distribution of $d^{\uparrow}$ in Fig. 16a); occupation of states can then be transferred to lower energies, yielding also a larger net gain. The width of the impurity band W is proportional to $\sqrt{x}$, yielding

$$\Delta E_{\mathrm{DExc}} \propto \sqrt{x}\ \ |t|. \tag{21}$$

Double exchange is a prominent interaction in diluted ferromagnetic II–VI semiconductors. The *p* states of the six-valent host anions that build the valence band lie lower than those of the five-valent anions in III–V semiconductors. Therefore, relevant impurity 3*d* states are always located in the bandgap of II–VI semiconductors. The spin-polarized DOS of $Zn_{0.95}Cr_{0.05}Te$ given in Fig. 17 shows the states of the $Cr^{2+}d^4(e^{2\uparrow}, {t_2}^{2\uparrow})$ electronic configuration: states of the two ${t_2}^{\uparrow}$ electrons occupy an approximately half-filled ${t_a}^{\uparrow}$ impurity band above the valence band of the ZnTe host; states of the two $e^{\uparrow}$ electrons and the bonding ${t_b}^{\uparrow}$ states lie in range of the host valence band and are completely filled. The exchange interaction arises from the half-filled ${t_a}^{\uparrow}$ band. The narrow energy distributions of the two sharp *e* state peaks reflect the strong localization of these nonbonding states, while t_a and t_b states are more delocalized.

Similar calculations for other semiconductors alloyed with magnetic 3*d* transition metals are reviewed in Sato et al. (2010) and Mahadevan and Zunger (2004).

2.3.3 *p–d* Exchange

The *p–d* exchange is a weaker but more extended ferromagnetic exchange interaction (Zener 1951b; Dietl et al. 2000). The mechanism is illustrated in Fig. 18;

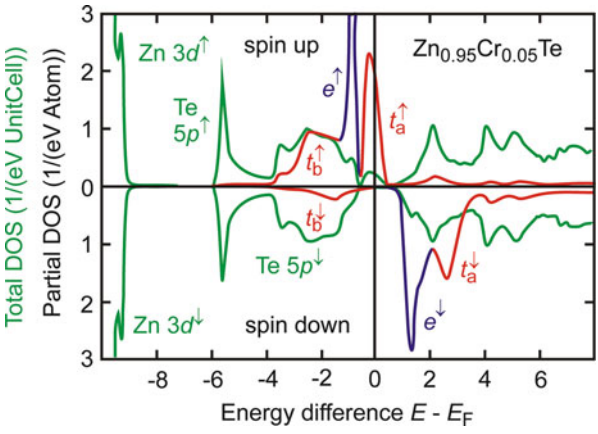


Fig. 17 Calculated density of spin-polarized states in $Zn_{0.95}Cr_{0.05}Te$. *Green lines* refer to the total averaged DOS per unit cell of the ZnTe host crystal; *blue* and *red lines* signify, respectively, states with *e* and t_2 symmetry of the partial DOS per Cr^{2+} 3*d* ion. Abscissa values are energies relative to the Fermi level E_F (After Sato et al. 2010)

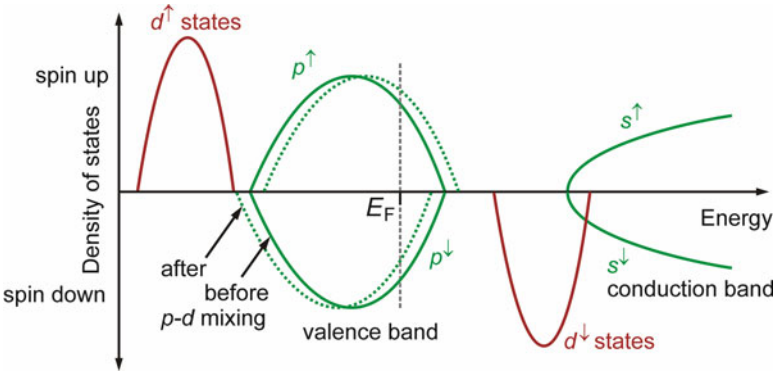


Fig. 18 Spin-polarized density of states for *p–d* exchange. *Solid* and *dotted lines* signify the host *p* band before and after hybridization with impurity *d* states of like spins

the low-energy $d^{\uparrow}$ impurity band lies below the host *p* band, and the $d^{\downarrow}$ impurity band lies above. The *p–d* hybridization enlarges the separation of *p* and *d* states with like spins: the host $p^{\uparrow}$ band moves to higher and the $p^{\downarrow}$ band to lower energy. If the Fermi energy E_F is located below the valence-band maximum as indicated in the figure, such shifts lead to a spin polarization in the host – the occupation of $p^{\uparrow}$ and $p^{\downarrow}$ bands is no longer balanced. Additional holes appear in the $p^{\uparrow}$ host band, while the $p^{\downarrow}$ band gets almost fully occupied; the host becomes polarized with a moment antiparallel to the local moment of the occupied $d^{\uparrow}$ states of the magnetic impurity. By configurational averaging, a homogeneous host polarization is obtained, which favors the ferromagnetic coupling of the magnetic impurities. The coupling energy is proportional to the host polarization and scales linearly with the alloying parameter *x*; this distinguishes the weaker, longer-ranged *p–d* exchange from the stronger, short-ranged double exchange with its $x^{1/2}$ dependence.

A level ordering like that of Fig. 18 is found in diluted magnetic III–V semiconductors with a narrow bandgap like GaMnSb or InMnSb. The density of spin-polarized states in $Ga_{0.95}Mn_{0.05}Sb$ (Fig. 19) shows that the large peak of the $d^{\uparrow}$ impurity band (basically originating from $e^{\uparrow}$ states) lies below the center of the host 5*p* band of Sb anions; the $d^{\downarrow}$ impurity band formed by $e^{\downarrow}$ and $t_a^{\downarrow}$ states lies above the Sb valence band and above E_F. The Mn^{2+} impurity has a localized magnetic moment of 5 magnetons μ_B; in the neutral state with respect to the host, i.e., Mn^{3+}, one electron per Mn ion is missing: ½ electron is transferred to the spin-up and ½ electron to the spin-down valence band. As pointed out above, the $p^{\downarrow}$ band is almost fully occupied; the *p* states of Sb hence carry about 1 μ_B per Mn ion, antiparallel oriented to the effective moment of 4 μ_B per Mn ion.

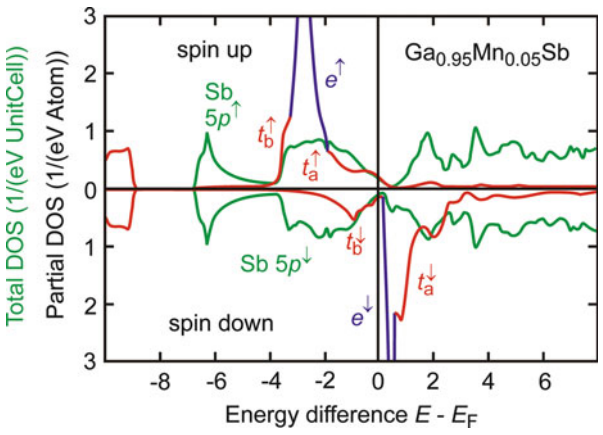


Fig. 19 Calculated density of spin-polarized states in $Ga_{0.95}Mn_{0.05}Sb$. *Green lines* refer to the total averaged DOS per unit cell of the GaSb host crystal; *blue* and *red lines* signify the partial DOS per Mn 3*d* ion. *Abscissa values* are energies relative to the Fermi level E_F (After Sato et al. 2010)

3 Spintronics

Electronic and optoelectronic devices are based on the charge transport of carriers. Including the carrier spin into current control may lead to solid-state devices with additional functionalities. Spintronics (i.e., spin-based electronics) deals with the use of carrier spin in (opto-)electronic materials. Established schemes of spintronics, also referred to as *magneto-electronics*, contain ferromagnetic metallic layers which are either separated by a nonmagnetic metal or by an insulator (e.g., Johnson 2004; Shinjo 2009). The resistance of such a layer stack is changed by keeping the magnetization in one layer fixed and changing the magnetization in the other. These changes, referred to as giant magnetoresistance or – for an insulating barrier – tunneling magnetoresistance, are widely used today in magnetic devices for high-density data storage. Both effects can also be employed in magnetoresistive random access memories (MRAMs), which represent nonvolatile solid-state devices. MRAM cells are programmed by changing the magnetization of one magnetic layer with respect to the other, using a magnetic field generated by two crossing currents. By employing the electron spin for digital switching instead of the external magnetic field, faster switching with less power consumption is feasable. Switching in this more efficient scheme is achieved by using a spin-polarized current through the memory cell.

More recent spintronic schemes also include ferromagnetic semiconductors and materials with a high degree of polarization of the conduction electrons, such as Heusler alloys, reviewed in Awschalom et al. (2002), Dietl et al. (2008), Xia et al. (2012), Felser and Fecher (2013), Schäpers (2016), Dyakonov (2017), and Tsymbal and Žutić (2019), or organic compounds as reviewed by Vardeny (2010). Strong spin polarization was recently observed in quantum spin Hall systems and topological insulaters; see Sect. 4 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors." This new state of matter has a significant potential for future spintronic devices.

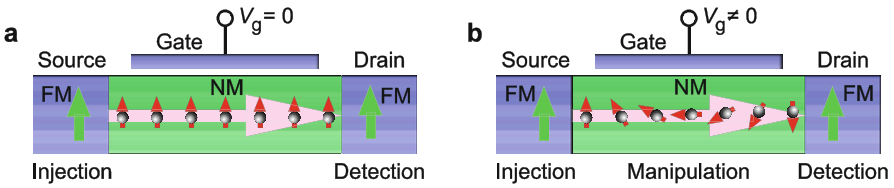


Fig. 20 Schematic of a spin field-effect transistor. Source and drain are ferromagnets (*FM*), *green arrows* indicate their parallel magnetic moments. (**a**) Electrons (*circles*) are injected from the source into the nonmagnetic (*NM*) semiconductor gate and move ballistically along the gate channel with their spins (*small red arrows*) aligned by the source magnetization. (**b**) With an applied gate voltage V_g, the electron spins precess about an axis out of the drawing plane; the current drops if the spins are not parallel to the drain magnetization

The term *spin* may refer to either the spin of a single carrier or to the average spin of an ensemble of carriers. In spin electronics, a spin polarization is created by generating a nonequilibrium of the spin polulation of the carriers. This can be achieved by optical excitation or by electrical spin injection. In optical excitation, the angular momentum of circularly polarized photons is transferred to the carriers; see below and Sect. 3.1 of ▶ Chap. 32, "Dynamic Processes." We will mainly discuss electrically created spin polarization, which is more convenient for device applications.

The Spin Field-Effect Transitor

The spin field-effect transitor (spin FET) is an instructive example for the mechanisms employed in a spintronic device. The structure of such a spin FET depicted in Fig. 20 is similar to that of a conventional FET, where the current through the gate is controlled by the electric field produced by a gate voltage. The current in the spin FET is also controlled by the gate voltage, but the pysical mechanism is different (Datta and Das 1990; Chuang et al. 2015). The spins of the carriers injected from the ferromagnetic source are spin polarized, and they keep this spin polarization in the semiconductor gate material. The structure of the gate must allow for a ballistic carrier transit to the drain without scattering; this can be achieved for electrons in a two-dimensional electron gas (2DEG, see Sect. 1.1 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors") and a short gate length. When a voltage is applied to the gate, the electrons propagating in the corresponding electric field experience an effective magnetic field arising from spin-orbit coupling (*Rashba effect*). The spins of the electrons then precess about this effective magnetic field, with the degree of spin precession controlled by the gate voltage and the travelled distance. When the electrons enter the drain, they experience minimum resistance if their spin is parallel to the magnetic moment of the ferromagnetic drain material; the current through the spin FET is then maximum. However, if their spin is misaligned, the electrons are scattered and the current is small. The drain thereby acts as a detector for the spin polarization.

From the operation principle of the spin FET outlined above, we infer the generic processes of spintronics:

- Generation of a nonequilibrium spin polarization
- Transport of spin-polarized carriers in the active medium
- Manipulation or storage of the polarized spin population
- Detection of the spin population

In the following, we discuss these processes in more detail.

3.1 Generation of a Spin Polarization

There are several means to create an unequilibrium in the spin population of an ensemble of particles carrying such spin. The ensemble may refer to carriers such as electrons or holes, but also to nuclei or any kind of excitation (Žutić et al. 2004). The alignment of the spins is generally defined with respect to a chosen axis of quantization. Such an axis is, e.g., given by an external magnetic field, or by the magnetization of a solid, or by the direction of carrier momentum. The spin-resolved component is usually indicated by ↑ (or +) for spin up and ↓ (or −) for spin down with respect to the chosen axis.

A general definition of the spin polarization is given by

$$P_x \equiv \frac{x_\uparrow - x_\downarrow}{x_\uparrow + x_\downarrow}. \tag{22}$$

Here x denotes the measured physical quantity; x may be, e.g., the conductivity, the charge current, or the carrier density. For an efficient manipulation of P_x, the active medium usually needs to be in a state of nonequilibrium. Such a state differs from the generally existing spin polarization of ferromagnets in equilibrium.

In a ferromagnetic metal or semiconductor, the magnetization generated by the magnetic atoms leads to a difference in the number of free cariers with their magnetic moment either parallel or antiparallel to the magnetization. In ferromagnets, carriers with a parallel spin orientation are referred to as *spin up* (↑), and they are called carriers with *majority spin*[17]. Spin-up carriers have a lower energy than spin-down carriers, and their population is correspondingly larger. This is shown for spin-polarized *d* states in Figs. 16, 17, 18, and 19.

Optical Generation of Spin Polarization

In semiconductors, a spin-polarization of free carriers can be generated by optical excitation with circularly polarized light. An optically generated spin polarization is also referred to as *optical spin orientation*. For illustration, we consider direct optical interband transitions in a zincblende semiconductor with a band structure depicted in

[17]In *semiconductors*, the terms *majority* and *minority* generally refer to the *charge* of the carriers; the symbols ↑ and ↓ then denote the quantum number m_j with respect to the axis of an external magnetic field, or – in optical measurements – the direction of light propagation.

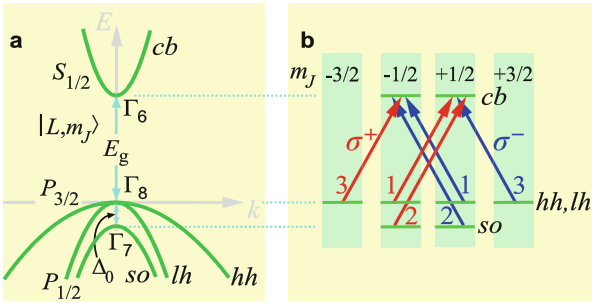


Fig. 21 Optical transitions in a direct-bandgap zincblende semiconductor. (**a**) Band structure with symmetries Γ_i at $k = 0$. (**b**) Selection rules for transitions between the m_J levels of valence and conduction band for circularly polarized light σ^+ (*red arrows*) and σ^- (*blue arrows*); *numbers* at the transitions denote relative dipole strengths

Fig. 21a. A photon inducing a transition in such a semiconductor carries a spin of 1 with respect to the projection of the angular momentum in the direction of propagation (in units of $\hbar$). The representation of the dipole operator of $\sigma^{+/-}$ circularly polarized light is proportional to the spherical harmonic $Y_{l=1}^{m=\pm1}$. For calculating, e.g., the intensity ratio between the σ^+ transition from a heavy-hole subband ($m_j = \pm 3/2$) to the conduction band ($m_j = \pm 1/2$) and that of a light-hole subband ($m_j = \pm 1/2$) to the conduction band, we insert the wavefunctions of S and P symmetry into the respective matrix elements (Žutić et al. 2004); we then obtain for direct optical transitions at $k = 0$

$$\frac{I_{\sigma^+}^{(hh)}}{I_{\sigma^+}^{(lh)}} = \frac{\left|\left\langle \tfrac{1}{2},\ -\tfrac{1}{2}\right| Y_1^{+1} \left|\tfrac{3}{2},\ -\tfrac{3}{2}\right\rangle\right|^2}{\left|\left\langle \tfrac{1}{2},\ +\tfrac{1}{2}\right| Y_1^{+1} \left|\tfrac{3}{2},\ -\tfrac{1}{2}\right\rangle\right|^2} = 3. \tag{23}$$

The intensity ratios of the other transitions are obtained accordingly; transitions complying with the selection rules of angular momentum conservation ($\Delta m_J = 1$) are indicated in Fig. 21b, along with their relative dipole strengths; the numbers apply for both optical absorption and emission.

The optical excitation produces a nonequilibrium in the spin population for both electrons and holes. This spin polarization relaxes rapidly for holes due to the strong spin-orbit coupling in the valence band, but for electrons, the spin-relaxation time is significantly larger – see Sect. 3 of ▶ Chap. 32, "Dynamic Processes." We note from the relative dipole strengths shown in Fig. 21b that the generated spin polarization depends on the photon energy. For an energy $E_g < h\nu < E_g + \Delta_0$, only transitions from the heavy and light hole subbands occur, while for $h\nu > E_g + \Delta_0$, also transitions from the split-off band contribute[18]. If we denote the density of electrons excited with σ^+photons and polarized parallel to the direction of light propagation ($m_J = +½$) with n_+, and those with antiparallel polarization ($m_J = -½$) with n_-, we obtain for σ^+ transitions involving only heavy and light holes a spin polarization

[18]For GaAs, e.g., the energy of the spin-orbit interaction in the valence band is $\Delta_0 = 0.34$ eV.

$$P_n^{(\sigma^+, hh+lh)} \equiv \frac{n_+ - n_-}{n_+ + n_-} = \frac{1-3}{1+3} = -\frac{2}{4} = -0.5; \tag{24}$$

the sign in the result of (24) indicates a spin oriented antiparallel to the light propagation. For σ^+photons with $h\nu > E_g + \Delta_0$, we analogously obtain

$$P_n^{(\sigma^+, hh+lh+so)} = \frac{(1+2)-3}{(1+2)+3} = \frac{0}{6} = 0. \tag{25}$$

With the additional contribution of the split-off subband, the ensemble of optically excited electrons is no longer spin polarized.

If the degeneracy of heavy and light holes in the zincblende semiconductor is lifted, e.g., by strain in a quantum well, the spin polarization can substially be inceased. If electrons are excited by σ^+ radiation from heavy-holes only, complete spin polarization $P_n^{(\sigma^+, hh)} = -1$ will result. Experimentally a degree of polarization over 0.8 was reported for resonant excitation of the strain-split hh_1-e_1 transition of an InGaAs/GaAs quantum well (Vasil'ev et al. 1993).

There are various other methods for optically inducing spin polarization including resonance techniques; see Sect. 3.3.

Spin Injection

In ferromagnets exists a significant degree of spin polarization *in equilibrium*. The current in such solids is carried by two independent and unequal parts with different spin projections: electrons of majority spin with their magnetic moment parallel to the magnetization and those of minority spin and antiparallel alignment. Transport experiments with junctions of a ferromagnet to a nonmagnetic solid such as a superconductor or an insulating tunnel barrier demonstrate that the spin polarization of the current is preserved after leaving the ferromagnet. This finding can be utilized to accumulate spin-polarized carriers in the nonmagnetic solid by electrical injection.

The principle of spin injection is illustrated in the scheme[19] Fig. 22. We consider a junction between a ferromagnetic and a nonmagnetic metal [20], and neglect any interface effects and spin depolarization. When a small bias is applied, both charge and spin are transported across the junction. The bands in the spin-resolved density of states of the ferromagnet are unequally filled, due to a much larger density of states $N_\uparrow$ at Fermi energy for majority carriers than $N_\downarrow$ for minority carriers[21]. More

[19]The schematic band structure of a typical transition-metal ferromagnet depicts separately the spatially extended 4*s* states with their large energy spread and the more localized 3*d* states, although they are hybridized in a solid. The itinerant 3*d* electrons have a much higher density of states, but they are substantially less mobile than the 4*s* electrons.

[20]Such a junction is usually made with metals, but it may also be made using a ferromagnetic diluted magnetic semiconductor and a degenerate semiconductor.

[21]The vertical displacement between spin-up and spin-down densities of states originates from the exchange-energy splitting between the spin-up and spin-down energy bands.

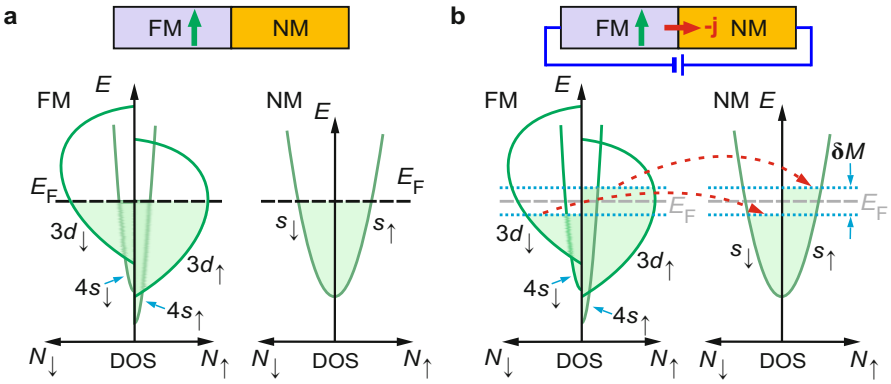


Fig. 22 Concept of spin injection from a ferromagnetic (*FM*) to a nonmagnetic (*NM*) solid with a common junction. The *green arrows* in the top schematics indicate majority spin in the ferromagnet. (**a**) FM and NM in equilibrium at zero bias. Spin-up states in the FM have a down-shifted density of states (*DOS*) and are more populated at Fermi energy E_F than spin-down states. (**b**) At applied bias, one spin population accumulates in the nonmagnetic solid and a nonequilibrium magnetization δM is generated; the *gray dashed line* denotes the Fermi level at equilibrium

spin-up than spin-down electrons are hence injected into the nonmagnetic metal. The magnetization current density J_m is proportional to the electric current density J_e:

$$J_m \propto \left(\frac{\mu_B}{e}\right) J_e, \tag{26}$$

where e the electron charge. The surplus of spin-up electrons leads to a magnetization δM in the nonmagnetic solid, given by the number of these electrons times the Bohr magneton μ_B.

Spin injection is usually not very efficient. Electrons in the ferromagnet are normally only partially spin polarized at the Fermi energy, the Fermi velocities in a ferro- and a nonmagnetic metal are different, and only a part of the spins is transferred across the interface. Many structures employ metal-ferromagnet/semiconductor interfaces (Adelmann et al. 2005; Jiang et al. 2005). The conductance matching can be improved applying an all-semiconductor junction with a diluted magnetic and a nonmagnetic semiconductor[22] (Schmidt 2005). Semiconductors doped with magnetic ions such as $Ga_{1-x}Mn_xAs$ or $Cd_{1-x}Mn_xTe$ offer versatile opportunities for tuning both magnetic and electrical properties. The exchange interaction and the onset of ferromagnetism can be controlled by the carrier density and hence by electric fields. It should be noted that different probes for spin polarization often

[22]A drawback of this approach is the usually low ferromagnetic transition temperature of ferromagnetic semiconductors, requiring operation below room temperature. In II–VI compounds, Mn^{2+} is isovalent with nonmagnetic cations; it provides only spin doping and can be introduced in high composition. In III–Vs, it yields both spin and carrier doping; the low solubility complicates growth due to cluster formation of metallic inclusions. Diluted magnetic semiconductors still serve as a valuable test bench for new and exploratory physics and ideas (Ohno 2010).

yield significantly different values; the measured polarization is affected by differences between bulk and surface densities of states of the ferromagnet and details of the interfacial properties.

3.2 Spin-Polarized Transport

The spin polarization generated in a nonmagnetic solid does not persist for long. The spin of electrons relaxes on a typical time scale of nanoseconds, and that of holes is typically two orders of magnitude faster. The spin-relaxing processes mostly involve spin-orbit coupling and momentum scattering by various mechanisms discussed in Sect. 3 of ▶ Chap. 32, "Dynamic Processes." In steady state, the spin polarization results from a balance between generated spin-polarized carriers and loss by spin relaxation. In nondegenerate semiconductors, the finite carrier lifetime $\tau_{n/p}$ leads to an additional depolarization, yielding an *effective* spin relaxation time $\tau^{-1} = \tau_s^{-1} + \tau_{n/p}^{-1}$, where τ_S denotes the pure spin-relaxation time. In the example illustrated in Fig. 22, the initial magnetization created by spin injection decreases with a characteristic spin-diffusion length L_s. The spatial distribution of the nonequilibrium magnetization δM injected into the nonmagnetic solid is depicted in Fig. 23.

The spin-diffusion length L_s in the nonmagnetic solid is determined by the spin-relaxation time τ_S,

$$L_s = \sqrt{\overline{D}\tau_s}, \tag{27}$$

where the spin-averaged diffusion coefficient is given by

$$\overline{D} = \frac{\sigma_\downarrow D_\uparrow + \sigma_\uparrow D_\downarrow}{\sigma_\uparrow + \sigma_\downarrow}. \tag{28}$$

Eq. 28 is based on the *two-current model* (Mott 1936; Fert and Campbell 1968). According to this model, the current in spin-polarized solids is carried by two independent and unequal parts with different spin projections of their electrons: electrons of majority spin with their magnetic moment parallel to the (permanent or nonequilibriuim) magnetization and those of minority spin with antiparallel

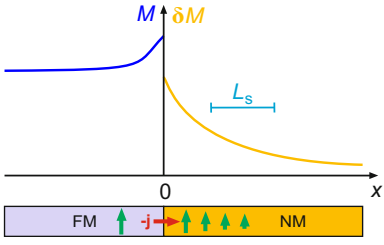


Fig. 23 Distribution of the magnetization M in a ferromagnet (*FM*) and the nonequilibriuim magnetization δM generated by spin injection into a nonmagnetic solid (*NM*) near the common interface at $x = 0$. L_s is the characteristic spin-diffusion length

alignment. The two channels experience different scattering processes and hence require different individual diffusion coefficients $D_\uparrow$ and $D_\downarrow$; the total conductivity is expressed by $\sigma = \sigma_\uparrow + \sigma_\downarrow$.

The two-current model provides a reasonabe description of the spin density in magnetic structures, but it does not consider the *phase* of the spin distribution; it can hence not interpret properties employing coherent spin phenoma. The two-current model provides an explanation for the behavior of magnetoresistive devices. Structures of particular importance are tunnel junctions and ferromagnetic layer stacks with nonmagnetic spacers, which both are implemented in read heads of hard-disc drives, nonvolatile solid-state memories, and sensing applications.

Magnetic Tunnel Junctions

A magnetic tunnel junction consists of two conductive ferromagnetic layers separated by a thin insulator of typically a few nm thickness. A tunneling of electrons through the insulator is more likely for a parallel orientation ($\uparrow\uparrow$) of the magnetizations than for antiparallel ($\uparrow\downarrow$) alignment. The magnetizations of the two ferromagnetic layers can be reversed individually by a small external magnetic field ($\sim$1 mT); the electrical resistance of such a junction can hence be switched between a low and a high resistance.

Tunneling magnetoresistance (TMR) The *tunneling magnetoresistance* is defined as the relative change of resistance R between antiparallel and parallel magnetizations of the two layers,

$$TMR \equiv \left(\frac{\Delta R}{R}\right)_{\text{TMR}} = \frac{R_{\uparrow\downarrow} - R_{\uparrow\uparrow}}{R_{\uparrow\uparrow}} = \frac{G_{\uparrow\uparrow} - G_{\uparrow\downarrow}}{G_{\uparrow\downarrow}}, \tag{29}$$

where $G = 1/R$ is the conductance.

The TMR effect can be explained by the two-current model (Jullière 1975). We consider in this model the spin polarization, given in the respective ferromagnet by the relative difference in the density of states at the Fermi energy for the majority and minority spins

$$P_N = \frac{N_\uparrow - N_\downarrow}{N_\uparrow + N_\downarrow}. \tag{30}$$

The tunneling current is composed of the contributions of majority and minority spins; the conductances for parallel and antiparallel aligment of the two ferromagnets FM(1) and FM(2) are therefor given by

$$G_{\uparrow\uparrow} \propto N_\uparrow^{(1)} N_\uparrow^{(2)} + N_\downarrow^{(1)} N_\downarrow^{(2)} \text{ and } G_{\uparrow\downarrow} \propto N_\uparrow^{(1)} N_\downarrow^{(2)} + N_\downarrow^{(1)} N_\uparrow^{(2)}. \tag{31}$$

Inserting (31) in (29) and using (30), we obtain

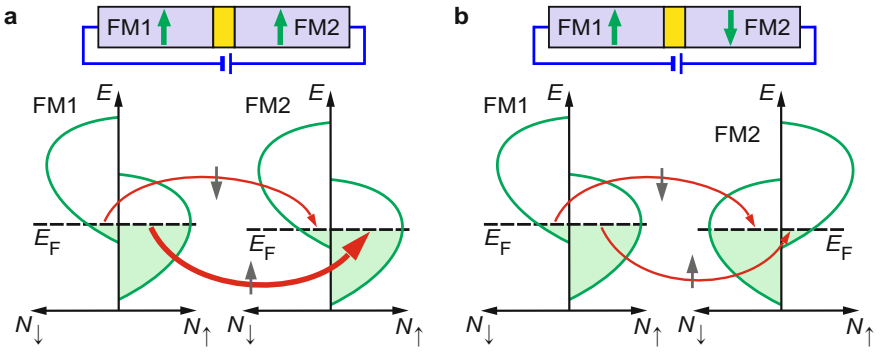


Fig. 24 Spin-resolved density of states for two ferromagnets with (**a**) parallel and (**b**) antiparallel magnetization, separated by a thin tunneling barrier (*yellow*). *Green arrows* indicate the orientations of the magnetizations, and *gray arrows* the spin of tunneling electrons. The *red arrows* designate the tunneling currents with the *thicker arrow* in (**a**) indicating strongest current

$$TMR = \frac{2\,P_N^{(1)} P_N^{(2)}}{1 - P_N^{(1)} P_N^{(2)}}. \tag{32}$$

The dependence of the tunneling current on the relative magnetization orientations can be inferred from the illustration of the spin-resolved density of states. The electron tunneling for parallel and antiparallel aligment of the ferromagnets is depicted in Fig. 24. A large conductance is only observed if both magnetizations are parallel, and the current is then mainly from spin-up electrons with large combined density of states at E_F in both ferromagnets (Fig. 24a).

Experimental magnetoresitance results of two ferrometallic films[23] separated by a thin insolating layer are shown in Fig. 25. In these structures, different ferromagnets with different coercive fields H_c were used on the two sides sides of the insulator (Moodera et al. 1995). At high external fields H,[24] the magnetizations of both ferromagnets are saturated and parallely aligned in the field direction, yielding a high tunneling probability and corresponding conductance (low resistance ΔR). As H decreases (sweep down), the magnetization in the film with lower H_c reverses (Co at $H < 0$ in Fig. 25a), whereas H_c in the film with higher H_c (CoFe) remains constant. The magnetizations in the films are now antiparallel, and the tunneling probability hence decreases, i.e., the resistance ΔR increases. When the sweep down continues, also the magnetization of the film with higher H_c reverses; with both magnetizations reversed and parallel again, the resistance decreases at further down sweep. In a sweep up of the external field H, we observe an analogous behavior at positive H values.

The typical zero-field conductance G of a comparable sample (junctions with the structure of Fig. 25a show identical results) is given in Fig. 25b. At low bias (up to

[23]Planar thermally evaporated films of ~100 Å thickness with bulk-like behavior were used.

[24]The magnetic field H is given in Oersted (Oe). The conversion to SI units is 1 Oe = 79.5775 A/m. Multiplication with the vacuum permeability yields the magnetic induction B: $\mu_0 \times 1$ Oe = 10^{-4}T.

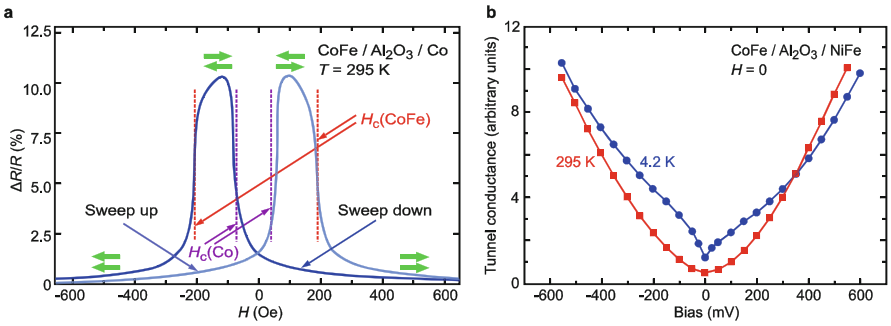


Fig. 25 Magnetoresistance in ferromagnetic tunnel junctions with a ~16 Å thick Al_2O_3 insulator. (**a**) Resistance depending on the magnetic field H oriented parallel to the interfaces between ferromagnets and insulator. *Green arrows* indicate the orientation of the magnetization in the two ferromagnets CoFe and Co; *dashed lines* mark their respective coercive fields H_c. (**b**) Zero-field conductance at varied bias for low and room temperature. (After Moodera et al. 1995)

~15 mV), G is nearly constant as expected for a tunnel junction. The increase at higher bias is attributed to magnon contributions, and the dip near zero bias in the low-temperature curve is attributed to spin flip scattering due to localization effects in the barrier or at the interface (Moodera and Mathon 1999).

First-principle calculations indicate that tunneling is more complicated than suggested by the simple two-current model presented above. Particularly the phase of the tunneling carriers must be taken into account, and the symmetry of the Bloch states in the ferromagnets and of the evanescent states in the barrier are important. The tunneling electrons need not only to get through the barrier, there must also be a state of the correct symmetry on the other side to accept them (Butler et al. 2001). This should at least partly account for the commonly observed decrease in the tunneling magnetoresistance with bias: as the bias increases, the states on the opposite side of the barrier for parallel alignment differ more.

The coherency of electron wavefunctions can be conserved during tunneling, if crystalline tunneling structures with low defect densities are used. Fabricating epitaxial heterostructures with single-crystal MgO insulating layers, an experimental evidence for the conservation of the wavefunction coherency across the tunnel barrier was provided (Yuasa et al. 2004). Results given in Fig. 26 show the expected increase of the tunneling magnetoresistance for thicker barriers[25]: at thicker barriers, tunneling is dominated by Fe states with momentum vectors normal to the Fe/MgO interface (perpendicular tunneling); these are Fe states of the Δ_1 band, which have highest spin polarization (Butler 2001; Mathon and Umerski 2001). The (temperature and sample independent) oscillatory shape of the magnetoresistance is assigned to the

[25]Simultaneously the tunneling resistance (not the *ratio*) increases exponentially; an increase over six orders of magnitude was measured as the MgO barrier thickness increases from 1.2 nm to 3.2 nm (Yuasa et al. 2004).

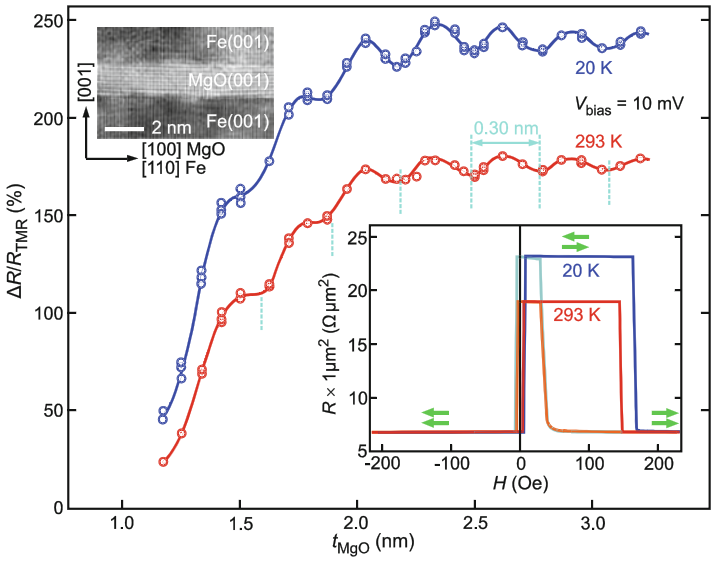


Fig. 26 Tunneling magnetoresistance of epitaxial Fe/MgO/Fe junctions with varied thickness t_{MgO}. The upper inset shows a high-resolution TEM image of a 1.8 nm thick MgO barrier (*bright contrast*) and the adjacent magnetic Fe layers. The *lower inset* shows magnetoresistance curves of junctions with a 2.3 nm thick barrier at low (*blue/light blue*) and room temperature (*red/orange*); *green arrows* indicate the magnetization orientation of the two Fe layers. (After Yuasa et al. 2004)

conservation of the coherency of electron wavefunctions across the tunnel barrier, where evanescent states in MgO originating from Δ_1 and Δ_5 bands of Fe interfere.

Giant Magnetoresistance (GMR) A magnetoresistance depending on the magnetization orientation of adjacent ferromagnetic layers exists also without a separating tunneling barrier. Multilayers composed of alternating ferromagnetic and nonmagnetic *conductive* layers show a large change in the electrical resistance, when the magnetization of the ferromagnetic layers is changed from parallel to antiparallel alignment. The discovery of this effect was recognized by the Physics Nobel Prize in 2007 due to the large impact on applications for digital storage today (Thompson 2008). The effect is based on a spin-dependent scattering in ferromagnetic materials and occurs for both, structures with a current in plane (CIP) and those with the current perpendicular to the plane (CPP) of the interfaces. The principal geometries are illustrated in Fig. 27.

For both geometries, the giant magnetoresistance is defined as a ratio similar to that of the tunneling magnetoresistance in Eq. 29,

$$\left(\frac{\Delta R}{R}\right)_{GMR} \equiv \frac{R_{\uparrow\downarrow} - R_{\uparrow\uparrow}}{R_{\uparrow\uparrow}}; \tag{33}$$

it must be noted that also another definition of GMR is in common, where $R(H)/R_{\uparrow\downarrow}(H = 0)$ is used instead. Both definitions were applied in the two original

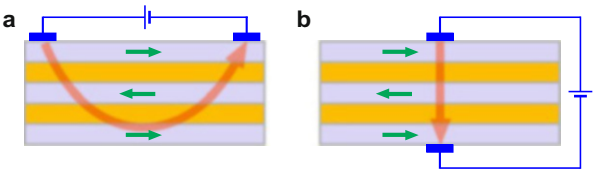


Fig. 27 Schematic of (**a**) the current-in-plane (CIP) and (**b**) the current-perpendicular-to-the-plane (CPP) geometries of multilayer structures displaying the GMR effect. Blue and orange layers denote respectively ferromagnetic and nonmagnetic layers; the green arrows indicate magnetizations. The net current flow (red arrows) of the CIP structure is in the plane of the layers and in the CPP structure perpendicular to the layers

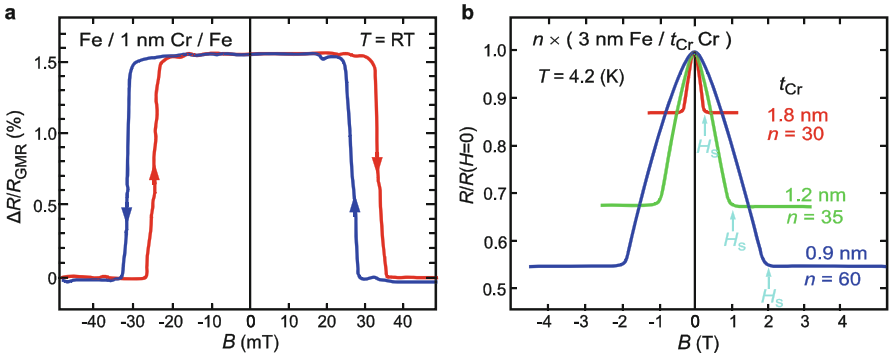


Fig. 28 GMR effect (**a**) in a Fe/Cr/Fe trilayer and room temperature and (**b**) in Fe/Cr/Fe multilayer CIP structures at low temperature. The Fe layers are antiferromagnetically coupled at zero field, H_s in (**b**) marks the external saturation field needed to align the ferromagnetic layers parallely. (After Binasch et al. 1989 and Baibich et al. 1988, respectively)

reports shown in Fig. 28. The measured structures consist of antiferromagentically coupled Fe layers separated by very thin nonmagnetic Cr layers. The electrical resistance of these structures is large[26] at zero external magnetic field. Large external fields overcome the antiferromagnetic coupling and create a parallel magnetization of the ferromagnetic layers, leading to a much lower resistance.

The origin of the spin-dependent scattering in metallic ferromagnets is the spin-split band structure illustrated, e.g., in Fig. 24. The ferromagnetic transition-metals Fe, Ni,

[26]The effect is termed *giant* due to the much smaller anisotropic magnetoresistance (AMR) discovered aready by Lord Kelvin 160 years ago. The AMR effect originates from a difference in the scattering cross-section when the electron current in a single magnetic sample is parallel or perpendicular to magnetically aligned atoms, an effect only on the order of a few per cent.

and Co have unoccupied states in the part-filled d bands for scattering into[27], in contrast to paramagnetic metals like noble metals; in the latter, the d bands are below the Fermi level and provide no unoccupied states for such scattering. The d bands in ferromagnets have different density of states at the Fermi energy and consequently a different scattering probability for spin-up and spin-down electrons, with generally a higher scattering probability for minority electrons. In the current-in-plane (CIP) structure, the mean free path of the electrons must exceed the total thickness of all layers (typically some tens of nanometres); electrons then sample all different layers. In current-perpendicular-to-the-plane (CPP) structures, the spin-diffusion length L_s (Eq. 27) must be larger than the thickness of the layers, cf. Fig. 27.

Using a proper nonmagnetic layer thickness, an antiferromagnetic coupling of the ferromagnetic layers at zero external field H can be accomplished. An itinerant electron is then alternating a majority and a minority carrier, yielding a high resistivity for the minority part. When an external field H exceeding the saturation field H_s is applied, the magetizations in the ferromagnetic layers are parallel and the current is basically carried by majority electrons with a smaller scattering probability.

Applying the two-current model introduced to explain the tunneling magnetoresistance, we can represent the different resistances of the majority and minority channels by resistors $R_\uparrow$ and $R_\downarrow$ with different values in a simple network. The schematic of Fig. 29 depicts the network for a three-layer structure with ferromagnets and either antiparallel or parallel orientation of their magnetization.

The spin-up channel in Fig. 29a is the majority spin channel with a low resistivity ($R_\uparrow$) in the first magnetic layer, but it is the minority-spin channel with a high

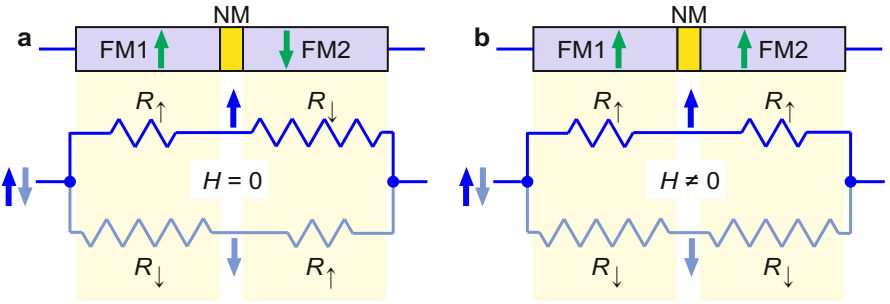


Fig. 29 Scheme with ferromagnetic layers (*FM*) represented by resistances $R_\uparrow$ and $R_\downarrow$, and a connecting conductive nonmagnetic layer (*yellow*); *blue and light blue* channels indicate paths of spin-up (majority) and spin-down (minority) carriers experiencing low ($R_\uparrow$) and high ($R_\downarrow$) resistivity, respectively. (**a**) Antiferromagnetically coupled layers at zero external field; (**b**) layers with parallel magnetization induced by an external field $H \neq 0$

[27]The ferromagnetic rare-earth metals have an analogous spin-split part-filled 4*f* band, but this is deep in the band structure and does not induce significant scattering.

resistivity ($R_\downarrow$) in the second magnetic layer; the spin-down channel has invers properties, being the minority channel in FM1 and the majority channel in FM2. For both spin channels, the serial resistance is high. This is different for the parallel magnetization depicted in Fig. 29b, where the spin-up channel is the majority spin channel with a low resistivity ($R_\uparrow$) in *both* magnetic layers; now the majority electron-channel dominates the conductivity and the combined structure has a low total resistance. The larger the difference between $R_\uparrow$ and $R_\downarrow$, the larger is the ratio in the magnetoresistance.

The basic question in GMR structures is how does the coupling of the ferromagentic layers depend on their mutual distance, i.e., on the thickness of the nonmagnetic spacer layer. It turns out that the underlying exchange coupling of the ferromagnetic layers is an extension of the coupling responsible for magnetic order within magnetic ions in metals[28]. This indirect exchange coupling of localized inner *d*- or *f*-shell electron-spins is mediated by conduction carriers and creates a distance-depending oscillating spin polarization in the vicinity of the magnetic ion. The same applies for thin magnetic layers separated by nonmagnetic spacers, where additionally interference effects occur. An oscillating magnetoresistance resulting from an alternating antiferromagnetic and ferromagnetic coupling for varied spacer thickness is shown in Fig. 30. The multilayer structure is composed of thin magnetic Co layers and separating nonmagnetic Cu layers. The damping of the oscillations for increased nonmagnetic spacer thickness is due to the increasing scattering of conduction electrons in the spacer layers. Similar dependences exist for other combinations such as Fe/Cr or Co/Ru multilayers. All such structures require very thin layers and a high integrity of the interfaces.

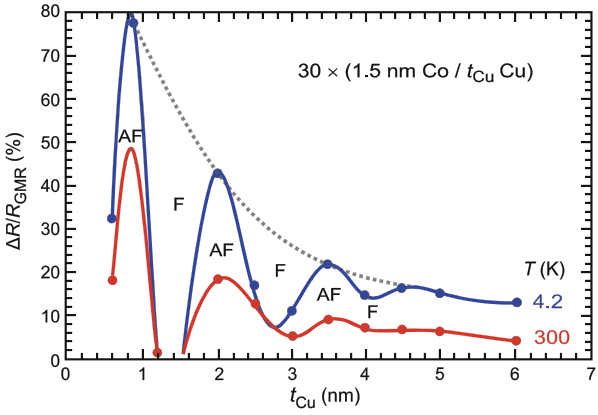


Fig. 30 Variation of the giant magnetoresistance as a function of the thickness of nonmagnetic copper layers for 30-fold (Co/Cu) CIP multilayers. Solid lines are guides to the eye; the dashed line represents the thickness dependence of the giant magnetoresistance at low temperature. (After Mosca et al. 1991)

[28]This Ruderman-Kittel-Kasuya-Yosida (RKKY) interaction was initially described for coupling of nuclear magnetic moments in a metal mediated by the interaction with conduction electrons (Ruderman and Kittel 1954).

Spin-Transfer Torque and Spin Valves

The relative orientation of magnetizations affects the spin-polarized current in the TMR and GMR effects as described above[29]. There exists also the reverse effect: a spin-polarized current can transfer an angular momentum from the carriers to a ferromagnet and alter the orientation of its magnetization, an effect referred to as *spin-transfer torque*; for an introductory review on this effect, see Ralph and Stiles (2008).

The ability to switch the magnetization by a spin-polarized current and thereby change the resistivity is utilized in a *spin valve*. A spin-valve device is a multilayer structure of alternating thin nonmagnetic and weakly coupled magnetic layers, where the magnetic layers can easily be switched from antiferromagnetic to ferromagnetic coupling. The switching is enabled by an asymmetric structure, achieved either by different thicknesses of the magnetic layers or due to two ferromagnetic materials of different coercivities. The thicker layer (or that with larger coercivity) serves as a pinned *reference layer*, and the thinner (or magnetically softer) layer is the *free layer* with variable orientation of its magnetization. The switching is performed either by a weak external magnetic field or by a current of spin-polarized carriers through the structure (Dieny 1994).

For a discussion of the spin-valve operation, we consider an FM1/NM/FM2 trilayer structure with ferromagnetic (FM) and nonmagnetic (NM) layers, such as that depicted in Fig. 29. The NM thickess is chosen to induce a merely *weak* coupling between FM1 and FM2, and FM1 (*reference layer*) is thicker than FM2 (*free layer*). We first assume an antiparallel magnetization of FM1 and FM2. When a current is spin-polarized by FM1 and enters FM2, the spin of its carriers rotates to become aligned with the magnetization of FM2, because FM2 exerts a torque on the

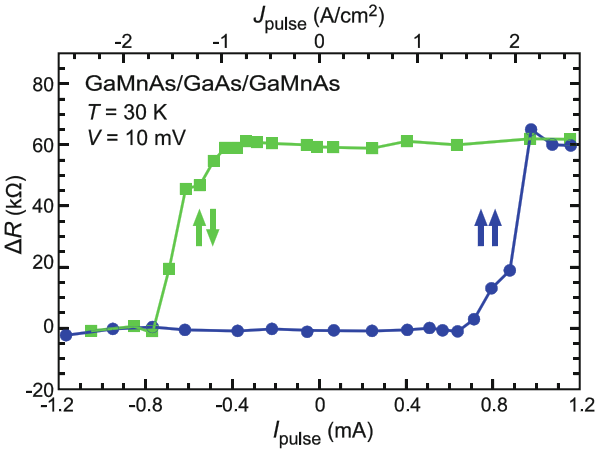


Fig. 31 Resistance change ΔR of a magnetic tunnel junction as a function of 1 ms long pulses I_{pulse}, plotted with respect to the resistance of the junction for parallel magnetization of the ferromagnetic $Ga_{0.953}$ $Mn_{0.047}As$ layers. (After Chiba et al. 2004)

[29]The effects of giant magnetoresistance (GMR) and tunneling magnetoresistance (TMR) are also referred to as spin-valve effects. A spin-valve device utilizing this effect has an asymmetric ferromagnetic structure.

flowing spins to reorient them. Due to the conservation of the total spin, there acts an equal and opposite torque on the FM2 layer by the flowing carriers, which – at sufficient current density – can switch the magnetization of the thin layer FM2 to become aligned with FM1. Now the magnetizations of FM1 and FM2 are parallel and we reverse the current direction; carriers then flow from the *thin* free FM2 layer toward the FM1 layer. The majority carriers pass through the FM2/NM interface, but the minority carriers are scattered back to the free FM2 layer. The transfer of their spin angular momentum to the FM2 magnetization switches the thin FM2 layer again into antiparallel alignment. These magnetization switches can be detected by the GMR or TMR effects, which result from the change in the relative magnetic orientation of the two FM layers.

The current-driven magnetization reversal of a ferromagnetic tunnel junction is illustrated in Fig. 31. It consists of a $Ga_{1x}Mn_xAs/GaAs$ double heterostructure with a 0.1 eV tunnel barrier for holes in the 6 nm thick GaAs barrier; the diluted magnetic semiconductors have $x = 4.7\%$ Mn compositition. The thicknesses of the reference and free layers are 80 and 15 nm, yielding ferromagnetic transition temperatures of 90 and 50 K, respectively. The total magnetic moment of the free layer is hence less than that of the thicker reference layer. Figure 31 clearly demonstrates magnetic switching with 1 ms long current pulses of sufficient density, measured by the difference in resistance between that of parallel magnetization (at $H_{ext} = 0$) and that after application of the pulse[30]; a critical current density for reversal on the order of $10^5 A/cm^2$ is obtained, two orders less than in metal structures (Chiba et al. 2004; Elsen et al. 2006). Still these current densities are comparable, if the difference in the magnitude of the magnetization between $Ga_{1x}Mn_xAs$ and a ferromagnet such as Co in a metallic tunnel junction is taken into account.

In practical spin valves, the ferromagnetic reference system is composed of *two* layers, which form an artificial antiferromagnet: a pinning layer and the antiferromagnetically coupled reference layer (Ennen et al. 2016). The magnetization of the reference layer can thereby be fixed into a certain direction and gets stable against external fields. Such a layer assembly is applied in GMR read heads of hard-disk drives; when two bits with magnetic moments oriented in opposite direction on the moving hard disk pass the spin-valve sensor, the magnetic moment of the free layer is reoriented according to the orientation of the next bit.

Current-induced switching is attractive for a magnetic random-access memory (MRAM); this is a fast, low-power nonvolatile memory with a nondestructive read-out operation, and it can be produced with a comparable density to the conventional DRAM (Khvalkovskiy et al. 2013); for a review, see Bhatti et al. (2017). Current research focuses on perpendicular instead of in-plane junctions and new materials with high anisotropy and low damping constant to achieve thermal stability and writability without sacrificing power consumption (Varghani et al. 2018).

[30]The measurement is performed in steps, because the TMR (ratio) is only large at small bias (<0.1 V), but switching requires a large bias (>0.5 V).

3.3 Optical Spin Detection

Spin detection is most easily performed electrically by the TMR or GMR effects discussed above. This is the standard method of all-metal devices employed in magneto-electronics. In *semiconductors*, the presence of spin-polarized carriers can also be detected by optical means: by absorption, reflection, or luminescence; we briefly point out a few methods.

In the *magneto-optical Kerr effect* (MOKE), a beam of linearly polarized light is reflected at the surface (Choi and Cahill 2014; Gort et al. 2020). The polarization of the reflected beam is rotated if spin-polarized carriers exist: the energy splitting of majority and minoriy carriers leads to a difference in absorption and hence to a difference of refractive indices of left- and right circularly polarized light; the ensuing different phase velocities of these components result in a rotation of the polarization axis and a slight ellipticity. The rotation angle is proportional to the magnetization in the semiconductor. This applies as well for the *transmitted* beam, where the phenomenon is referred to as *Faraday effect* (Sect. 3.2.2 of ▶ Chap. 12, "Photon–Free-Electron Interaction").

In a *spin light-emitting diode* (spin LED), the spin-polarized electrons are transferred from a ferromagnetic electrode into a *pin* diode structure with a quantum-well layer in the intrinsic (*i*) region; there the electrons recombine with unpolarized holes (Jiang et al. 2005). Circularly polarized light is emitted due to the spin polarization of the electrons and angular momentum conservation during the recombination process. Virtually pure circular polarization was demonstrated at room temperature with a GaAs/AlGaAs LED comprising a ferromagnetic Fe spin injector with an AlO_x tunnel barrier (Nishizawa et al. 2017). Since the optical selection rules illustrated in Fig. 21 apply for optical absorption *and* emission, the effect used for optical generation of a spin polarization can be reversed and also applied for optical spin detection.

A sensitive probe for carrier spins is the *ferromagnetic resonance* (FMR), where the loss of a microwave power due to absorbtion by the precessing magnetization (Larmor precession) in the sample is measured – analogous to the electron paramagnetic resonance outlined in Sect. 2.2 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients." Absorption occurs, if microwave and Larmor frequency are in resonance and the polarization of the wave matches the orientation of the magnetization. For a recent application to evaluate the spin-torque efficiency, see Karimeddiny et al. (2020).

Sensitive *spin-rectification effects* occur more generally in various resonances of spin dynamics. Spin rectification is the generation of a dc voltage due to the nonlinear coupling of an oscillating current to an oscillating magnetoresistance in magnetic structures; the effect is analogous to optical rectification outlined in Sect. 3.1.6 of ▶ Chap. 10, "Interaction of Light with Solids." For a review on spin rectification, see Harder et al. (2016).

4 Summary

Magnetic properties introduced into nonmagnetic semiconductors by substitutional paramagnetic 3*d* transition-metal cations or rare-earth cations are described by the magnetic susceptibility χ. Above a critical Curie (or Néel) temperature, thermal energy disturbs magnetic coupling and a paramagnetic behavior with an inverse dependence of χ on T is found; below this temperature spontaneous magnetic ordering of the magnetic moments located at the paramagnetic ions occurs, and χ depends on the orientation and the temporal evolution of the magnetic field. The magnetic moments are coupled by various, simultaneously acting exchange interactions. Superexchange mediated by *p* states of nonmagnetic anion ligands favors antiferromagnetism, while double exchange and *p–d* exchange favor ferromagnetism with a parallel alignment.

Diluted magnetic semiconductors represent interesting solids with the ability to control magnetic properties via electric fields and carriers introduced by injection or doping. The magnetic impurities represent substitutional *alloy* components on the cation sublattice and introduce both, locally strongly confined and more extended, hybridizing states into the host semiconductor. The ordering temperature depends on the alloy composition and the Fermi level, which controls the kind of prevailing interaction. The achievement of magnetic ordering well above ambient temperature proved challenging and is an active field of research.

In spin-based (opto-) electronics the carrier spin is utilized for current control by creating a nonequilibrium spin polarization. The spin polarization is then manipulated or stored and eventually electrically or optically detected. When phase effects can be neglected, the transport of spin-polarized carriers is well described by a two-current model assuming two independent channels of different spin projections, which experience different scattering processes. Typical spintronic structures comprise two ferromagnetic layers, which are separated by a thin nonmagnetic spacer layer. For insulating spacers and antiparallel aligned ferromagnets a large difference in resistance of the two spin-polarized channels in found, referred to as tunneling magnetoresistance. Similarly a giant magnetoresistance is observed for a thin *conductive* spacer. Both effects are widely applied in hard-disk heads for magnetic data storage. The orientation of the ferromagnetic magnetization can be altered by spin-transfer torque exerted by a spin-polarized current, an effect employed in nonvolatile magnetoresistive random access memories. Although current commercial devices are based on metal structures, diluted magnetic semiconductors promise potential for quantum-information processing due to a weak coupling of spin to charge and phonons, and a related relatively slow dephasing. Furthermore, these semiconductors serve as a valuable test bench for new and exploratory physics.

References

Adelmann C, Lou X, Strand J, Palmstrøm CJ, Crowell PA (2005) Spin injection and relaxation in ferromagnet-semiconductor heterostructures. Phys Rev B 71:121301

Awschalom DD, Loss D, Samarth N (eds) (2002) Semiconductor spintronics and quantum computation. Springer, Berlin

Balzarotti A, Czyzyk M, Kisiel A, Motta N, Podgòrny M, Zimnal-Starnawska M (1984) Local structure of ternary semiconducting random solid solutions: extended x-ray-absorption fine structure of $Cd_{1-x}Mn_xTe$. Phys Rev B 30:2295

Baibich MN, Broto JM, Fert A, Nguyen van Dau F, Petroff F, Eitenne P, Creuzet G, Friederich A, Chazelas J (1988) Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. Phys Rev Lett 61:2472

Bhatti S, Sbiaa R, Hirohata A, Ohno H, Fukami S, Piramanayagam SN (2017) Spintronics based random access memory: a review. Mater Today 20:530

Binasch G, Grünberg P, Saurenbach F, Zinn W (1989) Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange. Phys Rev B 39:4828

Butler WH, Zhang X-G, Schulthess TC, MacLaren JM (2001) Spin-dependent tunneling conductance of Fe|MgO|Fe sandwiches. Phys Rev B 63:054416

Chiba D, Sato Y, Kita T, Matsukura F, Ohno H (2004) Current-driven magnetization reversal in a ferromagnetic semiconductor (Ga,Mn)As/GaAs/(Ga,Mn)As tunnel junction. Phys Rev Lett 93:216602

Choi G-M, Cahill DG (2014) Kerr rotation in Cu, Ag, and Au driven by spin accumulation and spin-orbit coupling. Phys Rev B 90:214432

Chuang P, Ho S, Smith L et al (2015) All-electric all-semiconductor spin field-effect transistors. Nat Nanotech 10:35

de Seze L (1977) Antiferromagnetic dilute bond Ising model exhibiting a spin-glass phase transition. J Phys C: Sol State Phys 10:L353

Datta S, Das B (1990) Electronic analog of the electro-optic modulator. Appl Phys Lett 56:665

Dieny J (1994) Giant magnetoresistance in spin-valve multilayers. Magn Magn Mater 136:335

Dietl T, Ohno H (2014) Dilute ferromagnetic semiconductors: Physics and spintronic structures. Rev Mod Phys 86:187

Dietl T, Ohno H, Matsukura F, Cibert J, Ferrand D (2000) Zener model description of ferromagnetism in zinc-blende magnetic semiconductors. Science 287:1019

Dietl T, Awschalom DD, Kaminska M, Ohno H (eds) (2008) Spintronics, Semiconductors and semimetals, vol 82. Academic Press/Elsevier, Amsterdam

Dyakonov MI (ed) (2017) Spin physics in semiconductors, 2nd edn. Springer Nature, Cham

Elsen M, Boulle O, George J-M, Jaffrès H, Mattana R, Cros V, Fert A, Lemaitre A, Giraud R, Faini G (2006) Spin transfer experiments on (Ga,Mn)As/(In,Ga)As/(Ga,Mn)As tunnel junctions. Phys Rev B 73:035303

Ennen I, Kappe D, Rempel T, Glenske C, Hütten A (2016) Giant magnetoresistance: basic concepts, microstructure, magnetic interactions and applications. Sensors 16:904

Felser C, Fecher GH (eds) (2013) Spintronics, from materials to devices. Springer, Heidelberg

Fert A, Campbell IA (1968) Two-current conduction in nickel. Phys Rev Lett 21:1190

Furdyna JK, Kossut J (eds) (1988) Diluted magnetic semiconductors, Semiconductors and semimetals, vol 25. Academic Press, Boston

Giriat W, Furdyna JK (1988) Crystal structure, composition, and materials preparation of diluted magnetic semiconductors. In: Furdyna JK, Kossut J (eds) Diluted magnetic semiconductors, Semiconductors and semimetals, vol 25. Academic Press, Boston, pp 1–34

Goodenough JB (1955) Theory of the role of covalence in the perovskite-type manganites [La, *M* (II)]MnO_3. Phys Rev 100:564

Gort R, Bühlmann K, Saerens G, Däster S, Vaterlaus A, Acremann Y (2020) Ultrafast magnetism: the magneto-optical Kerr effect and conduction electrons. Appl Phys Lett 116:112404

Harder M, Gui Y, Hu C-M (2016) Electrical detection of magnetization dynamics via spin rectification effects. Phys Rep 661:1

Hirohata A, Sukegawa H, Yanagihara H, Žutić I, Seki T, Mizukami S, Swaminathan R (2015) Roadmap for emerging materials for spintronic device applications. IEEE Trans Magn 51:0800511

Jiang X, Wang R, Shelby RM, Macfarlane RM, Bank SR, Harris JS, Parkin SSP (2005) Phys Rev Lett 94:056601

Johnson M (ed) (2004) Magnetoelectronics. Elsevier, Amsterdam
Jullière M (1975) Tunneling between ferromagnetic films. Phys Lett 54A:225
Jungwirth T, Wang KY, Mašek J, Edmonds KW, König J, Sinova J, Polini M, Goncharuk NA, MacDonald AH, Sawicki M, Rushforth AW, Campion RP, Zhao LX, Foxon CT, Gallagher BL (2005) Prospects for high temperature ferromagnetism in (Ga, Mn)As semiconductors. Phys Rev B 72:165204
Kanamori J (1959) Superexchange interaction and symmetry properties of electron orbitals. J Phys Chem Solids 10:87
Karimeddiny S, Mittelstaedt JA, Buhrman RA, Ralph DC (2020) Transverse and longitudinal spin-torque ferromagnetic resonance for improved measurement of spin-orbit torque. Phys Rev Appl 14:024024
Khvalkovskiy AV, Apalkov D, Watts S, Chepulskii R, Beach RS, Ong A, Tang X, Driskill-Smith A, Butler WH, Visscher PB, Lottis D, Chen E, Nikitin V, Krounbi M (2013) Basic principles of STT-MRAM cell operation in memory arrays. J Phys D Appl Phys 46:074001
Kreitman MM, Barnet DL (1965) Probability tables for clusters of foreign atoms in simple lattices assuming next-nearest-neighbor interactions. J Chem Phys 43:364
Mahadevan P, Zunger A (2004) First-principles investigation of the assumptions underlying model-Hamiltonian approaches to ferromagnetism of 3*d* impurities in III-V semiconductors. Phys Rev B 69:115211
Mathon J, Umerski A (2001) Theory of tunneling magnetoresistance of an epitaxial Fe/MgO/Fe (001) junction. Phys Rev B 63:220403
Moodera J, Mathon G (1999) Spin polarized tunneling in ferromagnetic junctions. J Magn Magn Mater 200:248 (1999)
Moodera JS, Kinder LR, Wong TM, Meservey R (1995) Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions. Phys Rev Lett 74:3273
Mosca DH, Petroff F, Fert A, Schroeder PA, Pratt Jr WP, Laloee R (1991) Oscillatory interlayer coupling and giant magnetoresistance in Co/Cu multilayers. J Mag Mag Mater 94:L1
Mott NF (1936) The electrical conductivity of transition metals. Proc R Soc Lond A 153:699
Nagata S, Galazka RR, Mullin DP, Akbarzadeh H, Khattak GD, Furdyna JK, Keesom PH (1980) Magnetic susceptibility, specific heat, and the spin-glass transition in $Hg_{1-x}Mn_xTe$. Phys Rev B 22:3331
Nishizawa N, Nishibayashi K, Munekata H (2017) Pure circular polarization electroluminescence at room temperature with spin-polarized light-emitting diodes. PNAS 114:1783
Ohno H (2010) A window on the future of spintronics. Nat Mater 9:952
Ohno H, Chiba D, Matsukura F, Omiya T, Abe E, Dietl T, Ohno Y, Ohtani K (2000) Electric-field control of ferromagnetism. Nature 408:944
Ohno H, Shen A, Matsukura F, Oiwa A, Endo A, Katsumoto S, Iye Y (1996) (Ga, Mn)As: a new diluted magnetic semiconductor based on GaAs. Appl Phys Lett 69:363
Oseroff S (1982) Magnetic susceptibility and EPR measurements in concentrated spin-glasses: $Cd_{1-x}Mn_xTe$ and $Cd_{1-x}Mn_xSe$. Phys Rev B 25:6584
Oseroff S, Keesom PH (1988) Magnetic properties: macroscopic studies. In: Furdyna JK, Kossut J (eds) Diluted magnetic semiconductors, Semiconductors and semimetals, vol 25. Academic Press, Boston, pp 73–123
Pohl UW, Busse W (1989) Probability tables for small clusters of impurity atoms in sc, bcc and fcc lattices assuming long range interaction. J Chem Phys 90:6877
Ralph DC, Stiles MD (2008) Spin transfer torques. J Magn Magn Mater 320:1190
Ruderman MA, Kittel C (1954) Indirect exchange coupling of nuclear magnetic moments by conduction electrons. Phys Rev 96:99
Saito H, Zayets V, Yamagata S, Ando K (2003) Room-temperature ferromagnetism in a II-VI diluted magnetic semiconductor $Zn_{1-x}Cr_xTe$, Phys Rev Lett 90:207202
Sato K, Katayama-Yoshida H (2002) First principles materials design for semiconductor spintronics. Semicond Sci Technol 17:367

Sato K, Bergqvist L, Kudrnovský J, Dederichs PH, Eriksson O, Turek I, Sanyal B, Bouzerar G, Katayama-Yoshida H, Dinh VA, Fukushima T, Kizaki H, Zeller R (2010) First-principles theory of dilute magnetic semiconductors. Rev Mod Phys 82:1633

Schäpers T (2016) Semiconductor Spintronics. Walter de Gruyter, Berlin/Boston

Schmidt G (2005) Concepts for spin injection into semiconductors – a review. J Phys D Appl Phys 38:R107

Shinjo T (ed) (2009) Nanomagnetism and Spintronics. Elsevier, Amsterdam

Thompson SM (2008) The discovery, development and future of GMR: the Nobel Prize 2007. J Phys D Appl Phys 41:093001

Tsymbal EY, Žutić I (2019) (eds) Spintronics handbook, vol 2. Semiconductor spintronics, 2nd edn. CRC Press/Taylor & Francis, Boca Raton

Vardeny ZV (ed) (2010) Organic spintronics. CRC Press/Taylor and Francis, Boca Raton

Varghani A, Peiravi A, Moradi F (2018) Perpendicular STT-RAM cell in 8 nm technology node using $Co_1/Ni_3(1\ 1\ 1)||Gr_2||Co_1/Ni_3(1\ 1\ 1)$ structure as magnetic tunnel junction. J Magn Magn Mater 452:10

Vasil'ev AM, Daiminger F, Straka J, Forchel A, Kochereshko VP, Sandler GL, Uraltsev IN (1993) Optical orientation of holes and electrons in strained layer InGaAs/GaAs quantum-wells. Superlattices Microstruc 13:97

Wachter P (1979) Europium chalcogenides: EuO, EuS, EuSe and EuTe. In Gschneidner KA Jr; Eyring L (eds) Handbook on the physics and chemistry of rare earths, vol 2. Elsevier, Amsterdam, pp 507–574

Wolos A, Kaminska M (2008) Magnetic impurities in wide band-gap III-V semiconductors. In: Dietl T, Awschalom DD, Kaminska M, Ohno H (eds) Spintronics, Semiconductors and semimetals, vol 82. Academic Press/Elsevier, Amsterdam

Xia J, Ge W, Chang K (2012) Semiconductor spintronics. World Scientific, Singapore

Yuasa S, Nagahama T, Fukushima A, Suzuki Y, Ando K (2004) Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions. Nat Mater 3:868

Zener C (1951a) Interaction between the *d*-shells in the transition metals. II. Ferromagnetic compounds of manganese with perovskite structure. Phys Rev 82:403

Zener C (1951b) Interaction between the *d*-shells in the transition metals. Phys Rev 81:440

Zhao Y-J, Mahadevan P, Zunger A (2004) 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

Zunger A (1986) Electronic structure of 3d ransition-atom impurities in semiconductors. In: Zeits F, Ehrenreich H, Turnbull D (eds) Solid state physics, vol 39. Academic Press, New York, pp 275–464

Žutić I, Fabian J, Das Sarma S (2004) Spintronics: fundamentals and applications. Rev Mod Phys 76:323

Part IV

Photons

Interaction of Light with Solids

10

Karl W. Böer and Udo W. Pohl

Contents

1 Continuum Model of Solid-Light Interaction 370
1.1 Reflection, Transmission, and Absorption 371
1.2 Measurement of Optical Parameters 380
1.3 Frequency Dependence of the Dielectric Function and Dielectric Screening 394
2 Photonic Bandgap Structures 402
2.1 Photonic Crystals 402
2.2 Localized Defect Modes and Microcavities 409
3 Nonlinear Optical Effects 412
3.1 Electronic and Mixing Effects 413
3.2 Electro-Optical Effects 421
4 Summary 422
References 423

Abstract

The interaction of light with solids is described by Maxwell's equations, which treat the solid as a continuum and lead to its optical parameters as a function of the frequency of the electromagnetic radiation: the complex dielectric constant. The dielectric constant describes the ability of a solid to screen an electric field – with electronic and ionic contributions – and is one of the most important material parameters. This function is closely related to the index of refraction and the optical absorption (or extinction) coefficient. All these parameters are derived from measured quantities: the transmitted and reflected light as a function of wavelength, impinging angle, and polarization.

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_10

A periodic modulation of the dielectric constant along a spatial direction leads to a photonic bandgap for the propagation of specific modes along this direction, analogous to the electronic bandgap for electrons traveling in the periodic crystal potential. A complete bandgap for propagation along any direction can be created for three-dimensional periodicity; defects given by deviations from periodicity lead to localized states in such photonic crystals, similar to effects in the electronic counterpart, allowing for, e.g., waveguiding or suppresion of spontaneous emission.

At high field amplitudes, nonlinear optical effects occur due to the non-parabolicity of the lattice potential. These effects can be described by a field-dependent dielectric function. The resulting nonharmonic oscillations permit mixing of different signals with corresponding change in frequency and amplitude.

Keywords

Absorption · Dielectric constant · Dielectric function · Dielectric screening · Electro-optical effects · Ellipsometry · Fresnel coefficient · Fresnel equations · Index of refraction · Metamaterial · Microcavity · Nonlinear optical effects · Optical constants · Optical defect · Photonic bandgap structures · Photonic crystals · Reflectance · Reflection · TE mode · TM mode · Transmittance · Transmission · Upconversion

1 Continuum Model of Solid-Light Interaction

The interaction of electromagnetic radiation (photons) with semiconductors provides major insight into the electronic and phononic structure of these solids. Such interactions can be described as resonant and nonresonant. *Resonant absorption* related to ionic charge of crystal atoms is observed in the infrared part of the spectrum; resonant absorption dealing with bound electrons is observed at shorter wavelengths and is usually separated from the ionic component by a wavelength range with low optical absorption, except for semiconductors with almost zero bandgap. Both types of absorption describe *intrinsic* properties of the semiconductor. In addition, a large variety of *extrinsic* ionic and electronic resonance transitions of semiconductor defects can be distinguished. These optically induced transitions are extensively used for the identification of such defects. *Nonresonant interactions* can be observed between photons and free electrons, except at high densities of free electrons when they act jointly as plasmons.

The interaction between photons and elementary excitation processes in the semiconductor can be *elastic,* without absorption, or *inelastic,* with absorption. The latter requires damping of the excited state by either absorption or emission of phonons or by collisions with electrons or other quasiparticles. The resonant transitions will be described for ions in Sect. 2.1 of ▶ Chap. 11, "Photon–Phonon Interaction," and for electrons in ▶ Chap. 12, "Photon–Free-Electron Interaction." The description in this chapter is based on solutions of Maxwell's equations, which deal with the resonant transitions in a *phenomenological* way: the semiconductor is considered a continuum, disregarding its atomic structure. Comparison with the

experiment requires the transformation of experimentally accessible quantities, such as reflectance and optical transmissivity, into quantities obtained as a result of a theoretical analysis, such as the set of optical constants.[1] This relationship is summarized in Sect. 1.2.

1.1 Reflection, Transmission, and Absorption

In this section, the interaction of electromagnetic radiation with a semiconductor is described in a classical model. We begin with *Maxwell's equations,* which result in the dispersion equation, yielding the optical parameters of the semiconductor (its complex dielectric constant) as a function of the frequency of the electromagnetic radiation (the energy of the photons). For more information, see Palik (1985, 1991), Ward (1994).

Light impinging on a semiconductor is subject to a number of optical interactions before it is absorbed. First, a fraction of the light is reflected at the outer surface; another fraction is scattered by crystal imperfections, phonons, and other quasiparticles; then a fraction of the light within the semiconductor is absorbed by various elementary excitation processes. The unabsorbed fraction is transmitted and exits through the semiconductor's surfaces after partial reflection.

Initially, we regard the semiconductor as a *continuum*, represented by four parameters which can be measured macroscopically: the magnetic permeability μ, the dielectric constant ε, the space charge density ρ, and the electric conductivity σ. With these parameters, the relationship between absorption, reflection, and transmission can be obtained from *Maxwell's equations*[2]:

$$\frac{\partial}{\partial \mathbf{r}} \times \mathbf{E} = -\mu\mu_0 \frac{\partial \mathbf{H}}{\partial t} \tag{1}$$

$$\frac{\partial}{\partial \mathbf{r}} \times \mathbf{H} = \varepsilon\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E} \tag{2}$$

$$\frac{\partial}{\partial \mathbf{r}} \mathbf{E} = \frac{\rho}{\varepsilon\varepsilon_0} \tag{3}$$

$$\frac{\partial}{\partial \mathbf{r}} \mathbf{H} = 0. \tag{4}$$

[1] This chapter contains quite a few equations; the most important relations are pointed out by a shading.

[2] The operators $\frac{\partial}{\partial \mathbf{r}} \times$ and $\frac{\partial}{\partial \mathbf{r}} \cdot$ are also written as $\nabla\times$ or *rot* and $\nabla\cdot$ or *div*, respectively, with ∇ being the Nabla operator. The symbolic vector $\frac{\partial}{\partial \mathbf{r}}$ has the components $\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$, and the "$\times$" and "$\cdot$" operations yield consequently a vector and a scalar, respectively.

with the proper boundary conditions adapted to the given geometry. Here, **E** is chosen as the electric field vector rather than **F** as in other chapters of this book; **H** is the magnetic field vector.

1.1.1 Nonabsorbing Dielectrics

We consider light propagation and the corresponding energy flow in a homogeneous, nonmagnetic, and nonconductive dielectric; from Maxwell's equations, one obtains by putting $\rho = \sigma = 0$, $\mu = 1$, and $\varepsilon_0 \cdot \mu_0 = 1/c^2$ (with c being the speed of light in vacuum):

$$\frac{\partial}{\partial \mathbf{r}} \times \frac{\partial}{\partial \mathbf{r}} \times \mathbf{E} = \frac{\partial}{\partial \mathbf{r}} \left(\frac{\partial}{\partial \mathbf{r}} \mathbf{E} \right) - \frac{\partial^2}{\partial \mathbf{r}^2} \mathbf{E} = -\frac{\partial}{\partial \mathbf{r}} \times \left(\frac{1}{\varepsilon_0 c^2} \frac{\partial \mathbf{H}}{\partial t} \right)$$
$$= -\frac{1}{\varepsilon_0 c^2} \frac{\partial}{\partial t} \left(\frac{\partial}{\partial \mathbf{r}} \times \mathbf{H} \right), \tag{5}$$

which yields the *undamped wave equation* for the electric vector field

$$\frac{\partial^2}{\partial \mathbf{r}^2} \mathbf{E} = \frac{\varepsilon}{c^2} \frac{\partial^2}{\partial t^2} \mathbf{E} \tag{6}$$

and a similar one for the magnetic vector field. Assuming a plane wave entering the dielectric in the x-direction with linear polarization in the y-direction, one has with $\mathbf{E} = (0, E_y, 0)$ from Eq. 6

$$\frac{\partial^2 E_y}{\partial x^2} = \frac{\varepsilon}{c^2} \frac{\partial^2 E_y}{\partial t^2} \tag{7}$$

which can be solved with the trial solution

$$E_y = f(x) \exp(-i\,\omega t) \tag{8}$$

where $f(x)$ is the amplitude function and ω the oscillation frequency. Substitution of Eq. 8 into Eq. 7 yields

$$\frac{\partial^2 f}{\partial x^2} + \frac{\varepsilon \omega^2}{c^2} f = 0 \tag{9}$$

with the solution

$$f(x) = E_0 \exp\left(\pm i \left(\frac{\omega x}{v} \right) \right), \tag{10}$$

where

$$v = \frac{c}{\sqrt{\varepsilon}} = \frac{c}{n_r}. \tag{11}$$

Here, $n_r = \sqrt{\varepsilon}$ is the index of refraction[3] and c is the light velocity in vacuum. Thus, with the + sign in Eq. 10, Eq. 8 describes a plane wave traveling in + x-direction with a phase velocity v and amplitude E_0:

$$E_y = E_0 \exp\left(i\omega\left(\frac{x}{v} - t\right)\right) = E_0 \exp\left(i\omega\left(\frac{n_r}{c}x - t\right)\right). \tag{12}$$

The *energy flow* in this wave is given by the *Poynting vector* (a vector in the direction of the wave propagation)

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}. \tag{13}$$

Here, we assume $\mathbf{E} = (0, E_y, 0)$ and $\mathbf{H} = (0, 0, H_z)$. The *energy density* is given by

$$W = \frac{1}{2}\left(\varepsilon\varepsilon_0 \mathbf{E}^2 + \mu\mu_0 \mathbf{H}^2\right). \tag{14}$$

With an equal amount of energy in the electrical and magnetic component, one obtains for the total energy density twice the energy represented by the electrical vector:

$$W = \varepsilon\varepsilon_0 \mathbf{E}^2. \tag{15}$$

1.1.2 Metamaterials

With the assumptions made to obtain the undamped wave Eq. 6, we referred to the case $\mu = 1$, which is generally fulfilled for semiconductors and leads to the positive sign for the refractive index n_r and the phase velocity in Eq. 11. Both μ and ε depend, however, on the frequency ω of the electromagnetic radiation as discussed later, and a more general formulation of the wave equation reads:

$$\frac{\partial^2}{\partial \mathbf{r}^2}\mathbf{E} - \mu_0\mu(\omega)\,\varepsilon_0\,\varepsilon(\omega)\,\frac{\partial^2}{\partial t^2}\mathbf{E} = 0. \tag{16}$$

This leads to the complex index of refraction $\tilde{n} = \pm\sqrt{\varepsilon(\omega)\,\mu(\omega)}$ introduced for the case $\mu = 1$ below in Eq. 20. In *metamaterials*, there exists a frequency range where both $\varepsilon(\omega) < 0$ and $\mu(\omega) < 0$. In such a case $\tilde{n} = -\sqrt{\varepsilon(\omega)\,\mu(\omega)}$ applies, resulting in a negative real refractive index $n_r(\omega)$ at these frequencies.

A real $n_r(\omega) < 0$ leads to a number of counterintuitive consequences (Veselago 1968). The phase velocity v is negative according to Eq. 11. The vectors **E**, **H**, and **k** follow a *left*-hand rule; the energy flow is still given by the Poynting vector

[3] An index r is added to the widely used symbol n to distinguish this quantity from the carrier density n.

$\mathbf{S} = \mathbf{E} \times \mathbf{H}$, but $\mathbf{S}$ is antiparallel to the wavevector $\mathbf{k}$. Snell's law of refraction Eq. 38 remains valid, but the light beam is refracted on the *same* side of the normal since one of the two refractive indices is negative; consequently, a convex lens acts as a diverging lens and a concave lens focuses the light. A negative refractive index was experimentally first verified by scattering experiments at microwave frequencies (Shelby et al. 2001).

Metamaterials are fabricated using periodic stuctures with resonant building blocks of individual subwavelength size and periodicities below the wavelength, see Busch et al. (2007), Soukoulis and Wegener (2010). The structures resemble photonic crystals (Sect. 2), but the latter employ *wavelength-scale* building blocks which produce collective Bragg resonances. In metamaterials, the building blocks are ideally noninteracting and produce individual Mie resonances. Potential applications of metamaterials include superlenses with increased optical resolution beyond that of conventional lenses and lenses for high-gain antennas. For more details on metamaterials, see Engheta and Ziolkowski (2006), Zouhdi et al. (2009), and Cui et al. (2010).

In the following, we resume the treatment with a *positive* refractive index n_r.

1.1.3 Semiconductors with Optical Absorption

The introduction of a finite conductivity $\sigma = \sigma(\omega)$ (second term of Eq. 17) produces a damping contribution on the electromagnetic wave in a semiconductor, resulting in a finite optical absorption.[4] From Eqs. 2 and 5, we obtain the *damped wave equation* for the electric vector

$$\frac{\partial^2 \mathbf{E}}{\partial \mathbf{r}^2} = \frac{\varepsilon}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} + \frac{\sigma}{\varepsilon_0 c^2} \frac{\partial \mathbf{E}}{\partial t}. \tag{17}$$

Using the same trial solution as given in the previous section for $f(x)$, we obtain

$$\frac{\partial^2 f}{\partial x^2} + \frac{\omega^2}{c^2} \left(\varepsilon - i \left(\frac{\sigma}{\varepsilon_0 \omega} \right) \right) f = 0 \tag{18}$$

which has a solution that can be written exactly as Eq. 10. This yields a plane wave traveling in the x-direction

$$E_y = E_0 \exp\left(i\omega \left(\frac{\tilde{n}}{c} x - t \right) \right) \tag{19}$$

except that the index of refraction used to describe the ratio c/v (Eq. 11) is now complex and is identified as $\tilde{n}$. This *complex index of refraction*

[4] We can understand this by equating damping with transfer of energy into heat and optical absorption with extraction of this energy from the radiation field. Such an absorption occurs even outside a specific electronic or ionic resonance absorption – see ▶ Chaps. 11, "Photon–Phonon Interaction," and ▶ 12, "Photon–Free-Electron Interaction."

$$\tilde{n} = \frac{c}{v} = \sqrt{\tilde{\varepsilon}} = n_r + i\kappa \tag{20}$$

is related to the complex dielectric constant $\tilde{\varepsilon}$ in a similar fashion as given in Eq. 11 and contains as its real part the previously defined index of refraction n_r and as its imaginary part the *extinction coefficient* κ. The *complex dielectric constant* $\tilde{\varepsilon}$ is given by

$$\tilde{\varepsilon} = \varepsilon' + i\varepsilon'' \text{ with } \varepsilon' = \varepsilon \text{ and } \varepsilon'' = \frac{\sigma}{\varepsilon_0 \omega}, \tag{21}$$

with its imaginary part related to the conductivity. From Eqs. 20 and 21, we obtain the important relations for an optically absorbing (damping) homogeneous continuum:

or

$$\begin{aligned} \varepsilon' &= n_r^2 - \kappa^2 \\ \varepsilon'' &= \frac{\sigma}{\varepsilon_0 \omega} = 2 n_r \kappa \\ n_r^2 &= \tfrac{1}{2}\left(\varepsilon' + \sqrt{\varepsilon'^2 + \varepsilon''^2}\right) \\ \kappa^2 &= \tfrac{1}{2}\left(-\varepsilon' + \sqrt{\varepsilon'^2 + \varepsilon''^2}\right). \end{aligned} \tag{22}$$

The optical parameters ε', ε'', n_r, and κ, complemented by the related reflectance R_0 defined in Eq. 54, are listed in Table 1 for a number of typical semiconductors. Note that all optical constants depend strongly on the photon energy and also on temperature; moreover, crystals with a symmetry lower than cubic like, e.g., wurtzite semiconductors, show some dependence on the polarization. The spectral dependence of n_r is shown for some semiconductors in Fig. 1. For the wurtzite GaN, n_r is given for $\mathbf{E} \perp \mathbf{c}$; the difference to $\mathbf{E}//\mathbf{c}$ is $1.5 \pm 0.2\%$ at 2.48 eV (500 nm). The dielectric constant is discussed in more detail in Sect. 1.3.

Using Eq. 20, the propagating wave can be rewritten as

$$E_y = E_0 \exp\left(i\omega\left(\frac{n_r}{c}x - t\right)\right) \exp\left(-\frac{\omega\kappa}{c}x\right) \tag{23}$$

and shows the damping factor in the second exponential. Using a more conventional expression $\exp(-\alpha_o\, x)$ for the damping of the energy flux, with α_o as the *optical absorption coefficient* for the energy *density,* we obtain by comparison with the second exponential in Eq. 23

Table 1 Optical constants for element and III-V compound semiconductors for $E > E_g$: real and imaginary parts of the dielectric constant ε' and ε'', respectively, refractive index n_r, extinction coefficient κ, normal-incidence reflectance R_0, and absorption coefficient α_o for various photon energies $E = hv$. ε' and ε'' after Aspnes and Studna (1983); from these data, n_r and κ are calculated according to Eq. 22, R_0 according to Eq. 54 and α_o according to Eq. 24. For more extensive data, see Palik (1985, 1991)

E(eV)	ε'	ε''	n_r	κ	R_0	$\alpha_o(10^3\ \mathrm{cm}^{-1})$
Silicon(*Si*)						
1.5	13.488	0.038	3.673	0.005	0.327	0.78
2.0	15.254	0.172	3.906	0.022	0.351	4.47
2.5	18.661	0.630	4.320	0.073	0.390	18.48
3.0	27.197	2.807	5.222	0.269	0.461	81.73
3.5	22.394	33.818	5.610	3.014	0.575	1,069.19
4.0	12.240	35.939	5.010	3.586	0.591	1,454.11
4.5	−19.815	24.919	2.452	5.082	0.740	2,317.99
5.0	−10.242	11.195	1.570	3.565	0.675	1,806.67
5.5	−9.106	8.846	1.340	3.302	0.673	1,840.59
6.0	−7.443	5.877	1.010	2.909	0.677	1,769.27
Germanium(*Ge*)						
1.5	21.560	2.772	4.653	0.298	0.419	45.30
2.0	30.361	10.427	5.588	0.933	0.495	189.12
2.5	13.153	20.695	4.340	2.384	0.492	604.15
3.0	12.065	17.514	4.082	2.145	0.463	652.25
3.5	9.052	21.442	4.020	2.667	0.502	946.01
4.0	4.123	26.056	3.905	3.336	0.556	1,352.55
4.5	−14.655	16.782	1.953	4.297	0.713	1,960.14
5.0	−8.277	8.911	1.394	3.197	0.650	1,620.15
5.5	−6.176	7.842	1.380	2.842	0.598	1,584.57
6.0	−6.648	5.672	1.023	2.774	0.653	1,686.84
Aluminum Arsenide(*AlAs*)						
2.5			3.394	0.001	0.297	0.25
3.0			3.77	0.053	0.337	16.11
3.5			4.48	0.334	0.405	118.47
4.0			4.92	2.49	0.523	1,009.37
4.5			4.18	2.63	0.505	1,199.38
5.0			2.48	3.95	0.642	2,001.50
5.5			1.36	2.89	0.609	1,610.83
Gallium Arsenide(*GaAs*)						
1.5	13.435	0.589	3.666	0.080	0.327	12.21
2.0	14.991	1.637	3.878	0.211	0.349	42.79
2.5	18.579	3.821	4.333	0.441	0.395	111.74
3.0	16.536	17.571	4.509	1.948	0.472	592.48
3.5	8.413	14.216	3.531	2.013	0.425	714.20
4.0	9.279	13.832	3.601	1.920	0.421	778.65
4.5	6.797	22.845	3.913	2.919	0.521	1,331.28
5.0	−11.515	18.563	2.273	4.084	0.668	2,069.81

(continued)

Table 1 (continued)

E(eV)	ε'	ε''	n_r	κ	R_0	$\alpha_o(10^3\ cm^{-1})$
5.5	−6.705	8.123	1.383	2.936	0.613	1,636.68
6.0	−4.511	6.250	1.264	2.472	0.550	1,503.20
Indium Arsenide(*InAs*)						
1.5	13.605	3.209	3.714	0.432	0.337	65.69
2.0	15.558	5.062	3.995	0.634	0.370	128.43
2.5	15.856	15.592	4.364	1.786	0.454	452.64
3.0	6.083	13.003	3.197	2.034	0.412	618.46
3.5	5.973	10.550	3.008	1.754	0.371	622.13
4.0	7.744	11.919	3.313	1.799	0.393	729.23
4.5	−1.663	22.006	3.194	3.445	0.566	1,571.19
5.0	−5.923	8.752	1.524	2.871	0.583	1,455.26
5.5	−3.851	6.008	1.282	2.344	0.521	1,306.62
6.0	−2.403	6.005	1.434	2.112	0.448	1,284.15
Indium Phosphide(*InP*)						
1.5	11.904	1.400	3.456	0.203	0.305	30.79
2.0	12.493	2.252	3.549	0.317	0.317	64.32
2.5	14.313	3.904	3.818	0.511	0.349	129.56
3.0	17.759	10.962	4.395	1.247	0.427	379.23
3.5	6.400	12.443	3.193	1.948	0.403	691.21
4.0	6.874	10.871	3.141	1.730	0.376	701.54
4.5	8.891	16.161	3.697	2.186	0.449	996.95
5.0	−7.678	14.896	2.131	3.495	0.613	1,771.52
5.5	−4.528	7.308	1.426	2.562	0.542	1,428.14
6.0	−2.681	5.644	1.336	2.113	0.461	1,285.10

$$\alpha_o = \frac{2\omega\kappa}{c} = \frac{4\pi}{\lambda}\kappa; \tag{24}$$

here the energy flow is given by the product of the electric and magnetic vectors, thus producing a factor of 2 in the exponent. However, H_z is phase shifted by δ with $\tan\delta = \kappa/n_r$. From Eqs. 22 and 23, we also obtain

$$\alpha_o = \frac{\sigma}{\varepsilon_0 n_r c}, \tag{25}$$

which shows the direct connection of α_o with the electrical conductivity. This conductivity is to be taken at the optical frequency and needs further explanation.

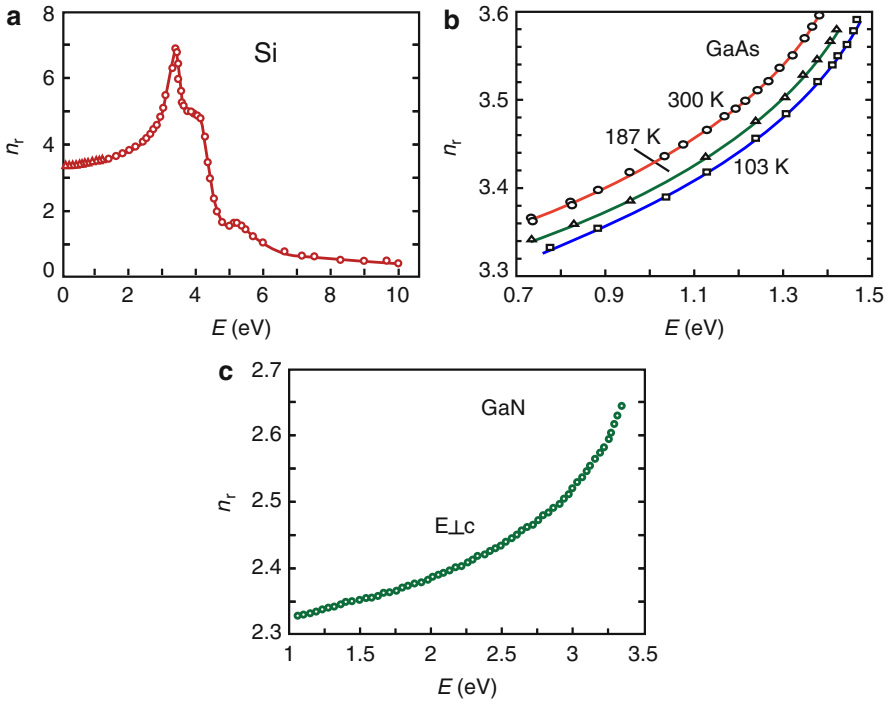


Fig. 1 Refractive index n_r of **(a)** Si, **(b)** GaAs, and **(c)** GaN at 300 K as a function of the photon energy; (**a**) illustrates the frequency dependence for a wide range, curves in (**b**) demonstrate the temperature dependence; the refractive index of GaN is given for $\mathbf{E}\perp\mathbf{c}$; after Philipp and Taft (1960), Marple (1964), and Ejder (1971), respectively

1.1.4 The Complex Electrical Conductivity

The right-hand side of Maxwell's Eq. 2 is considered as a total current density **j**; this quantity includes both the current induced by a *change* of the electric field and that originating from the electric field in a conductive solid:

$$\mathbf{j} = \varepsilon\,\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \sigma\,\mathbf{E}.$$

After applying a Fourier transformation from the time into the frequency domain, this current density can be written

$$\mathbf{j}(\omega) = -i\,\omega\,\varepsilon(\omega)\,\varepsilon_0 \mathbf{E} + \sigma(\omega)\,\mathbf{E}(\omega); \tag{26}$$

when introducing a *complex conductivity* $\widetilde{\sigma} = \sigma' + i\,\sigma''$ or using the *complex* dielectric constant $\widetilde{\varepsilon}$ according to Eq. 21, Eq. 26 reads

$$\mathbf{j}(\omega) = \widetilde{\sigma}(\omega)\,\mathbf{E}(\omega) = -i\,\omega\widetilde{\varepsilon}(\omega)\,\varepsilon_0\,\mathbf{E}(\omega). \tag{27}$$

The real and imaginary parts of the complex conductivity are consequently

$$\begin{aligned}\sigma' &= \varepsilon'' \varepsilon_0 \omega \\ \sigma'' &= -\varepsilon' \varepsilon_0 \omega = -\left(n_\mathrm{r}^2 - \kappa^2\right) \varepsilon_0 \omega.\end{aligned} \tag{28}$$

Observe that the real part of $\widetilde{\sigma}$ is proportional to the imaginary part of $\widetilde{\varepsilon}$ and vice versa.

The total current density can now be expressed in terms of the real and imaginary parts of the dielectric constant; Eq. 26 then reads

$$\begin{aligned}\mathbf{j}(\omega) = \widetilde{\sigma}(\omega)\ \mathbf{E}(\omega) &= \varepsilon_0 \omega \left(\varepsilon''(\omega) - i\varepsilon'(\omega)\right) \mathbf{E}(\omega) \\ &= \left(\sigma'(\omega) - i\omega\varepsilon_0 \left(n_\mathrm{r}^2 - \kappa^2\right)\right) \mathbf{E}(\omega).\end{aligned} \tag{29}$$

The first term in Eq. 29 comprising σ' is proportional to the displacement current; this part of the total current is caused by bound electrons surrounding each atom core, which oscillate out of phase with the applied electric field.

1.1.5 Dielectric Polarization

For later discussions of microscopic models, it is advantageous to introduce the dielectric polarization as the response of the solid on an applied electric field **E**. The electric displacement **D**, field strength **E**, and polarization **P** are related by

$$\mathbf{D} = \varepsilon_0 \widetilde{\varepsilon}\ \mathbf{E} = \varepsilon_0 \mathbf{E} + \mathbf{P} \tag{30}$$

and

$$\mathbf{P} = \varepsilon_0 \widetilde{\chi} \mathbf{E} = \varepsilon_0 \left(\widetilde{\varepsilon} - 1\right) \mathbf{E} \tag{31}$$

with $\widetilde{\varepsilon}$ as the complex dielectric constant. Here, only a linear relationship between **P** and **E** is given. Higher-order terms in **E** are of importance at higher fields and give rise to nonlinear optical effects, which are discussed in Sect. 3. Equation 31 defines the susceptibility $\widetilde{\chi}$ as the proportionality constant between the polarization **P** and the electric field **E**. From Eqs. 30 and 31, we have an electric displacement

$$\mathbf{D} = \varepsilon_0 \left(1 + \widetilde{\chi}\right) \mathbf{E}. \tag{32}$$

In an anisotropic material, $\widetilde{\varepsilon}$ and $\widetilde{\chi}$ have tensor form with components

Table 2 Dielectric constant tensor

Crystal system	Characteristic symmetry	Number of independent coefficients	Form of tensor showing independent coefficients
Cubic	Four 3-fold axes	1	$\begin{pmatrix} \tilde{\varepsilon} & 0 & 0 \\ 0 & \tilde{\varepsilon} & 0 \\ 0 & 0 & \tilde{\varepsilon} \end{pmatrix}$
Tetragonal	One 4-fold axis	2	$\begin{pmatrix} \tilde{\varepsilon}_1 & 0 & 0 \\ 0 & \tilde{\varepsilon}_1 & 0 \\ 0 & 0 & \tilde{\varepsilon}_3 \end{pmatrix}$
Hexagonal	One 6-fold axis		
Trigonal	One 3-fold axis		
Orthorhombic	Three mutually perpendicular twofold axes of higher order	3	$\begin{pmatrix} \tilde{\varepsilon}_1 & 0 & 0 \\ 0 & \tilde{\varepsilon}_2 & 0 \\ 0 & 0 & \tilde{\varepsilon}_3 \end{pmatrix}$
Monoclinic	One 2-fold axis	4	$\begin{pmatrix} \tilde{\varepsilon}_{11} & 0 & \tilde{\varepsilon}_{13} \\ 0 & \tilde{\varepsilon}_{22} & 0 \\ \tilde{\varepsilon}_{13} & 0 & \tilde{\varepsilon}_{33} \end{pmatrix}$
Triclinic	A center of symmetry or no symmetry	6	$\begin{pmatrix} \tilde{\varepsilon}_{11} & \tilde{\varepsilon}_{12} & \tilde{\varepsilon}_{13} \\ \tilde{\varepsilon}_{12} & \tilde{\varepsilon}_{22} & \tilde{\varepsilon}_{23} \\ \tilde{\varepsilon}_{13} & \tilde{\varepsilon}_{23} & \tilde{\varepsilon}_{33} \end{pmatrix}$

$$P_i = \varepsilon_0 \tilde{\chi}_{ij} E_j \quad \text{and} \quad D_i = \varepsilon_0 \tilde{\varepsilon}_{ij} E_j. \tag{33}$$

This matrix relationship must be used when applying Maxwell's equations to an anisotropic medium. For instance, from *Poisson's equation* $\frac{\partial}{\partial \mathbf{r}} \mathbf{E} = \rho / (\tilde{\varepsilon}\, \varepsilon_0)$, we have for the *ij* component

$$\varepsilon_0 \tilde{\varepsilon}_{ij} \frac{\partial E_j}{\partial x_i} = \rho. \tag{34}$$

Depending on the crystal symmetry, the dielectric constant tensor can be reduced and contains at most six independent coefficients – see Table 2. This dielectric constant will be discussed in more detail in Sect. 1.3.

1.2 Measurement of Optical Parameters

The parameters experimentally measured are the intensities of reflected and transmitted radiation and its change of the state of polarization. On the other hand, the parameters obtained from an analysis of the optical phenomena are the complex dielectric constant (or index of refraction) and the complex conductivity as well as the amplitude and polarization of electromagnetic waves after interacting with the semiconductor. It is the purpose of this section to summarize briefly the most relevant interrelation of these two sets of parameters, thereby permitting a quantitative comparison between theory and experiment. For more details, see Palik (1985, 1991) or Ward (1994).

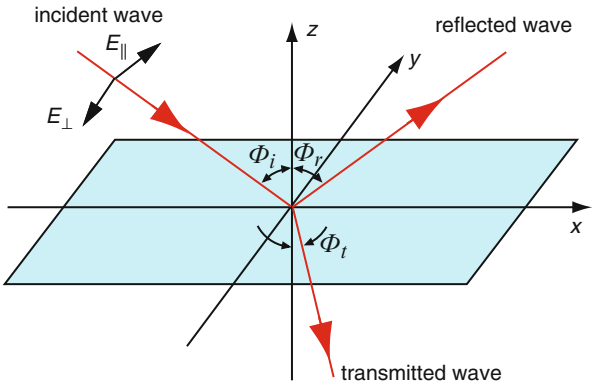


Fig. 2 Coordinate system with light beams and the interface between the two different media above and below $z = 0$

1.2.1 Reflectance and Transmittance in Dielectrics

The index of refraction as well as the extinction and absorption coefficients can be related to the amplitude and polarization of reflected and transmitted optical waves which can be measured directly. The relationship between these waves can be obtained from the wave equation and the boundary conditions at the interface between two media (see Palik 1985; Balkanski and Moss 1994; Born and Wolf 2002).

The electric vector of the incident wave (denoted by subscript i) can be described by its two components normal and parallel to the plane of incidence at $z = 0$ (Fig. 2):

$$\left(E_y, E_z\right)_{\mathrm{i}} = \left(E_{\perp}, E_{||}\right)_{\mathrm{i}} \exp\left(i\omega\left(\frac{n_{\mathrm{r1}}}{c}(x \sin \Phi_{\mathrm{i}} + z \cos \Phi_{\mathrm{i}}) - t\right)\right) \tag{35}$$

with the corresponding components of the reflected (index r) and transmitted (index t) waves

$$\left(E_y, E_z\right)_{\mathrm{r}} = \left(E_{\perp}, E_{||}\right)_{\mathrm{r}} \exp\left(i\omega\left(\frac{n_{\mathrm{r1}}}{c}(x \sin \Phi_{\mathrm{r}} + z \cos \Phi_{\mathrm{r}}) - t\right)\right) \tag{36}$$

$$\left(E_y, E_z\right)_{\mathrm{t}} = \left(E_{\perp}, E_{||}\right)_{\mathrm{t}} \exp\left(i\omega\left(\frac{n_{\mathrm{r2}}}{c}(x \sin \Phi_{\mathrm{t}} + z \cos \Phi_{\mathrm{t}}) - t\right)\right). \tag{37}$$

The coordinate system is chosen so that the interface between the two media is normal to the z-axis and cuts the interface at $z = 0$. The x- and y-axes are chosen so that the plane of incidence lies in the xz-plane; the angles are identified in Fig. 2.[5]

[5]The components of the electric field vector parallel and perpendicular to the plane of incidence, $E_{||}$ and $E_{\perp}$, are also referred to as p- (or π-) *polarized* and s- (or σ-) polarized (p from German for *parallel* and s from German for *senkrecht*) respectively.

From the condition that at the plane of incidence ($z = 0$) the tangential component of the electrical field of all three waves must be the same, we require that the corresponding exponents must be equal: $n_{r1} \sin \Phi_i = n_{r1} \sin \Phi_r = n_{r2} \sin \Phi_t$. From here, *Snell's law* can be deduced:

$$\begin{aligned} \Phi_i &= \Phi_r \\ n_{r1} \sin \Phi_i &= n_{r2} \sin \Phi_t. \end{aligned} \tag{38}$$

For $n_{r1} > n_{r2}$, this yields an angle of total reflection (i.e., for $\Phi_t = 90°$), with $\Phi_i = \Phi_c$ the critical angle,[6] given by

$$\sin \Phi_c = \frac{n_{r2}}{n_{r1}}. \tag{40}$$

Total reflection is sometimes used in optical fibers and in photosensing devices for light trapping by properly shaping the surface in order to increase the optical path within the device for more optical absorption and thereby increasing the photosensitivity.

The amplitudes of the tangential components of the reflected and transmitted waves must be continuous when passing through $z = 0$; hence,

$$\left(E_{||i} - E_{||r}\right) \cos \Phi_i = E_{||t} \cos \Phi_t \tag{41}$$

$$\left(E_{\perp i} + E_{\perp r}\right) = E_{\perp t} \tag{42}$$

$$n_{r1} \left(E_{\perp i} - E_{\perp r}\right) \cos \Phi_i = n_{r2} E_{\perp t} \cos \Phi_t \tag{43}$$

$$n_{r1} \left(E_{||i} + E_{||r}\right) = n_{r2} E_{||t}. \tag{44}$$

The first set of two equations is obtained from the electric vector components; the second set is obtained from the magnetic vector components.

After solving this set of four equations for the four components of the electric vector, we obtain *Fresnel's equations:*

[6] A Brewster angle Φ_B is defined as the angle under which no component $E_{||}$ is reflected. This happens for $\Phi_t \perp \Phi_r$, or $\Phi_t + \Phi_r = 90°$; from this, we obtain with Eq. 38 $n_{r1} \sin \Phi_i = n_{r2} \cos \Phi_r$, and eventually

$$\tan \Phi_B = n_{r2}/n_{r1}. \tag{39}$$

$$
\begin{aligned}
E_{\|\mathrm{r}} &= E_{\|\mathrm{i}} \frac{n_{\mathrm{r}2}\cos\Phi_{\mathrm{i}} - n_{\mathrm{r}1}\cos\Phi_{\mathrm{t}}}{n_{\mathrm{r}2}\cos\Phi_{\mathrm{i}} + n_{\mathrm{r}1}\cos\Phi_{\mathrm{t}}} \\
E_{\perp\mathrm{r}} &= E_{\perp\mathrm{i}} \frac{n_{\mathrm{r}1}\cos\Phi_{\mathrm{i}} - n_{\mathrm{r}2}\cos\Phi_{\mathrm{t}}}{n_{\mathrm{r}1}\cos\Phi_{\mathrm{i}} + n_{\mathrm{r}2}\cos\Phi_{\mathrm{t}}} \\
E_{\|\mathrm{t}} &= E_{\|\mathrm{i}} \frac{2n_{\mathrm{r}1}\cos\Phi_{\mathrm{i}}}{n_{\mathrm{r}1}\cos\Phi_{\mathrm{t}} + n_{\mathrm{r}2}\cos\Phi_{\mathrm{i}}} \\
E_{\perp\mathrm{t}} &= E_{\perp\mathrm{i}} \frac{2n_{\mathrm{r}1}\cos\Phi_{\mathrm{i}}}{n_{\mathrm{r}2}\cos\Phi_{\mathrm{t}} + n_{\mathrm{r}1}\cos\Phi_{\mathrm{i}}}
\end{aligned}
\tag{45}
$$

which are generally valid and are the basis for all following discussions.

All measurable quantities are related to the *energy flux*, i.e., for the incident and reflected waves of the Poynting vector. For the incident and reflected waves, we have

$$
W_a = \varepsilon_0 n_{\mathrm{r}1}^2 E_a^2 \quad \text{for} \quad a = \mathrm{i}, \mathrm{r}. \tag{46}
$$

For the transmitted wave, we have

$$
W_{\mathrm{t}} = \varepsilon_0 n_{\mathrm{r}2}^2 E_{\mathrm{t}}^2. \tag{47}
$$

The reflectance[7] R and transmittance T are defined by the ratios of the reflected and transmitted energy flux to the incident energy flux normal to the interface:

$$
R = \left(\frac{E_{\mathrm{r}}}{E_{\mathrm{i}}}\right)^2 \quad \text{and} \quad T = \left(\frac{E_{\mathrm{t}}}{E_{\mathrm{i}}}\right)^2 \frac{n_{\mathrm{r}2}\cos\Phi_{\mathrm{t}}}{n_{\mathrm{r}1}\cos\Phi_{\mathrm{i}}}, \tag{48}
$$

which can be computed from the normal and parallel components given by the Fresnel equations. The resulting formulae become rather lengthy and confusing.

Some simplified cases better demonstrate the typical behavior. These simplifications include

1. Air as the first medium with $n_{\mathrm{r}1} = 1$ and $\sigma_1 = 0$
2. A nonabsorbing second medium with $\sigma_2 = 0$
3. An incident wave that is normal to the interface, i.e., $\phi_{\mathrm{i}} = 0$

With assumptions (1) and (2), we obtain from Eqs. 48 and 45 for the two components of the reflected beam

[7]Reflectance (etc.) is used rather than reflectivity since it is not normalized to the unit area; this is similar to the use of the word resistance (not normalized) versus resistivity, distinguishing between the suffixes *-ance* and *-ivity*.

$$R_\perp = \frac{\sin^2(\Phi_t - \Phi_i)}{\sin^2(\Phi_t + \Phi_i)} \quad \text{and} \quad R_{||} = \frac{\tan^2(\Phi_t - \Phi_i)}{\tan^2(\Phi_t + \Phi_i)} \tag{49}$$

and the transmitted beam

$$T_\perp = \frac{\sin 2\Phi_i \sin 2\Phi_t}{\sin^2(\Phi_i + \Phi_t)} \quad \text{and} \quad T_{||} = \frac{\sin 2\Phi_i \sin 2\Phi_t}{\sin^2(\Phi_i + \Phi_t)\cos^2(\Phi_i - \Phi_t)}. \tag{50}$$

When assumptions (2) and (3) hold, we obtain the well-known relations for the reflected and transmitted beams at normal incidence:

$$\begin{aligned} R &= R_\perp = R_{||} = \left(\frac{n_{r1} - n_{r2}}{n_{r1} + n_{r2}}\right)^2 \\ T &= T_\perp = T_{||} = \frac{4\, n_{r1}\, n_{r2}}{(n_{r1} + n_{r2})^2}. \end{aligned} \tag{51}$$

1.2.2 Reflectance and Transmittance in Semiconductors

In semiconductors, condition (2) of the previous section (no absorption) no longer holds since there *is* absorption. In the following, we drop the subscript 2 in the optical constants of the semiconductor and just use n_r, while outside vacuum is assumed. In spectral ranges in which n_r is much larger than κ, we still obtain rather simple approximations, which for the reflected wave are

$$R_\perp \cong \frac{(n_r - \cos\Phi_i)^2 + \kappa^2}{(n_r + \cos\Phi_i)^2 + \kappa^2} \quad \text{and} \quad R_{||} \cong R_\perp \frac{(n_r - \sin\Phi_i \tan\Phi_i)^2 + \kappa^2}{(n_r + \sin\Phi_i \tan\Phi_i)^2 + \kappa^2}. \tag{52}$$

These reflectance[7] components are shown for a set of n_r and κ in Fig. 3. There is a substantial amplitude difference for low angles of incidence between the parallel and perpendicular polarized components. The ratio of the components $R_{||}$ and $R_\perp$ shows a maximum at the *pseudo-Brewster angle*[6] given by

$$\tan^2\Phi_B \cong n_r^2 + \kappa^2. \tag{53}$$

Reflection at this angle can be used to obtain nearly linearly polarized light from the normal component (the amplitude of the parallel component is negligible) – see Fig. 3.

For normal incidence, parallel or normal components can no longer be distinguished in reflectance, and we obtain for an absorbing medium

$$R_0 = R_\perp = R_{||} = \frac{(n_r - 1)^2 + \kappa^2}{(n_r + 1)^2 + \kappa^2}. \tag{54}$$

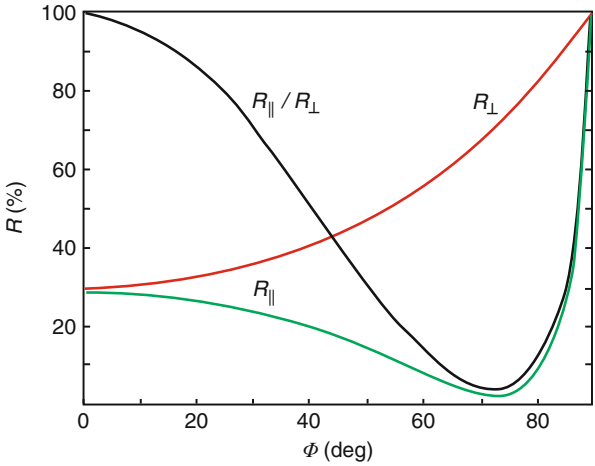


Fig. 3 Polarized components of the reflectance parallel and perpendicular to the semiconductor surface as a function of the incident angle, computed from Eq. 49 for $n_r = 3$ and $\kappa = 1$

Equation 54 can be rewritten as

$$\kappa^2 + \left(n_r - \frac{1+R_0}{1-R_0}\right)^2 = \frac{4R_0}{(1-R_0)^2}. \tag{55}$$

This is the equation of a circle centered at $n = (1 + R_0)/(1 - R_0)$ with a radius of $2\sqrt{R_0}/(1 - R_0)$, as shown in Fig. 4.

For a semiconductor plate with two planar surfaces separated by a distance d, we must consider reflection from *both* surfaces, as shown in Fig. 5. The result of the *series* of sequential reflections indicated by the subscript Σ is

$$R_\Sigma = R_0\left(1 + \frac{T_0^2(1-R_0)^2}{1-R_0^2T_0^2}\right) \quad \text{and} \quad T_\Sigma = T_0\frac{(1-R_0)^2}{1-R_0^2T_0^2}, \tag{56}$$

with T_0 as the transmittance through the slab at its first pass under normal incidence without reflection. For a perfectly transmitting slab ($T_0 = 1$), we consequently obtain

$$R_\Sigma = \frac{2R_0}{1-R_0} \cong 2R_0. \tag{57}$$

In a more precise analysis, a phase shift δ between both reflected waves must be included, which, for normal incidence, is given by

$$\delta_0 = \frac{2\pi n_r d}{\lambda}, \tag{58}$$

where λ is the wavelength of the light in vacuum and d is the slab thickness. The total reflectance of this plate at normal incidence is given by

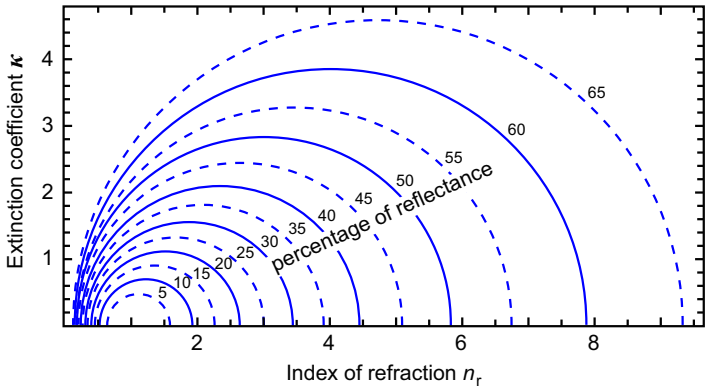


Fig. 4 Relation between reflectance of a single surface at normal incidence and the optical constants n_r and κ for a single surface using Eq. 55

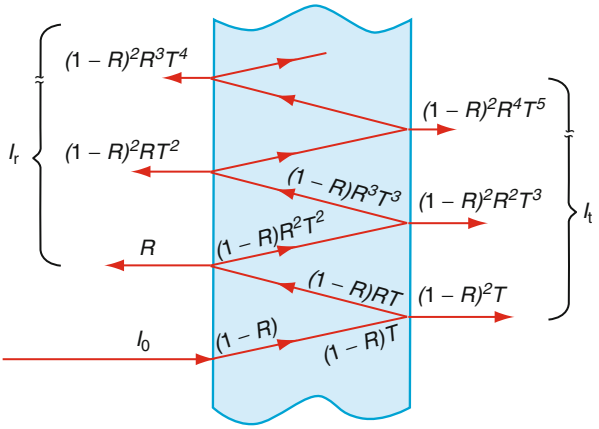


Fig. 5 Multireflection in a semitransparent slab with normal-incidence single-pass transmittance T and reflectance R

$$R_\Sigma = \frac{\left(\exp(\alpha_o d/2) - \exp(-\alpha_o d/2)\right)^2 + 4R_0 \sin^2\delta_0}{\left(\exp(\alpha_o d/2) - R_0 \exp(-\alpha_o d/2)\right)^2 + 4R_0 \sin^2(\delta_0 + \psi_0)}, \tag{59}$$

where α_o is the optical absorption coefficient (Eq. 24), R_0 is the reflectance from a single surface from Eq. 54, and $\psi_0 = \tan^{-1}\left(2\kappa/\left(n_r^2 + \kappa^2 - 1\right)\right)$. The transmittance through this plate is given by

$$T_\Sigma = \frac{(1-R_0)^2 + 4R_0 \sin^2\psi_0}{\left(\exp(\alpha_o d/2) - R_0 \exp(-\alpha_o d/2)\right)^2 + 4R_0 \sin^2(\delta_0 + \psi_0)}. \tag{60}$$

For vanishing absorption ($\kappa = \alpha_o = 0, \psi_0 = 0$), we confirm from Eqs. 59 and 60 that the sum of transmittance and reflectance is 1.

The interference pattern from the superposition of front and back reflections makes an evaluation of κ (or α_o) in thin planar layers difficult. However, one can deduce the index of refraction from the difference in the wavelength of adjacent maxima (or minima) of the transmittance. T_Σ has extrema for

$$2n_r d = N\lambda_1 \quad \text{and} \quad 2n_r d = (N+1)\lambda_2, \tag{61}$$

where N is the order of the fringe. By elimination of N, we obtain the refractive index from

$$2n_r d = \left(\lambda_2^{-1} - \lambda_1^{-1}\right)^{-1}. \tag{62}$$

This method is applicable if the refractive index n_r does not vary significantly between the two extrema. The useful Eq. 62 may also be applied to evaluate the thickness d of a transparent slab from a known refractive index n_r.

In order to determine the absorption coefficient, compromises must be made to average out the interference pattern by making either the surfaces slightly non-planar (rough) or the light slightly polychromatic. This yields an "average" reflectance

$$\overline{R} = R_0\left(1 + \overline{T}\exp(-\alpha_o d)\right) \tag{63}$$

and an "average" transmittance

$$\overline{T} = \frac{(1-R_0)^2\exp(-\alpha_o d)}{1 - R_0^2\exp(-2\alpha_o d)} \cong (1-R_0)^2\exp(-\alpha_o d), \tag{64}$$

with R_0 given by Eq. 54. From the measured values of $\overline{T}$ and $\overline{R}$, the more relevant values of α_o and R_0, and hence of κ and n_r, can be obtained. This is done most easily with an Abac chart (Nazarewicz et al. 1962), as shown in Fig. 6.

The reflectance and absorption coefficients of some typical semiconductors are given in Table 1.

1.2.3 Modulation Spectroscopy

For measuring optical constants, modulating one of the parameters of a semiconductor while analyzing the optical response is a very powerful technique. The light itself (frequency or intensity), temperature, pressure, mechanical stress (uniaxial), and electric or magnetic field have all been employed as parameters to be modulated. As optical response, the reflectance is most often used by measuring its relative change $\Delta R/R_0$ or its higher derivatives (see Sect. 2.3.3 of ▶ Chap. 29, "Carrier Generation") as a function of the wavelength; absorption within the band is too strong to be observed except for extremely thin platelets. The modulation is detected

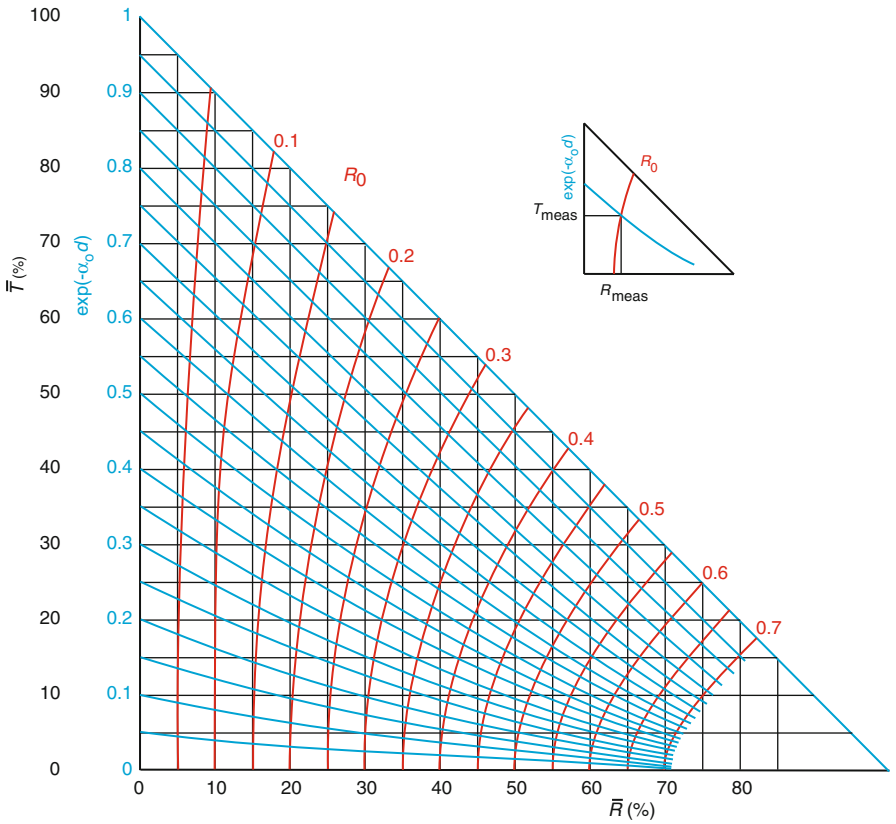


Fig. 6 Abac chart for obtaining the optical absorption coefficient α_o and the normal-incidence reflectance R_0 from measured values of the average transmittance $\overline{T}$ and the average reflectance $\overline{R}$

with a phase-sensitive lock-in technique, which is extremely sensitive and permits detection of rather small signals ($<10^{-5}$) – see Cardona (1969), Seraphin (1973), Aspnes (1980), and Pollak and Shen (1993). With this technique, it is relatively easy to detect changes in spectral ranges of high optical absorption. For instance, we can measure the energy difference between the critical points within the valence and conduction bands. This method is analyzed below.

From the reflectance equation (Eq. 51)

$$R_0 = \left(\frac{n_r - n_{ra}}{n_r + n_{ra}}\right)^2, \tag{65}$$

we obtain the modulated signal of the reflected light beam, which is proportional to the changes in the complex dielectric constant; n_{ra} is the index of refraction for the

ambient, i.e., an external medium in contact with the semiconductor. Using the *Maxwell relation*

$$n_r^2 = \varepsilon_{opt} = \varepsilon \text{ and } n_{ra}^2 = \varepsilon_{a,opt} = \varepsilon_a, \tag{66}$$

such changes can be expressed as (Re = real part)

$$\frac{\Delta R}{R_0} = \mathrm{Re}\left(\frac{2n_{ra}}{n_r(\varepsilon/\varepsilon_a)}\Delta\varepsilon\right), \tag{67}$$

Equation 67 can also be written as

$$\frac{\Delta R}{R_0} = \mathrm{Re}((\alpha_S - i\beta_S)\Delta\tilde{\varepsilon}) = \alpha_S(\varepsilon', \varepsilon'')\Delta\varepsilon' + \beta_S(\varepsilon', \varepsilon'')\Delta\varepsilon'', \tag{68}$$

where α_s and β_s are the *Seraphin coefficients* (Seraphin and Bottka 1965) and ε', ε'' and $\Delta\varepsilon'$, $\Delta\varepsilon''$ are, respectively, the real and imaginary parts of the dielectric constants and their changes due to the modulation.

A more useful equation for comparing the experimental results with a theoretical analysis is the inverse relation

$$\Delta\varepsilon = \frac{\dfrac{\Delta R}{R_0} + 2i\Delta\theta}{\alpha_S - i\beta_S} \tag{69}$$

with

$$\Delta\theta(\omega) = -\frac{\omega}{\pi}P\int_0^\infty \frac{\Delta R(\omega')}{R_0(\omega')}\frac{d\omega'}{\omega'^2 - \omega^2}, \tag{70}$$

where P indicates the principal value of the integral. The integral can be evaluated by a Kramers-Kronig analysis – see Sect. 1.2 of ▶ Chap. 11, "Photon–Phonon Interaction," and Basu (1998).

The analysis of the modulation reflection spectrum is complicated. It permits identification of the energy and type of critical points. As an example, we will give below a short outline of how the energy difference from the valence band to an extremum of an upper conduction band can be obtained from typical reflection signals.

The complex dielectric constant for band-to-band excitation in a one-electron approximation is given by (see Sect. 1.2 of ▶ Chap. 13, "Band-to-Band Transitions," and, e.g., Aspnes 1980)

$$\varepsilon(E,\gamma) = 1 + \frac{e^2\hbar^2}{\varepsilon_0 m_0^2 E^2} \sum_{k,c,v} |\mathbf{e}\,\mathbf{M}_{cv}(\mathbf{k})|^2 \left(\frac{1}{E_c(\mathbf{k}) - E_v(\mathbf{k}) - hv - i\gamma} + \frac{1}{E_c(\mathbf{k}) - E_v(\mathbf{k}) + hv + i\gamma} \right), \tag{71}$$

where $\mathbf{M}_{cv}$ is the momentum matrix element, $\mathbf{e}$ is the unit polarization vector of the impinging electromagnetic wave, $E = hv$ is the photon energy, E_c and E_v are energies of conduction and valence bands, and γ is an empirical damping parameter. Equation 71 can be simplified for a direct transition into one parabolic band to

$$\varepsilon(E,\gamma) \cong \frac{e^2\hbar^2}{\varepsilon_0 m_0^2 E^2} |\mathbf{e}\,\mathbf{M}_{\mathrm{cv}}(\mathbf{k})|^2 \int \frac{d^3k}{E_{\mathrm{c}}(\mathbf{k}) - E_{\mathrm{v}}(\mathbf{k}) - hv - i\gamma}, \tag{72}$$

with

$$E_{\mathrm{c}}(\mathbf{k}) - E_{\mathrm{v}}(\mathbf{k}) = E_{\mathrm{g}} + \frac{\hbar^2}{2} \left(\frac{k_x^2}{m_x} + \frac{k_y^2}{m_y} + \frac{k_z^2}{m_z} \right). \tag{73}$$

From Eq. 72, we deduce that an isolated critical point (of the type M_0 or M_3 – Sect. 1.1 of ▶ Chap. 13, "Band-to-Band Transitions") can be described by

$$\varepsilon(E,\gamma) \cong A\gamma^{-n} \exp(i\theta) \ \left(E - E_{\mathrm{g}} + i\gamma\right)^n, \tag{74}$$

where A is an amplitude factor, θ is a phase factor for the critical-point transition, and n is an exponent which is $-1/2$, 0 (logarithmic), or 1/2 for one-, two-, or three-dimensional critical points. The energy gap E_{g} (i.e., the energy difference at the critical point in any of the valence and conduction bands related to the initiated transition) and the damping γ can be determined from the line shape, which typically shows three extrema near the critical point (Fig. 7). With subscripts A and B identifying the two largest extrema, E_{g} and γ can be estimated according to (Aspnes 1980)

$$E_{\mathrm{g}} = E_A + (E_B - E_A) f(q) \tag{75}$$

and

$$\gamma = (E_B - E_A) g(q) \cong E_B - E_A, \tag{76}$$

with $0.8 < g(q) < 1.2$ and $f(q)$ monotonically increasing from 0 to 1 with increasing $q = (\Delta R/R_0)_A/(\Delta R/R_0)_B$. Hence, E_{g} always lies closest to the largest peak in the reflection spectrum: the closer to the peak, the larger the ratio of their magnitude.

1.2.4 Ellipsometry

Ellipsometry is widely used to determine the optical constants of metals, semiconductors, and insulators. The technique is sensitive to the *phase* of the radiation

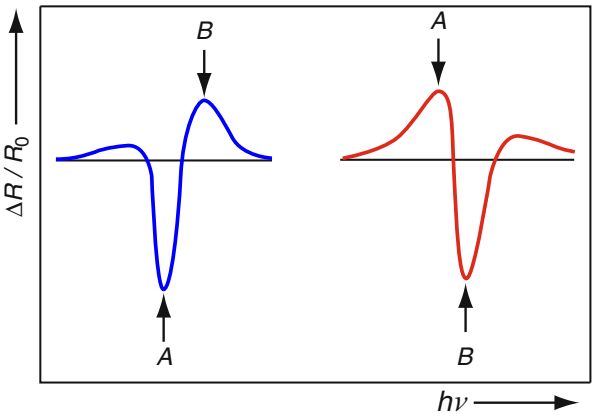


Fig. 7 Typical line shape of a reflection signal caused by a simple band-to-band transition

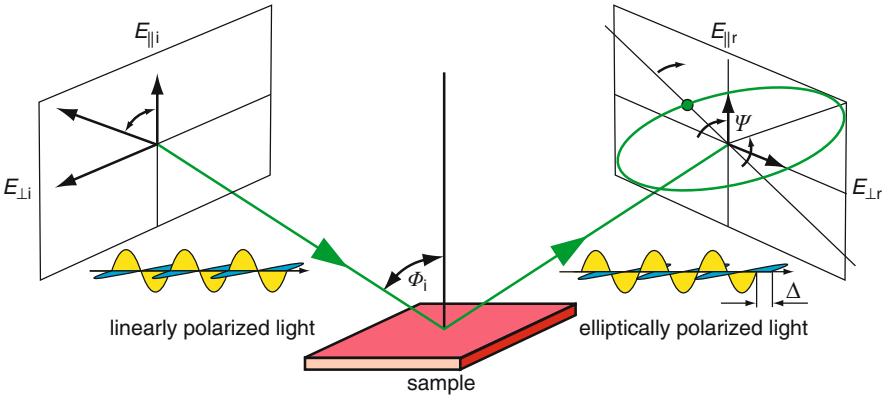


Fig. 8 Principle of a ellipsometry measurement. Linearly polarized incident light is reflected from the sample surface and experiences a change of its polarization state; the generally elliptically polarized reflected light is characterized by the ellipsometric angles Ψ and Δ

unlike reflectometry, which evaluates the intensity and hence the power flow. Ellipsometry measures the change in polarization of an incident radiation (in a known state) by the interaction with the solid; the interaction process may involve reflection or refraction, transmission, or scattering (Azzam and Bashara 1987). We focus on the usually applied reflection ellipsometry illustrated in Fig. 8. The linearly polarized beam of a light source with components $E_{\perp \mathrm{i}}$ and $E_{||\mathrm{i}}$ of the electric field normal and parallel to the plane of incidence meets the sample surface at an incident angle Φ_i near the Brewster angle Φ_B. The electric field components are in phase. After reflection, the amplitudes and the mutual phase of the components are altered,

yielding an elliptically polarized beam.[8] The optical constants of the sample are obtained from the parameters of the ellipse.

From Fresnel's Eq. 45, we read the *Fresnel coefficients*

$$\begin{aligned} r_{||} &= E_{||\mathrm{r}}/E_{||\mathrm{i}} = \frac{n_{\mathrm{r}2}\cos\Phi_{\mathrm{i}} - n_{\mathrm{r}1}\cos\Phi_{\mathrm{t}}}{n_{\mathrm{r}2}\cos\Phi_{\mathrm{i}} + n_{\mathrm{r}1}\cos\Phi_{\mathrm{t}}}, \\ r_{\perp} &= E_{\perp\mathrm{r}}/E_{\perp\mathrm{i}} = \frac{n_{\mathrm{r}1}\cos\Phi_{\mathrm{i}} - n_{\mathrm{r}2}\cos\Phi_{\mathrm{t}}}{n_{\mathrm{r}1}\cos\Phi_{\mathrm{i}} + n_{\mathrm{r}2}\cos\Phi_{\mathrm{t}}}. \end{aligned} \tag{77}$$

If media 1 and 2 are transparent, their refractive indices and also the angles Φ_{i} and Φ_{t} are real numbers, yielding real Fresnel coefficients; however, if one medium (or both) is absorbing, all quantities are complex quantities. In the following, we follow the conventional notation and omit the tilde (~) indicating complex quantities. The complex Fresnel coefficients are often written in terms of their amplitude $|r|$ and the phase shift δ_r experienced upon reflection

$$\begin{aligned} r_{||} &= |r_{||}|\ \exp\left(i\delta_{r_{||}}\right), \\ r_{\perp} &= |r_{\perp}|\ \exp\left(i\delta_{r_{\perp}}\right). \end{aligned} \tag{78}$$

The measurement of the polarization state in ellipsometry yields the ratio ρ of the Fresnel coefficients, which is also expressed in terms of the ellipsometric angles Ψ and Δ,

$$\rho = r_{||}/r_{\perp} = \tan\psi \exp(i\Delta). \tag{79}$$

We note that the reflectance R in Eq. 48 equals $|\rho|^2$. From Eq. 79 follows

$$\tan\psi = |r_{||}|/|r_{\perp}|, \text{ and} \tag{80}$$

$$\Delta = \delta_{r_{||}} - \delta_{r_{\perp}}. \tag{81}$$

The ellipsometric angles in Eqs. 80 and 81 express which change the incident oscillating electric field vector experiences in amplitude and phase upon reflection in its components parallel and perpendicular to the plane of incidence. Use of Eq. 79 and Snell's law Eq. 38 yields a relation of the refractive indices in terms of ρ and ϕ_{i} (Azzam and Bashara 1987)

[8] The reflected light is elliptically polarized if there is absorption in the sample. Without absorption, the reflected light remains linearly polarized, but with an altered angle of polarization with respect to the plane of incidence.

$$n_2 = n_1 \tan\phi_i \left(1 - \frac{4\rho}{(1+\rho)^2}\sin^2\phi_i\right)^{1/2}$$
$$= n_1 \sin\phi_i \left(1 + \left(\frac{1-\rho}{1+\rho}\right)^2 \tan^2\phi_i\right)^{1/2}. \quad (82)$$

According to Eq. 82, the complex index of refraction in the sample n_2 can be determined if that in the medium of incidence n_1 is known and the ratio ρ of the Fresnel coefficients is measured at an angle ϕ_i. Since $n = \sqrt{\varepsilon}$ (Eq. 20), a measurement at varied photon energies (*spectroscopic ellipsometry*) also yields the complex dielectric function $\varepsilon(\hbar\omega)$.

Equation 82 describes the ideal two-phase case with homogeneous media (e.g., ambient and bulk) and a sharp interface. Real samples are, however, generally covered with oxides and adsorbed contaminants and are usually microscopically rough; even atomically sharp surfaces have a relaxed and usually also reconstructed structure which differs from the bulk. Such a transition region alters the measured data with respect to that of a pure bulk by averaging the dielectric responses of overlayer and bulk. In the analysis of ellipsometric data, usually still the perfect case is assumed, and the measured quantity is referred to as *pseudodielectric function* $\langle\varepsilon\rangle = \langle\varepsilon_1\rangle + i\langle\varepsilon_2\rangle$. The true bulk dielectric function ε can be extracted from $\langle\varepsilon\rangle$ by assuming an ideal three-phase model consisting of ambient/overlayer/bulk. A rough surface is typically modeled using the Bruggemann effective-medium model (Bruggemann 1935), where the (homogeneous) overlayer is assumed to consist of a given fraction of bulk material and the remaining fraction of voids (i.e., vacuum). The effect of a 1 nm thick oxide overlayer on $\langle\varepsilon\rangle$ for GaAs is shown in Fig. 9a. The same effect was found for a 1.4 nm thick microscopic roughness modeled as a uniform film consisting of 60% GaAs and 40% voids (Aspnes and Studna 1983).

In the two-phase model, the pseudodielectric function is calculated from the ellipsometrically measured complex reflectance ratio ρ according to Eqs. 82 and 20

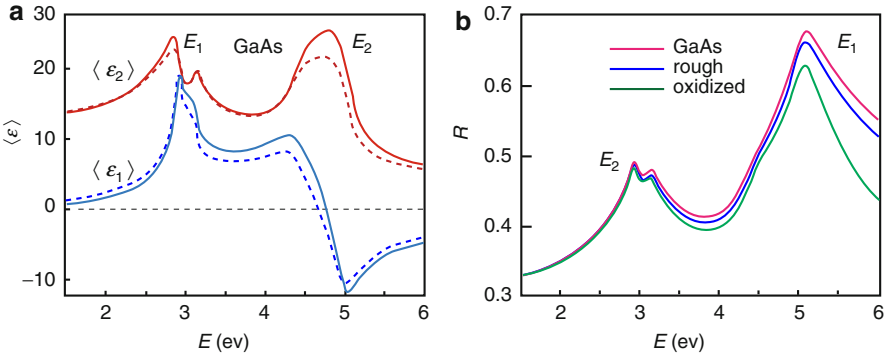


Fig. 9 **(a)** Pseudodielectric function $\langle\varepsilon\rangle$ of a GaAs bulk crystal without (*solid line*) and with a 1 nm thick oxide overlayer (*dashed line*). **(b)** Reflectance of GaAs with an abrupt surface and surfaces covered with 3 nm microscopic roughness or oxide; from Aspnes and Studna 1983

$$\langle\varepsilon\rangle = \varepsilon_a \left(\sin^2\Phi_i + \sin^2\Phi_i \tan^2\Phi_i \left(\frac{1-\rho}{1+\rho}\right)^2 \right), \tag{83}$$

where ε_a is the dielectric function of the ambient. If an overlayer with a dielectric function ε_o is inserted between ambient and bulk sample, and the reasonable assumption $|\varepsilon| \gg |\varepsilon_o| \gg |\varepsilon_a|$ applies, then the resulting $\langle\varepsilon\rangle$ can be approximately expressed in the three-layer model as (Burge and Bennett 1964; Aspnes 1976)

$$\langle\varepsilon\rangle = \varepsilon + \frac{4\pi i}{\lambda} \frac{d}{\varepsilon_o(\varepsilon - \varepsilon_a)} \left(\varepsilon - \varepsilon_a \sin^2\Phi_i\right)^{1/2}. \tag{84}$$

Here, d is the thickness of the overlayer and ε is the dielectric function of the solid without overlayer.

The reflectance is also affected by overlayer effects, albeit to a smaller extent as illustrated in Fig. 9b. Both, a 3 nm thick roughness composed as in panel (a) but 3 nm thick and an oxide layer of equal thickness reduce the reflectance. The effect is more pronounced at higher photon energy, where the penetration depth into the sample is smaller.

1.3 Frequency Dependence of the Dielectric Function and Dielectric Screening

The interrelated optical constants (summarized in Eqs. 22, 24, and 45 leading to Eq. 54) show a strong dependence on the photon energy as illustrated in Table 1 and Figs. 1a and 9. In this section, we outline the origin of this dependence for the dielectric constant; a more detailed treatement on the underlying microscopic interactions is given in the following chapters.

The dielectric constant $\tilde{\varepsilon}$ descibes the static and dynamic response of the solid on an electric field $\mathbf{E}$ by a polarization $\mathbf{P}$ according to $\mathbf{P} = \varepsilon_0(\tilde{\varepsilon} - 1)\mathbf{E}$. In this response, we can distinguish a transverse dielectric constant related to interaction with electromagnetic radiation and a longitudinal one related to charge screening; the dielectric screening can be divided into electronic and ionic parts. The transverse dielectric constant is influenced by resonance transitions. It becomes a function of the optical excitation frequency, as shown for the real part of $\tilde{\varepsilon}$ in Fig. 9. It is no longer a constant and therefore better called the *dielectric function.*

1.3.1 Longitudinal and Transverse Dielectric Constants

Most of the discussion in the literature deals with the longitudinal dielectric constant, which is responsible for the screening of free and bound charges and is important for the field distribution seen by carriers near impurities, for plasmon interaction, etc. Much less attention has been given to the transverse dielectric constant which couples with transverse phonons even though it is responsible for the screening seen by an interacting electromagnetic field, i.e., for optical properties. One of the reasons is the fact that most of this interaction takes place near the center of the Brillouin zone with negligible change of

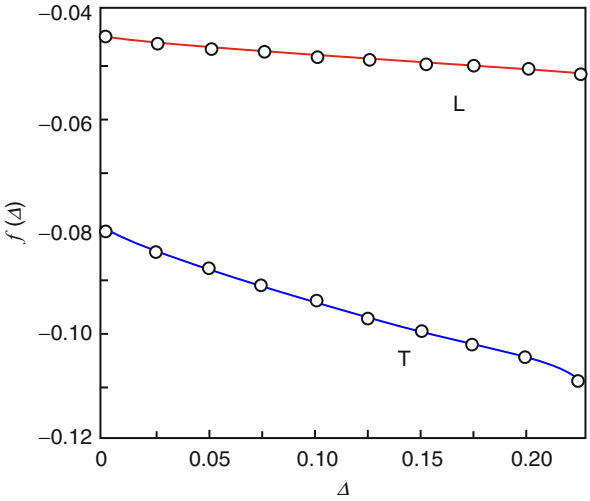


Fig. 10 Correction function f (Δ) (Eq. 86) for the longitudinal (L) and transverse (T) dielectric constant near $\omega = 0$ (After Sharma and Auluck 1981)

the momentum vector, and it can be shown that near $\mathbf{q} = 0$, the longitudinal and transverse dielectric constants are the same (Sharma and Auluck 1981):

$$\varepsilon_{\mathrm{L}}(\mathbf{q} = 0, \omega) = \varepsilon_{\mathrm{T}}(\mathbf{q} = 0, \omega). \tag{85}$$

This justifies the use of the common $\varepsilon = \varepsilon_{\mathrm{L}}$ for the analysis of most optical data.

When the wavevector deviates from $\mathbf{q} = 0$, however, both dielectric constants become different. For the static dielectric constant, Sharma and Auluck (1981) have evaluated ε_{L} and ε_{T} and obtained for the real part

$$\varepsilon'_{\mathrm{L,T}} = \varepsilon'(\mathbf{q} = 0, \omega = 0) - 24\left(\frac{E_{\mathrm{p}} E_{\mathrm{F}}}{E_{\mathrm{g}}^2 k_{\mathrm{F}}}\right)^2 q^2 \left| f_{\mathrm{L,T}}(\Delta) \right|, \tag{86}$$

where E_{p} is the plasmon energy (see Sect. 1.1 of ▶ Chap. 12, "Photon–Free-Electron Interaction"), E_{F} is the Fermi energy ($\hbar^2 k_{\mathrm{F}}^2/(2m)$), and k_{F} is the wavenumber at the Fermi surface assumed to be spherical. The bandgap is E_{g}, $\Delta = E_{\mathrm{g}}/(4E_{\mathrm{F}})$, and $f_{\mathrm{L,T}}(\Delta)$ is the correction function for transverse or longitudinal interaction, shown in Fig. 10. Both ε_{L} and ε_{T} decrease proportional to q^2, however, with ε_{T} becoming smaller than ε_{L}.

1.3.2 Spectral Ranges of the Dielectric Function

The interaction of the solid with electromagnetic radiation with a frequency $\nu \neq 0$ can be described classically by oscillations with resonance frequencies. There are several types of oscillators with respective resonances. *Ionic* oscillations occur in lattices with (at least partial) ionic forces and refer to the motion of the ions. *Electronic* oscillations comprise the motion of electrons jointly acting as an entity with respect to ions of the lattice; these may be the ensemble of free conduction-band electrons or that of all valence electrons oscillating with respect to the ion cores. The effect on the dielectric

function $\widetilde{\varepsilon}(\omega)$ is given by the sum of all resonances of these oscillators, leading to the Kramers-Heisenberg dielectric function

$$\widetilde{\varepsilon}(\omega) = 1 + \sum_j \frac{\omega_{0j}^2}{\omega_{0j}^2 - \omega^2 - i\gamma_j\omega}, \tag{87}$$

where ω_{0j} are the resonance frequencies and γ_j are the corresponding damping coefficients. Near resonance, the states of the photon and the respective oscillator interact strongly and cannot be distinguished; these mixing states are referred to as *polariton* states. The resonance with ionic lattice oscillations is labeled *phonon-polariton* and described in Sect. 3.1 of ▶ Chap. 11, "Photon–Phonon Interaction," and the *plasmon-polariton* referring to oscillations of electron ensembles is treated in Sect. 1 of ▶ Chap. 12, "Photon–Free-Electron Interaction." Further effects occur at band-to-band transitions considered in ▶ Chap. 13, "Band-to-Band Transitions." In the nonresonant regions between resonances, the dielectric function is approximately constant. The general form of $\widetilde{\varepsilon}(\omega)$ can be concluded from the energies of the contributing resonances as illustrated in Fig. 11.

At frequencies above the bandgap energy ($h\nu > h\nu_g \equiv E_g$), electronic band-to-band transitions take energy from the electric (electromagnetic) field and are therefore responsible for the screening, which, dependent on the excitation spectrum, is a rather complicated function of the frequency. Calculated data of ε for ZnS, ZnSe, and ZnTe in this energy range obtained from density-polarisation functional theory is discussed by Grüning and Attaccalite (2016).

For $h\nu < E_g$, only few electrons are free; their influence on the dielectric constant is usually negligible. Impurity atoms with electronic resonances in this range are too few to markedly influence $\widetilde{\varepsilon}$. Hence, for $\hbar\omega_{TO} < h\nu < E_g$, we have a nearly constant dielectric constant: $\varepsilon' = \varepsilon_{opt} = \text{const}$, and $\varepsilon'' = 0$; ε_{opt} is also referred to as ε_∞ in literature.

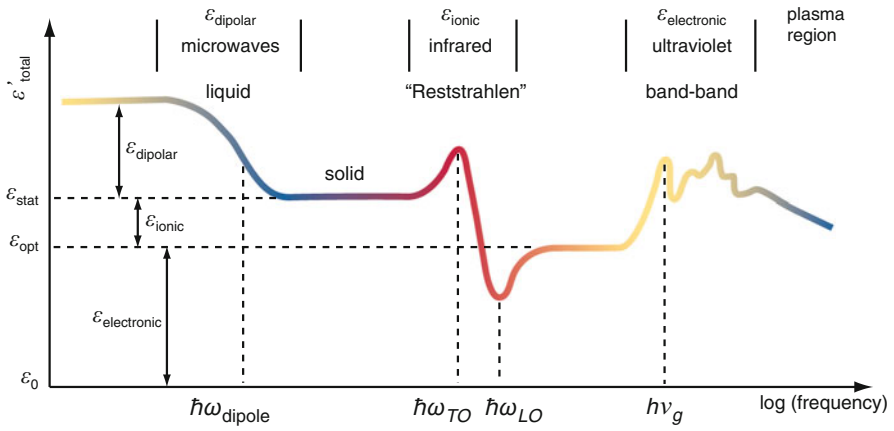


Fig. 11 Frequency dependence of the real part ε' of the dielectric constant, with ionic and electronic resonances indicated. The structure within the band-to-band transition is indicative of the multilevel transition

At lower frequencies, ionic motion can follow the ac-electric field with resonances in the *Reststrahl* range (Sect. 2.2 of ▶ Chap. 11, "Photon–Phonon Interaction"). For $hv < \hbar\omega_{TO}$, the real part of the dielectric constant ε' is given by the static dielectric constant ε_{stat}. The major change in ε occurs near the energy of the transversal optical phonon $\hbar\omega_{TO}$, see Sect. 2.1 of ▶ Chap. 11, "Photon–Phonon Interaction" (Fig. 2 of ▶ Chap. 11, "Photon–Phonon Interaction").

When melting, the semiconductor's molecular dipoles become free to rotate at rather low frequencies, causing another step in the dispersion relation shown in Fig. 11.

In summary, the tabulated values of the dielectric constants are restricted to the nonresonance frequency ranges of the spectrum: the optical dielectric constant between band edge and *Reststrahl* frequencies and the static dielectric constant below the *Reststrahl* frequency but above melting-related transitions. The static and optical dielectric constants of various crystals at room temperature are listed in Table 3. Near resonances, the dielectric function changes in a complex fashion with frequency and polarization of the interacting field.

The values of the dielectric constants given in Table 3 depend on ambient conditions. At 300 K temperature, coefficients $(\partial \ln \varepsilon_{stat,opt}/\partial T)_P$ for typical semiconductors are on the order of 10^{-4} K^{-1}; hydrostatic pressure coefficients $(\partial \ln \varepsilon_{stat,\,opt}/\partial P)_T$ are on the order of -10^{-2} GPa^{-1} (Samara 1987). Over a wide range, the temperature dependence is nonlinear and reasonably described by a similar function like that used for the bandgap E_g,

$$\varepsilon_{stat,opt} = \varepsilon_{stat,opt}(T=0) + \frac{\alpha T^2}{\beta + T}. \tag{88}$$

Table 3 Static and optical dielectric constants[a] of various crystals at $T = 300$ K in the nonresonant ranges: ε_{stat} for $hv < \hbar\omega_{TO}$ and ε_{opt} for $\hbar\omega_{LO} < hv < E_g$. Two entries denote respective different quantities parallel and perpendicular to the c axis of wurtzite crystals

Crystal	ε_{stat}	ε_{opt}	Crystal	ε_{stat}	ε_{opt}	Crystal	ε_{stat}	ε_{opt}
C	5.7	5.7	InN	13.1, 14.4	8.4, 8.4	PbS	19	18.5
Si	11.9	11.9	InP	12.6	10.9	PbSe	280	25.2
Ge	16.0	16.0	InAs	15.2	12.4	PbTe	450	36.9
SiC	10.0	6.7	InSb	16.8	15.7	CuCl	6.3	3.7
AlN	8.3, 8.9	4.4, 4.8	ZnO	7.8, 8.8	3.7, 3.8	CuBr	7.0	4.4
AlP	9.8	7.5	ZnS	8.6	5.2	CuI	7.1	5.5
AlAs	10.1	8.2	ZnSe	8.6	5.7	CuO_2	7.1	6.2
AlSb	12.1	10.2	ZnTe	10.3	7.3	LiF	9.3	1.9
GaN	9.6, 10.6	5.4, 5.4	CdO	21.9	2.1	KCl	4.5	2.2
GaP	11	9.1	CdS	10.2, 9.0	5.4, 5.3	KBr	4.5	2.4
GaAs	12.8	10.9	CdSe	10.2, 9.3	6.2, 6.3	KI	4.7	2.7
GaSb	15.7	14.4	CdTe	10.4	7.1	CsI	6.3	3.1

[a]For many applications such a simple representation of ε is much too coarse. For a listing of $n(\omega)$ and $k(\omega)$ in a very large frequency range from vacuum UV to IR, see Palik (1985), Adachi (2012)

In contrast, the dependence on hydostatic pressure is found to be linear at least up to several GPa (Goñi et al. 1989).

The imaginary part of the dielectric function ε'' cannot be described in the classical model indicated for ε' above; it requires a quantum-mechanical analysis considered in ▶ Chap. 13, "Band-to-Band Transitions." If the real part $\varepsilon'(\omega)$ is known in the entire frequency range, the imaginary part can also be obtained by using Kramers-Kronig relations, see Sect. 1.2 of ▶ Chap. 11, "Photon–Phonon Interaction." For a general overview, see Stern (1963).

1.3.3 Dielectric Screening as Function of Wavevector

There are occasions where it is more appropriate to consider dielectric screening as a function of the wavevector instead of the frequency. In the neighborhood of a lattice defect, one uses the dielectric constant to describe the influence of the screening of the surrounding lattice on the electric potential extending from such a defect. Only when the polarization of this lattice, farther away from the center by at least a few lattice constants is considered, the static dielectric constant $\varepsilon_{\text{stat}}$ may be well used.

When the interaction in closer vicinity of the center is concerned so that less screening from the lattice occurs, the dielectric constant decreases. For an estimate of such a dielectric function, one considers the eigenstates between which such interaction takes place, yielding as the screening function an expression similar to the expression for optical band-to-band absorption Eq. 13 of ▶ Chap. 13, "Band-to-Band Transitions" (Cohen 1963):

$$\varepsilon'(\mathbf{k},\mathbf{k}') = 1 + \frac{e^2}{\varepsilon_0 k^2} \sum_{v,c,k'} \frac{\left|\mathbf{e}\cdot\mathbf{M}^{cv}_{kk'}\right|}{E_c(\mathbf{k}+\mathbf{k}') - E_v(\mathbf{k})}. \tag{89}$$

where $\mathbf{M}^{cv}_{kk'}$ is the matrix element for transitions from $\mathbf{k}$ in the valence band to $\mathbf{k}'$ in the conduction band. This expression shows more clearly the decrease of $\varepsilon \propto k^{-2}$ for larger values of $\mathbf{k}$ – see Fig. 12.

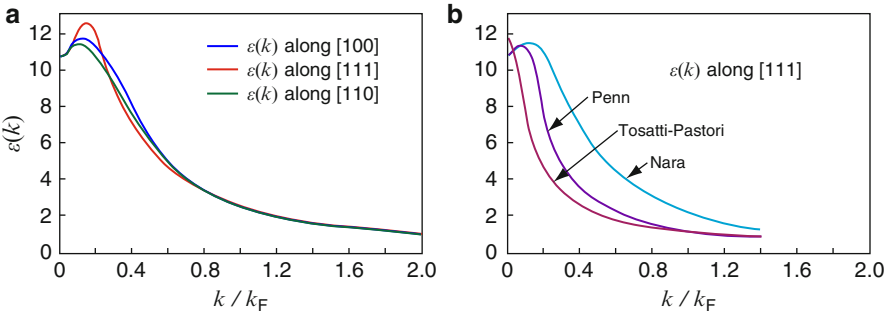


Fig. 12 Dielectric constant of Si as a function of the wavenumber **(a)** according to numerical computations by Nara (1965) and **(b)** in comparison to analytical approximations by Penn (1962) and Tosatti and Pastori Parravicini (1971). k_F is given by Eq. 91

The screening function was computed for Si first by Nara (1965) and for other semiconductors by Walter and Cohen (1970), Brust (1972), and Wang (1997). An approximation was given by Penn (1962):

$$\varepsilon'(k) = 1 + \frac{\left(\frac{\hbar\omega_p}{E_g^*}\right)^2 \left(1 - \frac{E_g^*}{4E_F}\right)}{\left(1 + \frac{E_F}{E_g^*}\left(\frac{k}{k_F}\right)^2 \sqrt{1 - \frac{E_g^*}{4E_F}}\right)^2} \tag{90}$$

where ω_p is the plasma frequency, E_g^* is the energy difference between bonding and antibonding states in the valence and conduction bands (Table 4), E_F is the Fermi energy for valence electrons, and $k_F = k(E_F)$; for screening by valence electrons in an isotropic crystal, k_F is given by

$$k_F^3 = 3\pi^2 n_v, \tag{91}$$

Table 4 Electronic crystal parameters: Fermi energy of valence electrons $E_{F,v}$ plasmon energy of valence electrons, fraction of ionicity[a], dielectric average bandgap E_g^*, and Thomas-Fermi screening length of valence electrons (After van Vechten 1980)

Crystal	$E_{F,v}$ (eV)	$\hbar\omega_p$ (eV)	f_i (%)	E_g^* (eV)	λ_{TF} (Å)	Crystal	$E_{F,v}$ (eV)	$\hbar\omega_p$ (eV)	f_i (%)	E_g^* (eV)	λ_{TF} (Å)
C	28.9	31.2	0	13.5	0.39	BeO	25.3	28.3	62.0	18.8	0.40
Si	12.5	16.6	0	4.77	0.48	BeS	15.6	19.7	28.5	7.47	0.45
Ge	11.5	15.6	0	4.31	0.49	BeSe	14.3	18.4	26.1	6.57	0.46
α-Sn	8.72	12.7	0	3.06	0.52	BeTe	12.0	16.1	16.9	4.98	0.49
SiC	19.4	23.2	17.7	9.12	0.43	MgTe	9.04	13.0	55.6	4.80	0.52
BN	28.1	30.6	25.7	15.2	0.39	ZnO	17.6	21.5	65.5	12.5	0.44
BP	17.8	21.7	5.8	7.66	0.44	ZnS	12.6	16.7	62.3	7.85	0.48
BAs	16.1	20.1	2.6	6.47	0.45	ZnSe	11.4	15.6	62.3	6.98	0.49
AlN	19.3	23.0	44.6	11.0	0.45	ZnTe	9.91	14.0	59.9	5.66	0.51
AlP	23.4	16.5	38.8	6.03	0.47	CdS	10.8	14.9	67.9	7.01	0.50
AlAs	11.4	15.5	43.2	5.81	0.47	CdSe	9.94	14.0	68.4	6.42	0.51
AlSb	9.77	13.8	43.3	4.68	0.51	CdTe	8.75	12.7	67.5	5.40	0.53
GaN	18.2	22.1	45.2	10.3	0.44	HgS	10.8	14.9	79.0	8.20	0.50
GaP	12.4	16.5	32.8	5.76	0.48	HgSe	9.91	14.0	68.0	6.06	0.51
GaAs	11.5	15.6	31.0	5.20	0.48	HgTe	8.75	12.6	65.2	4.95	0.52
GaSb	9.82	13.9	26.0	4.12	0.51	CuF	20.3	23.9	76.6	18.1	0.42
InN	14.9	18.9	49.6	8.36	0.46	CuCl	12.6	16.7	74.6	9.60	0.48
InP	10.7	14.8	42.1	5.16	0.50	CuBr	11.1	15.2	73.5	7.35	0.49
InAs	10.1	14.2	35.9	4.58	0.51	CuI	10.1	14.1	69.2	6.61	0.51
InSb	8.76	12.7	32.7	3.76	0.52	AgI	8.77	12.8	77.3	6.48	0.52

[a]With f defined by $f_i = (C/E_g^*)^2$ with $C = 1.5e^2(Z_b/r_b - Z_a/r_a)\exp(-r_{ab}/\lambda_{TF})$, Z_i and r_i the valency and covalent radii, and E_g^* the distance of bonding and antibonding band states in valence and conduction bands

where n_v is the density of valence electrons. See also Tosatti and Pastori Parravicini (1971).

The behavior of the dielectric constant for Si as a function of the wavenumber was computed by several groups and is shown in Fig. 12. Aside from a small maximum near $k = 0$, it shows a rapid decrease toward $\varepsilon = 1$ for higher k values, indicating that the capability of the lattice to screen decreases $\propto k^{-2}$, that is, with higher excitation into the band. Since $k \propto 1/r$, it also indicates that as expected, the screening decreases rapidly as one approaches the immediate proximity of the defect center.

1.3.4 Empirical Screening Parameters

An empirical formula for the dielectric function was given by Nara and Morita (1966)

$$\frac{1}{\varepsilon(k)} = \frac{A k^2}{k^2 + \alpha^2} + \frac{B k^2}{k^2 + \beta^2} + \frac{1}{\varepsilon_{stat}} \frac{\gamma^2}{k^2 + \gamma^2} \tag{92}$$

with adjustable parameters which, for Si, have the values $A = 1.175$, $B = -0.175$, $\alpha = 0.7572$, $\beta = 0.3123$, and $\gamma = 2.044$.

Often, instead of the dielectric function, one uses an *effective dielectric constant* given by

$$\frac{1}{\varepsilon^*} = \frac{1}{\varepsilon_{opt}} - \frac{1}{\varepsilon_{stat}}. \tag{93}$$

when the screening close to a defect center is concerned. This effective dielectric constant expresses the difference between the electronic polarization already accounted for in a Bloch electron picture and the lattice polarization including ionic displacements.

1.3.5 Screened Coulomb Potential

When the Coulomb potential screened by the dielectric function extends too far from a center to give reasonable agreement with the experiment because of other interfering charges, one introduces a stronger screened Coulomb potential. Such a screening can be caused by additional charged carriers that are interfering within the reach of the Coulomb potential. The long-range part of the Coulomb potential can be reduced by an empirical exponential factor

$$V(r) = \frac{e}{4\pi \varepsilon_{stat} \varepsilon_0 r} \exp\left(-\frac{r}{\lambda_s}\right). \tag{94}$$

This potential is called the *Yukawa potential,* with λ_s the screening radius, and is shown in Fig. 13 compared with the unscreened Coulomb potential.

When the screening is caused by a relatively small density of free electrons surrounding the center, e.g., a charged defect, λ_s is defined as a *Debye-Hückel*-type[9] *length*:

[9]In semiconductors, mostly free electrons rather than ions (as in the Debye-Hückel theory) provide the screening. This characteristic length in semiconductor physics is more commonly called the *Debye length*.

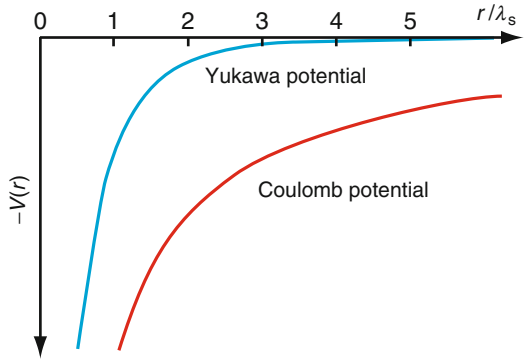


Fig. 13 Coulomb and screened Coulomb (Yukawa) potentials. The radius r is scaled in units of the screening radius λ_s

$$\lambda_D = \sqrt{\frac{\varepsilon_{stat}\,\varepsilon_0\,kT}{e^2\,n}} = 381\sqrt{\frac{\varepsilon_{stat}}{10}\,\frac{T(\mathrm{K})}{100}\,\frac{10^{16}}{n\,(\mathrm{cm}^{-3})}}\ (\text{Å}). \tag{95}$$

Here, the Boltzmann statistic is applied for the screening electrons. In contrast, λ_s is defined as a *Thomas-Fermi length* when a higher carrier density requires the application of the Fermi-Dirac statistics:

$$\lambda_{TF} = \pi^{2/3}\sqrt{\frac{\varepsilon_{stat}\,\varepsilon_0\,\hbar^2}{3^{2/3} m_n\,e^2\,n^{1/3}}} = 11.6\,\sqrt{\frac{\varepsilon_{stat}}{10}\,\frac{m_0}{m_n}\left(\frac{10^{18}}{n\,(\mathrm{cm}^{-3})}\right)^{1/3}}\ (\text{Å}). \tag{96}$$

Finally, a combination of dielectric and electronic screening was proposed by Haken (1963), who introduced as an effective dielectric constant

$$\frac{1}{\varepsilon_{eff}} = \frac{1}{\varepsilon_{stat}} + \frac{1}{2}\left(\frac{1}{\varepsilon_{opt}} - \frac{1}{\varepsilon_{stat}}\right)\left(\exp\left(\frac{r_e}{r_{pe}}\right) + \exp\left(\frac{r_h}{r_{ph}}\right)\right). \tag{97}$$

with

$$r_{pe} = \sqrt{\frac{\hbar}{2m_n\,\omega_{LO}}}\ \text{and}\ r_{ph} = \sqrt{\frac{\hbar}{2m_p\,\omega_{LO}}} \tag{98}$$

the polaron radii for electrons and holes; r_e and r_h in Eq. 97 are the relative distances of the electron and hole from their respective center of gravity within an exciton – see Sect. 1 of ▶ Chap. 14, "Excitons."

By introducing empirical parameters in simple relations, an improved agreement between theory and experiment is obtained, and the results can be given in an analytical form. More recently Thomas-Fermi screening was introduced into exchange hybrid functionals in total-energy calculations, yielding realistic band structures for a wide range of semiconductors (Clark and Robertson 2010).

2 Photonic Bandgap Structures

The solids considered in Sect. 1 were homogeneous media characterized by a unique dielectric function.[10] If the dielectric properties in the solid are periodically altered, we obtain a *photonic crystal* (Russell 1986; John 1987; Yablonovitch et al. 1991). Photonic crystals can be considered an analogue to semiconductors where the atoms of the crystal lattice are replaced by a periodic arrangment of mesoscopic media (periodicity on the order of the optical wavelegth) with differing dielectric constants: periodic dielectric structures modify the refractive index in spatial dimensions and thereby create a bandgap for photons at optical frequencies similar to the periodic atomic lattice creating a bandgap for electrons. Such a photonic bandgap gives rise to distinct optical phenomena, which can be utilized for, e.g., low-loss waveguiding, highly reflecting and even omnidirectional mirrors, suppresion or enhancement of spontaneous radiation of an emitter in a solid, efficient light outcoupling, and slow light with a low group velocity; for reviews, see Villeneuve and Piché (1994), Joannopoulos et al. (2008), Sakoda (2001), Busch and Wehrspohn (2003), Busch et al. (2007), and Baba (2008).

2.1 Photonic Crystals

The basic concepts of Brillouin zones, dispersion relations, or Bloch waves developed for electrons in ▶ Chaps. 6, "The Origin of Band Structure," to ▶ 8, "Bands and Bandgaps in Solids," also apply for photons, i.e., optical modes, propagating through a medium with periodically varying dielectric properties. Some differences arise from the scalar waves used for electrons, while the electromagnetic wave of an optical mode is described by *vectors* **H** and **E**. In addition, the time-dependent Schrödinger equation allows also for negative eigenvalues, while only *squares* of eigenfrequencies and thus positive values appear in the wave equation of electrodynamics. Still, the conclusions are similar: the propagation of some energies is inhibited by interference effects (Bragg reflections), and defects representing deviations from periodicity create localized states.

Photonic crystals can be fabricated for one, two, or three dimensions as indicated in Fig. 14. A one-dimensional (1D) photonic crystal is made of an alternating stack of two layers with different dielectric constants ε_A and ε_B. In a 2D photonic crystal, periodicity of ε_A and ε_B occurs in two directions (achieved, e.g., by etching parallel holes into a homogeneous medium), and a 3D photonic crystal has periodicity of ε_A and ε_B in all directions. Fabrication of 3D structures for optical frequencies is particularly challenging due to the requirement of high precision on a submicron scale; a total bandgap for propagation in *any* direction can only be achieved for a 3D photonic crystal.

A number of approximations is usually assumed for the dielectric materials of the photonic crystal for calculating the dispersion relation. The *field strengths are small*,

[10] An exception was the inclusion of an overlayer considered in ellipsometry (Eq. 84), leading to the description by an *effective* yet spatially not varying dielectric function $\langle\varepsilon\rangle$.

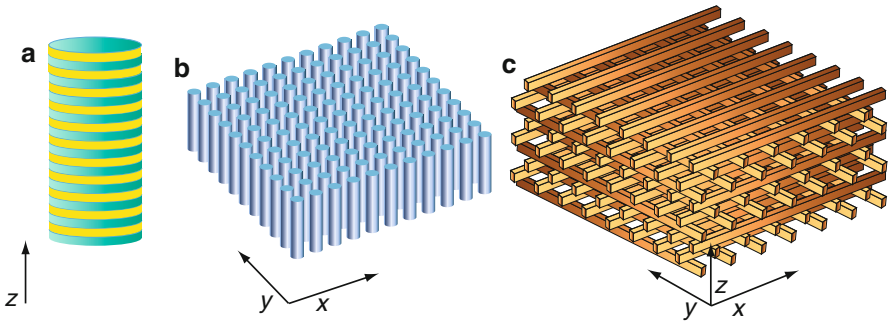


Fig. 14 Schematic of photonic crystals in **(a)** one, **(b)** two, and **(c)** three dimensions. Arrows indicate axes of periodicity, providing photonic bandgaps for specific modes along these directions

yielding linear equations and allowing for ignoring higher-order terms considered in Sect. 3. The *dielectric media are isotropic*, leading to a scalar $\varepsilon(\mathbf{r}, \omega)$ instead of a tensor $\underline{\boldsymbol{\varepsilon}}(\mathbf{r}, \omega)$. The frequency dependence of the dielectric function is neglected, and an appropriate *dielectric constant* for the considered frequency range is used. Losses in the dielectrics are also neglected, yielding a *real* dielectric constant $\varepsilon(\mathbf{r})$, and the media are assumed to be nonmagnetic and nonconductive, i.e., $\mu = 1$ and $\sigma = 0$. With the above approximations, $n = n_\mathrm{r} = \sqrt{\varepsilon}$ applies.

The optical properties of a photonic crystal are described by the dispersion relations for modes traveling through the medium; these follow from the eigenwert equation for the vector field. Generally, the eigenwert problem of photonic crystals is solved for the magnetic field $\mathbf{H}(\mathbf{r})$ and not for the electric field $\mathbf{E}(\mathbf{r})$ due to a significant simplification in the eigenwert evaluation[11]; $\mathbf{E}(\mathbf{r})$ is then determined via Eq. 2.

Combining Maxwell's curl equations 2 and 1 we obtain a wave equation for the magnetic field:

$$\frac{\partial}{\partial \mathbf{r}} \times \left(\frac{1}{\varepsilon(\mathbf{r})} \frac{\partial}{\partial \mathbf{r}} \times \mathbf{H} \right) - \frac{\omega^2}{c^2} \mathbf{H} = 0, \tag{99}$$

where $c = 1/\sqrt{\varepsilon_0 \mu_0}$ is the vacuum speed of light. The dielectric constant $\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r} + \mathbf{R}_i)$ with primitive lattice vectors $\mathbf{R}_i$ contains the entire information on the spatially periodic structure of the photonic crystal. Due to this periodicity, the solutions for $\mathbf{H}$ (and also those for $\mathbf{E}$) can be written in terms of the Bloch form, i.e., a plane wave (or, more generally, linear combinations of plane waves) times a vector function with the periodicity of the lattice,

$$\begin{aligned} \mathbf{H}_{n\mathbf{k}}(\mathbf{r}, t) &= \exp(i(\mathbf{k}\mathbf{r} - \omega t))\ \ \mathbf{u}_{n\mathbf{k}}(\mathbf{r}), \\ \mathbf{u}_{n\mathbf{k}}(\mathbf{r}) &= \mathbf{u}_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}). \end{aligned} \tag{100}$$

[11] Equation 99 expressed in terms of $\mathbf{E}$ is correct but not hermitian, see Joannopoulos et al. (2008).

Here, $\mathbf{R}$ is a lattice vector for the Bravais lattice over which $\varepsilon(\mathbf{r})$ is periodic in space, $\mathbf{k}$ is a wavevector in the first Brillouin zone defined by the primitive components of the reciprocal lattice vector $\mathbf{G}$ which fulfills $\mathbf{G}\,\mathbf{R} = m\,2\pi$, and n is the band index that labels the multiple solutions corresponding to $\mathbf{k}$ in the reduced zone scheme. Substituting the solution Eq. 100 into the wave equation 99 with periodic boundary conditions, a (basically infinite) family of solutions with discrete eigenvalues $(\omega_n/c)^2$ labeled by their band index is found; such a problem corresponds to higher harmonics of a vibrating string. The eigenvalue problem is solved numerically using matrices typically of size 1,000 × 1,000.

2.1.1 Scaling of Solutions

The solutions obtained from the eigenvalue problem are scalable due to the linearity of the Maxwell equations (Joannopoulos et al. 2008). If we know an electromagnetic eigenmode $\mathbf{H}(\mathbf{r})$ for a given frequency ω in a periodic dielectric assembly $\varepsilon(\mathbf{r})$, we also know the solution of another dielectric configuration which is scaled by a factor s (e.g., enlarged) with respect to the former, i.e., $\varepsilon_{\text{structure2}}(\mathbf{r}) = \varepsilon(\mathbf{r}/s)$; the new mode profile and its frequency are obtained by scaling the old mode and the old eigenfrequency by the same factor: $\mathbf{H}_{\text{structure2}}(\mathbf{r}_{\text{structure2}}) = \mathbf{H}(\mathbf{r}_{\text{structure2}}/s)$ and $\omega_{\text{structure2}} = \omega/s$. This scaling rule is important for applications: models mechanically fabricated on a length scale of centimeters to be tested in the microwave regime have the same electromagnetic properties as downscaled microstructures operating with light.

Scaling can also be applied to the dielectric constant. If the dielectric constant is changed by a constant scaling factor everywhere in a new structure such that $\varepsilon_{\text{structure2}}(\mathbf{r}) = \varepsilon(\mathbf{r})/s^2$, the mode profiles remain unchanged, but their corresponding frequencies are changed to $\omega_{\text{structure2}} = s\,\omega$.

Due to the scaling property of photonic structures, often wavenumbers and frequencies with dimensionless units are used defined by

$$k' = \frac{k\ a}{2\pi}, \quad \omega' = \frac{\omega\ a}{2\pi c}, \tag{101}$$

where c is the speed of light. The primed dimensionless quantity k' measures the wavenumber in units of $2\pi/a$ (a is the spatial periodicity of ε), and the boundary of the first Brillouin zone is just given by $k' = ½$. Spatial coordinates are measured in units of a, yielding for one dimension $z' = z/a$. Photonic band schemes are usually discussed in terms of these scaled quantities.

2.1.2 One-Dimensional Photonic Crystal

A one-dimensional photonic crystal consists of periodically alternating layers with different dielectric constants as shown in Fig. 15. Such multilayer films are widely used since long as distributed Bragg reflectors (DBR mirrors) and interference filters.

The dispersion relation of the 1D photonic crystal for the modes traveling along the z direction follows from the eigenwert equation for the vector field, analogous to the periodicity approach for electron waves discussed in Sect. 1.2 of ▶ Chap. 6, "The Origin of Band Structure." In the 1D problem, $\varepsilon(\mathbf{r}) = \varepsilon(z)$ applies, given by $\varepsilon_<$ in the layer with low refractive index and $\varepsilon_>$ in the other.

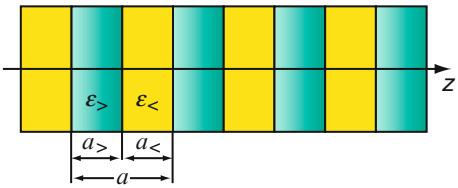


Fig. 15 One-dimensional photonic crystal realized by a stack comprising layers with larger and smaller dielectric constants $\varepsilon_>$ and $\varepsilon_<$ in a repetition with periodicity $a = a_> + a_<$

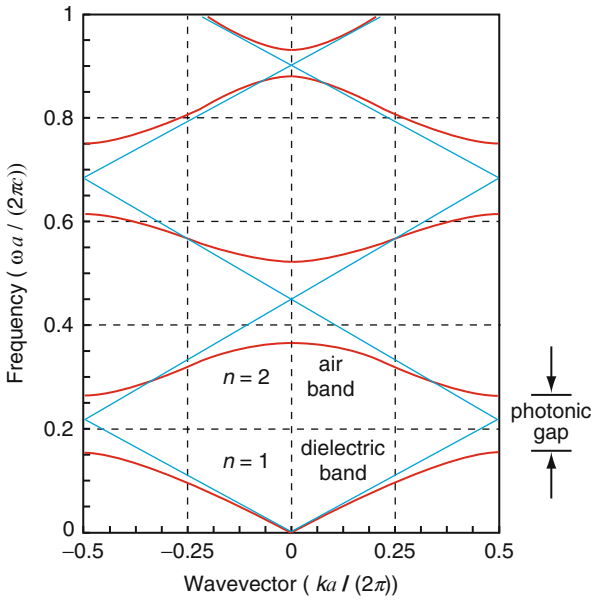


Fig. 16 Band structure for on-axis propagation in a one-dimensional photonic crystal, realized by a stack comprising layers A and B in a repetition with periodicity $a = a_A + a_B \equiv a_> + a_<$. *Red lines* refer to $a_A = a_B = a/2$ and $\varepsilon_A = 12$, $\varepsilon_B = 1$; *blue lines* apply for a uniform $\varepsilon = \varepsilon_> = \varepsilon_<$

If the 1D system has a *uniform* dielectric constant $\varepsilon_< = \varepsilon_> = \varepsilon$, then the eigenvalues representing the frequency spectrum are given by the light line $\omega(k) = ck/\sqrt{\varepsilon}$; this linear dispersion relation is depicted by the blue lines in Fig. 16. The artificial periodicity folds the bands to the first Brillouin zone, a scenario corresponding to the *empty lattice* discussed in Sect. 4.1 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals." There are many eigenvalues ω_n for each k, n being the band index.[12] The first band with lowest frequencies has a degeneracy at the zone boundary: the mode $k = \pi/a$ lies at an equivalent wavevector to the

[12]Here, n is just a label not to be confused with the refractive index; the number of eigenvalues ω_n is given by the size of the (generally truncated) matrix in the secular equation of the eigenvalue problem for a periodically perturbed ε. Considering more plane waves enlarges the secular equation and thereby the number of solutions ω_n for a given k including increasingly more contributions of shorter waves with higher frequencies.

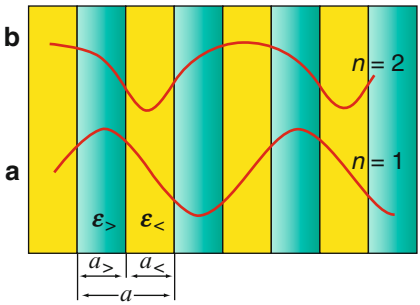


Fig. 17 Squared electric field of modes related to **(a)** the top of band labelled 1 and **(b)** the bottom of band 2 for the 1D photonic crystal shown in Fig. 15 with parameters used in Fig. 16. Bright regions represent layers with the lower dielectric constant (After Inoue and Ohtaka 2004)

$k = -\pi/a$ mode and at the same frequency ω_1; the same applies for higher bands at higher frequencies.

Now, the periodic variation of $\varepsilon(z)$ is introduced as indicated in Fig. 15. Values applied for the calculated results presented in Fig. 16 are $\varepsilon_> = \varepsilon(\text{Si}) = 12$, $\varepsilon_< = \varepsilon(\text{air}) = 1$, and $a_> = a_< = a/2$. The dispersion relations $\omega_n(k)$ show features also found for photonic crystals in higher dimensions. The first band with lowest frequency ($n = 1$) has a linear dispersion near $k = 0$ similar to the unperturbed case. Since the wavelength $\lambda = 2\pi/k$ of the corresponding waves is large (much larger than the periodicity a of ε), such waves experience an *averaged*, largely uniform dielectric constant. Near the boundary of the Brillouin zone, the dispersion curves bend and have zero slope at $k = \pi/a$ (or, $k' = 0.5$). Here, the group velocity $\partial\omega(k)/\partial k$ vanishes. The solutions are standing waves with a short wavelength $\lambda = 2a$. We note that the degeneracy of the empty lattice is now lifted: the two waves have different frequencies separated by a gap.

It is instructive to study the intensity of the two modes at the zone boundary, given by the square of the electric field. Fig. 17 shows that the mode of the lower band ($n = 1$) has maximum intensity in the layers with high dielectric constant $\varepsilon_> = \varepsilon_{\text{Si}}$, while the mode of the upper band[13] ($n = 2$) has maximum intensity at $\varepsilon_< = \varepsilon_{\text{air}}$; the band below the gap is therefore referred to as the *dielectric band* and a band above a gap as the *air band* analogous to the valence and conduction bands in a semiconductor. A similar behavior is also found in the pair of bands at higher frequencies at the zone boundary; lower-order modes generally tend to concentrate its displacement field in regions of high dielectric constant. Even if – for large dielectric contrast – *both* bands of a pair concentrate in $\varepsilon_>$ layers, the bottom band always concentrates more; the photonic bandgap originates from this difference in field energy location (Joannopoulos et al. 2008).

In the frequency range of the photonic bandgap, there are no solutions for on-axis propagating waves with a real wavevector k. Any spatially periodic dielectric variation in one dimension leads to a bandgap for propagation in this direction, and the gap widens for larger dielectric contrast. The reflectivity spectrum of such a layer stack exhibits a frequency range with nearly perfect reflectance $R = 1$ (compare

[13]There are only two ways to center such standing waves: the nodes can either be centered in each $\varepsilon_>$ layer or in each $\varepsilon_<$ layer; other positions violate the symmetry of the unit cell about its center.

Fig. 19); this range is given by the optical bandgap and is referred to as stop band. Such *distributed Bragg reflectors* are widely used in optical devices as dielectric mirrors (DBR mirrors).

2.1.3 Two-Dimensional Photonic Crystals

A two-dimensional photonic crystal with periodicity in $\varepsilon(\mathbf{r})$ along two axes and constant ε along the third shows some additional features, which are not observed in 1D photonic crystals. The electromagnetic vector field decouples into two fields of different polarization directions; their modes are referred to as TM (transverse-magnetic) and TE (transverse-electric). TM modes have **H** in the photonic-crystal plane and **E** normal to the plane, while TE modes have the reverse: **E** in the plane and **H** normal. The photonic bandgaps of TE and TM modes may occur at different frequencies such that there may be a gap for one but not the other or that the gap of one is larger than that of the other. Examples are shown in Fig. 18 for a hexagonal 2D lattice near $\omega a/(2\pi c)$ equals 0.84 and 0.44, respectively. Furthermore, there occur many narrow bands with zero group velocity in the interior of the Brillouin zone and not only at the zone boundary.

The band structure of a two-dimensional photonic crystal composed of a triangular airhole pattern ($\varepsilon = 1$) in Si ($\varepsilon = 12$) is shown in Fig. 18. The difference in ε of the two media is sufficient to open a complete photonic bandgap for in-plane propagation in such structures. A useful quantity for characterizing such gaps is the gap-midgap ratio defined by the frequency width of the gap $\Delta\omega$ over the

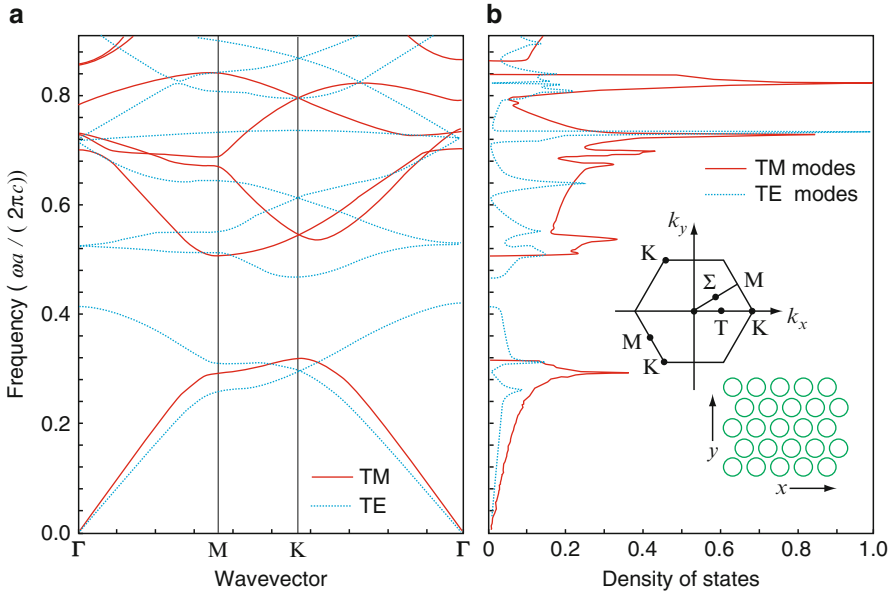


Fig. 18 **(a)** Band structure and **(b)** density of states for a two-dimensional photonic crystal with a triangular pattern of cylindrical holes (*bottom inset*), the *top inset* shows the corresponding reciprocal lattice (After Birner et al. 1999)

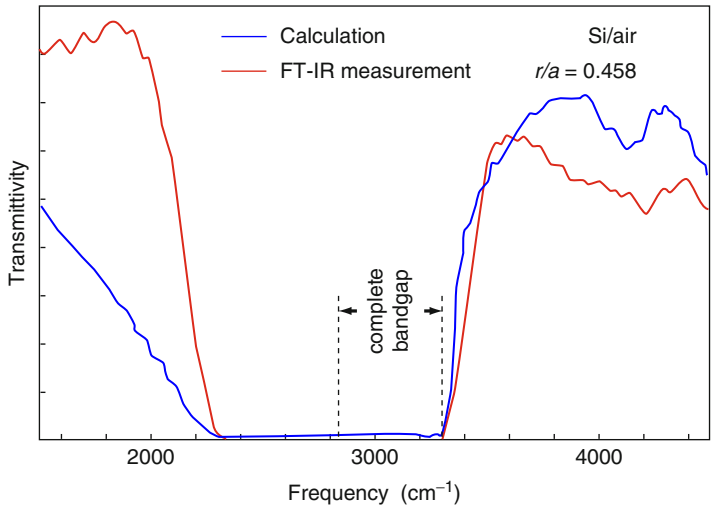


Fig. 19 Calculated and measured transmission of a triangular Si air-hole structure for TM modes. *Dashed lines* mark the frequency interval of the complete bandgap for in-plane propagation (After Birner et al. 1999)

frequency ω_0 in the middle of the gap; this quantity remains constant in scaled systems. Besides the dielectric contrast $\Delta\omega/\omega_0$ also depends, however, on the ratio of hole radius to hole distance. The effective dielectric constant in the long-wavelength limit discussed for the first band at the zone boundary is given for TE modes by the arithmetic volume average of $\varepsilon_>$ and $\varepsilon_<$; for TM modes, however, no such explicit expression can be given (Kirchner et al. 1998), yielding different bandgaps despite equal periodicity of the two mode types.

Photonic bandgaps appear as spectral regions where the photonic density of states $N(\omega)$ equals zero. Analogous to the electronic density of states, $N(\omega)$ is obtained by counting all allowed states of a given frequency

$$N(\omega) = \sum_n \int_{\mathrm{BZ}} d^D k \;\; \delta(\omega - \omega_n(\mathbf{k})). \tag{102}$$

The superscript D denotes the dimensions being 2 in 2D photonic crystals. $N(\omega)$ of the airhole structure with the band structure of Fig. 18a is given in panel *b* of the figure. The triangular hole arrangement forms a hexagonal lattice with the hexagonal Brillouin zone shown in the inset of Fig. 18b. Comparing band structure and corresponding density of state, we find that most peaks in $N(\omega)$ are related to the critical points Γ, M, or K of the Brillouin zone[14] but points of zero group velocity also

[14]Instead of the notations M and K, also the labels X and J, respectively, are used in literature for critical points of a two-dimensional hexagonal Brillouin zone.

occur in the interior of the zone. Bandgaps are easily seen and obviously different for TM and TE modes, though there is a clear complete gap near $\omega a/(2\pi c) = 0.44$.

Airhole structures like that illustrated in Fig. 18b were realized with many semiconductors using electron-beam lithography and anistropic etching. The photonic bandgap of a Si structure for in-plane TM modes propagating in the Γ-M direction is shown in Fig. 19. The ratio of hole radius to hole distance is $r/a = 0.458$ with $a = 1.55$ μm; the transmission was measured using Fourier transform IR spectroscopy. Within the range marked by dashed lines, no propagation occurs for any polarization and any lateral direction.

2.1.4 Three-Dimensional Photonic Crystals

A complete photonic bandgap for mode propagation in *any* direction can only be achieved for 3D photonic crystals. Such a bandgap provides an omnidirectional mirror for light in this frequency range and thus the ability to localize light at a single point. Dispersion relations are similar to those of 2D photonic crystals illustrated in Fig. 18, but complete photonic bandgaps are rarer since they must apply for all modes along all directions in the Brillouin zone and not just for a single plane. Still structures exist and were experimentally demonstrated, albeit fabrication remains challenging; for a review, see von Freymann et al. (2010).

A variety of approaches was applied for 3D photonic-crystal fabrication. A *woodpile* (or logpile) *structure* as shown in Fig. 14c was made by a repeated deposition and either patterning of a dielectric film or wafer bonding followed by substrate removal. Such structures consist of a pile of straight rods with adjacent layers rotated by 90° (Ho et al. 1994). With a lateral rod distance of a and every second layer shifted by $a/2$, the sequence repeats after four layers with a vertical lattice constant b; this leads for $b/a = \sqrt{2}$ to a face-centered (*fcc*) unit cell with a two-rod basis. Theoretically, a gap-midgap ratio of 0.18 is achieved for 30% volume filling with Si and a ratio of 0.25 for the inverse structure at 82% Si volume filling. Experimentally, up to 40 dB suppression of transmittance was achieved (Noda et al. 2000). An *inverse-opal structure* was fabricated employing self-assembly of spheres. An opal structure can be made via colloidal assembly of silica or polystyrene spheres; neither the *fcc* nor the hexagonally close-packed (*hcp*) structure exhibits a complete bandgap, but the inverse *fcc* structure does for $n_r > 2.8$ and the inverse *hcp* structure for $n_r > 3.1$ (Busch et al. 2007). They can be made by infiltration of the air voids by the dielectric and removal of the spheres using wet etching. A gap-midgap ratio of up to 0.086 can theoretically be achieved for partial infiltration for inverse *fcc* opals; experimentally, a ratio of 0.05 was reported applying this approach (Blanco et al. 2000). Invers opals with tuned position of the photonic bandgap by appropriate solvent infiltration are reported by Lonergan et al. (2020).

2.2 Localized Defect Modes and Microcavities

2.2.1 Optical Defects

The photonic crystals considered in the previous section have perfect periodicity and consequently bandgaps at specific frequencies; modes with frequencies inside the

bandgap have a complex wavevector $k + i\kappa$. The imaginary component leads to an exponential decay of these evanescent modes on a length scale $1/\kappa$. The decay constant κ grows as the frequency approaches the midgap and disappears at the gap edges.

Evanescent modes cannot be excited in perfect photonic crystals due to their lack of translational symmetry. However, if the translational symmetry of the photonic crystal is broken by some optical defect, one evanescent mode may be compatible with the symmetry of this defect (the locally altered period) and thereby be sustained (see, e.g., Maksimovic et al. 2008). This – still evanescent – mode is then localized at the defect, and its frequency lies within the gap. Since κ gets maximum near midgap, states with such a frequency can be localized more tightly than those near the gap edges.

Defects can be created by various means. If a 1D photonic crystal (Fig. 15) is considered, a single layer of the sequence can be altered by using another thickness or another dielectric constant. The same applies for 2D and 3D photonic crystals, where, e.g., in a 2D airhole structure (inset Fig. 18b), one hole is omitted or altered in radius; such an irregularity creates a *point defect*. In 2D and 3D photonic crystals also *line defects* can be created, e.g., by omitting an entire row of holes in a 2D airhole structure. Eventually also a planar defect can be created in a 3D photonic crystal just by altering an entire plane in the crystal. The integration of defects to create waveguides and cavities (point defects) in 3D photonic crystals is reviewed by Rinne et al. (2008). According to a recent numerical study a cavity with large Q factor can be created with only 3 unit cells of an inverse woodpile structure with Si-like dielectric function (Devashish et al. 2019).

The introduction of optical defects is the basis for photonic-crystal cavities and photonic-crystal waveguides. The microcavity of a vertical-cavity surface-emitting laser (VCSEL) is an optical defect of a 1D photonic crystal formed by the top and bottom DBR mirrors; the $\lambda/(4n_>) + \lambda/(4n_<)$ periodicity of the distributed Bragg reflectors is broken at the location of the active zone which comprises electrically pumped quantum wells in the antinode of the localized defect mode. The electronic transition of the quantum wells is designed to be resonant to the optical defect, which sustains this specific mode and weakens the suppression of propagation within the stop band (the optical bandgap) at lasing energy. Line defects are used to guide light along a line in a photonic crystal for modes that lie in the photonic bandgap; thereby low-loss waveguides even with sharp bends can be created. A photonic-crystal fiber is also a 2D photonic crystal which guides light not only by refractive index differences; in a Bragg fiber, a lateral photonic bandgap is formed by concentric rings of multilayer films, and a holey fiber uses airholes in cross section to confine the light to the core representing the optical defect. Details are discussed by Hu et al. (2019).

2.2.2 Microcavity Effects

In presence of an optical defect, the global density of states $N(\omega)$ given in Eq. 102 is locally altered, yielding the *local density of states*

$$N(\mathbf{r},\omega) = \sum_n \int_{\mathrm{BZ}} d^D k \;\; |\mathbf{E}_{n\mathbf{k}}(\mathbf{r})|^2 \;\; \delta(\omega - \omega_n(\mathbf{k})). \tag{103}$$

The local density of states strongly modifies the electromagnetic environment of an embedded optically active material. The spontaneous emission of, e.g., a quantum dot (Sects. 2.2 of ▶ Chap. 3, "The Structure of Semiconductors," and 3.4 of ▶ Chap. 8, "Bands and Bandgaps in Solids") located in a 3D photonic crystal can completely be suppressed for frequencies inside the complete photonic bandgap (Yablonovitch 1987). On the other hand, the spontaneous emission can also be strongly enhanced, an effect referred to as *Purcell effect* (Purcell et al. 1946; Gérard et al. 1998). This effect is particularly pronounced in cavities with simultaneously a high quality factor $Q = \omega_0/\Delta\omega$ (the inverse of the gap-midgap ratio) *and* a small volume of the confined mode V; the latter condition leads to *microcavities* with dimensions in the range of λ.

At large Q/V ratio, the interaction of the emitter and the mode in the cavity is enhanced. The effect of this interaction was widely studied with quantum dots placed in microcavities providing 3D confinement for the light. In the *regime of weak coupling* between the emitter and the cavity mode, the emission rate of the emitter is increased over that of a simple planar cavity (1D confinement) for on-resonance conditions; off-resonance the emission rate can be strongly suppressed (Bayer et al. 2001). The change of the emission rate from the value without cavity $\Gamma_0 = 1/\tau_0$ with respect to the value for cavity coupling Γ_c is described by the *Purcell factor* $F_P = \Gamma_c/\Gamma_0$. The coupling to the cavity mode also affects the emission characteristics of the emitter. This effect is employed to strongly enhance the outcoupling efficiency of light generated in a semiconductor since otherwise most light is lost by internal total reflection due to the large refractive index step to air.[15]

In the *strong-coupling regime*, a reversible exchange of energy between the emitter and the cavity mode occurs: the photon is emitted into the mode of the cavity and reabsorbed by the emitter. The characteristic frequency of this oscillation is the *Rabi frequency* ω_R (Rabi et al. 1939). As a consequence, a Rabi splitting of the initial mode frequency into two frequencies is observed. The effect is shown in Fig. 20, where the emission of a quantum dot embedded in a microcavity was thermally shifted over the cavity resonance[16] (Reithmaier et al. 2004). The two corresponding emissions show a characteristic no-crossing behavior and a double peak of the mixed coherent state in resonance (at $T = 21$ K), split by the Rabi energy $\hbar\omega_R$. The mixed half-light/half-matter state is described in the field of cavity quantum electrodynamics by a quasiparticle called *cavity polariton*. A discussion on the physics and potential application was given by Kockum et al. (2019). There are various types of polaritons in semiconductors all resulting from a strong coupling

[15]Light generated, e.g., by luminescence within a GaAs crystal ($n_r \cong 3.5$) is coupled out to air by only ~4% due to the small critical angle of total reflection (~17°, Eq. 40) and the consequential small fraction of the solid angle. The resulting poor external quantum efficiency is strongly increased if the light is emitted into the defect mode of a cavity cladding the light emitter. This is of particular importance for the performance of single-photon sources.

[16]The emission of the quantum dot shifts to lower energy at increased temperature according to the change of the bandgap described by the Varshni dependence Eq. 44 of ▶ Chap. 8, "Bands and Bandgaps in Solids"; the energy of the cavity resonance also experiences a redshift by the T-dependence of the refractive index but at a much lower rate.

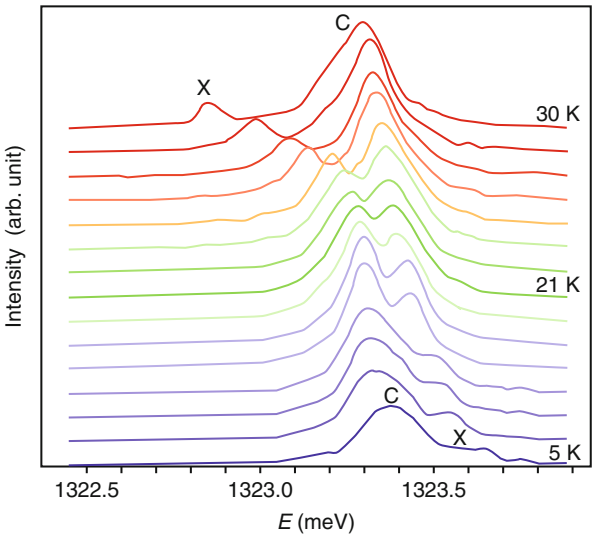


Fig. 20 Emision spectra of a quantum dot in a microcavity in the strong-coupling regime. X and C denote the excitonic quantum-dot emission and the cavity mode, respectively. The temperature is varied to tune the dot emission through the cavity resonance (After Reithmaier et al. 2004)

of electromagnetic waves with an elementary excitation like, e.g., a phonon or a plasmon; they will be discussed along with the respective excitations.

3 Nonlinear Optical Effects

Nonlinear optical effects are based on the fact that at sufficiently high amplitudes, all solid-state oscillations become anharmonic (Rabin and Tang 1975). This amplitude range is readily accessible with high light intensities (i.e., values of the electric field comparable to interatomic electric fields, typically 10^8 V/m, like those provided by pulsed lasers,[17] Bloembergen 1982) or with excitation near resonances where high-amplitude oscillations can easily be excited. Anharmonic oscillations cause the displacement **D** and the polarization **P** to respond no longer linearly to the electric field **E**; the description of the dependence requires higher-order terms in the dielectric constant or dielectric susceptibility.

These nonlinearities permit active interaction between two or more photons with a large variety of technically interesting phenomena of light mixing, rectification, and amplification. It opens a host of opportunities in the field of nonlinear optics to develop new devices and to design new experiments for analytical purposes. We will

[17]Material destruction is avoided by monochromatic irradiation in a wavelength range of little absorption and by the use of short pulses. Material destruction can occur by simple lattice heating, by dielectric breakdown (10^9 W/cm^2 is equivalent to 10^6 V/cm oscillation amplitude and achieved by focusing a 10^3 W laser on a spot of 10 μm diameter), by stimulated Brillouin scattering with intense multiphonon absorption, or by self-focusing. Typical destruction thresholds are 100MW/cm^2 for a 100 ns pulse in a low-absorbing range and 1 kW/cm^2 for a cw laser beam (In comparison, sunlight on the earth's surface transmits 100 mW/cm^2). For more on damage, see Kildal and Iseler (1976).

discuss here the basic elements of nonlinear optical effects. For more information, see reviews by Zernike and Midwinter (1973), Chemla and Jerphagnon (1980), Boyd (2008), and Stegeman and Stegeman (2012).

3.1 Electronic and Mixing Effects

Nonlinear effects are mostly investigated in the frequency range between atomic and electronic resonance absorption, i.e., between *Reststrahl* absorption in the infrared range (Sect. 2.2 of ▶ Chap. 11, "Photon–Phonon Interaction") and the band edge. Here, the electronic contribution can come from either the valence electrons or free conduction-band electrons if their density is high enough. We will discuss the influence of valence electrons first.

3.1.1 Nonresonant Effects of Valence Electrons

When the amplitude of the external electric field **E** is small, the induced polarization **P** is linear in **E**:

$$\mathbf{P} = \varepsilon_0 \chi \mathbf{E} = \varepsilon_0 (\varepsilon - 1) \mathbf{E}, \tag{104}$$

where χ is the linear susceptibility tensor (assumed isotropic and real for simplicity). At higher amplitudes, the corresponding polarization has higher terms in **E** and is known as *hyperpolarization*. It is expressed as an expansion in **E** containing the higher harmonics:

$$\begin{aligned} \frac{1}{\varepsilon_0}\mathbf{P}(\omega_i) = {} & \chi^{(1)}(\omega_i)\,\mathbf{E}(\omega_i) + \sum_{j,k} \chi^{(2)}\left(-\omega_i : \omega_j, \omega_k\right) \mathbf{E}\left(\omega_j\right) \mathbf{E}(\omega_k) \\ & + \sum_{j,k,l} \chi^{(3)}\left(-\omega_i : \omega_j, \omega_k, \omega_l\right) \mathbf{E}\left(\omega_j\right) \mathbf{E}(\omega_k)\,\mathbf{E}(\omega_l) + \ldots \end{aligned} \tag{105}$$

where the i-th order susceptibility $\chi^{(i)}$ is an $i + 1$-rank tensor containing waves of possibly the same or different frequencies ω_i, ω_j, ω_k, etc[18]; in the notation of the frequencies, the resulting value is given before the colon and the incident stimulating frequencies thereafter. The first-order second-rank susceptibility tensor $\chi_{ij}^{(1)}$

[18]Sometimes, an inverted relation between E and P is used; for instance, for the second-order term, one obtains

$$E_i(\omega_3) = \frac{1}{\varepsilon_0} \sum_{i,j,k} \delta_{ijk}(-\omega_3 : \omega_1, \omega_2)\, P_j(\omega_1) P_k(\omega_2) \tag{106}$$

with

$$\chi_{ijk}^{(2)}(-\omega_3 : \omega_1, \omega_2) = \varepsilon_0 \sum_{l,m,n} \chi_{il}^{(1)}(\omega_3) \chi_{jm}^{(1)}(\omega_1) \chi_{kn}^{(1)}(\omega_2) \delta_{lmn}(-\omega_3 : \omega_1, \omega_2). \tag{107}$$

corresponds to χ in Eq. 104. Often, the second-order third-rank susceptibility tensor $\chi^{(2)}_{ijk}$ is replaced by the tensor d_{im} with the convention

$$i = 1, \ldots, 3 \text{ for } x, y, z \text{ and } m = 1, \ldots, 6 \text{ for } xx, yy, zz, yz, zx, xy, \tag{108}$$

respectively, and $\chi^{(2)}_{ijk} \equiv 2d_{im}$. The dependence on the nonlinear second-order part then reads

$$\begin{pmatrix} P_1 \\ P_2 \\ P_3 \end{pmatrix} = \varepsilon_0 \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} \end{pmatrix} \begin{pmatrix} E_1^2 \\ E_2^2 \\ E_3^2 \\ 2E_2E_3 \\ 2E_3E_1 \\ 2E_1E_2 \end{pmatrix}. \tag{109}$$

All $d_{ij} = 0$ for centrosymmetric crystals such as Si and Ge, which both show no second-order nonlinear effect (but they do show *third*-harmonic generation – see, e.g., Attaccalite and Grüning 2013; Lan 2016). Cubic zincblende semiconductors such as GaAs or ZnSe have only single-valued non-zero components: $d_{14} = d_{25} = d_{36}$; wurtzite semiconductors such as GaN or ZnO have components with three different values: $d_{15} = d_{24}$, $d_{31} = d_{32}$, and d_{33}.

The values of the nonlinear susceptibility show dispersion like other optical parameters of semiconductors. Due to a substantial effect of dispersion (e.g., calculated data for AlAs and CdTe reported by Attaccalite and Grüning 2013) values for some semiconductors are given in Table 5 for a single wavelength only; it should be noted that published data show a large scatter. For evaluation of $\chi^{(n)}$, see Moss et al. (1973), Choy and Byer (1976), Kurtz et al. (1979), Luppi and Véniard (2016).

Such nonlinearities represent the ability of interaction of electromagnetic waves of different frequencies with each other. The influence of two electromagnetic waves (ω_1 and ω_2) acting simultaneously on a crystal with anharmonic oscillations can be expressed as

$$m_0 \left(\frac{\partial^2 u}{\partial t^2} - \gamma \frac{\partial u}{\partial t} + \omega_0^2 u + q_2 u^2 \right) = eE_x^{(1)} \exp(i\omega_1 t) + eE_x^{(2)} \exp(i\omega_2 t), \tag{110}$$

Table 5 Second-order nonlinear optical susceptibilities for some semiconductors at $\lambda = 1.06$ μm

Crystal	d_{36}	d_{15}	d_{31}	d_{33}
GaAs	210			
ZnSe	27			
ZnO		2.1	2.0	−6.2
CdSe		31	29	56
GaN		8.0	8.2	−16.5

where u is the displacement, m_0 the oscillating mass, and a polarization of both waves in x direction is assumed; ω_0 is the appropriate resonance frequency, and q_2 is assumed as the nonharmonicity factor. The displacement $u(\omega)$ now contains components of $\pm\,\omega_1 \pm \omega_2$, the sum and differences of the original frequencies:

$$u(\omega) = A(\pm\omega_1,\ \pm\omega_2)\exp[i(\pm\omega_1 \pm \omega_2)t]. \tag{111}$$

When using a specific combination of ω_1 and ω_2, e.g., $\omega_1 + \omega_2$ with $\omega_1 = \omega_2$, we obtain for the amplitude of the second harmonic

$$A(2\omega_1) = -q_2 \frac{e^2\left(E_x^{(1)}\right)^2}{2m_0^2} \frac{1}{\left[\omega_0^2 - (2\omega_1)^2 + 2i\gamma\omega_1\right]} \frac{1}{\left[\omega_0^2 - \omega_1^2 + i\gamma\omega_1\right]^2}. \tag{112}$$

From the polarization relation $P(2\omega_1) = eNu(2\omega_1)$, one obtains the appropriate equation for the susceptibility. In general, one has

$$\chi^{(2)}_{xxx}(-\omega_3 : \omega_1, \omega_2) = q_2 \frac{e^3 N}{\varepsilon_0 m_0} \frac{1}{D(\omega_3)} \frac{1}{D(\omega_1)} \frac{1}{D(\omega_2)} \tag{113}$$

with $D(\omega_i) = \omega_0^2 - \omega_i^2 + i\gamma\omega_i$. For higher-order susceptibilities $\chi^{(n)}$, the $D(\omega_i)$ term is repeated $n + 1$ times. In the extrinsic range between band edge and Reststrahl energy (Sect. 2.2 of ▶ Chap. 11, "Photon–Phonon Interaction"), one has with (ω, ω_1, ω_2) $\ll \omega_0$ a simple relationship for the polarization: $\chi^{(2)}_{xxx} = q_2\, e^3 N/\left(\varepsilon_0 m_0 \omega_0^6\right)$. This relation can be transformed by using Eq. 45 for (κ,$\gamma \to 0$) into

$$\chi^{(2)}_{xxx} = \frac{\left(n_r^2 - 1\right) e}{m_0 \omega_0^4} q_2. \tag{114}$$

In order to estimate the anharmonicity parameter q_2, we can use the displacement u which is obtained from

$$m_0\, \omega_0^2\, u = -m_0\, q_2\, u^2. \tag{115}$$

Here, restoring force and nonlinear force are made equal to each other – see Eq. 110. This yields $q_2 = \omega_0^2/u$ and, consequently,

$$\chi^{(2)}_{xxx} = \frac{\left(n_r^2 - 1\right) e}{m_0 \omega_0^2\, u} = 1.76 \cdot 10^{-9}\left(n_r^2 - 1\right)\left(\frac{10^{16}}{\omega_0}\right)^2 \left(\frac{0.1\,\text{nm}}{u}\right)\ (\text{cm/V}). \tag{116}$$

Typical displacements for observing unharmonicities are on the order of the covalent radii (l Å). It should be noted, however, that restoring forces $\propto u^2$ imply that crystals with inversion symmetry cannot produce second(even)-order

harmonics.[19] Still second-order harmonics can be produced in strained such crystals as discussed for Si by Prussel and Véniard (2018).

3.1.2 Nonlinear Polarization of Free Electrons

A free-carrier gas has a statistical center of symmetry, thereby excluding second-order effects in polarization. Free carriers contribute, however, to third-order effects in the susceptibility. These can be distinguished from other effects by their dependency on the carrier density.

Nonlinear polarization of free carriers is caused by nonparabolicity of conduction or valence bands and by the field dependence of some of the carrier scattering mechanisms (Rustagi 1970; Wang and Ressler 1969, 1970).

The third-order susceptibility due to free carriers (fc) can be given as

$$\chi_{\mathrm{fc}}^{(3)} = \chi_{\mathrm{np}}^{(3)} + \chi_{\mathrm{cs}}^{(3)} + \chi_{\mathrm{if}}^{(3)} \tag{117}$$

with the subscripts "np," "sc," and "if" standing for nonparabolicity of the respective band, carrier scattering, and interference between the two processes. In a simple (Kane) model, the nonparabolic contribution can be estimated (Wolf and Pearson 1966) as

$$\chi_{\mathrm{np}}^{(3)}(-\omega_3;\omega_1,\omega_2) = n\frac{e^4}{\varepsilon_0 m_n^2 E_{\mathrm{g}}\omega_1\,\omega_2\,\omega_3^2}\,\frac{1+\frac{8}{5}E_{\mathrm{F}}/E_{\mathrm{g}}}{\left(1+4E_{\mathrm{F}}/E_{\mathrm{g}}\right)^{5/2}} \tag{118}$$

where E_{F} is the Fermi energy. For $E_{\mathrm{F}} \ll E_{\mathrm{g}}$ and semiconductors of $m_n = 0.1\ m_0$ and $E_{\mathrm{g}} \cong 1$ eV, we have $\chi_{\mathrm{np}}^{(3)} \cong 10^{-33} n$ (cm^2/V^2), which, for moderate carrier densities, can approach the values of $\chi^{(3)}$ due to valence electrons.

3.1.3 The Different Mixing Effects

In semiconductors with no inversion symmetry, the second (or higher)-order susceptibility is large enough for relatively efficient (>10%) interaction of two or more photons (or phonons) with each other, often without invoking resonance transitions in the solid. These interactions can be used to generate light of a different wavelength; others can be used for light amplifiers or detectors of light beams. Such interactions consist of

1. Second-harmonic generation (SHG) by two identical photons $\propto\chi^{(2)}$ with $\omega_i = \omega_j$
2. Third-harmonic generation (THG) by three identical photons $\propto\chi^{(3)}$ with $\omega_j = \omega_k = \omega_l$
3. Sum and difference frequency mixing $\propto\chi^{(2)}$ with $\omega_i = \omega_j \pm \omega_k$

[19] When the restoring potential is nonlinear but symmetrical with respect to the equilibrium position $\mathbf{r}_0$, then the nonlinear oscillatory motion on both sides of $\mathbf{r}_0$ is equal, and the Fourier analysis of the ensuing oscillation does not contain even coefficients. Still, such a potential can produce odd harmonics (3rd, 5th, ...). For producing even-order harmonics, the nonlinear medium must not have inversion symmetry.

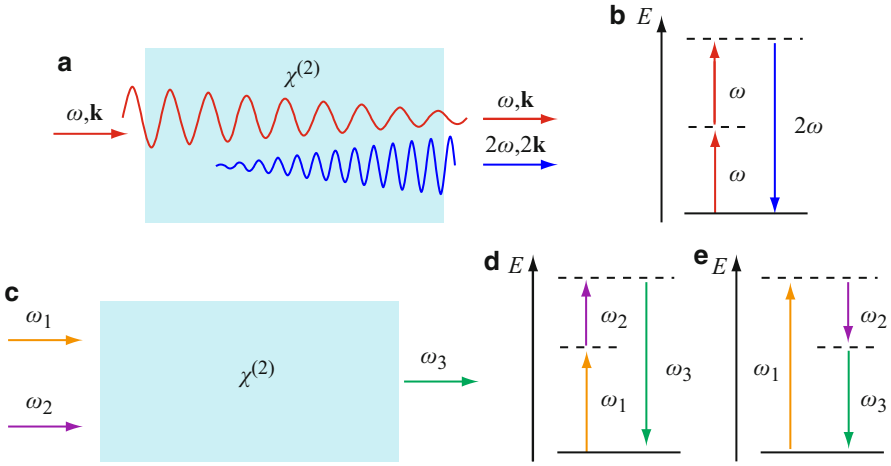


Fig. 21 Various nonlinear optical mixing effects. **(a)** and **(b)** second harmonic generation, **(c)** schematic set-up for sum-frequency generation $\omega_3 = \omega_1 + \omega_2$ shown in the energy diagram **(d)** and for difference-frequency generation $\omega_3 = \omega_1 - \omega_2$ shown in **(e)**

4. Parametric amplification, closely related to mixing[20] $\propto\chi^{(3)}$
5. Optical rectification $\propto\chi^{(2)}$ with $\omega_i = -\omega_j$
6. Self-focusing of light $\propto\chi^{(3)}$ with $\omega_j = -\omega_k = \omega_l$

Some of these multiphoton interactions can be used as analytical tools and are important for gaining new insight about the oscillatory behavior of solids, which is not otherwise obtainable. For example, multiphonon absorption can yield excited states which cannot be excited otherwise because of selection rules. A few examples are given below and illustrated in Fig. 21. For a review, see Byer and Herbst (1977), Warner (1975), or Mandel (2010).

Most of these experiments are based on the mixing of several photons or of photons and phonons in order to probe levels with different symmetries. As an example, we can compare a typical linear absorption experiment in which the parity of the excited state must be opposite to that of the ground state (since the dipole moment is odd and causes the final state to change its parity) with a nonlinear two-photon interaction in which the parity of the excited state remains the same since the dipole matrix element appears twice. Consequently, one can observe the same parity transitions compared to the ground state with nonlinear mixing experiments.

[20] Amplification is achieved when a small signal at ω_s is mixed with a strong laser pump beam at ω_p and results in the creation of an additional beam at ω_i, the *idler frequency,* according to $\omega_p = \omega_s + \omega_i$. In this process, energy from ω_p is pumped into ω_s and ω_i; consequently, ω_s is amplified.

There is a very wide variety of experiments in which nonlinear optical effects are used whenever two or more quasi-particles such as photons, phonons, polaritons, or electrons are involved. Several of these will be discussed in other sections of this book, e.g., multiphonon absorption (Sect. 3.2 of ▶ Chap. 11, "Photon–Phonon Interaction") and the Burstein-Moss shift (Sect. 3.2 of ▶ Chap. 17, "Optical Properties of Defects"). In each of these cases and sometimes implicit, a nonlinearity of the dielectric polarization is involved, i.e., ε and χ are a function of the incident field.

An important topic is the nonlinear *solid-state spectroscopy,* which is an extension of the well-known one-photon spectroscopy. It deals with all types of interactions accessible for single photons and probes the higher amplitude behavior; it also makes bulk regions of crystals accessible for excitation in ranges of very high optical absorption (see the following section) or permits stimulations of transitions which are forbidden for simple photon or phonon processes as discussed before.

An interesting range for nonlinear spectroscopy is that in which the sum or difference of energy of two photons equals a specific *resonance* transition. Here, high amplitudes can be reached with rather modest input signals (see below). For details on three-photon mixing, see Lee and Ahmad (1996).

3.1.4 Upconversion and Difference Mixing

Since the polarization is low when far from a resonance transition and higher-order susceptibilities rapidly decrease with increasing order, we need a high-intensity laser for most experiments to create signals above noise. Near resonances, however, the efficiency for higher-order interaction dramatically increases. The absorption of single photons in this range is usually too high to reach regions beyond the near-surface layer. These difficulties can be avoided by *upconversion,* i.e., by mixing two low-energy photons (Fig. 21d). These photons enter the crystal bulk without appreciable absorption, while the effect of an upconverted photon with energy near or at the electron resonance with large absorption can now be studied within the bulk. Such an upconversion is commonly used to reach exciton-polaritons (see Sect. 1.3 of ▶ Chap. 14, "Excitons") or characteristic points in band-to-band transitions.

Difference mixing (Fig. 21e) can bring the resulting photon into the phonon absorption range, i.e., the *Reststrahl* range (Sect. 2.2 of ▶ Chap. 11, "Photon–Phonon Interaction"), where the bulk photon-phonon or phonon-polariton interactions can be studied (see Kildal and Mikkelsen 1973 and Sect. 1.3 of ▶ Chap. 14, "Excitons").

The transfer of energy between the two beams of light for mixing depends on the coherence length of the interaction. The energy transfer follows the conservation-of-photon relation given by the Manley-Rowe equations (Manley and Rowe 1959):

$$\frac{1}{\omega_1}\frac{dI_1}{dz} = \frac{1}{\omega_2}\frac{dI_2}{dz} = \frac{1}{\omega_3}\frac{dI_3}{dz}. \tag{119}$$

In *sum frequency generation* ($\omega_3 = \omega_1 + \omega_2$), both lower energy laser beams lose power, the sum of which is gained by the higher-frequency beam. In *difference frequency generation* ($\omega_3 = \omega_1 - \omega_2$), the higher-frequency beam (ω_1) loses power while both lower-frequency beams (ω_2 and ω_3) gain the equivalent amount; such a scenario is shown in Fig. 21e; the corresponding setup has *one* beam ω_1 entering the crystal to produce *two* emerging beams ω_2 and ω_3. This effect can be used to amplify the amplitude of a low-intensity beam. The frequency ω_2 can be generated with high efficiency (approaching 50%) and can be amplified when a cavity (Sect. 2.2) is provided and tuned to ω_2. This constitutes a parametric oscillator (see Byer and Herbst 1977), Marty et al. (2021).

3.1.5 Phase Matching

In order for the different beams to remain in phase throughout the crystal, the index of refraction for the involved frequencies must remain the same. Even away from resonances, however, this is usually not the case. As a result, the beams with frequency ω_1 and ω_3 will *walk off* in slightly different directions, limiting the *coherence length* (the length along which two light beams can interact when they remain parallel to each other and in phase) to

$$l_{\text{coh}} = \frac{\lambda}{4(n_{\text{r}}(\omega_2) - n_{\text{r}}(\omega_3))}, \tag{120}$$

where λ is the fundamental wavelength for second-harmonic generation ($\omega_1 = \omega_2$ and $\omega_3 = 2\ \omega_2$ in Eq. 119, compare Fig. 21b). However, by carefully adjusting the incident angle θ with respect to the optical axis, one can match the refractive index for the fundamental beam to that of the extraordinary beam if birefringence is strong enough to compensate dispersion (Fig. 22): $n_{\text{ordinary}}(\omega) = n_{\text{extraordinary}}(2\omega)$. This *birefringence phase matching* utilizes the angle dependence of the extraordinary beam and has been analyzed by Midwinter and Warner (1965) for uniaxial crystals

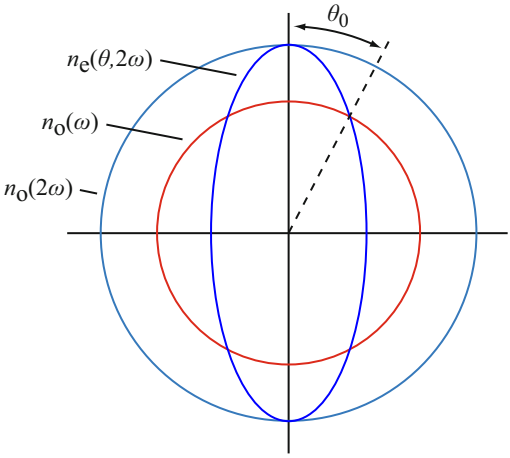


Fig. 22 Phase matching achieved by varying the angle of propagation θ in the crystal such that $n_o(\omega) = n_e(2\omega)$. At an angle θ_0 maximum second-harmonic output is obtained

and by Hobden (1967) for biaxial crystals. For perfect phase matching also momentum conservation $\mathbf{k}(2\omega) = 2\,\mathbf{k}(\omega)$ must apply.

Other methods of phase matching include matching in optically active media, Faraday rotation, and using anomalous dispersion; see the literature listed in the review by Chemla and Jerphagnon (1980). Phase matching is important for efficient nonlinear optical devices.

3.1.6 Mixing and dc Fields

Nonparabolic effects on the polarization can be tested by an external dc field that causes a sufficient relative shift of the bound valence electrons relative to the ion cores. This could be regarded as a prestressing of the lattice atoms in an external electric field and testing these prestressed atoms with the electromagnetic irradiation. In the presence of the dc field, the nonlinear susceptibilities increase. As a result, one observes an increase in the power of second-harmonic generation with applied external field:

$$P(2\omega) \propto E_{\mathrm{dc}}^2. \tag{121}$$

The inverse of this process – namely, the frequency difference generation of two identical photons –

$$\omega_1 - \omega_2 = 0, \tag{122}$$

is called *optical rectification* and generates a dc field (from an ac field) or, in an external circuit, a dc current. For optical rectification in zincblende crystals, see Rice et al. (1994).

3.1.7 Conversion Efficiencies

Depending on the material, rank of susceptibility, the frequencies' proximity to a resonance, and the acting field amplitude, the conversion efficiencies for nonlinear processes can vary from less than 1% to values in excess of 80%. The efficiency of a phase-matched harmonic generation was evaluated by Armstrong et al. (1962) and is given by

$$\eta(\omega \to 2\omega) = \frac{I(2\omega)}{I(\omega)} = \tanh^2(\varepsilon A_0\, l_{\mathrm{coh}}), \tag{123}$$

where A_0 is the amplitude of the incident laser and l_{coh} is the coherence length. Total power transfer into the upconverted beam is theoretically possible; experimentally, well above 50% efficiency is achieved.

For transitions close to resonances, giant oscillator strengths are sometimes observed, e.g., when creating excitonic molecules (Sect. 1.4 of ▶ Chap. 14, "Excitons"). In CuCl, up to fourth-order excitation and cascade recombination was observed for such multiphoton processes (Maruani et al. 1978).

3.2 Electro-Optical Effects

Electro-optical effects refer to a change in the dielectric function with an applied electric field. Such changes can be derived from the anharmonic oscillator model. When such a change is linear with the electric field, it is referred to as the *Pockels effect.* When it is quadratic, it is called the *Kerr effect.*

3.2.1 The Pockels Effect

When in Eq. 110 a constant electric field E_0 is added on top of the electromagnetic radiation, we have similar conditions as indicated earlier for nonharmonic lattice oscillations – see Sect. 1.2 of ▶ Chap. 4, "Elasticity and Phonons." Even for small-amplitude oscillations (i.e., low light intensities), the range of anharmonic oscillations can be reached by sufficient prestressing – here by a sufficient dc field. This lowers the symmetry of centrosymmetric crystals so that these also show second-harmonic generation (e.g., seen in Si). Such a prestressing causes a shift in the resonance frequency

$$\Delta\left(\omega_0^2\right) = \omega_0^2 - \left(\omega_0'\right)^2 = 2q_2 \frac{eE_0}{m\omega_0^2} \tag{124}$$

and results in a change in the index of refraction

$$\Delta n_\mathrm{r} = \frac{\left(n_\mathrm{r}^2 - 1\right)\Delta\left(\omega_0^2\right)}{2n_\mathrm{r}\left(\omega_0^2 - \omega^2\right)}. \tag{125}$$

This change in n_r is anisotropic because of the external dc field, which renders the optical properties of isotropic semiconductors anisotropic. Such an anisotropy is expressed by the index of refraction ellipsoid

$$\frac{x^2}{n_1^2} + \frac{y^2}{n_2^2} + \frac{z^2}{n_3^2} + \frac{2yz}{n_4^2} + \frac{2zx}{n_5^2} + \frac{2xy}{n_6^2} = 1 \tag{126}$$

with indices as identified in relation Eq. 108. Conventionally, the anisotropic index of refraction relates to the field as

$$\frac{1}{n_m^2} = \frac{1}{n_\mathrm{r}^2} + \sum_i r_{mi} E_i \tag{127}$$

with the electro-optical coefficient r_{mi}. We can show easily that in x direction, r_{xx} and χ_{xxx} are related as $n_m^4 r_{xx} = -2\chi_{xxx}$ (using Eqs. 114, 124, and 125) and in a general form (Franken and Ward 1963)

$$n_m^4 r_{mi} = -2d_{im}. \tag{128}$$

This anisotropy results in different light velocities dependent on the relative orientation of the dc field and light-beam polarization (with propagation perpendicular to the dc field). It causes a rotation of the plane of polarization of the light while traveling through the medium (Meintjes and Raab 1999). When rotation by π/2 is achieved, a light beam of linear polarization is fully switched on when exiting through a polarizing filter with orientation normal to the polarization of the incident beam. The field necessary to achieve this switching is

$$E_{\pi/2} = \frac{\lambda}{n_r^3 r_{41} L} \tag{129}$$

The field to accomplish this switching is on the order of 10^4 V/cm for $\lambda = 1$ μm, $n_r \cong 3$, $L = 3$ cm, and $r_{41} \cong 10^{-10}$ cm/V.

3.2.2 The Kerr Effect

The Kerr effect depends on introducing a third-order term in displacement $m_0 q_3 u^3$ into Eq. 110. Here, the index of refraction changes proportional to quadratic terms in the dc field. Such an effect is the only electro-optic nonlinear effect in centrosymmetric semiconductors. The corresponding change in the index of refraction is given by

$$\Delta n_r = -q_3 \frac{3e^2\left(n_r^2 - 1\right)}{2n_r m_0^2 \omega_0^4 \left(\omega_0^2 - \omega^2\right)} E_0^2. \tag{130}$$

obtained similarly as in Eq. 125. Centrosymmetric semiconductors in which the Kerr effect can be observed are the element semiconductors such as Si and Ge. More on the Kerr effect in seminconductors is reported in Sheik-Bahae et al. (1994). The Kerr effect is also well known in certain organic liquids such as nitrobenzene and CS_2. Its field-induced birefringence is often used for modulation or fast switching of a light beam.

4 Summary

The interaction of electromagnetic radiation with a semiconductor can be described by two parameters: the complex dielectric constant and conductivity. They can be derived from Maxwell's equations applied to the semiconductor as a continuum without considering its atomic structure. These parameters are closely associated with the index of refraction, the extinction (or optical absorption) coefficient, and the displacement current. These in turn can be related to directly measured quantities such as the transmitted and reflected light as a function of the wavelength, impinging angle, and polarization. The so-obtained optical constants are connected to the microscopic theory of optically induced elementary excitation processes and involve a wide variety of quasiparticles such as electrons, holes, excitons, phonons, or polaritons – particles which will be discussed in later chapters.

Light propagation and light-matter interaction can be controlled by an artificially introduced periodic modulation of the dielectric constant on a scale of the wavelength. Photonic band structures are thereby created similar to the band structure for electrons originating from the periodic lattice potential. Deviations from periodicity create optical defects, which localize optical modes; line defects allow for guiding optical waves; point defects form microcavities like omnidirectional mirrors allowing to enhance or suppress radiation of an enclosed emitter.

Nonlinear optical effects originate from anharmonic terms in the potential, which couple electromagnetically induced oscillations. The effects emerge at high field amplitudes or by prestressing with an applied dc field and are described by higher-order terms in a field-dependent dielectric function (or polarization). Nonlinear effects lead to a mixing of two or more excitation processes (of, e.g., photons, phonons, polarons) and result in new states with the sum or difference of incident energies. They can involve transitions far from resonance (virtual states) or near resonance, the latter with much increased conversion efficiencies.

References

Adachi S (2012) The handbook on optical constants of semiconductors. World Scientific

Armstrong JA, Bloembergen N, Ducuing J, Pershan PS (1962) Interactions between light waves in a nonlinear dielectric. Phys Rev 127:1918

Aspnes DE (1976) Spectroscopic ellipsometry of solids. In: Seraphin BO (ed) Optical properties of solids: new developments. North-Holland, Amsterdam, p 799

Aspnes DE (1980) Modulation spectroscopy/electric field effects on the dielectric function of semiconductors. In: Moss TS, Balkanski M (eds) Optical properties of solids, vol 2, North-Holland, Amsterdam, p 109

Aspnes DE, Studna AA (1983) Dielectric functions and optical parameters of Si, Ge, GaP, GaAs, GaSb, InP, InAs, and InSb from 1.5 to 6.0 eV. Phys Rev B 27:985

Attaccalite C, Grüning M (2013) Nonlinear optics from an ab initio approach by means of the dynamical Berry phase: application to second- and third-harmonic generation in semiconductors. Phys Rev B 88:235113

Azzam RMA, Bashara NM (1987) Ellipsometry and polarized light. Elsevier Science BV, Amsterdam

Baba T (2008) Slow light in photonic crystals. Nat Photon 2:465

Balkanski M, Moss TS (1994) Optical properties of semiconductors. North-Holland, Amsterdam

Basu PK (1998) Theory of optical processes in semiconductors. Clarendon Press, Oxford

Bayer M, Reinecke TL, Weidner F, Larionov A, McDonald A, Forchel A (2001) Inhibition and enhancement of the spontaneous emission of quantum dots in structured microresonators. Phys Rev Lett 86:3168

Birner A, Busch K, Müller F (1999) Photonische Kristalle. Phys Bl 55(4):27 (Photonic crystals, in German)

Blanco A, Chomski E, Grabtchak S, Ibisate M, John S, Leonard SW, Lopez C, Meseguer F, Miguez H, Mondia JP, Ozin GA, Toader O, van Driel HM (2000) Large-scale synthesis of a silicon photonic crystal with a complete three-dimensional bandgap near 1.5 micrometres. Nature 405:437

Bloembergen N (1982) Nonlinear optics and spectroscopy. Rev Mod Phys 54:685

Born M, Wolf E (2002) Principles of optics, 7th edn. Cambridge University Press, Cambridge

Boyd RW (2008) Nonlinear optics, 3rd edn. Academic Press, London

Bruggemann DAG (1935) Berechnung verschiedener physikalischer Konstanten von heterogenen Substanzen. Ann Phys (Leipzig) 24:636 (Calculation of various physical constants of heterogeneous substances, in German)
Brust D (1972) Model calculation of the q-dependent dielectric function of some zinc-blende semiconductors. Phys Rev B 5:435
Burge DK, Bennett HE (1964) Effect of a thin surface film on the ellipsometric determination of optical constants. J Opt Soc Am 54:1428
Busch K, Wehrspohn R (2003) Special issue on Photonic Crystals: optical materials for the 21th century. Phys Stat Sol A 197(3):593–702
Busch K, von Freymann G, Linden S, Mingaleev SF, Tkeshelashvili L, Wegener M (2007) Periodic nanostructures for photonics. Phys Rep 444:101
Byer RL, Herbst RL (1977) Parametric oscillation and mixing. Top Appl Phys 16:81
Cardona M (1969) Modulation spectroscopy. In: Seitz F, Turnbull D, Ehrenreich H (eds) Solid state physics, suppl 11. Academic Press, New York
Chemla DS, Jerphagnon J (1980) Nonlinear optical properties. In: Moss TS, Balkanski M (eds) Handbook on semiconductors, vol 2. North-Holland, Amsterdam, pp 545–598
Choy MM, Byer RL (1976) Accurate second-order susceptibility measurements of visible and infrared nonlinear crystals. Phys Rev B 14:1693
Clark SJ, Robertson J (2010) Screened exchange density functional applied to solids. Phys Rev B 82:085208
Cohen MH (1963) Generalized self-consistent field theory and the dielectric formulation of the many-body problem. Phys Rev 130:1301
Cui TJ, Smith DR, Liu R (eds) (2010) Metamaterials: theory, design, and applications. Springer, New York
Devashish D, Ojambati OS, Hasan SB, van der Vegt JJW, Vos WL (2019) Three-dimensional photonic band gap cavity with finite support: enhanced energy density and optical absorption. Phys Rev B 99:075112
Ejder E (1971) Refractive index of GaN. Phys Stat Sol A 6:445
Engheta N, Ziolkowski RW (eds) (2006) Metamaterials: physics and engineering explorations. Wiley, Hoboken
Franken PA, Ward JF (1963) Optical harmonics and nonlinear phenomena. Rev Mod Phys 35:23
Gérard JM, Sermage B, Gayral B, Legrand B, Costard E, Thierry-Mieg V (1998) Enhanced spontaneous emission by quantum boxes in a monolithic optical microcavity. Phys Rev Lett 81:1110
Goñi AR, Syassen K, Strossner K, Cardona M (1989) Pressure dependence of the direct optical gap and refractive index of Ge and GaAs. Semicond Sci Technol 4:246
Grüning M, Attaccalite C (2016) Performance of polarisation functionals for linear and nonlinear optical properties of bulk zinc chalcogenides ZnX (X = S, Se, and Te). Phys Chem Chem Phys 18:21179
Haken H (1963) Theory of excitons II. In: Kuper CG, Whitfield GD (eds) Polarons and excitons. Oliver & Boyd, Edinburgh, p 294
Ho K-M, Chan CT, Soukoulis CM, Biswas R, Sigalas M (1994) Photonic band gaps in three dimensions: new layer-by-layer periodic structures. Sol State Commun 89:413
Hobden MV (1967) Phase-matched second-harmonic generation in biaxial crystals. J Appl Phys 38:4365
Hu DJJ, Xu Z, Shum PP (2019) Review on photonic crystal fibers with hybrid guiding mechanisms. IEEE Access 7: 67469
Inoue K, Ohtaka K (eds) (2004) Photonic crystals. Springer, Berlin
Joannopoulos JD, Johnson SG, Winn JN, Meade RD (2008) Photonic crystals – molding the flow of light, 2nd edn. Princeton University Press, Princeton
John S (1987) Strong localization of photons in certain disordered dielectric superlattices. Phys Rev Lett 58:2486

Kildal H, Iseler GW (1976) Laser-induced surface damage of infrared nonlinear materials. Appl Optics 15:3062
Kildal H, Mikkelsen JC (1973) The nonlinear optical coefficient, phasematching, and optical damage in the chalcopyrite $AgGaSe_2$. Opt Commun 9:315
Kirchner A, Busch K, Soukoulis CM (1998) Transport properties of random arrays of dielectric cylinders. Phys Rev B 57:277
Kockum AF, Miranowicz A, De Liberato S, Savasta S, Nori F (2019) Ultrastrong coupling between light and matter. Nat Rev Phys 1:19
Kurtz SK, Jerphagnon J, Choi MM (1979) Nonlinear dielectric susceptibilities. In: Hellwege KH, Hellwege AM (eds) Landolt-Börnstein new series III/11. Springer, Berlin, p 671
Lan Y-Z (2016) New understandings of third harmonic generation coefficient of bulk silicon: implementation and application of the full ab initio band structure sum-over-states method. Comput Cond Matter 8:22
Lee KF, Ahmad HB (1996) Fractional photon model of three-photon mixing in the non-linear interaction process. Opt Laser Technol 28:35
Lonergan A, Hu C, O'Dwyer C (2020) Filling in the gaps: the nature of light transmission through solvent-filled inverse opal photonic crystals. Phys Rev Mater 4:065201
Luppi E, Véniard V (2016) A review of recent theoretical studies in nonlinear crystals: towards the design of new materials. Semicond Sci Technol 31:123002
Maksimovic M, Hammer M, van Groesen E (2008) Field representation for optical defect resonances in multilayer microcavities using quasi-normal modes. Opt Commun 281:1401
Mandel P (2010) Nonlinear optics. Wiley-VCH, Weinheim
Manley JM, Rowe HE (1959) General energy relations in nonlinear reactances. Proc IRE 47:2115
Marple DTF (1964) Refractive index of GaAs. J Appl Phys 35:1241
Marty G, Combrié S, Raineri F, De Rossi A (2021) Photonic crystal optical parametric oscillator. Nat Photon 15:53
Maruani A, Oudar JL, Batifol E, Chemla DS (1978) Nonlinear spectroscopy of biexcitons in CuCl by resonant coherent scattering. Phys Rev Lett 41:1372
Meintjes EM, Raab RE (1999) A new theory of Pockels birefringence in non-magnetic crystals. J Opt A 1:146
Midwinter JE, Warner J (1965) The effects of phase matching method and of uniaxial crystal symmetry on the polar distribution of second-order non-linear optical polarization. Brit J Appl Phys 16:1135
Moss TS, Burrell GJ, Ellis B (1973) Semiconductor Optoelectronics. Wiley, New York
Nara H (1965) Screened impurity potential in Si. J Phys Soc Jpn 20:778
Nara H, Morita A (1966) Shallow donor potential in silicon. J Phys Soc Jpn 21:1852
Nazarewicz P, Rolland P, da Silva E, Balkanski M (1962) Abac chart for fast calculation of the absorption and reflection coefficients. Appl Optics 1:369
Noda S, Tomoda K, Yamamoto N, Chutinan A (2000) Full three-dimensional photonic bandgap crystals at near-infrared wavelengths. Science 289:604
Palik ED (ed) (1985) Handbook of optical constants of solids I. Academic Press, New York
Palik ED (ed) (1991) Handbook of optical constants of solids II. Academic Press, San Diego
Penn DR (1962) Wave-number-dependent dielectric function of semiconductors. Phys Rev 128:2093
Philipp HR, Taft EA (1960) Optical constants of silicon in the region 1 to 10 eV. Phys Rev 120:37
Pollak FH, Shen H (1993) Modulation spectroscopy of semiconductors: bulk/thin film, microstructures, surfaces/interfaces and devices. Mater Sci Eng R10:275
Purcell EM, Torrey HC, Pound RV (1946) Resonance absorption by nuclear magnetic moments in a solid. Phys Rev 69:37
Prussel L, Véniard V (2018) Linear electro-optic effect in semiconductors: ab initio description of the electronic contribution. Phys Rev B 97:205201

Rabi II, Millman S, Kusch P, Zacharias JR (1939) The molecular beam resonance method for measuring nuclear magnetic moments; the magnetic moments of $_3Li^6$, $_3Li^7$ and $_9 F^{19}$. Phys Rev 55:526

Rabin H, Tang CL (eds) (1975) Quantum electronics: a treatise, vol 1, Nonlinear optics. Academic Press, New York

Reithmaier JP, Sek G, Löffler A, Hofmann C, Kuhn S, Reitzenstein S, Keldysh LV, Kulakovskii VD, Reinecke TL, Forchel A (2004) Strong coupling in a single quantum dot-semiconductor microcavity system. Nature 432:197

Rice A, Jin Y, Ma XF, Zhang X-C, Bliss D, Larkin J, Alexander M (1994) Terahertz optical rectification from <110> zinc-blende crystals. Appl Phys Lett 64:1324

Rinne SA, García-Santamaría F, Braun PV (2008) Embedded cavities and waveguides in three-dimensional silicon photonic crystals. Nat Photon 2:52

Russell PSJ (1986) Optics of Floquet-Bloch waves in dielectric gratings. Appl Phys B 39(4):231

Rustagi KC (1970) Effect of carrier scattering on nonlinear optical susceptibility due to mobile carriers in InSb, InAs, and GaAs. Phys Rev B 2:4053

Sakoda K (2001) Optical properties of photonic crystals. Springer, Berlin

Samara GA (1987) Temperature and pressure dependences of the dielectric constants of semiconductors. Phys Rev B 27:3494

Seraphin BO (1973) Proceedings of the first international conference on modulation spectroscopy. Surf Sci 37:1–1011

Seraphin BO, Bottka N (1965) Franz-Keldysh effect of the refractive index in semiconductors. Phys Rev 139:A560

Sharma AC, Auluck S (1981) Transverse dielectric function for a model semiconductor. Phys Rev B 24:4729

Sheik-Bahae M, Wang J, Van Stryland EW (1994) Nondegenerate optical Kerr effect in semiconductors. IEEE J Quantum Electron 30:249

Shelby RA, Smith DR, Schultz S (2001) Experimental verification of a negative index of refraction. Science 292:77

Soukoulis CM, Wegener M (2010) Optical metamaterials – more bulky and less lossy. Science 330:1633

Stegeman GI, Stegeman RA (2012) Nonlinear optics: phenomena, materials and devices. Wiley, Hoboken

Stern F (1963) Elementary theory of the optical properties of solids. In: Seitz F, Turnbull D (eds) Sol state physics, vol 15. Academic Press, New York, p 299

Tosatti E, Pastori Parravicini G (1971) Model semiconductor dielectric function. J Phys Chem Sol 32:623

van Vechten JA (1980) A simple man's view of the thermochemistry of semiconductors. In: Moss TS, Keller SP (eds) Handbook on semiconductors, vol 3. North-Holland, Amsterdam, pp 1–111

Veselago VG (1968) The electrodynamics of substances with simultaneously negative values of ε and μ. Sov PhysUspekhi 10:509

Villeneuve PR, Piché M (1994) Photonic bandgaps in periodic dielectric structures. Progr Quantum Electr 18:153

von Freymann G, Ledermann A, Thiel M, Staude I, Essig S, Busch K, Wegener M (2010) Three-dimensional nanostructures for photonics. Adv Funct Mater 20:1038

Walter JP, Cohen ML (1970) Wave-vector-dependent dielectric function for Si, Ge, GaAs, and ZnSe. Phys Rev B 2:1821

Wang CC, Ressler NW (1969) Nonlinear optical effects of conduction electrons in semiconductors. Phys Rev 188:1291

Wang CC, Ressler NW (1970) Observation of optical mixing due to conduction electrons in *n*-type germanium. Phys Rev B 2:1827

Wang R, Ruden PP, Kolnik J, Oguzman I, Brennan KF (1997) Dielectric properties of wurtzite and zincblende structure gallium nitride. J Phys Chem Sol 58:913

Ward L (1994) The optical constants of bulk materials and films, 2nd edn. Institute of Physics Publishing, Bristol
Warner J (1975) Quantum electronics: a treatise, vol 1, Nonlinear optics. Academic Press, New York, p 103
Wolf PA, Pearson GA (1966) Theory of optical mixing by mobile carriers in semiconductors. Phys Rev Lett 17:1015
Yablonovitch E (1987) Inhibited spontaneous emission in solid-state physics and electronics. Phys Rev Lett 58:2059
Yablonovitch E, Gmitter TJ, Leung KM (1991) Photonic band structure: the face-centered-cubic case employing nonspherical atoms. Phys Rev Lett 67:2295
Zernike F, Midwinter JE (1973) Applied non-linear optics. Wiley, New York
Zouhdi S, Sihvola A, Vinogradov AP (2009) Metamaterials and plasmonics: fundamentals, modelling, applications. Springer, New York

Photon–Phonon Interaction

11

Karl W. Böer and Udo W. Pohl

Contents

1 Lattice Polarization 430
1.1 Electric Fields and Polarizability 430
1.2 Dielectric Response and Kramers-Kronig Relations 433
2 The Dielectric Function in the IR Range 438
2.1 Elementary Oscillators 438
2.2 IR Reflection and Reststrahlen 444
3 Scattering of Photons with Phonons 445
3.1 The Phonon-Polariton 445
3.2 One- and Multiphonon Absorption 448
3.3 Brillouin and Raman Scattering 449
4 Summary 463
References 464

Abstract

The interaction of photons with solids comprises ionic and electronic oscillations; this chapter focuses on lattice vibrations. The dielectric polarization is related to the atomic polarizability. The dynamic response of the dielectric function on electromagnetic radiation can be described classically by elementary oscillators, yielding strong interaction of photons and TO phonons with a resulting large Reststrahl absorption in the IR range. The dispersion is described by a phonon-polariton, which is observed in inelastic scattering processes. Brillouin scattering at acoustic phonons and Raman scattering at optical phonons provide direct information about the spectrum and symmetry of vibrations in a semiconductor.

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_11

Keywords

Brillouin scattering · Dielectric Function · Elastic and Inelastic scattering · Kramers-Kronig Relations · Lattice Polarization · Phonon dispersion · Phonon-Polariton · Photon-Phonon Interaction · Raman Scattering · Reststrahl Absorption

The continuum model presented in ▶ Chap. 10, "Interaction of Light with Solids," will now be refined by introducing the atomic microstructure of the semiconductor. The interplay of the electromagnetic radiation with the solid can be described classically by interaction with atomic oscillations. In this chapter, we will deal with the interaction of photons with lattice oscillation, i.e., with the *vibronic* part; a description of the relation between electric fields and the lattice polarization provides the foundation for the understanding of this interaction. Respective interactions of photons with *electrons* are discussed in ▶ Chaps. 12, "Photon–Free-Electron Interaction," and ▶ 13, "Band-to-Band Transitions."

In the following we discuss how an external field is related to the polarization of the lattice and introduce the time dependence of the dielectric response. Then we focus on the coupling of electromagnetic radiation to lattice oscillations in ionic solids.

1 Lattice Polarization

1.1 Electric Fields and Polarizability

The dielectric polarization enhances a local electric field $\mathbf{E}_{\text{loc}}$ in the presence of an external field $\mathbf{E}_{\text{ext}}$. Neglecting space-charge effects the local electric field $\mathbf{E}_{\text{loc}}$ can be expressed as

$$\mathbf{E}_{\text{loc}} = \mathbf{E}_{\text{ext}} + \mathbf{E}_{\text{surf}}, \tag{1}$$

where $\mathbf{E}_{\text{surf}}$ is the field at the center of a cavity produced by the induced charges at the sample surface. This field can be evaluated in an isotropic crystal by a spherical cavity, yielding

$$\mathbf{E}_{\text{surf}} = \frac{1}{3\varepsilon_0}\mathbf{P}. \tag{2}$$

For anisotropic crystals the derivation of the local field is more complex (Jackson 1999). From the definition $\mathbf{P} = \varepsilon_0(\varepsilon - 1)\,\mathbf{E}$ (see Eq. 31 in ▶ Chap. 10, "Interaction of Light with Solids" and Eqs. 1 and 2) we obtain the relation between the local and externally applied fields:

$$\mathbf{E}_{\text{loc}} = \frac{\varepsilon + 2}{3}\mathbf{E}_{\text{ext}}. \tag{3}$$

When using the fact that the total polarization per unit volume is composed of N_j atoms with individual polarizability α_j, we obtain

$$\mathbf{P} = \varepsilon_0 \mathbf{E}_{\text{loc}} \sum_j N_j \alpha_j = \varepsilon_0 \frac{\varepsilon + 2}{3} \mathbf{E}_{\text{ext}} \sum_j N_j \alpha_j. \tag{4}$$

Since $\mathbf{P} = \varepsilon_0(\varepsilon - 1)\,\mathbf{E}$, we obtain the *Clausius-Mossotti relation*

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{1}{3} \sum_j N_j \alpha_j = \frac{1}{3V} \sum_j \alpha_j, \tag{5}$$

which relates the dielectric constant with the atomic polarizabilities.

The polarizability in ionic crystals is additive. For instance, in a binary compound *AB*, we have

$$\alpha_{AB} = \alpha_A + \alpha_B. \tag{6}$$

When one atomic polarizability is known (e.g., for Li obtained by Pauling 1927), one can derive all other ionic polarizabilities from Eqs. 5 and 6. For a review see Shanker et al. (1986).

1.1.1 Ionic and Electronic Polarizability

In (at least partially) ionic crystals, the atomic polarizability α_j can be divided into a ionic part α_i, determined by the relative shift of oppositely charged ions, and an electronic part α_e, determined by the relative shift of the electrons. The static dielectric constant $\varepsilon_{\text{stat}}$ contains both parts and can be obtained from Eq. 5:

$$\frac{\varepsilon_{\text{stat}} - 1}{\varepsilon_{\text{stat}} + 2} = \frac{1}{3}(N_\text{i}\,\alpha_\text{i} + N_\text{e}\,\alpha_\text{e}). \tag{7}$$

At high frequencies only the electrons can follow a changing external field; we therefore obtain another dielectric constant referred to as the *optical dielectric constant* ε_{opt}, yielding

$$\frac{\varepsilon_{\text{opt}} - 1}{\varepsilon_{\text{opt}} + 2} = \frac{1}{3} N_\text{e}\,\alpha_\text{e}; \tag{8}$$

consequently, we relate the *ionic polarizability* to the difference of Eqs. 7 and 8:

$$\alpha_\text{i} = \frac{3}{N_\text{i}} \left(\frac{\varepsilon_{\text{stat}} - 1}{\varepsilon_{\text{stat}} + 2} - \frac{\varepsilon_{\text{opt}} - 1}{\varepsilon_{\text{opt}} + 2} \right), \tag{9}$$

a quantity which is on the order of $1/N_\text{i} \cong 10^{-24}$ cm^3.

Some ionic polarizabilities are given in Table 1. It shows an increase of these polarizabilities with increasing number of electrons within each column of elements (see also Fig. 1). For the same electronic shell (e.g., F^- and Na^+, or Cl^-, K^+, and Ca^{2+}) it also increases with increasing nuclear charge. The additivity rule for polarizabilities applied for calculating the listed values has been found to be satisfied surprisingly well even for the predominantly covalent III-V crystals.

Table 1 Crystalline state polarizability of ions in units of 10^{-24} cm^3 (After Shanker et al. (1986), and Pandey et al. (1977))

Ion	α_i	Ion	α_i
Li^+	0.029	F^-	0.867
Na^+	0.314	Cl^-	3.063
K^+	1.136	Br^-	4.276
Rb^+	1.785	I^-	6.517
Cs^+	3.015		
Mg^{2+}	0.094	O^{2-}	1.657
Ca^{2+}	1.157	S^{2-}	4.497
Sr^{2+}	1.795	Se^{2-}	5.686
Ba^{2+}	3.188	Te^{2-}	9.375
Al^{3+}	3.831	P^{3-}	2.659
Ga^{3+}	4.435	As^{3-}	3.786
In^{3+}	6.489	Sb^{3-}	6.737

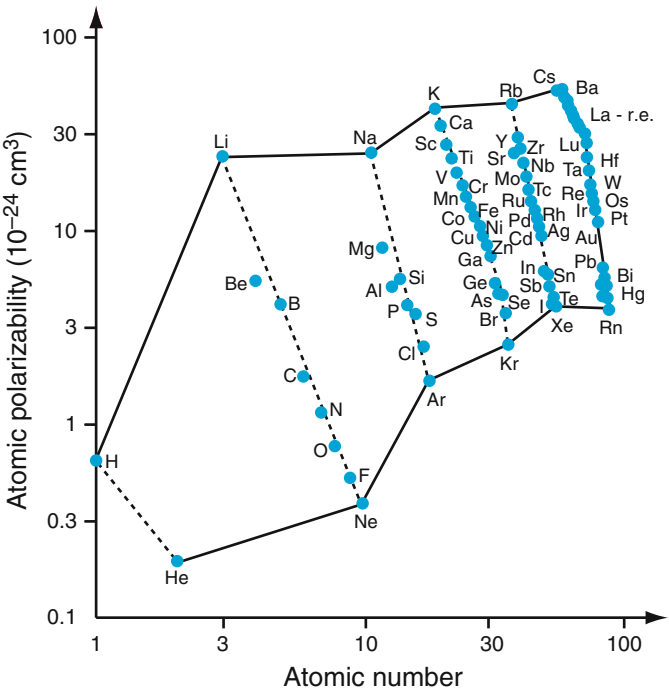


Fig. 1 Atomic polarizabilities as a function of the atomic number (After Jonscher 1983)

1.1.2 Piezoelectricity and Electrostriction

Crystals lacking inversion symmetry become electrically polarized when they are elastically strained (*piezoelectricity*), or they change their lattice constant when exposed to an external electric field (*electrostriction*). This effect is used in electromechanical transducers. Examples for such materials are SiO_2 (α-quartz), ADP ($NH_4H_2PO_4$), and KDP (KH_2PO_4). A much smaller quadratic

(in **E**) effect is observed in other ionic crystals due to the nonharmonicity of the lattice forces.

If the centers of positive and negative charges in a piezoelectric crystal do not coincide under zero external field, the crystal may show a spontaneous electric polarization and is called *ferroelectric*. The spontaneous polarization may be compensated by charges from the atmosphere absorbed at the surface. In some crystals such a compensation may be relieved after heating, when spontaneous polarization again becomes observable. These crystals, of which Turmalin is an example, are called *pyroelectric*. In other materials (e.g., organic waxes), such a net dipole moment may be frozen-in during solidification. These are called *electretes* (Gutman 1948).

Such a state of ordered molecular dipoles is stabilized by its lower total energy. The order is destroyed at temperatures exceeding the critical temperature T_C (*Curie temperature*). Examples of ferroelectric crystals with Curie temperatures (K) and saturation polarization (μCoul/cm^2) – separated by a comma in parentheses – are potassium dihydrogen phosphate KDP (213, 9), barium titanate (393, 26), and potassium niobate (712, 30). For more information see Lines and Glass (1979).

1.2 Dielectric Response and Kramers-Kronig Relations

The previous discussion of the dielectric polarization indicates a time-delayed response of the semiconductor to changes in the electric field. Such a delay is caused by the time it takes for the different dielectric displacements to adjust; inertia requires that any response cannot be instantaneous. Phenomenologically, we can express the total response as

$$\mathbf{D}(t) = \varepsilon_0 \mathbf{E}(t) + \mathbf{P}(t), \tag{10}$$

with $\mathbf{D}(t)$ the dielectric displacement. The relation between the changing electric field $\mathbf{E}(t)$ causing a change in polarization $\mathbf{P}(t)$ can be related by a dielectric response function $f(t)$. The principle of causality requires the output $\mathbf{P}(t)$ is zero for $t < 0$ if the system is excited by an electric field $\mathbf{E}(t)$ turned on at $t = 0$. Thus, $\mathbf{P}(t)$ can be written as

$$\mathbf{P}(t) = \int_0^\infty f(\tau)\, \mathbf{E}(t-\tau)\ d\tau, \tag{11}$$

with 0 as the lower integration limit. With respect to optical phenomena, it is more appropriate to transform Eq. 11 from the time into the frequency domain by means of a Fourier transformation in order to introduce the susceptibility:

$$P(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} P(t) \exp\big(-i\omega t\big)\, dt = \widetilde{\chi}(\omega) E(\omega). \tag{12}$$

The frequency-dependent polarization $P(\omega)$ is determined by the frequency-dependent susceptibility $\widetilde{\chi}(\omega)$ which can be expressed in terms of the response function

$$\widetilde{\chi}(\omega) = \chi'(\omega) + \chi''(\omega) = \int_0^{\infty} f(t) \exp\big(-i\omega t\big)\, dt \tag{13}$$

with

$$\begin{aligned} \chi'(\omega) &= \int_0^{\infty} f(t)\, \cos\big(\omega t\big)\, dt \;\text{ and }\; \chi'(\omega) = \chi'(-\omega) \quad \text{(even)} \\ \chi''(\omega) &= \int_0^{\infty} f(t)\, \sin\big(\omega t\big)\, dt \;\text{ and }\; \chi''(\omega) = -\chi''(-\omega) \quad \text{(odd)} \end{aligned} \tag{14}$$

or similar relations between $\widetilde{\varepsilon}(\omega)$ and $f(t)$. It is the purpose of a microscopic theory to provide such a response function as will be shown for a simple example in Sect. 2.1.

1.2.1 Kramers-Kronig Relations

As shown above, $\chi'(\omega)$ and $\chi''(\omega)$, or $\varepsilon'(\omega)$ and $\varepsilon''(\omega)$ are related to each other. With the help of the Hilbert transformation

$$P\int_{-\infty}^{\infty} \frac{\sin(\omega t)}{\omega - \omega_a} d\omega = \pi \cos(\omega_a t) \tag{15}$$

such a relation can be further developed, as shown below. P is the *Cauchy principal value* of the integral: in order to avoid the singularity at $\omega = \omega_a$, $P\int_{-\infty}^{\infty} \ldots d\omega$ is split according to $\lim_{\delta \to 0} \int_{-\infty}^{\omega_a - \delta} \ldots d\omega \; + \int_{\omega_a + \delta}^{\infty} \ldots d\omega$.

We can eliminate the unknown dielectric response function in Eq. 13 by inserting Eq. 15 into the first equation of Eq. 14:

$$\begin{aligned} \chi'(\omega_a) &= \int_0^{\infty} f(t) \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\sin(\omega t)}{\omega - \omega_a} d\omega\, dt \\ &= \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{d\omega}{\omega - \omega_a} \int_{-\infty}^{\infty} f(t) \sin(\omega t)\, dt; \end{aligned} \tag{16}$$

the extension of the lower integration limit over ω from 0 to $-\infty$ is permitted for causality reasons: its contribution (for $t < 0$) must be equal to zero. When inserting $\chi''(\omega_a)$ for the second integral in Eq. 16 we obtain

$$\chi'(\omega_a) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\chi''(\omega)}{\omega - \omega_a} d\omega, \tag{17}$$

and in a similar fashion

$$\chi''(\omega_a) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\chi'(\omega)}{\omega - \omega_a} d\omega. \tag{18}$$

These equations can easily be transformed[1] into the more familiar forms of the Kramers-Kronig integrals

$$\chi'(\omega_a) = \frac{2}{\pi} P \int_0^{\infty} \frac{\omega \chi''(\omega)}{\omega^2 - \omega_a^2} d\omega \quad \text{and} \quad \chi''(\omega_a) = \frac{2\omega_a}{\pi} P \int_0^{\infty} \frac{\chi'(\omega)}{\omega^2 - \omega_a^2} d\omega. \tag{19}$$

When using $\widetilde{\varepsilon}(\omega)$, we obtain the corresponding *Kramers-Kronig relations* (Kramers (1929), Kronig (1926); see, e.g., also Cardona 1969; Yu and Cardona 1999):

$$\varepsilon'(\omega_a) = 1 + \frac{2}{\pi} P \int_0^{\infty} \frac{\omega \varepsilon''(\omega)}{\omega^2 - \omega_a^2} d\omega \tag{20}$$

and

$$\varepsilon''(\omega_a) = \frac{-2\omega_a}{\pi} P \int_0^{\infty} \frac{\varepsilon'(\omega)}{\omega^2 - \omega_a^2} d\omega, \tag{21}$$

where ω_a is the arbitrary frequency at which ε' and ε'' are evaluated. Equations 20 and 21 show that the values of ε' and ε'' at a given frequency ω_a depend on the behavior of ε'' and ε', respectively, in the *entire frequency range*; thus the extinction coefficient and the dielectric constant depend on each other. This interdependence is amplified[2] when the contributing transition at ω is a resonance transition and lies close to the evaluated transition ω_a.

The dispersion relation given here relates the dispersion process to absorption processes in a single integral formula. For instance, it permits us to determine the dispersion at any frequency if we know the absorption in the entire frequency range. However, frequency ranges far away from the range to be evaluated, i.e., far away from ω_a, have little influence. It also makes immediately clear that there is no dispersion ($\varepsilon' = 1$) if there is no absorption ($\varepsilon'' = 0$). The shape of $\varepsilon'(\omega)$ depends in a sensitive way on the shape of $\varepsilon''(\omega)$: for instance, an edge in the absorption

[1] This follows from $P \int_{-\infty}^{\infty} \frac{f(x)}{x - a} dx = P \int_0^{\infty} \frac{x(f(x) - f(-x)) + a(f(x) + f(-x))}{x^2 - a^2} \, dx.$

[2] Such an interdependence can be visualized by considering a row of coupled pendula and forcing one of them to oscillate according to a given driving force. All other pendula will influence the motion, the more so, the closer the forced oscillation is to the resonance frequency of the others.

corresponds to a maximum in the dispersion, while an isolated maximum in absorption corresponds to a decline of ε' with ω.

Kramers-Kronig relations also exist between other related optical parameters. The real and imaginary parts of the complex refractive index $\tilde{n}$ are connected by

$$n_r(\omega_a) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{\omega \kappa(\omega)}{\omega^2 - \omega_a^2} d\omega \tag{22}$$

and

$$\kappa(\omega_a) = \frac{-2\omega_a}{\pi} P \int_0^\infty \frac{n_r(\omega) - 1}{\omega^2 - \omega_a^2} d\omega. \tag{23}$$

Another Kramers-Kronig relation relates amplitude and phase of reflection. The complex amplitude of reflection at a single surface can be written as

$$\tilde{r} = \rho \exp(i\Delta)$$

where ρ is the modulus of the reflection amplitude at normal incidence, which is connected to the measured reflectance $R(\omega)$ by $R = |\rho|^2$; Δ is the phase difference between the incident and reflected wave. The phase difference is related to $R(\omega)$ by

$$\Delta(\omega_a) = -\frac{\omega_a}{\pi} P \int_0^\infty \frac{\log(R(\omega))}{\omega^2 - \omega_a^2} d\omega. \tag{24}$$

Since

$$\rho \exp(i\Delta) = \rho \cos\Delta + i\rho \sin\Delta = \frac{n_r + i\kappa - 1}{n_r + i\kappa + 1} \tag{25}$$

(see Sect. 1.2 of ▶ Chap. 10, "Interaction of Light with Solids") we can relate the reflectance R to n_r and κ by equating real and imaginary parts of Eq. 25, yielding

$$\begin{aligned} n_r(\omega) &= \frac{1 - R(\omega)}{1 + R(\omega) - 2\sqrt{R(\omega)}\cos\Delta} \\ \kappa(\omega) &= \frac{2\sqrt{R(\omega)}\sin\Delta}{1 + R(\omega) - 2\sqrt{R(\omega)}\cos\Delta}. \end{aligned} \tag{26}$$

Using Eq. 26 both n_r and κ, and hence also the real and imaginary parts of the dielectric function, can be determined for any frequency by measuring $R(\omega)$ over the entire frequency range and calculating $\Delta(\omega)$ by Eq. 26. Kramers-Kronig

relations can also be applied to derive *changes* in the dielectric function from measured *changes* in the reflectance, see Sect. 1.2 of ▶ Chap. 10, "Interaction of Light with Solids."

1.2.2 Sum Rules

The interrelation of $\varepsilon'(\omega)$ and $\varepsilon''(\omega)$ in the entire frequency range leads to a number of interesting sum rules. These rules are helpful in checking the consistency of the approximation used, e.g., to check whether all important transitions are included in the Kramers-Kronig relation. Equation 20 yields the following dependence for $\omega_a = 0$:

$$\varepsilon'(0) = \varepsilon_{stat} = 1 + \frac{2}{\pi} P \int_0^\infty \frac{\varepsilon''(\omega)}{\omega} d\omega. \tag{27}$$

The static dielectric constant ε_{stat} is hence determined by the total area under the curve of $\varepsilon''(\omega)$, independent on the spectral range where the absorption occurs.

At very high frequencies, when all valence electrons behave like free electrons, the dielectric function becomes that of an ensemble of free electrons (see Sect. 1 of ▶ Chap. 12, "Photon–Free-Electron Interaction") given by

$$\varepsilon'(\omega) \cong 1 - \frac{\omega_p^2}{\omega^2} \quad \text{with} \quad \omega_p^2 = \frac{n\ e^2}{\varepsilon_{opt}\, \varepsilon_0\, m_0}, \tag{28}$$

where n is the valence-electron density and m_0 is the free-electron mass. Equating this result with Eq. 20 for $\omega_a \gg \omega$ we obtain the dipole sum rule

$$\int_0^\infty \omega\ \varepsilon''(\omega)\, d\omega = \frac{\pi}{2} \omega_p^2. \tag{29}$$

This sum rule is equivalent to the *f*-sum rule for atoms (Kronig 1926). It indicates that strong emission in one part of the spectrum must be compensated by additional absorption in the same or other parts of the spectrum (Stern 1963). The corresponding sum rule for the absorption coefficient is

$$\int_0^\infty \omega \kappa(\omega)\, d\omega = \frac{\pi}{4} \omega_p^2. \tag{30}$$

Another important sum rule is (Lifshits et al. 1985)

$$\int_0^\infty (n_r(\omega) - 1)\, d\omega = 0, \tag{31}$$

i.e., the index of refraction averaged over all frequencies must be equal to one.

A review over further sum rules which are helpful for absorption and dispersive processes is given by Smith (1985).

2 The Dielectric Function in the IR Range

The dielectric function was introduced as a frequency-dependent material parameter in the effective-medium model of solid-light interaction discussed in ▶ Chap. 10, "Interaction of Light with Solids." The frequency dependence originates microscopically from the interaction with ionic and electronic oscillations in the solid typically occurring in the IR and UV ranges, respectively (Fig. 11 of ▶ Chap. 10, "Interaction of Light with Solids"). In the following we focus on the lattice vibrations. Their interaction with the electromagnetic radiation can be described classically, providing a rather simple illustration of connecting the microscopic lattice parameters with optical constants.

2.1 Elementary Oscillators

In semiconductors with ionic bonding forces, certain lattice oscillations couple strongly with electromagnetic radiation. These are the transverse optical oscillations near the center of the Brillouin zone at $q = 0$, or, equivalently, $\lambda \gg a$. Such lattice oscillations were discussed in Sect. 3.1 of ▶ Chap. 4, "Elasticity and Phonons," by analyzing the equation of motion.

Now an external force, caused by an electric field E, is introduced which interacts with the ionic charges of the lattice. Since most semiconductors are partially ionic and partially covalent, the effective charge e_{eff} instead of the electron charge e is used here – see end of this section. Counteracting the external force are the elastic restoring forces assumed to be $\propto\beta$, and a damping friction term $\propto\gamma$ which prevents unchecked energy extraction from the external field at resonance. As in classical mechanics the friction is assumed to be proportional to the velocity. The friction constant γ is related to absorption (inelastic scattering events) and is equal to $1/\tau$, where τ is the lifetime of the given state, or, as will be discussed later, the appropriate relaxation time (see Sect. 4 of ▶ Chap. 22, "Carrier-Transport Equations"). The equation of motion for the relative movement $u = u_1 - u_2$ of the ions in a diatomic crystal under the influence of an electromagnetic wave of frequency ω then reads

$$M_{\text{r}}\left(\frac{d^2u}{dt^2} + \gamma\frac{du}{dt} + \frac{\beta}{M_{\text{r}}}u\right) = e_{\text{eff}}E = e_{\text{eff}}E_0\exp(-i\omega t). \tag{32}$$

M_{r} is the reduced mass of an ion pair in the diatomic crystal given by[3]

[3]Sometimes the *effective atomic weight* is used, related to the mass M_{H} of the hydrogen atom:

$$M_{\text{r}}^{*} = M_{\text{r}}/M_{\text{H}}. \tag{33}$$

$$\frac{1}{M_r} = \frac{1}{M_1} + \frac{1}{M_2}. \tag{34}$$

Equation 32 has the solution

$$u(t) = \tilde{A}\, E_0 \exp(-i\omega t) \tag{35}$$

with the complex amplitude factor

$$\tilde{A}(\omega) = A' + iA'' = \frac{e_{\text{eff}}}{M_r} \quad \frac{1}{\omega_{\text{TO}}^2 - \omega^2 - i\gamma\omega}. \tag{36}$$

The derivation of Eq. 36 is straightforward and can be found in many textbooks. The relevant eigenfrequency ($\omega_{\text{TO}} \cong \sqrt{\beta/M_r}$) is the transverse optical frequency at the center of the Brillouin zone (see Sect. 3.1.3 in ▶ Chap. 4, "Elasticity and Phonons") and is the frequency where the interaction with the electromagnetic wave is strongest.

Introducing Eq. 36 into Eq. 35, we obtain for the ionic displacement

$$u = \frac{e_{\text{eff}}}{M_r} \quad \frac{1}{\omega_{\text{TO}}^2 - \omega^2 - i\gamma\omega} \quad E. \tag{37}$$

Equation 37 describes the displacement in a microscopic lattice oscillator.

In the macroscopic approach, the interaction with light is described by the dielectric polarization. A crystal containing N lattice oscillators per unit volume shows a lattice polarization of

$$P_{\text{latt}} = N\, e_{\text{eff}} u = \frac{N\, e_{\text{eff}}^2}{M_r} \quad \frac{1}{\omega_{\text{TO}}^2 - \omega^2 - i\gamma\omega} \quad E. \tag{38}$$

The electronic polarization is

$$P_{\text{el}} = \varepsilon_0 \left(\varepsilon_{\text{opt}} - 1\right) \; E. \tag{39}$$

The total polarization thus becomes

$$P = P_{\text{latt}} + P_{\text{el}} = N\, e_{\text{eff}} u = \left(\frac{N\, e_{\text{eff}}^2}{M_r} \quad \frac{1}{\omega_{\text{TO}}^2 - \omega^2 - i\gamma\omega} + \varepsilon_0 \left(\varepsilon_{\text{opt}} - 1\right) \right) \; E. \tag{40}$$

When using the definition of the complex dielectric constant, P is also written as

$$P = \varepsilon_0 (\tilde{\varepsilon} - 1) \; E. \tag{41}$$

By comparing this relation with Eq. 40, we obtain the complex dielectric constant

$$\widetilde{\varepsilon}(\omega) = \varepsilon' + i\varepsilon'' = \varepsilon_{\text{opt}} + \frac{\omega_{\text{p}}^2}{\omega_{\text{TO}}^2 - \omega^2 - i\gamma\omega}, \tag{42}$$

where ω_{p} is introduced as a characteristic frequency, the *ionic plasma frequency*[4]

$$\omega_{\text{p}} = \sqrt{\frac{N\, e_{\text{eff}}^2}{M_{\text{r}}\, \varepsilon_0}} \tag{43}$$

or

$$\omega_{\text{p}} = 2.94 \times 10^{13}\, \frac{e_{\text{eff}}}{e} \sqrt{\frac{N}{10^{22}}\, \frac{20}{M_{\text{r}}^*}}\, (\text{s}^{-1});$$

M_{r}^* is the effective atomic weight (Eq. 33).

We can now separate the real and imaginary parts of Eq. 42, and obtain

$$\varepsilon'(\omega) = n_{\text{r}}^2 - \kappa^2 = \varepsilon_{\text{opt}} + \omega_{\text{p}}^2\, \frac{\omega_{\text{TO}}^2 - \omega^2}{\left(\omega_{\text{TO}}^2 - \omega^2\right)^2 + \gamma^2\omega^2}, \tag{44}$$

which is proportional to the squared index of refraction (for $\kappa^2 \ll n_{\text{r}}^2$), and

$$\varepsilon''(\omega) = 2n_{\text{r}}\kappa = \omega_{\text{p}}^2\, \frac{\omega\ \gamma}{\left(\omega_{\text{TO}}^2 - \omega^2\right)^2 + \gamma^2\omega^2}, \tag{45}$$

which is proportional to the optical absorption coefficient (for $n_{\text{r}} \cong$ const).

Both ε' and ε'' are given as functions of the frequency of the impinging optical wave in Fig. 2. It shows an absorption peak at the transverse optical frequency ω_{TO}, with a half-width equal to the damping factor γ; other useful relations for γ are:

$$\Delta\omega_{1/2} = \gamma \quad \text{or} \quad \frac{\Delta\omega}{\omega} \cong \tan\delta \cong \frac{\gamma}{\omega_{\text{TO}}}.$$

The real part of the dielectric function shows a transition from the optical dielectric constant at high frequencies ($\omega \gg \omega_{\text{TO}}$) to the static dielectric constant for low frequencies ($\omega \ll \omega_{\text{TO}}$). The amplitudes at the extrema and their interrelationship are given in Fig. 2.

The relations derived here are rather general and show a strong absorption at the eigenfrequency of dipole oscillators. The half-width of the absorption peak is proportional to the damping. With $1/\gamma$ given by the lifetime of the state, the half-width of the line is also referred to as its *natural line width*. Often the *Lorentzian*

[4] ω_{p} has the same form as the plasma frequency for electrons (Eq. 4 in ▶ Chap. 12, "Photon–Free-Electron Interaction"), except N is the density and M_{r} the mass of phonons. Sometimes the definition of ω_{p} includes an additional factor ε_{opt} in the denominator.

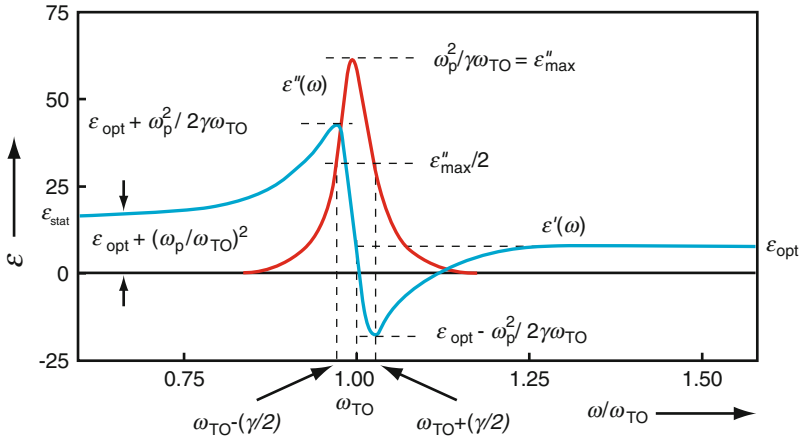


Fig. 2 Real (*blue curve*) and imaginary (*red curve*) parts of the dielectric function near a single resonance of lattice-atom oscillators, calculated from Eqs. 44 and 45 with $\varepsilon_{opt} = 12$, $\omega_p/\omega_{TO} = 1.7$, and $\gamma/\omega_{TO} = 0.05$

line-shape function $g(\omega)$ is used to describe the spectral distribution of the optical absorption near ω_{TO} (see Sect. 2 of ▶ Chap. 17, "Optical Properties of Defects"):

$$g(\omega) = \frac{\gamma}{2\pi} \frac{1}{(\Delta\omega)^2 + (\gamma/2)^2}, \tag{46}$$

where, $\Delta\omega = (\omega_{TO} - \omega)$. Using $g(\omega)$ as band-shape function, Eqs. 44 and 45 can be approximated for $\omega \cong \omega_{TO}$ as

$$\varepsilon' = \varepsilon_{opt} + \frac{\pi\omega_p^2}{\omega_{TO}} \frac{\Delta\omega}{\gamma} g(\omega) \quad \text{and} \quad \varepsilon'' = \frac{\pi\omega_p^2}{2\omega_{TO}} g(\omega). \tag{47}$$

A few additional relations are instructive. For low frequencies, we obtain from Eq. 44 the relation

$$\varepsilon'(0) = \varepsilon_{stat} = \varepsilon_{opt} + \frac{\omega_p^2}{\omega_{TO}^2}. \tag{48}$$

Therefore, Eq. 42 can also be written as

$$\tilde{\varepsilon}(\omega) = \varepsilon_{opt} + \frac{(\varepsilon_{stat} - \varepsilon_{opt})}{(\omega_{TO}^2 - \omega^2 - i\gamma\omega)/\omega_{TO}^2}. \tag{49}$$

The corresponding relation for a *longitudinal wave* (which is not excited by a transverse electromagnetic wave) can be obtained by setting $\tilde{\varepsilon}(\omega) = 0$ in Eq. 42 or

Eq. 49. Neglecting damping ($\gamma = 0$), we obtain for the frequency ω_{LO} of the longitudinal optical oscillation at $q = 0$ the useful *Lyddane-Sachs-Teller* relation (Lyddane et al. 1959)

$$\frac{\omega_{LO}}{\omega_{TO}} = \sqrt{\frac{\varepsilon_{stat}}{\varepsilon_{opt}}}. \tag{50}$$

For resonances including damping, the Lyddane-Sachs-Teller relation requires some modification – see Chang and Mitra (1968). Using Eq. 50 the complex dielectric function in Eq. 49 can be written as

$$\tilde{\varepsilon}(\omega) = \varepsilon_{opt} \frac{\left(\omega_{LO}^2 - \omega^2 - i\gamma\omega\right)}{\left(\omega_{TO}^2 - \omega^2 - i\gamma\omega\right)}. \tag{51}$$

Both the LO and TO frequencies of diatomic crystals are obtainable from the reflection spectra by fitting the parameters of a simple harmonic lattice oscillator to the measured spectra as shown in Fig. 3.

In covalent monatomic crystals, with vanishing dipole moment, $\varepsilon_{stat} = \varepsilon_{opt}$; hence $\omega_{TO} = \omega_{LO}$, which for these materials is usually denoted as ω_0. For the description of crystals with a (partial) ionic bonding the concept of effective charge, which was already used in Eq. 32, is introduced.

2.1.1 Effective Charges

Crystals with partially covalent and partially ionic binding forces can be described as crystals composed of ions, but with a fractional (i.e., an *effective*) charge. Since kinetic effects are involved, this charge is slightly different from the static effective charge introduced in Sect. 3.1 of ▶ Chap. 2, "Crystal Bonding." Three types of dynamic effective charges are used, which are all equal for nonpolarizable (rigid) ions.

The effective charge e_{eff} that appeared in Eq. 32 is called the *macroscopic transverse effective charge* denoted here e_T; it is also referred to as *Born effective charge*. It is the apparent charge of a dipole under the influence of an external macroscopic field and the charge measured directly in experiment. Using the definition of the ionic plasma frequency Eq. 43, we obtain from Eq. 48

$$e_T = \omega_{TO} \sqrt{\frac{\varepsilon_0 \left(\varepsilon_{stat} - \varepsilon_{opt}\right) M_r}{N}}. \tag{52}$$

The charge e_T does, however, not reproduce the real situation, in which a dipole inside the crystal is subjected to a *local* field, which is the sum of the macroscopic field minus the depolarization field induced by it (see Eq. 3). This effect is included in the *real effective charge* introduced by Szigeti (1949); this charge e_S is defined for diatomic cubic crystals by

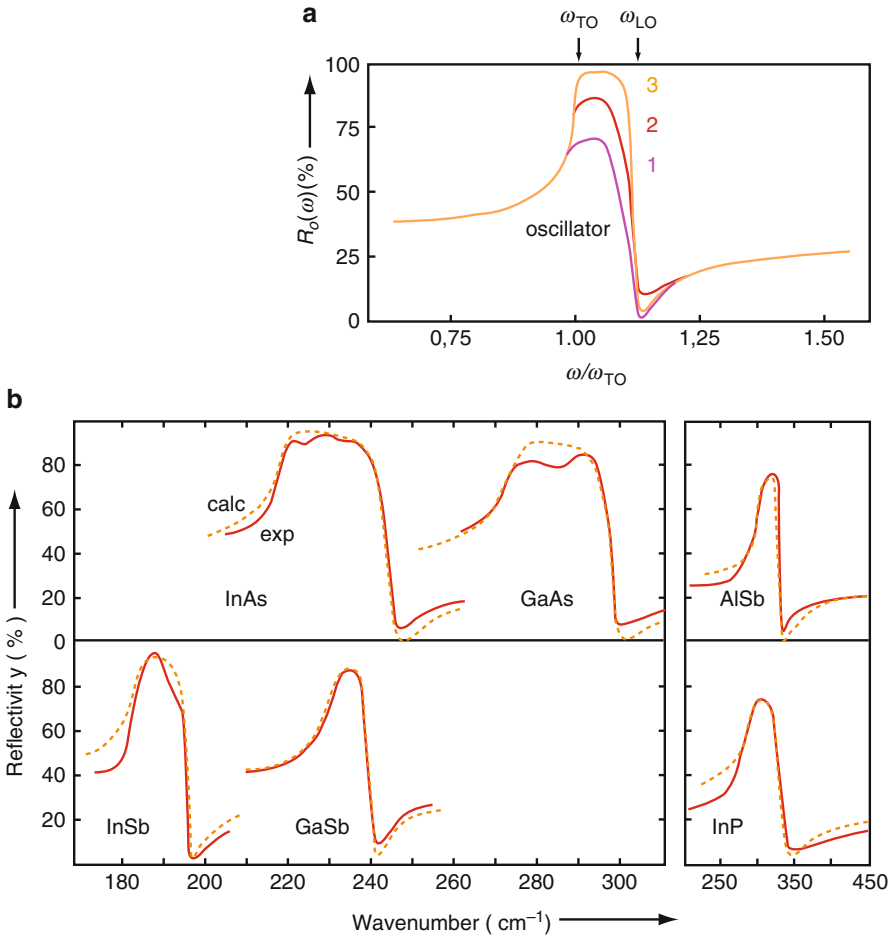


Fig. 3 (**a**) Reflectance of a semi-infinite plate in the range of an absorption maximum, calculated for a simple harmonic oscillator with $\varepsilon_{opt} = 12$, $\varepsilon_{stat} = 15$, $\omega_p/\omega_{TO} = 1.73$, and $\gamma/\omega_{TO} = 0.05$, 0.02, and 0.004 for *curves 1–3*, respectively. (**b**) *Solid lines*: Experimental IR reflection spectra for InAs, GaAs, InSb, GaSb (measured at helium temperature) and AlSb, InP (measured at room temperature). *Dashed lines*: Spectra calculated for simple harmonic oscillators with frequencies ω_{TO}, ω_{LO}, and γ adjusted to fit the measured spectra (After Hass 1967)

$$e_S = \frac{3}{\varepsilon_{opt} + 2} e_T. \tag{53}$$

There is another dynamic effective charge which relates a *longitudinal* oscillation to its induced longitudinal polarization. It was introduced by Callen (1949) and is called *Callen effective charge* e_C defined by

$$e_C = \frac{\omega_{TO}}{\varepsilon_{opt}} \sqrt{\frac{\varepsilon_0 \left(\varepsilon_{stat} - \varepsilon_{opt}\right) M_r}{N}} = \frac{e_T}{\varepsilon_{opt}}. \tag{54}$$

For nonpolarizable ions with $\varepsilon_{opt} = 1$ we have $e_T = e_S = e_C$. A list of effective charges for some III-V semiconductors is given in Table 2.

2.2 IR Reflection and Reststrahlen

The optical eigenfrequencies of typical semiconductor lattices are on the order of 10^{13} s^{-1}; that is, they are in the 20–60 μm IR range. As discussed in the previous section, only the long-wavelength TO oscillations of a diatomic lattice have a nonvanishing dipole moment; they are *infrared active* and absorb light.

Since the absorption coefficient is rather large, it is difficult to measure this coefficient in transmission: it requires a very thin layer. The change in the reflectance is more easily accessible. For a sufficiently thick sample, so that neglecting the reflection at the back surface is permissible, the reflectance is given by Eq. 54 of ▶ Chap. 10, "Interaction of Light with Solids." The reflectance shows a steep rise near ω_{TO}, followed by a sharp drop-off with a minimum at $\omega = \omega_{LO}$ (Fig. 3). The reflectance spectrum, although similar to $\varepsilon'(\omega)$, shows distinct differences in its quantitative behavior, as seen by comparing the blue curve in Fig. 2 with Fig. 3; the observed reflectivity spectra for typical diatomic (III-V) semiconductors follows much of the simple theory given above, as shown in Fig. 3b.

The range of high optical reflection can be used to obtain nearly monochromatic IR light when employing multiple reflections (Fig. 4). In each reflection, one loses a fraction of the nonreflected (i.e., the transmitted) light leaving very little of it after several of these steps. This method is called the *Reststrahlen method* (from "residual rays" in German). The Reststrahlen wavelength is identified by λ_{res} (μm) $= 10^4/\omega_{TO}$ (cm^{-1}), see Fig. 3, since for reasonable damping factors the

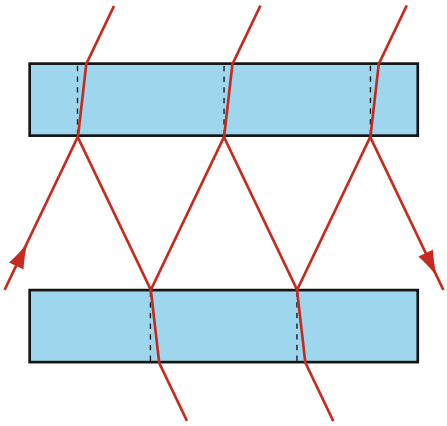


Fig. 4 Reststrahlen setup to obtain monochromatic light near ω_{TO}. The unreflected part of the spectrum is absorbed at the *black* wall of the *box* after penetration through each slab

Table 2 Reststrahlen wavelength λ_{res}, transverse and longitudinal optical frequencies ω_{TO} and ω_{LO}, optic and static dielectric constants, and effective charges for some III-V semiconductors and SiC

Crystal	λ_{res} (μm)	ω_{TO} (cm^{-1})	ω_{LO} (cm^{-1})	ε_{opt}	ε_{stat}	e_T/e	e_C/e	e_S/e
AlSb	31	319	340	10.240	11.633	1.9	0.19	0.48
GaP	27	367	403	9.102	11.051	2.0	0.24	0.58
GaAs	37.3	268	292	10.890	12.861	2.2	0.2	0.51
GaSb	43.5	230	240	14.440	15.694	1.8	0.13	0.33
InP	33	303	345	9.610	12.401	2.5	0.26	0.66
InAs	45	219	243	12.250	15.133	2.6	0.22	0.56
InSb	55	180	191	15.682	17.723	2.5	0.16	0.42
SiC	12.6	796	972	6.708	10.002	2.5	0.56	1.14

reflection maximum lies only slightly above ω_{TO}. Herein, ω_{TO} can be approximated by

$$\omega_{TO} \cong \sqrt{\frac{\beta}{M_r}}. \tag{55}$$

Here β can be estimated for an ionic crystal from the Coulomb force,

$$\beta \cong \frac{Z^2 e^2}{4\pi \varepsilon_0 a_0^3}. \tag{56}$$

with Z as the ionic charge and a_0 as the interionic equilibrium distance. M_r is given by Eq. 34.

Table 2 lists the Reststrahlen wavelengths for a number of semiconductors. As can be seen from Eqs. 55 and 56, the Reststrahlen frequency decreases (i.e., its wavelength increases) with increasing ionic mass (or, reduced mass), with increasing interionic distance, and with decreasing ionic charge.

3 Scattering of Photons with Phonons

3.1 The Phonon-Polariton

In the previous section, the interaction of an electromagnetic wave with TO phonons at $q = 0$ was discussed. We will now analyze this interaction more precisely. In actuality, photons have a finite momentum $\hbar q = \hbar \ 2\pi/\lambda = h/\lambda$ which is very small compared to the extent of the Brillouin zone $\hbar \ \pi/a$. For IR light, the ratio $(\hbar \ 2\pi/\lambda)/(\hbar\pi/a) = 2a/\lambda$ is on the order of 10^{-5}. When enlarging the $E(q)$ diagram, shown for phonons in Fig. 13 of ▶ Chap. 4, "Elasticity and Phonons," for the TO branch near

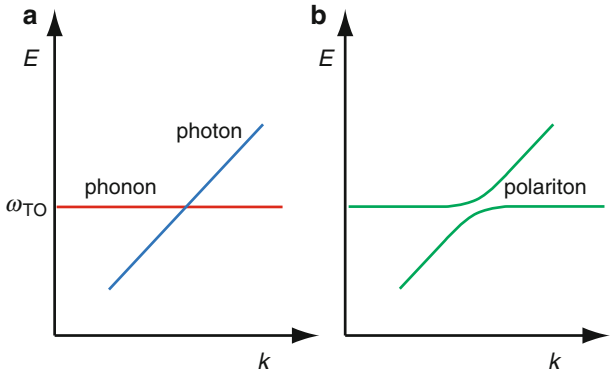


Fig. 5 Dispersion relation $E(k)$ of simple oscillatory phenomena (**a**) without and (**b**) with interaction

$q = 0$, we obtain a horizontal line. For the photon we have another straight line according to

$$p_{\text{phot}} = \frac{h}{\lambda} = \frac{h\nu}{c} = \hbar k. \tag{57}$$

Both lines cross as shown in Fig. 5a (in Fig. 5, k is used instead of q to emphasize the interaction with light).

Since there *is* interaction between the photon and the phonon, we observe a characteristic split in the dispersion spectrum according to the *von Neumann noncrossing principle.* This split occurs near the $E(k)$ value where the dispersion curves of the two quasi-particles would cross if they did not interact with each other. This is shown schematically in Fig. 5b, and will become more transparent from a quantitative analysis.

Since photons and phonons are bosons, the interaction between the two leads to a state that can no longer be distinguished as a photon or a phonon. To emphasize this, a new name is used to describe this state: the *polariton* or, more precisely, the *phonon-polariton,* when mixing between a *phonon* and a photon state occurs, to set it apart from the *exciton-polariton* (Sect. 1.3 of ▶ Chap. 14, "Excitons") or *plasmon-polariton* (Sect. 1.1 of ▶ Chap. 12, "Photon–Free-Electron Interaction").

The set of governing equations for such polaritons, yielding the dispersion relation, can be obtained from Maxwell's equations and the equation of motion in the polarization field

$$\frac{\partial}{\partial \boldsymbol{r}} \times \boldsymbol{E} = -\mu\mu_0 \frac{\partial \boldsymbol{H}}{\partial t} \tag{58}$$

$$\frac{\partial}{\partial \boldsymbol{r}} \times \boldsymbol{H} = \varepsilon\varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} + \sigma \boldsymbol{E} = \varepsilon\varepsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} + \frac{\partial \boldsymbol{P}}{\partial t} \tag{59}$$

$$\frac{\partial^2 \mathbf{P}}{\partial t^2} + \omega_0^2 \mathbf{P} = \chi \varepsilon_0 \mathbf{E}, \tag{60}$$

with ω_0 as the eigenfrequency of the free oscillator (here a phonon, later in our discussion it could be an exciton or a plasmon). These equations define the electric, magnetic, and polarization fields. Assuming linear polarized planar waves, propagating in the x direction with $\mathbf{E} = (0, E_y, 0)$, $\mathbf{H} = (0, 0, H_z)$, and $\mathbf{P} = (0, P_y, 0)$, we obtain after insertion of

$$E_y = E_{y0} \exp(i(kx - \omega t)) \tag{61}$$

$$H_z = H_{z0} \exp(i(kx - \omega t)) \tag{62}$$

$$P_y = P_{y0} \exp(i(kx - \omega t)) \tag{63}$$

into Eqs. 58, 59, and 60, the following governing equations for $\omega(k)$:

$$\omega \varepsilon \varepsilon_0 E_y + \omega P_y - k H_z = 0 \tag{64}$$

$$k E_y - \omega \mu \mu_0 H_z = 0 \tag{65}$$

$$\chi \varepsilon_0 E_y + \left(\omega^2 - \omega_0^2\right) P_y = 0. \tag{66}$$

These equations yield the dispersion equation for polaritons after eliminating E_y, P_y, and H_z:

$$\omega^4 - \left(\omega_0^2 + \frac{\chi}{\varepsilon} + \frac{k^2}{\varepsilon \varepsilon_0 \mu \mu_0}\right) \omega^2 + \frac{k^2 \omega_0^2}{\varepsilon \varepsilon_0 \mu \mu_0} = 0. \tag{67}$$

Entering the resonance frequencies for phonons (Eq. 49), neglecting damping, and using the Lyddane-Sachs-Teller relation (Eq. 50), we obtain for semiconductors with $\mu = 1$ the relation $\chi = \varepsilon_{\text{opt}}\left(\omega_{\text{LO}}^2 - \omega_{\text{TO}}^2\right)$ and further the *phonon-polariton equation* (after using $\varepsilon_0 \mu_0 = 1/c^2$):

$$\omega^4 - \omega^2 \left(\omega_{\text{LO}}^2 + \frac{c^2 k^2}{\varepsilon_{\text{opt}}}\right) + \frac{c^2 k^2}{\varepsilon_{\text{opt}}} \omega_{\text{TO}}^2 = 0. \tag{68}$$

The resulting dispersion curves are shown in Fig. 6. The slope of the longitudinal branch approaches $c/\sqrt{\varepsilon_{\text{opt}}}$, i.e., it has a slightly larger slope than the slope of the transverse branch near $k = 0$, which is $c/\sqrt{\varepsilon_{\text{stat}}}$.

The energy of the phonon-polariton is typically on the order of 50 meV. The first direct observation of the lower branch of this dispersion spectrum was published by Henry and Hopfield (1965) and is shown for GaP in Fig. 7. The measurement involves Raman scattering and will be discussed in more detail in Sect. 3.3. The upper branch was measured for GaP by Fornari and Pagannone (1978).

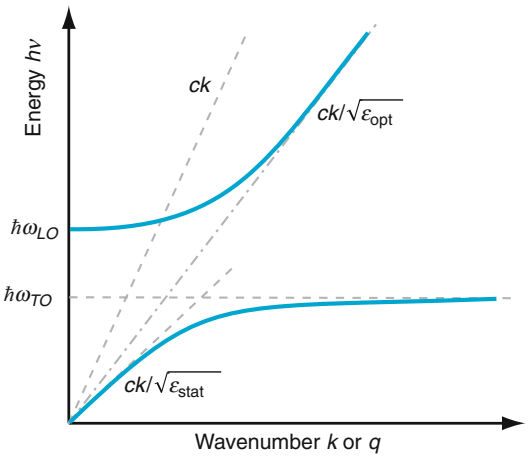


Fig. 6 Dispersion relation for the phonon-polariton with appropriate *slopes* indicated. The *dashed line* $h\nu = ck$ is for photons in vacuum

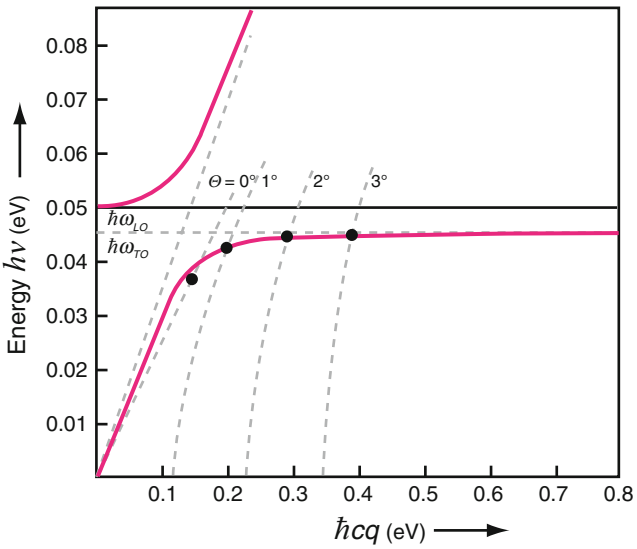


Fig. 7 Dispersion curves of the LO and TO phonons of GaP and the 6328 Å (≅1.96 eV) photon (*dashed*) at long-wavelengths (small q values), and of the polariton (*solid curves*). Raman scattering *angles* are indicated (After Henry and Hopfield 1965). Observe that the edge of the Brillouin zone is at ($\pi/a = 5.76.10^7$ cm^{-1}) $\hbar cq = 1137$ eV, i.e., far to the right in the scale given for the abscissa

3.2 One- and Multiphonon Absorption

The *one-phonon spectrum* provides information only very near to the Γ point, i.e., the center of the Brillouin zone (see previous section). A *multiphonon* interaction, however, yields a wealth of information about the phonon spectrum in the entire Brillouin zone, since the second phonon can provide the necessary momentum shift. The second phonon interacts because of the anharmonicity of the lattice potential,

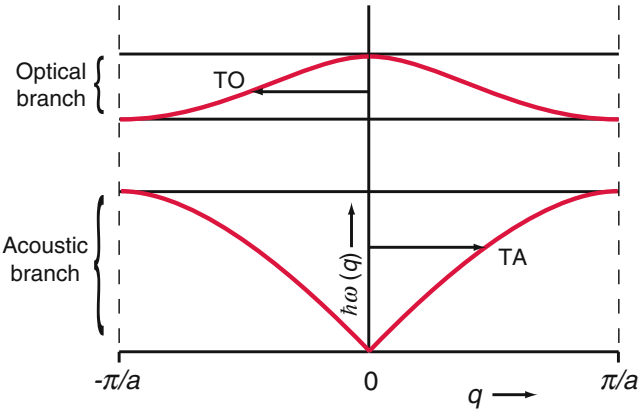


Fig. 8 Schematics of a two-phonon absorption from a photon (of $k \cong 0$) with conservation of momentum

which results in a *weak* optical absorption. Such an interaction takes place when all three (or more) particles fulfill energy and momentum conservation:

$$h\nu = \hbar\omega(\mathbf{q}_1) \pm \hbar\omega(\mathbf{q}_2) \ (\pm \ \ldots) \text{ and } \hbar k = \hbar\mathbf{q}_1 \pm \hbar\mathbf{q}_2 \ (\pm \ \ldots), \qquad (69)$$

where the positive signs in the sums represent simultaneous absorption of two (or more) phonons and negative signs, in a two-phonon process, a simultaneous absorption of photon and an emission of another phonon. An example for a two-phonon absorption with momentum conservation is shown in Fig. 8.

Stronger absorption peaks are obtained at critical points – that is, where the slopes of the two (or more) $\omega(q)$ branches of the involved phonons are nearly horizontal; in other words, when the density of such phonons is relatively high – see Sect. 3.2.1 of ▶ Chap. 4, "Elasticity and Phonons."

The multiphonon processes are observable in materials in which the much stronger one-phonon absorption does not occur because of selection rules, e.g., in Si and Ge, as the latter would hide the weaker multiphonon absorptions. In Fig. 9 a typical multiphonon absorption spectrum is given for Si. The correlation with the phonon dispersion spectrum is indicated by horizontal arrows. In the lower and upper parts of the figure are two-phonon and three-phonon processes, respectively. The corresponding positions are indicated in the dispersion curves at the left and right sides of the absorption spectrum.

3.3 Brillouin and Raman Scattering

Scattering of photons is based on the modulation of the polarizability by phonons or other quasi-particles. There is a multitude of possibilities for such scattering processes; in the following we focus on the interaction with phonons.

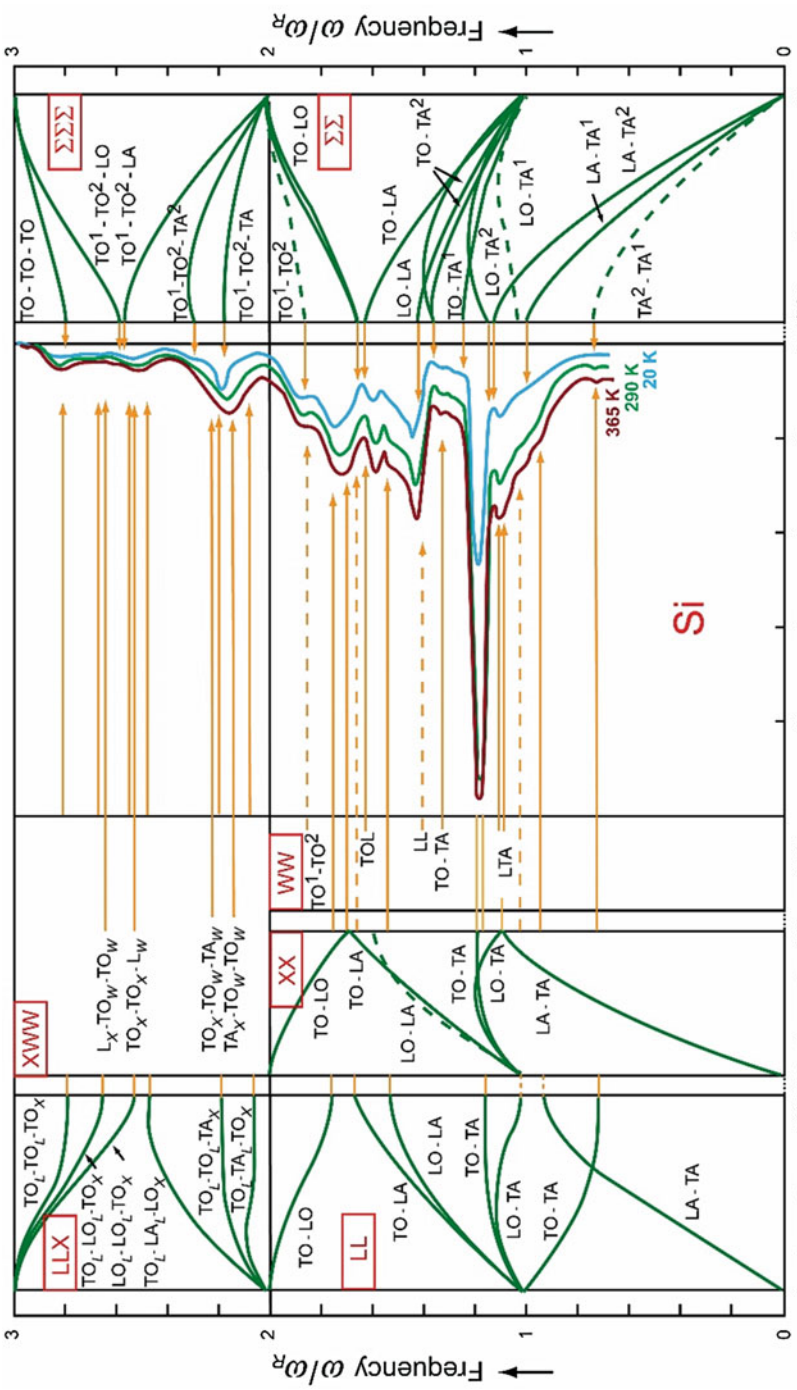


Fig. 9 Multiphonon absorption in Si correlated to the combined density of states along the Δ, Λ, and Σ axes, showing maxima near the critical points *L*, *X*, *W*, and Σ, measured at different temperatures; $\omega_R = 15.5 \cdot 10^{12}\ s^{-1}$ used for normalizing the ordinate, is the TO frequency (≡ LO frequency) of Si (After Mitra 1985)

3.3.1 Elastic and Inelastic Scattering

During a scattering event, energy and momentum must be conserved. If the scattering particle absorbs only a small amount of energy compared to the photon energy, while a substantial amount of momentum is transferred, we refer to an *elastic scattering* event. When a significant amount of energy and momentum are transferred, we speak of an *inelastic scattering* event. When this energy transfer occurs to or from the phonon, we speak of Stokes or anti-Stokes scattering, respectively, with

$$h\nu_\mathrm{s} = h\nu_\mathrm{i} \pm \hbar\omega(\boldsymbol{q}) \tag{70}$$

and

$$\hbar\mathbf{k}_\mathrm{s} = \hbar\mathbf{k}_\mathrm{i} \pm \hbar\mathbf{q}. \tag{71}$$

where ($h\nu_\mathrm{s}$, $\mathbf{k}_\mathrm{s}$) and ($h\nu_\mathrm{i}$, $\mathbf{k}_\mathrm{i}$) are the energy and wavevector of the scattered and incident photon, and $\hbar\omega(\mathbf{q})$ and $\mathbf{q}$ are the wavevector and energy of a phonon. The angle between $\mathbf{k}_\mathrm{s}$ and $\mathbf{k}_\mathrm{i}$ is determined by the experimental setup: the direction of incoming light and the offset position of the scattered-light detector. The selection of $\mathbf{q}$ is then automatic fulfilling Eq. 72. The vector diagram in Fig. 10 indicates the influence of the scattering angle θ on the selection of the phonon with appropriate wavevector $\mathbf{q}$ to maintain momentum conservation in the scattering process:

$$q = \sqrt{k_\mathrm{s}^2 + k_\mathrm{i}^2 - 2k_\mathrm{s}k_\mathrm{i}\cos\theta}. \tag{72}$$

This indicates an important peculiarity of the scattering process. Since the energies or momenta of the scattering particles can be vastly different, substantially different paths in the Brillouin zone are traversed. With

$$|\mathbf{k}| = \frac{2\pi}{\lambda} \text{ and } 0 < q < \frac{\pi}{a} \cong 10^8 \ \mathrm{cm}^{-1}, \tag{73}$$

we see that photons of energy comparable to phonons (IR), or of visible light, have a $\mathbf{k}$ vector of 10^3–10^5 cm^{-1}; this is very small compared to the extent of the Brillouin zone. These photons therefore can probe only the range very close to the center of the Brillouin zone near $q = 0$, i.e., the long-wavelength part of the phonon spectrum (see Sect. 3.1). *X-ray photons* can provide significantly higher momenta up to the

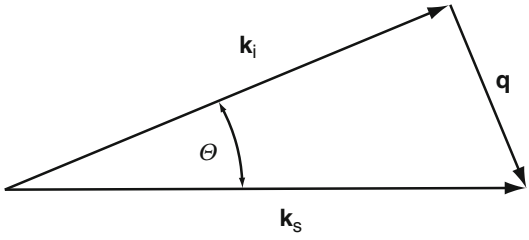


Fig. 10 Vector diagram of momentum conservation during the scattering process

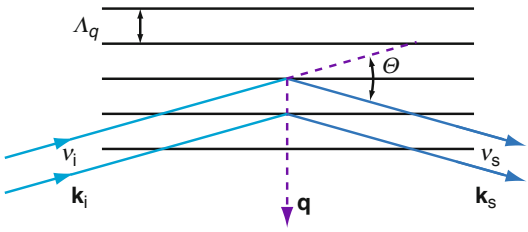


Fig. 11 Brillouin scattering by a moving acoustic wave, with Bragg reflection at an angle θ. Λ is the spatial periodicity of the density wave

zone boundary for $\lambda = 2a$, and inelastic X-ray scattering is applied particularly for studying vitreous solids (Baldi et al. 2010, 2016). However, the *energy* transmitted by the phonon is only a very small fraction of the x-ray energy and poses difficulties in detecting the small relative Stokes or anti-Stokes shift of such an interaction. On the other hand, substantially larger shifts in the phonon momentum can be obtained by scattering with neutrons, as discussed in Sect. 3.3.5 of ▶ Chap. 4, "Elasticity and Phonons."

Elastic scattering of light on phonons, i.e., scattering on density fluctuation, resulting in fluctuation of n_r and κ, is called *Rayleigh scattering*. As it represents a very small effect in solids, it will not be considered further.[5] *Inelastic scattering* caused by *acoustic phonons* is identified as *Brillouin scattering*, and inelastic scattering caused by *optical phonons* is called *Raman scattering.* In most of these scattering events only a small amount of light is scattered by phonons when an intense light beam impinges on the solid.

3.3.2 Brillouin Scattering

Brillouin scattering (Brillouin 1922) – the inelastic scattering at acoustic phonons – can be understood on the basis of classical arguments. The propagation of acoustic phonons as sound waves causes local density waves, creating waves of small changes in the index of refraction propagating with the speed of sound. The scattered light is a Bragg reflection (Fig. 11) from the grating of these density waves, which has a spacing of $\Lambda_q = 2\pi/q$ and moves with velocity v_q, resulting in a Doppler shift of the incident light.

[5]Rayleigh scattering is a well-known effect in media with large density fluctuations, such as gasses. The elastic scattering proceeds without changes in frequency of the scattered photon. Rayleigh scattering is responsible for the blue light of the sky by scattering the short-wavelength component of the sunlight on density fluctuations of the earth's atmosphere. The scattering amplitude – and consequently the absorption coefficient α – increases with decreasing wavelength: $\alpha \propto (n_r - 1)^2/(N\ \lambda^4)$, where N is the density of air molecules and λ is the wavelength. In solids, the Rayleigh component can usually be neglected, except near critical points where density fluctuations can become rather large, e.g., when electron–hole condensation starts to occur. Frozen-in density fluctuations in glasses, although very small, provide transparency limitations for fiber optics because of such Rayleigh scattering.

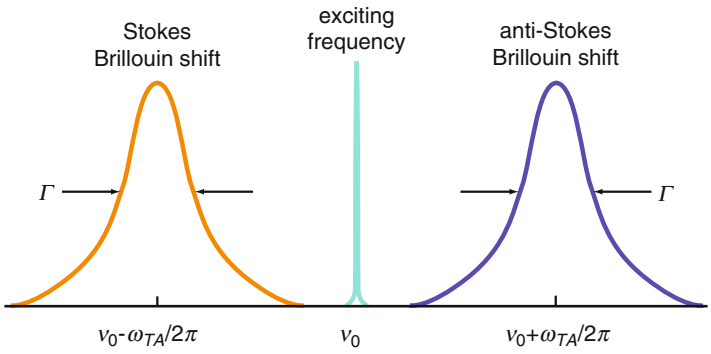


Fig. 12 The Brillouin-scattering spectrum with Stokes and anti-Stokes scattering components

The light beam is scattered at the angle θ (Fig. 10[6]) and contains two components: one photon with the phonon frequency added (anti-Stokes) and another with this phonon frequency subtracted (Stokes):

$$h\nu_\mathrm{i} \pm \hbar\omega_\mathbf{q} = h\nu_\mathrm{i} \pm 2h\nu_\mathrm{i}\, n_\mathrm{r} \frac{v_\mathrm{sound}}{\mathrm{c}} \sin\frac{\theta}{2}. \tag{74}$$

Here v_sound is the sound velocity and c is the light velocity.[7] Such a Brillouin shift is shown in Fig. 12 and is a direct measure of the phonon frequency and the sound velocity. Thus, it yields information on elastic constants, their anisotropy, and various other properties related to the interaction of acoustic phonons with other low-energy excitation phenomena (Pine 1972, 1983). The width Γ of the two Brillouin components yields information on a variety of relevant damping processes, such as carrier-induced damping or structural relaxation (Balkanski and Lallemand 1973).

A typical spectrum for Brillouin scattering is shown in Fig. 13a. The Stokes and anti-Stokes shifts are generally very small (usually below 1 meV); furthermore, the coupling to acoustic phonons is weaker than to optical phonons, requiring a high-resolution setup with strong stray-light suppression for detecting Brillouin scattering. The Brillouin peaks of scattering at TA and LA phonons of CdS are superimposed on the strong signal of the elastically scattered Rayleigh line at zero energy shift.

Resonant Brillouin Scattering When the wavelength of monochromatic light approaches an electronic transition, *resonant Brillouin scattering* can be observed

[6]The scattering angle θ defined in Fig. 10 is twice the Bragg angle θ_B, which is the angle between the diffracting planes and the incident or diffracted beam - cf. Fig. 11.

[7]The very small difference between the refractive indices at $h\nu_\mathrm{i} + \hbar\omega_\mathbf{q}$ and $h\nu_\mathrm{i} - \hbar\omega_\mathbf{q}$ is neglected in Eq. 74.

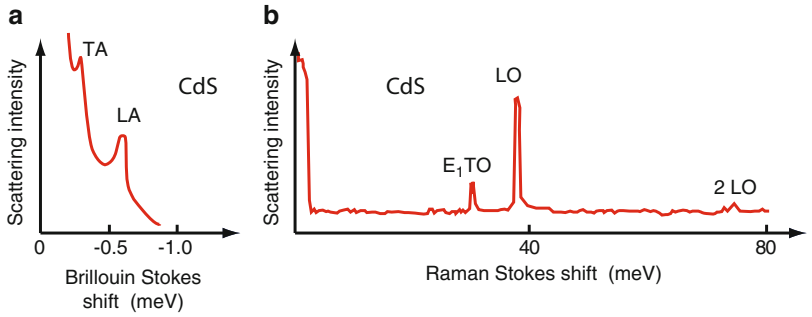


Fig. 13 Typical anti-Stokes spectra for (**a**) Brillouin and (**b**) Raman scattering of CdS measured at $T = 10$ K (After Broser and Rosenzweig (1980) and Martin and Damen (1971), respectively)

(Ulbrich and Weisbuch 1978). Scattering with a phonon of sufficient energy to permit such an electronic transition is then enhanced. Fine tuning to achieve this transition can be achieved by changing the laser frequency or the bandgap by slight temperature variation (Pine 1972). Such an enhancement is usually found for energy of the incident light below the bandgap of semiconductors; for a review, see Yu (1979). Resonant Brillouin scattering also occurs in scattering processes on polaritons, where the phonon supplies missing energy to match a resonance (Jusserand 2015). Intense acoustical waves can also be produced by acoustoelectric domain generation (Conwell 1967). Strongly enhanced scattering, expressed by a large scattering cross-section, is observed when the band edge or other critical points are approached – see, e.g., Hamaguchi et al. (1978). Another means to enhance Brillouin scattering is the application of a metallic nanoantenna (Freeman et al. 2020).

Stimulated Brillouin Scattering Stimulated Brillouin scattering can be achieved at high optical intensities, where both the Brillouin scatterd Stokes light and the acoustic waves can be amplified by a parametric process (Chiao et al. 1964). The energy for the parametric amplification (see Sects. 3.1.3 and 3.1.4 of ▶ Chap. 10, "Interaction of Light with Solids") of the acoustic wave is supplied from the Stokes component of the Brillouin scattering. Usually longitudinal acoustic waves are most strongly excited, see Hasegawa and Hotate (1999). Stimulated Brillouin scattering has recently attracted interest for photonic applications such as slow-light generation (Pant et al. 2012) and light storage (Dong et al. 2015, Merklein et al. 2018).

3.3.3 Raman Scattering

Raman scattering occurs with emission or absorption of optical phonons – for a review, see Balkanski (1980), Jones et al. (2019). While Brillouin scattering is very sensitive to the scattering angle (the acoustic phonon energy changes linearly with $\mathbf{q}$ near $\mathbf{q} = 0$), the classical Raman scattering is not. At a well-defined optical phonon energy, added to or subtracted from the incident photon, scattering occurs

independent of the scattering angle. The absence of spatial dispersion is due to the weak q-dependence of the optical phonon frequencies near $q = 0$ (see Sect. 3.1).

The selection rules for light *scattering* are different from the selection rules for optical *absorption*. Therefore, a different set of optical phonons is *Raman active* at the center of the Brillouin zone – see Mitra (1969). Specifically, for crystals with inversion symmetry, even-parity excitations are Raman active, while odd-parity transitions are observed in IR absorption. The set of Raman active oscillations can be identified by group-theoretical rules according to the crystal symmetry. Consequently, the intensity distribution and polarization of different Raman lines, observed in anisotropic semiconductors, depend on the angle and polarization of the impinging light *relative to the crystal orientation*; see Birman (1974) and Poulet and Mathieu (1970).

Raman scattering is a relatively rare process. Its probability can be estimated from the polarizability tensor (Born and Huang 1954). The efficiency for scattering with TO phonons is estimated as

$$\eta_s = \frac{3\,h^4 \nu_s^4 L d\Omega}{\rho\ c^4 \omega_{TO}} |\alpha_R|^2 \quad \times \quad \begin{cases} f_{BE}(\omega_{TO}) + 1 & \text{for Stokes processes} \\ f_{BE}(\omega_{TO}) & \text{for anti} - \text{Stokes processes} \end{cases}, \tag{75}$$

where ρ is the density, L is the sample length for radiation of the scattered light of frequency ν_s, emitted into the solid angle $d\Omega$; α_R is the first derivative of polarizability (with respect to a ionic displacement Δx) $\partial\chi^{(1)}/\partial\Delta x$, or *Raman polarizability*, which relates to the polarization by $P = \chi^{(1)}E + N\alpha_R E\,u$ (Wynne 1974); $f_{BE}(\omega_{TO})$ is the phonon population given by the Bose-Einstein distribution function. This efficiency is strongly dependent on the frequency of the exciting light and, through $f_{BE}(\omega_{TO})$, on the temperature. Typically, it is on the order of 10^{-6}–10^{-7}; thus, one needs a strong monochromatic light source to generate an observable (faint) scattering signal. An example for Stokes and anti-Stokes spectra is given in Fig. 14a.

Typical Raman spectra are shown for CdS in Fig. 13b. By choosing the polarization with respect to the lattice orientation, discrimination between TO and LO phonons is possible. The ratio of scattering efficiencies for parallel and normal polarization of the scattered beam with respect to the polarization of the exciting beam assists in analyzing more complex Raman spectra (Loudon 1964). Another example, given in Fig. 14b, shows the typical development of the scattering spectra when the composition of a ternary alloy is varied. These spectra can be explained by a system of two coupled modes of lattice oscillation (Jahne 1976) and provide valuable information on crystal structure and composition.

In an alloy of the type $A_xB_{1-x}C$ like $Al_xGa_{1-x}Sb$ phonon spectra show an either *one-mode* or *two-mode* behavior; also a mixed one-two-mode may occur. A one-mode system appears if $M_A < \left(M_B^{-1} + M_C^{-1}\right)^{-1}$ applies for the masses in the alloy (Chang and Mitra 1971). In such a case, realized for I-VII alloys, a single set of long-wavelength phonon appears with a wavelength continuously shifting with alloy composition. The two-mode behavior, which is more common for III-V and II-VI

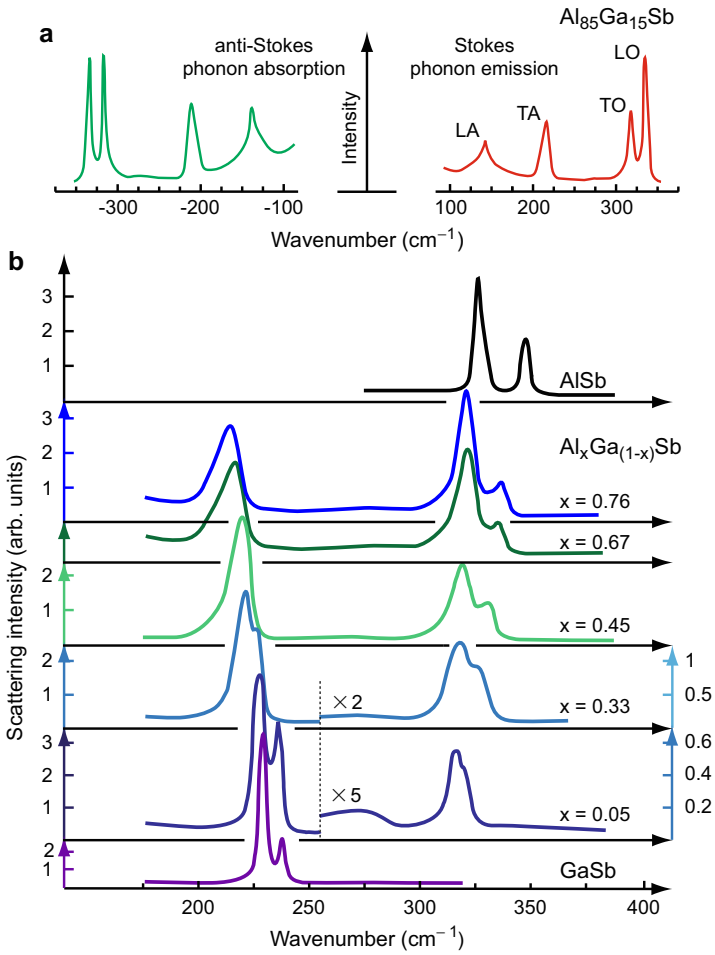


Fig. 14 (**a**) Raman Stokes and anti-Stokes spectra of $Al_{0.85}Ga_{0.15}Sb$. (**b**) Dependence of the Stokes branch on the composition of $Al_xGa_{1-x}Sb$ (After Charfi et al. 1977)

semiconductors, appears if the mass relation is not fulfilled. Here two sets of optical modes being characteristic for the binary (not alloyed) solids appear with intensities roughly proportional to the respective alloying fraction, see Fig. 14b. For more information see Chang and Mitra (1968, 1971). A recent calculation on randomly alloyed zincblende arsenides and nitrides based on density functional theory is reported by Santos et al. (2019).

Polar and Nonpolar Raman Scattering We must distinguish scattering on lattice vibrations which are associated with a dipole moment (*polar modes*) and such on *nonpolar modes*. The latter are: (a) vibrations in covalent crystals, and (b) vibrations

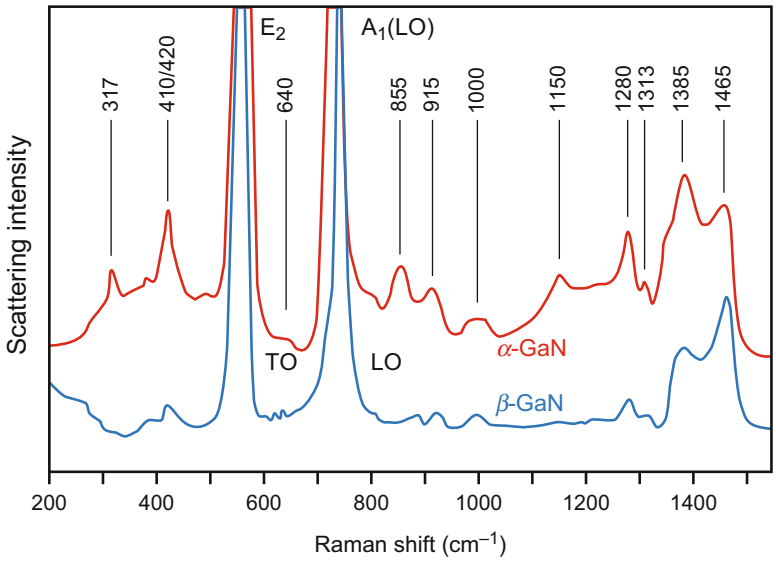


Fig. 15 Second-order Raman spectra of wurtzite (*upper curve*) and zincblende (*bottom curve*) GaN at 300 K with laser excitation at 632.8 nm (After Siegle et al. 1997)

in crystals with an ionic bonding fraction that do not possess a dipole moment, i.e., long-wavelength acoustic modes with $q \ll \pi/a$.

For a discussion of scattering on nonpolar phonons, see Born and Huang (1954) or Cochran (1973). The discussion of polar-mode scattering is more involved in Poulet (1955). Both treatments are substantially different from each other (Hayes and Loudon 1978).

First- and Higher-Order Raman Scattering First- and second-order Raman scattering are distinguished by the emission or absorption of one or two phonons. In second-order Raman scattering, both phonons may be emitted or absorbed, or one may be emitted and another one absorbed, giving a Stokes and an anti-Stokes component. In addition, we distinguish sequential and simultaneous scattering events involving two phonons. Second-order Raman scattering provides access to the entire Brillouin zone, since the second phonon can deliver the necessary momentum. In nonpolar modes, it has been used to obtain information on the deformation potential (Carles et al. 1978), which is important for electron–phonon scattering. A review of second- and higher-order Raman scattering is given by Spitzer and Fan (1957), for ab initio calculation of second-order Raman spectra see Windl et al. (1995), Gillet et al. (2017). Second-order Raman spectra of GaN are shown in Fig. 15. Besides the two strong first-order LO and TO modes, a large number of weaker second-order peaks is found; from symmetry considerations and phonon-dispersion curves the features with energies below the TO mode were assigned to overtone processes of acoustic phonons, while features up to

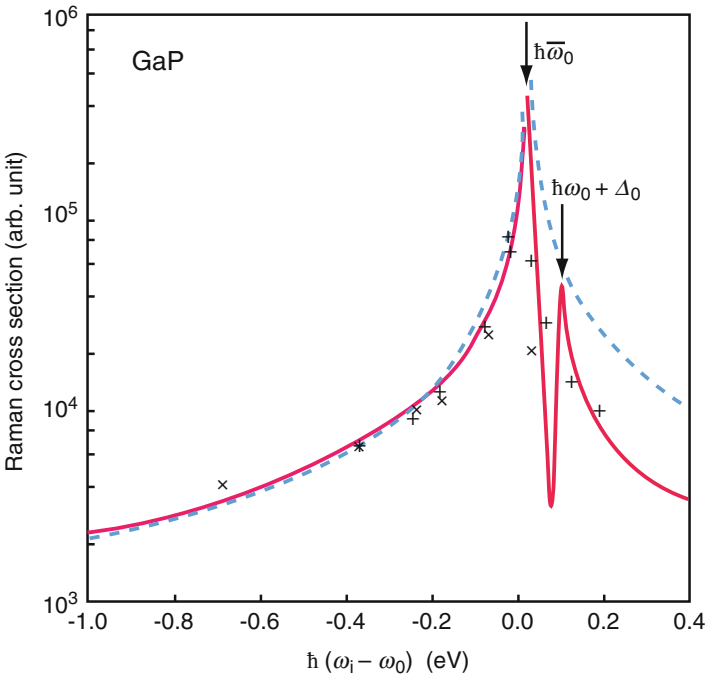


Fig. 16 Resonance enhancement of the first-order TO(Γ) Raman line in GaP; abscissa values are the difference between the energy of the incident photon $\hbar\omega_i$ and the bandgap energy $\hbar\omega_g$. *Crosses* are experimental values, the *solid* and *dashed curves* give calculated scattering intensities with and without spin-orbit splitting, respectively (After Weinstein and Cardona 1973)

1100 cm^{-1} belong to combinations between acoustic and optical phonons, and features at higher energies to combinations of optical phonons and optical overtones (Siegle et al. 1997).

Resonant Raman Scattering Resonant Raman scattering – like resonant Brillouin scattering – is observed with substantial enhancement of the scattering cross-section for band-to-band transitions, free and bound excitons, and for polaritons when the energy of the initiating light is slightly below the energy for the transition which then is enabled by the additional phonon. In Fig. 16 such a resonant enhancement is shown for the first-order TO (Γ) mode in the region of the direct bandgap of GaP. The bandgap energy can be considered as being modulated by the phonon oscillation by a small amount δE_g, and hence to a corresponding proportional Raman polarizability α_R in Eq. 75. The dispersion of the Raman scattering efficiency is proportional to $|\partial\chi/\partial E|^2$. The band-to-band transition of the zincblende semiconductor GaP involves two transitions E_0 and $E_0 + \Delta_0$ split by spin-orbit interaction (see Fig. 11 of ► Chap. 8, "Bands and Bandgaps in Solids"). The solid and dashed curves show the

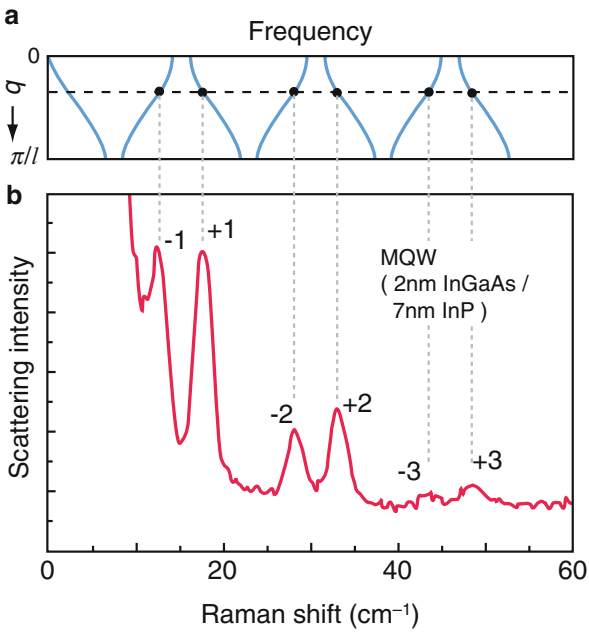


Fig. 17 (**a**) Schematic of folded acoustic phonon dispersion-curve in the Brillouin zone; the *horizontal dashed line* denotes the wavevector transfer in the Raman scattering process (**b**) pairs of folded acoustic phonons in the Raman spectrum of a 10 × (2 nm $In_{0.53}Ga_{0.47}As$/7 nm InP) multi quantum well (After Geurts et al. 1995)

calculated dispersion $|\partial\chi/\partial E|^2$ either with inclusion of $E_0 + \Delta_0$ transitions or considering only the E_0 transition, respectively (Weinstein and Cardona 1973). In the calculations, the bandgap energy was replaced by $\hbar\overline{\omega}_0 = \hbar\omega_0 + \omega_{TO}/2$, where ω_{TO} is the TO(Γ) phonon frequency; thereby the theoretical curves are set to peak at $\hbar\overline{\omega}_0$, since the separate resonances due to the incident and scattered photon energies were not expected to be resolved.

Raman Scattering from Local Modes Raman scattering from local modes of optical phonons yields the frequency of these modes related to certain crystal defects. Thereby, it has given a great deal of information about these defects, including the symmetry of their nearest lattice environment, the mass and bonding force of these impurities (sensitive even for isotope distinction), and their tendency to form defect associates (Barker and Sievers 1975; Hayes and Loudon 1978).

Raman scattering can also be modified or enhanced by surface interaction, e.g., by reduced bonding of surface plasmons (see Burstein et al. 1979).

3.3.4 Raman Scattering in Superlattices

Dynamics in a superlattice of alternating layers *A* and *B* depend on whether phonon modes of one material can also exist in the other material or not. While acoustic modes propagating in both materials can be described by an averaged dispersion, the dispersion of optical phonons may not overlap in the two materials and the modes get confined. For a review on Raman scattering in superlattices, see Abstreiter (1986), Abstreiter et al. (1986), Fasolino and Molinari (1990), and Ruf (1998).

Folded Acoustic Phonons The long-wavelength character of acoustic phonons and the common propagation in both materials of a superlattice yield an effective sound velocity, which reflects the relative thicknesses of layers A and B. Normal to the interfaces,[8] the superlattice periodicity leads to a backfolding of this averaged bulk-like dispersion to the mini-Brillouin zone, see Sects. 3.3.1 of ▶ Chap. 4, "Elasticity and Phonons," and 3.1.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids." Raman scattering provides direct evidence for such a folding: new doublets are seen, as presented in Fig. 17. These doublets appear for the second and higher folded branches of the phonon-dispersion curves, as shown in the insert of this figure. The dispersion relation within the folded lower branches for TA or LA modes can be approximated by (Rytov 1956)

$$\omega_m = \nu_{\text{sound}}^{\perp} \left| \frac{2\pi m}{l} \right| \pm q, \tag{76}$$

where m is the order of the branch, $\nu_{\text{sound}}^{\perp}$ is the averaged sound velocity normal to the layers, and l is the superlattice periodicity; here, the small splittings at the boundaries of the mini Brillouin zone are neglected.

The frequency shift of the Raman signal depends on the period l of the superlattice and the wavevector k_i of the incident photon (i.e., the wavelength of the excitation laser). A change of l alters the width of the mini-zone; this breaks up the bulk phonon branch at different points (Jusserand et al. 1984). The folded acoustic phonon peaks in the Raman spectrum occur at those frequencies where the horizontal line in Fig. 17a marking the wavevector k_i crosses the folded phonon dispersion curve.

Confined Optical Phonons Optical phonons have usually flat dispersion relations. Their frequency depends strongly on atomic bonding and may be distinctly different in layers A and B constituting the superlattice. Their dispersion then does not overlap in the two materials, yielding nonpropagating modes, which are separately confined in the layers A and B; this is analogous to electronic states in (multi) quantum wells with thick barrier layers.

The oscillation amplitude of a confined optical mode is restricted to the respective layer, yielding a series of harmonics with a longest wavelength $\lambda_{\max}/2$. If the confining layer has a thickness of n lattice constants a_0, the effective wavevector of the modes is given by

$$q_m = \frac{m\pi}{(n+\delta)a_0}, \quad m = 1, 2, \ \ldots \ , n, \tag{77}$$

where the correction δ ($\cong 1$) accounts for a short penetration of the oscillation into the cladding layers of different material. These wavevectors lead to a series of discrete

[8]There is no such folding of the branches *parallel* to the superlattice layers, i.e., in in-plane directions. Thus, phonons propagating in this direction do not show the additional Raman doublets, as shown in the lower curve in Fig. 18 of ▶ Chap. 4, "Elasticity and Phonons."

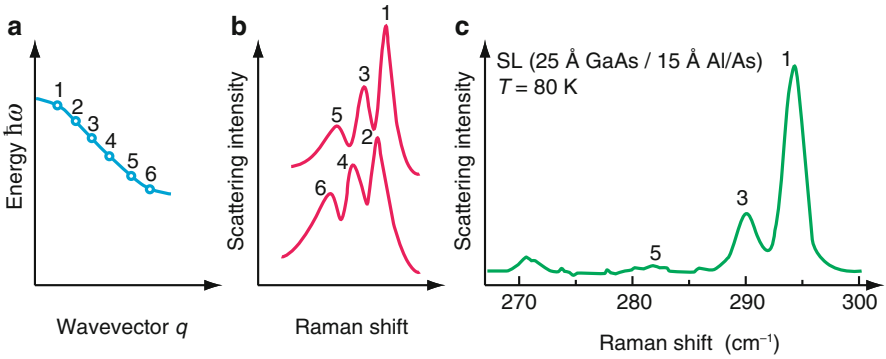


Fig. 18 (**a**) Schematic of an optical phonon dispersion with frequencies selected by discrete wavevectors as given by Eq. 77. (**b**) Raman spectra (schematic) corresponding to (**a**) a series with either even or odd index, which are distinguished by selection rules. (**c**) LO modes of a GaAs/AlAs superlattice (After Jusserand et al. 1985)

mode frequencies as depicted in Fig. 18a and corresponding Raman modes shown in Fig. 18b; even and odd indexed modes may be selected due to symmetry selection rules. Figure 18c gives a Raman spectrum of a GaAs/AlAs superlattice showing LO modes confined in the GaAs layers.

Confined Acoustic Phonons Confinement effects like those found in the periodic structures discussed above are also observed in an individual nanostructure. Studies on nanowires show that the phonon dispersion is already modified at a structure size well above the dominant thermal phonon wavelength $\lambda \cong (v_{\text{sound}}h)/(kT)$, which is typically 1.5 nm at room temperature (RT), and also substantially above the average phonon mean free path given by the Casimir length $L_C = 3\kappa/(C_V v_{\text{sound}})$ (see Sect. 4.1 of ▶ Chap. 5, "Phonon-Induced Thermal Properties"). Several branches of confined acoustic phonon modes were found near the center of the Brillounin zone for GaAs nanowires of ~120 nm diameter, much above $L_C \cong 20$ nm at RT (Kargar et al. 2016). Phonons near the center of the Brillouin zone have a substantial effect on electron relaxation (Bannov et al. 1995). The phonon confinement may affect carrier transport in electronic devices with comparable size, since it alters electron-phonon scattering rates (Pokatilov et al. 2007).

3.3.5 Raman Scattering in Glasses

Raman scattering in glasses does not follow the selection rules for anisotropic crystals.[9] Hence, a wider spectrum of Raman transitions is observed, although some of the gross features are similar to those obtained in the same material in crystalline form. A typical Raman spectrum is shown in Fig. 19 for amorphous

[9]In glasses, one cannot plot Brillouin zones; there is a breakdown of q conservation, i.e., all momenta can contribute during scattering, causing substantial broadening; for a study of this effect applying Brillouin scattering see Vacher et al. (2006).

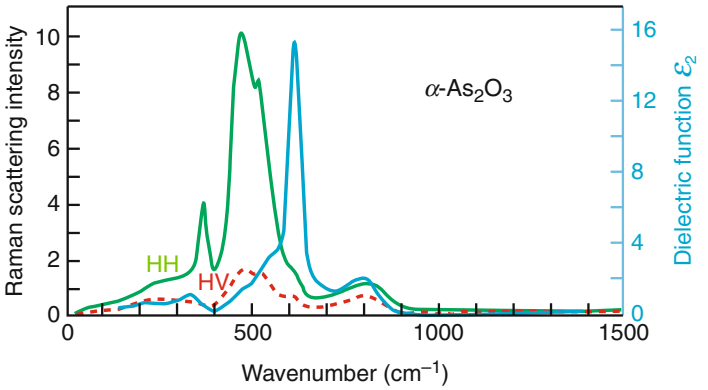


Fig. 19 Reduced Raman spectra (*green* and *red*) and optical IR spectrum (*blue*) for amorphous As_2O_3 (After Galeener et al. 1979)

As_2O_3 and compared to the IR optical dispersion distribution. It is obvious that both spectra expose different features which can be associated with the various oscillatory modes of the different As-O bonds: rocking, stretching, and bending of an As_4O_6 cluster molecule. Some of these are optically active, while others are Raman active. For more detail, see the review of Galeener et al. (1983).

Spectra of Raman-active phonon modes can be used to translate measured peak positions to strain in studied samples: compressive strain induces a blue shift, tensile strain a red shift. The respective strain coefficients for crystalline solids are known since long[10] and were, e.g., used to determine the critical thickness of coherently strained epitaxial layers (Kontos et al. 1999). More recently, a calibration for the Raman/strain relation was also provided for hydrogenated amorphous Si:H (α-Si:H), an important electronic material (for the structure, see Sect. 3.2.2 of ▶ Chap. 3, "The Structure of Semiconductors"). The Raman spectrum of a 1.1 μm thick α-Si:H layer on a (100) crystalline c-Si (crystalline) wafer (Fig. 20) shows a dominating TO peak (sharp for c-Si, broad for α-Si:H) and a weak LO peak; it agrees reasonably with a spectrum computed using density-functional (perturbation) theory (Strubbe et al. 2015). Bending the wafer yields a shift of the TO peak of

$$\Delta\omega = s\left(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}\right) = s\mathrm{Tr}\,\underline{\boldsymbol{\varepsilon}} \tag{78}$$

where s is the strain coefficient and Tr the trace of the strain tensor.[11] The experimental value $s = -510 \pm 120$ cm^{-1} for α-Si:H is consistent with the value computed

[10]Strain coefficients for c-Si are reported by Anastassakis et al. (1990).

[11]Generally, the strain coefficient is a rank-2 tensor like $\underline{\boldsymbol{\varepsilon}}$. For an isotropic solid such as α-Si:H, only diagonal components exist due to symmetry. Since the TO phonon band of α-Si:H transforms as a scalar rather than a vector as for c-Si, only a single value s remains and the trace of the strain tensor yields the total acting strain.

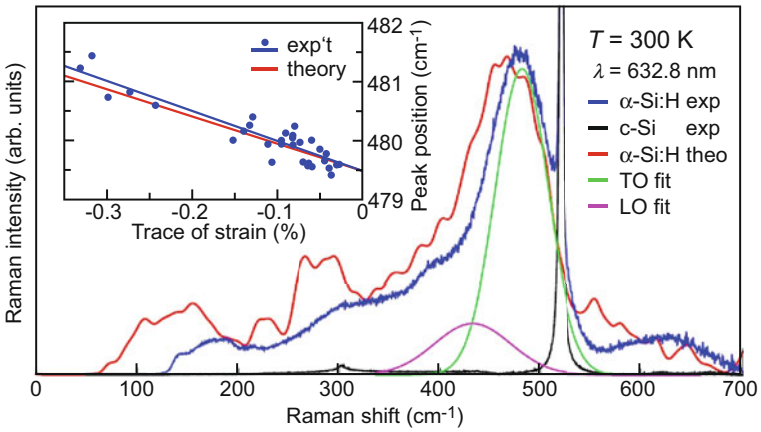


Fig. 20 Raman spectra for α-Si:H on c-Si (*blue curve*) and c-Si (*black*). The *red curve* is an ab initio calculated spectrum. TO and LO phonon peaks are obtained from a fit to the measured α-Si:H spectrum. Inset: Measured (*blue dots*) and calculated (*red*) TO peak shifts induced by compressive strain; *lines* are linear regressions (After Strubbe et al. 2015)

for a structure with randomly broken Si-Si bonds and a typical 10% hydrogen content (Strubbe et al. 2015).

4 Summary

Photons and TO phonons interact strongly near $q = 0$, resulting in a large absorption, which is characteristic for the atomic mass, bonding force, and effective ionic charge of the specific semiconductor. This absorption is located in the 10–100 μm range and is referred to as Reststrahl absorption. In semiconductors with vanishing ionic charges, photons cannot excite single phonons; however, a much weaker optical absorption is observed by simultaneous interaction with two or more phonons, thereby offering for such materials the opportunity to obtain information about the entire phonon dispersion spectrum – in addition to neutron spectroscopy.

Phonon spectroscopy offers access to direct information about the vibrational spectrum of a semiconductor and many of its defects. Such an information is helpful in judging bonding forces, effective ionic charges of crystal lattices, and the local structure of crystal defects which have modes in the gap between acoustic and optical modes of the host lattice. Photon scattering with acoustic (Brillouin) and optical (Raman) phonons provides data on the anisotropic spectrum of sound velocities, effective mass, phonon dispersion, and – from the line shape – the lifetimes of the involved states. With a different set of selection rules, spectra become accessible which cannot be observed by single-phonon absorption experiments.

References

Abstreiter G (1986) Light scattering in novel layered semiconductor structures. In: Grosse P (ed) Festkörperprobleme/Adv Solid State Phys 26:41

Abstreiter G, Merlin R, Pinczuk A (1986) Inelastic light scattering by electronic excitations in semiconductor heterostructures. IEEE J Quantum Electron 22:1771

Anastassakis E, Cantarero A, Cardona M (1990) Piezo-Raman measurements and anharmonic parameters in silicon and diamond. Phys Rev B 41:7529

Baldi G, Giordano VM, Monaco G, Ruta B, (2010) Sound attenuation at terahertz frequencies and the boson peak of vitreous silica. Phys Rev Lett 104:195501

Baldi G, Giordano VM, Ruta B, Monaco G (2016) On the nontrivial wave-vector dependence of the elastic modulus of glasses. Phys Rev B 93:144204

Balkanski M (ed) (1980) Handbook of semiconductors, vol 2. North-Holland, Amsterdam

Balkanski M, Lallemand P (1973) Photonics. Gauthiers-Villard, Paris

Bannov N, Aristov V, Mitin V, Stroscio MA (1995) Electron relaxation times due to the deformation-potential interaction of electrons with confined acoustic phonons in a free-standing quantum well. Phys Rev B 51:9930

Barker AS Jr, Sievers AJ (1975) Optical studies of the vibrational properties of disordered solids. Rev Mod Phys 47(Suppl 2):S1

Birman JL (1974) Infra-Red and Raman Optical Processes of Insulating Crystals; Infra-Red and Raman Optical Processes of Insulating Crystals. Vol. 25/2b, Springer, Berlin Heidelberg

Born M, Huang K (1954) Dynamical theory of crystal lattices. Oxford University Press, London

Brillouin L (1922) Diffusion de la lumiere et des rayonnes X par un corps transparent homogene; influence de l'agitation thermique. Ann Phys (Leipzig) 17:88 (Diffusion of light and X-rays through transparent homogeneous bulk; influence of thermic motion, in French)

Broser I, Rosenzweig M (1980) Magneto-Brillouin scattering of polaritons in CdS. Solid State Commun 36:1027

Burstein E, Chen CY, Lundquist S (1979) Light scattering in solids. In: Birman J, Cummins HZ, Rebane KK (eds) 2nd joint USA-USSR symposium on light scattering in condensed matter. Plenum Press, New York, p 479

Callen HB (1949) Electric breakdown in ionic crystals. Phys Rev 76:1394

Cardona M (1969) Optical constants of insulators: dispersion relations. In: Nudelman S, Mitra MM (eds) Optical properties of solids. Plenum Press, New York, pp 137–151

Carles RN, Saint-Cricq N, Renucci MA, Bennucci BJ (1978) In: Balkanski M (ed) Lattice dynamics. Flammarion, Paris, p 195

Chang IF, Mitra SS (1968) Application of a modified random-element-isodisplacement model to long-wavelength optic phonons of mixed crystals. Phys Rev 172:924

Chang IF, Mitra SS (1971) Long wavelength optical phonons in mixed crystals. Adv Phys 20:359

Charfi F, Zuoaghi M, Llinares C, Balkanski M, Hirlimann C, Joullie A (1977) Small wave vector modes in $Al_{1-x}Ga_xSb$. In: Balkanski M (ed) Lattice dynamics. Flammarion, Paris, p 438

Chiao RY, Townes CH, Stiocheff BP (1964) Stimulated Brillouin scattering and coherent generation of intense hypersonic waves. Phys Rev Lett 12:592

Cochran W (1973) The dynamics of atoms in crystals. Edward Arnold, London

Conwell EM (1967) High-field transport in semiconductors. Academic Press, New York

Dong CH, Shen Z, Zou CL, Zhang YL, Fu W, Guo G-C (2015) Brillouin-scattering-induced transparency and non-reciprocal light storage. Nature Commun 6:6193

Fasolino A, Molinari E (1990) Calculations of phonon spectra in III-V and Si-Ge superlattices: a tool for structural characterization. Surf Sci 228:112

Fornari B, Pagannone M (1978) Experimental observation of the upper polariton branch in isotropic crystals. Phys Rev B 17:3047

Freeman R, Lemasters R, Kalejaiye T, Wang F, Chen G, Ding J, Wu M, Demidov VE, Demokritov SO, Harutyunyan H, Urazhdin S (2020) Brillouin light scattering of spin waves inaccessible with free-space light. Phys Rev Research 2:033427

Galeener FL, Lucovsky G, Geils RH (1979) Raman and infrared spectra of vitreous As_2O_3. Phys Rev B 19:4251
Galeener FL, Leadbetter AJ, Stringfellow MW (1983) Comparison of the neutron, Raman, and infrared vibrational spectra of vitreous SiO_2, GeO_2, and BeF_2. Phys Rev B 27:1052
Geurts J, Gnoth D, Finders J, Kohl A, Heime K (1995) Interfaces of InGaAs/InP multi quantum wells studied by Raman spectroscopy. Phys Status Solidi A 152:211
Gillet Y, Kontur S, Giantomassi M, Draxl C, Gonze X (2017) Ab initio approach to second-order resonant Raman scattering including exciton-phonon interaction. Sci Rep 7:7344
Gutman F (1948) The electret. Rev Mod Phys 20:457
Hamaguchi C, Adachi S, Itoh Y (1978) Resonant Brillouin scattering phenomena in some II–VI compounds. Solid State Electron 21:1585
Hasegawa T, Hotate K (1999) Measurement of Brillouin gain spectrum distribution along an optical fiber by direct frequency modulation of a laser diode. Proc SPIE 3860:306
Hass M (1967) Lattice reflection. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 3. Academic Press, New York, p 3
Hayes W, Loudon R (1978) Scattering of light by crystals. Wiley, New York
Henry CH, Hopfield JJ (1965) Raman scattering by polaritons. Phys Rev Lett 15:964
Jackson JD (1999) Classical electrodynamis, 2nd edn. Wiley, New York
Jahne E (1976) Long-wavelength optical phonons in mixed crystals: I. A system of two coupled modes. Phys Status Solidi B 74:275. And: Long-wavelength optical phonons in mixed crystals. II. The persistence of common gaps. Phys Status Solidi B 75:221
Jonscher AK (1983) Dielectric relaxation in solids. Chelsea Dielectric Press, London
Jones RR, Hooper DC, Zhang L, Wolverson D, Valev VK (2019) Raman techniques: Fundamentals and frontiers. Nanoscale Res Lett 14:231
Jusserand B, Paquet D, Kervarec J, Regreny A (1984) Raman scattering study of acoustical and optical folded modes in GaAs/GaAlAs superlattices. J Physique Colloq 45:145
Jusserand B, Alexandre F, Paquet D, Le Roux G (1985) Raman scattering characterization of interface broadening in GaAs/ AlAs short period superlattices grown by molecular beam epitaxy. Appl Phys Lett 47:301
Jusserand B, Poddubny AN, Poshakinskiy AV, Fainstein A, Lemaitre A (2015) Polariton resonances for ultrastrong coupling cavity optomechanics in GaAs/AlAs multiple quantum wells. Phys Rev Lett 115:267402
Kargar F, Debnath B, Joona-Pekko K, Säynätjoki A, Lipsanen H, Nika DL, Lake RK, Balandin AA (2016) Direct observation of confined acoustic phonon polarization branches in freestanding semiconductor nanowires. Nature Commun 7:13400
Kramers HA (1929) Die Dispersion und Absorption von Röntgenstrahlen. Phys Z 30:522 (The dispersion and absorption of X-rays, in German)
Kronig R d L (1926) On the theory of dispersion of X-rays. J Opt Soc Am 12:547
Kontos AG, Anastassakis E, Chrysanthakopoulos N, Calamiotou M, Pohl UW (1999) Strain profiles in overcritical (001) ZnSe/GaAs heteroepitaxial layers. J Appl Phys 86:412
Lifshits E, Pitaevski LP, Landau LD (1985) Electrodynamics of continuous media. Elsevier, Amsterdam
Lines ME, Glass AM (1979) Principles and applications of ferroelectrics and related materials. Oxford University Press, London
Loudon R (1964) The Raman effect in crystals. Adv Phys 13:423
Lyddane RH, Sachs RG, Teller E (1959) On the polar vibrations of alkali halides. Phys Rev 59:673
Martin RM, Damen TC (1971) Breakdown of selection rules in resonance Raman scattering. Phys Rev Lett 26:86
Merklein M, Stiller B, Eggleton BJ (2018) Brillouin-based light storage and delay techniques. J Opt 20:083003
Mitra SS (1969) Infrared and Raman spectra due to lattice vibrations. In: Nudelman S, Mitra MM (eds) Optical properties of solids. Plenum Press, New York, pp 333–451

Mitra SS (1985) Optical properties of nonmetallic solids for photon energies below the fundamental band gap. In: Palik ED (ed) Handbook of optical constants of solids. Academic Press, New York, pp 213–270

Pandey RN, Sharma TP, Dayal B (1977) Electronic polarisabilities of ions in group III-V crystals. J Phys Chem Solids 38:329

Pant R, Byrnes A, Poulton CG, Li E, Choi DY, Madden S, Luther-Davies B, Eggleton BJ (2012) Photonic-chip-based tunable slow and fast light via stimulated Brillouin scattering. Opt Lett 37:969

Pauling L (1927) The theoretical prediction of the physical properties of many-electron atoms and ions. Mole refraction, diamagnetic susceptibility, and extension in space. Proc Roy Soc Lond A 114:181

Pine AS (1972) Resonance Brillouin scattering in cadmium sulfide. Phys Rev B 5:3003

Pine AS (1983) Brillouin scattering in semiconductors. In: Cardona M (ed) Light scattering in solids I. Springer, Berlin, pp 253–273

Pokatilov E P, Nika DL, Askerov AS, Balandin AA (2007) Size-quantized oscillations of the electron mobility limited by the optical and confined acoustic phonons in the nanoscale heterostructures. J Appl Phys 102:54304

Poulet H (1955) Sur certaines anomalies de l'effet Raman dans les cristaux. Ann Phys (Paris) 10:908. (On certain anomalies of the Raman effect in crystals, in French)

Poulet H, Mathieu JP (1970) Spectres des Vibration et Symétrie des Cristeaux. Gordon & Breach, London (Vibration spectra and symmetry of crystals, in French)

Ruf T (1998) Phonon scattering in semiconductors, quantum wells and superlattices. Springer, Berlin

Rytov SM (1956) Electromagnetic properties of a finely stratified medium. Sov Phys -JETP 2:466

Santos AM, Alves HWL, Ataide CA, Guilhon I, Marques M, Teles LK (2019) Rigorous statistical thermodynamical model for lattice dynamics in alloys. Phys Rev B 100:144205

Shanker J, Agrawal GG, Dutt N (1986) Electronic polarizabilities and photoelastic behaviour of ionic crystals. Phys Status Solidi B 138:9

Siegle H, Kaczmarczyk G, Filippidis L, Litvinchuk AP, Hoffmann A, Thomsen C (1997) Zone-boundary phonons in hexagonal and cubic GaN. Phys Rev B 55:7000

Smith DY (1985) Dispersion theory, sum rules, and their application to the analysis of optical data. In: Palik ED (ed) Handbook of optical constants of solids. Academic Press, New York, pp 35–68

Spitzer WG, Fan HY (1957) Determination of optical constants and carrier effective mass of semiconductors. Phys Rev 106:882

Stern F (1963) Elementary theory of the optical properties of solids. In: Seitz F, Turnbull D (eds) Solid state physics, vol 15. Academic Press, New York, p 299

Strubbe DA, Johlin EC, Kirkpatrick TR, Buonassisi T, Grossman JC (2015) Stress effects on the Raman spectrum of an amorphous material: Theory and experiment on α-Si:H. Phys Rev B 92:241202

Szigeti B (1949) Polarizability and dielectric constant of ionic crystals. Trans Faraday Soc 45:155

Ulbrich RG, Weisbuch C (1978) Resonant Brillouin scattering in semiconductors. In: Treusch J (ed) Festkörperprobleme/Adv Solid State Phys 18:217. Vieweg, Braunschweig

Vacher R, Ayrinhac S, Foret M, Rufflé B, Courtens E (2006) Finite size effects in Brillouin scattering from silica glass. Phys Rev B 74:012203

Weinstein BA, Cardona M (1973) Resonant first- and second-order Raman scattering in GaP. Phys Rev B 8:2795

Windl W, Karch K, Pavone P, Schütt O, Strauch D (1995) Full ab initio calculation of second-order Raman spectra of semiconductors. Int J Quantum Chem 56:787

Wynne JJ (1974) Spectroscopy of third-order optical nonlinear susceptibilities I. Comments Solid State Phys 6:31

Yu PY (1979) Resonant Brillouin scattering of exciton polaritons. Comments Solid State Phys 9:37

Yu PY, Cardona M (1999) Fundamentals of semiconductors: physics and materials properties, 2nd edn. Springer, Berlin

Photon–Free-Electron Interaction

12

Karl W. Böer and Udo W. Pohl

Contents

1 Free-Electron Resonance Absorption 468
1.1 Electron-Plasma Absorption 469
1.2 Valence-Electron Plasma Absorption 472
1.3 Charge-Density Waves 474
2 Nonresonant Free-Carrier Absorption 475
2.1 Dispersion Relation for Free Carriers 475
2.2 Free-Electron Absorption 478
3 Carrier Dispersion in Electric and Magnetic Fields 484
3.1 Magnetoplasma Reflection 487
3.2 Cyclotron-Resonance Absorption and Faraday Effect 489
4 Plasmon Dispersion in 2D Semiconductors 491
5 Summary 494
References 495

Abstract

The interaction of photons with free electrons or holes in the respective bands strongly influences optical absorption and reflection in the spectral region between the absorption edge at bandgap energy and the Reststrahlen wavelength. In a driving external field, the ensemble of free electrons (or, at higher frequency, of valence-electrons) oscillates with respect to the ion cores on the whole, leading to a plasma resonance absorption. The plasmon dispersion has two branches with frequencies depending on the carrier density. Nonresonant carrier absorption occurs away from the resonance, with a free-electron

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_12

contribution predominantly from indirect transitions within the conduction band, and prevalent direct transitions for holes in the valence band. The spectra and underlying dispersion relations provide valuable information about the effective masses of electrons and holes, carrier concentrations, and carrier-relaxation times.

Keywords

Cyclotron-resonance absorption · Electron-plasma absorption · Faraday effect · Free-carrier absorption · Free-electron dispersion · Magnetoplasma reflection · Photon-free-electron interaction · Plasma frequency · Plasmon · Plasmon dispersion · Valence-electron plasma absorption

1 Free-Electron Resonance Absorption

The optical excitation of electrons or holes within their respective bands may involve a large variety of excitation mechanisms and multiparticle influences during the excitation process, such as:

- Direct (vertical in **k** space) transition from one into another branch of the same band
- Indirect transition with phonon assistance
- Transitions followed by other inelastic scattering processes
- Transitions involving collective effects of electrons (plasmons)
- Transitions between subbands created by a magnetic field

These processes will be discussed in the following sections. The treatment is quite similar to that applied to lattice vibrations discussed in ▶ Chap. 11, "Photon–Phonon Interaction." Collective electronic resonances and single-particle excitations can be well described in such a classical model. The dielectric function of free electrons is similar to that of optical lattice vibrations, except that here the eigenfrequency is zero and the damping processes of the related optical absorption require a quantum-mechanical treatment.

The interaction of photons with free electrons in the conduction band or holes in the valence band can result in *resonant* or *nonresonant* absorption. Resonant absorption is discussed first because of the similarities in the mathematical treatment with the ionic oscillations. The logical extension of the field-induced lattice oscillation is the polarization of the *electronic shell* surrounding each lattice atom in an external electromagnetic field. In a classical model, the shift of the electron cloud with respect to the nucleus is proportional to the electric field and given by the electronic polarizability α_{el}. The resulting dispersion formula is similar to Eq. 42 in ▶ Chap. 11, "Photon–Phonon Interaction," except for a different eigenfrequency and damping factor relating to the electronic shell. When the electron density is large enough, such electron-collective effects cause resonance absorption, which for semiconductors lies in the IR range beyond the band edge.

1.1 Electron-Plasma Absorption

In an external electric field, the electrons act jointly when they are shifted as an entity with respect to the ionized donors. The field E_x exerts a force $-eE_x$ and causes a shift by Δx from their on-the-average neutral position, creating a net charge at the two outer surfaces of the solid, which can be calculated by integrating the Poisson equation

$$\delta E_x = \frac{\rho}{\varepsilon \varepsilon_0} \Delta x = \frac{ne}{\varepsilon \varepsilon_0} \Delta x, \tag{1}$$

where n is the concentration of free electrons. The induced field δE_x counteracts the driving field E_x; that is, $e\delta E_x$ acts as restoring force, causing oscillations of the electron collective, following the equation of motion (Pines 1999):

$$n m_n \frac{d^2 \Delta x}{dt^2} + n m_n \gamma \frac{d\Delta x}{dt} = -n \ e \, \delta E_x = -\frac{n^2 \ e^2}{\varepsilon_{\text{opt}} \varepsilon_0} \Delta x; \tag{2}$$

m_n is the effective mass of a free electron, and γ is a damping parameter which can be related to an energy-relaxation time of electrons by $\gamma = 1/\tau_{\text{el}}$. We have introduced for ε the optical dielectric constant ε_{opt}, since we expect oscillations between the band edge and the Reststrahl frequency. With a driving external field, this equation can be modified as a damped harmonic oscillator equation (similar to Eq. 32 in ▶ Chap. 11, "Photon–Phonon Interaction")

$$m_n \left(\frac{d^2 \Delta x}{dt^2} + \gamma \frac{d\Delta x}{dt} + \omega_p^2 \Delta x \right) = eE_x \exp(-i\omega t). \tag{3}$$

ω_{p} is the *plasma frequency* obtained by comparing Eqs. 2 and 3:

$$\omega_{\text{p}} = \sqrt{\frac{n e^2}{\varepsilon_{\text{opt}} \varepsilon_0 m_n}}. \tag{4}$$

ω_{p} is the frequency at which an undamped plasma of electrons and positive ions oscillate on the whole in a longitudinal mode; this mode interacts strongly with electrons. There is also a transverse oscillation which interacts with electromagnetic radiation.

The restoring force in the motion of the electron plasma is proportional to the square of the plasma frequency. As a quantum of quantized energy, $\hbar\omega_{\text{p}}$ is referred to as a *plasmon*:

$$\hbar\omega_{\text{p}} = 38.9 \times \sqrt{\frac{n\,(\text{cm}^{-3})}{10^{16}} \frac{10}{\varepsilon_{\text{opt}}} \frac{m_0}{m_n}} (\text{meV}). \tag{5}$$

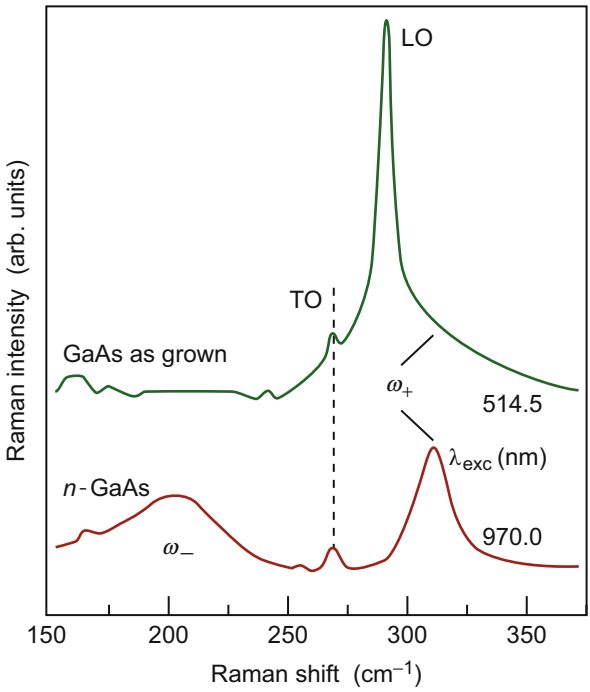


Fig. 1 Plasmon modes ω_- and ω_+ in the Raman spectrum of n-type GaAs. The nominally undoped as-grown sample is dominated by the LO resonance (After Tiginyanu et al. 1997)

Plasmons can be observed directly by IR spectroscopy or Raman scattering, when the density of free carriers exceeds $\sim 10^{15}$ cm^{-3}. Raman spectra of undoped GaAs show for the backscattering geometry applied for Fig. 1 a strong LO phonon resonance[1] (at 292 cm^{-1}) and some weak resonance of the forbidden TO phonon (at 268 cm^{-1}), while n-doped GaAs ($n = 3.5 \times 10^{17}$ cm^{-3}) shows the strong coupled plasmon modes ω_- and ω_+. Longitudinal plasmon modes can scatter with other quasi-particles such as LO phonons and electrons (Harper et al. 1973). The frequency of these plasmons increases with the square root of the electron density according to Eq. 4 and is on the order of the optical phonon frequencies at a carrier density between 10^{15} and 10^{17} cm^{-3}. Resonance effects occur which are similar to those of the photon – TO-phonon interaction leading to polaritons. However, the completely longitudinal plasma oscillation cannot interact with *transverse* electromagnetic radiation, leaving the TO phonon unaffected.

The occurrence of *two* Plasmon modes (Fig. 2a) originates from the photon-phonon interaction and a resulting typical split of the dispersion relation (shown in Fig. 2b) according to the von Neumann noncrosing rule, similar to the phonon-polariton discussed in Sect. 3.1 of ▶ Chap. 11, "Photon–Phonon Interaction." At low carrier density, the plasmon energy shows the square-root dependence on the carrier density n according to Eq. 4. As the plasmon energy approaches the energy of LO

[1]In optical spectroscopy energies are often given in units of cm^{-1}, which are related to eV units by 10 cm^{-1} ≅ 1.24 meV.

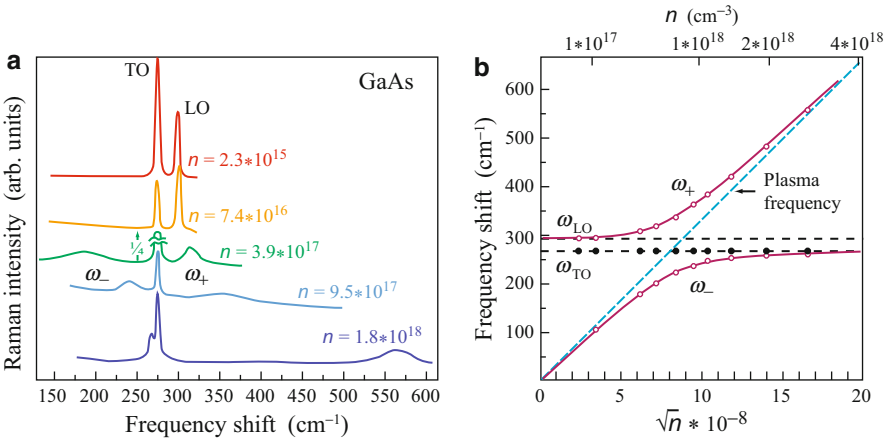


Fig. 2 (**a**) Anti-Stokes Raman scattering from plasmons coupled with LO phonons in *n*-type GaAs. The high-energy peak ω_+ shifts to higher energies, while the lower-energy peak ω_- approaches the TO peak with increasing electron density *n*. The TO peak remains at the same energy (After Mooradian and Wright 1966). (**b**) Eigenenergies of the mixed plasmon-phonon state in GaAs samples of different electron densities *n*; the TO phonon is unaffected (After Mooradian and McWhorter 1967)

phonons at higher carrier concentration, a mixed plasmon-phonon state is formed with two branches such that the dispersion curves of phonons and plasmon do not cross, see Fig. 2b.

This set of dispersion curves can be obtained from the polariton equation (Eq. 68 in ▶ Chap. 11, "Photon–Phonon Interaction"), considering plasmons instead. The LO resonance splits into two branches $\hbar\omega_+$ and $\hbar\omega_-$, given by

$$\omega_{\pm}^2 = \frac{1}{2}\left(\omega_{\mathrm{LO}}^2 + \omega_{\mathrm{p}}^2\right) \pm \frac{1}{2}\sqrt{\left(\omega_{\mathrm{LO}}^2 + \omega_{\mathrm{p}}^2\right)^2 - 4\omega_{\mathrm{p}}^2\omega_{\mathrm{TO}}^2} \tag{6}$$

and shown in Fig. 2b. The coupled mode energies can be obtained from the Raman spectra shown in Fig. 2a. The dots show the unchanged TO branch, and the circles show the two measured branches of the LO phonon as a function of the electron density, which causes the change of the plasmon frequency. The multiparticle interaction involves a phonon, a photon, and a plasmon (Patel and Slusher 1968). For more information, see Platzman and Wolff (1973); a review emphasizing recent advances is given by Foteinopoulou et al. (2019). Crystals with *wurtzite* structure are optically anisotropic; in Raman spectra of GaN substantial differences in polariton spectra for different polarizations were reported (Irmer et al. 2013).

The dielectric function of the plasmon follows from the oscillatory solution of Eq. 3; the oscillation amplitude is similar to Eq. 37 in ▶ Chap. 11, "Photon–Phonon Interaction,"

$$\Delta x = \frac{eE_x}{m_n} \quad \frac{1}{\omega_p^2 - \omega^2 - i\gamma\omega}. \tag{7}$$

For $\omega = \omega_p$ the amplitude is resonant and inverse to the damping parameter. This amplitude can be related to the polarization by

$$P_x = \frac{ne}{\varepsilon_{opt}} \Delta x. \tag{8}$$

From Eqs. 7 and 8, and Eq. 39 in ▶ Chap. 11, "Photon–Phonon Interaction," with Eq. 4, we obtain for the complex dielectric constant

$$\tilde{\varepsilon}(\omega) = \varepsilon' + i\varepsilon'' = \varepsilon_{opt} + \frac{\omega_p^2}{\omega_p^2 - \omega^2 - i\gamma\omega} \tag{9}$$

which can be separated into real and imaginary parts as

$$\varepsilon' = \varepsilon_{opt} + \omega_p^2 \left(\frac{\left(\omega_p^2 - \omega^2\right)}{\left(\omega_p^2 - \omega^2\right)^2 + \gamma^2\omega^2} \right) \tag{10}$$

and

$$\varepsilon'' = \omega_p^2 \frac{\omega\gamma}{\left(\omega_p^2 - \omega^2\right)^2 + \gamma^2\omega^2}. \tag{11}$$

Equations 10 and 11 have a form similar to Eqs. 44 and 45 in ▶ Chap. 11, "Photon–Phonon Interaction," except that the resonance frequency is given by the plasma frequency ω_p.

1.2 Valence-Electron Plasma Absorption

The plasmon absorption discussed before is caused by free electrons in the conduction band. In addition, we observe *valence-electron* plasmons when all the *valence*-electrons oscillate with respect to the cores. The mathematical theory is quite similar to that given in the previous section, except that in the expression for the plasmon frequency the density of *all* valence-electrons (typically 10^{23} cm^{-3}) is entered, m_n equals the electron rest mass, and the optical dielectric constant $\varepsilon \cong 1$. This results in a plasmon frequency of $\sim 10^{16}$ s^{-1} and a plasmon energy of $\sim$10 eV.

The valence plasmon absorption can be measured as distinct losses when electrons penetrate through a thin layer of the semiconductor. The inelastic scattering of electrons is analyzed in electron energy-loss spectroscopy (EELS)

typically carried out in a transmission-electron microscope; for a review see Egerton (2009). Figure 3 gives EELS spectra of GaN and $In_{0.5}Ga_{0.5}N$ regions in a nanowire, showing the composition dependence of the plasma peak within a selected small spot of 5–10 nm diameter. Typical plasmon energies are listed in Table 1.

Ab initio theoretical calculation of plasmon energies proved computational demanding due to the complex frequency dependence of the dynamical screened interaction potential $W(\omega)$ in the commonly applied *GW* approximation. The frequency dependence of the polarizability is hence usually simplified through an analytical model with a single pole in the so-called plasmon-pole approximation (Hybertsen and Louie 1986). Even an accurately computed bandgap energy does, however, not necessarily yield a correct description of electronic screening (Larson et al. 2013). A recent multipole model approach applying an *effective* representation of the frequency dependence of W presently offers a high level of accuracy at reasonable computational cost (Leon et al. 2021).

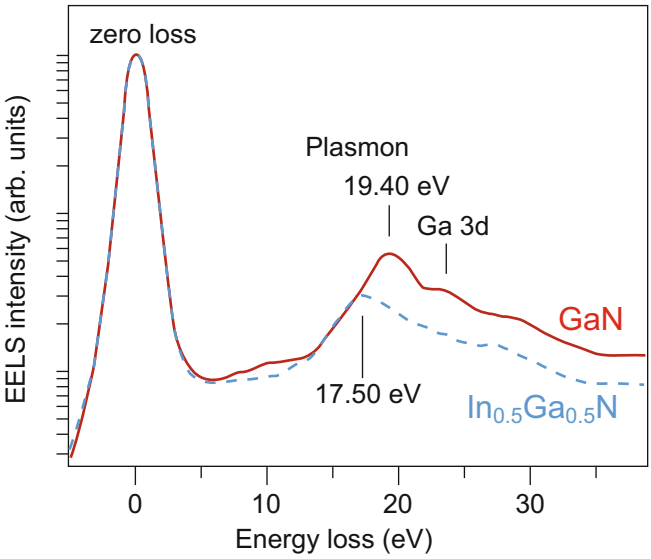


Fig. 3 Plasmon peaks in electron energy-loss spectra of GaN and $In_{0.5}Ga_{0.5}N$ (After Kong et al. 2012)

Table 1 Valence-electron plasmon energies of some semiconductors

Semiconductor	Si	Ge	GaP	GaAs	InP	InSb
$\hbar\omega_p$ (eV)	16.9	16.2	16.6	15.8	14.8	12.8

1.3 Charge-Density Waves

In materials with a high density of free electrons (metals or degenerate semiconductors) the conduction-electron charge-density, which is usually constant in a homogeneous material, can undergo a wave instability. Then the charge density becomes sinusoidally modulated in space (Overhauser 1978) with an extra periodicity not related to the lattice periodicity:

$$\rho(\mathbf{r}) = \rho_0(\mathbf{r})(1 + A\cos(\mathbf{qr} + \phi)), \tag{12}$$

where A is the amplitude (typically ~0.1 Å) and ϕ is the phase; $\mathbf{q}$ is the wavevector of the charge-density wave

$$|\mathbf{q}| \cong 2p_F/\hbar, \tag{13}$$

with p_F as the momentum at the (here assumed to be spherical) Fermi surface. This standing wave has a wavelength $\lambda = 2\pi/|\mathbf{q}|$, causing neighboring ions in the solid to experience a different potential. The lattice ions are therefore slightly displaced with an average amplitude of ~0.01 . . . 0.1 Å and with the periodicity of the charge-density wave (periodic lattice distortion, Overhauser 1978), albeit λ may be incommensurate with the ion lattice. Such waves are observed below a (sometimes high) transition temperature, e.g., in TaS_2 and $TaSe_2$, as satellites to the Bragg reflection caused by the slight lattice displacement (Wilson et al. 1975, Eichberger et al. 2010). These layered crystals consist of three-atom-thick chalcogen–transition metal–chalcogen sandwiches, but even after decades of research, the mechanism of wave formation is not well understood (Rossnagel 2011).[2]

In a scheme of strong electron–phonon coupling (a kind of *Peierls model*), the potential created by the charge-density wave and the related periodic lattice distortion opens an energy gap at the Fermi surface. The gain by lowering the energy of occupied electron states may then overcompensate the cost of increased Coulomb repulsion and elastic energy. The criterion for wave formation (Chan and Heine 1973) reads

$$\frac{4g_{\mathbf{q}}^2}{\hbar\omega_{\mathbf{q}}} - 2U_{\mathbf{q}} + V_{\mathbf{q}} \geq \chi_e^{-1}, \text{ where } \frac{4g_{\mathbf{q}}^2}{\hbar\omega_{\mathbf{q}}} > 2U_{\mathbf{q}} > V_{\mathbf{q}} > 0. \tag{14}$$

Here $g_{\mathbf{q}}$ is the electron-phonon coupling constant, $\hbar\omega_{\mathbf{q}}$ is the phonon mode $\mathbf{q}$ related to the wave, $U_{\mathbf{q}}$ is the Coulomb interaction, $V_{\mathbf{q}}$ is the screened exchange interaction, and χ_e is the electronic (non-interacting) susceptibility; a serious implied simplification is that $g_{\mathbf{q}}$, $U_{\mathbf{q}}$, and $V_{\mathbf{q}}$ are assumed to be independent of $\mathbf{k}$. According to

[2] A variety of models describe important aspects of the phase transition to charge-density waves, named Fermi surface nesting, excitonic insulator, Peierls instability, giant Kohn anomaly, or band Jahn-Teller effect.

Eq. 14, wave formation occurs for strong electron-phonon and electron-electron exchange interaction (large $g_{\mathbf{q}}$ and $V_{\mathbf{q}}$), weak Coulomb interaction and small lattice strain energy (small $U_{\mathbf{q}}$ and $\omega_{\mathbf{q}}$), and for large susceptibility (large $\chi_e(\mathbf{q})$). The criterion captures essential aspects; e.g., the ground state of charge-density waves generally occurs in systems with strong electron–phonon interaction. However, in real solids, the energy gain is spread over a substantial part of the Brillouin zone, and Eq. 14 is basically not valid.

In the *excitonic insulator model*, the driving force for wave formation is the electron-hole attraction, inducing a spontaneous formation of excitons if their binding energy equals the narrow bandgap of an indirect semiconductor. The result of Peierls and excitonic insulator instabilities is qualitatively the same and mathematically similar; in either case, a novel ground state with broken translational symmetry is created, exhibiting a charge-density wave, a phonon lattice distortion, and a gap in the one-electron spectrum. For a detailed discussion, see Rossnagel (2011). The dynamics of charge-density waves is discussed by Storeck et al. (2020).

Charge-density waves occur in various phase transitions. There is an intimate relation between charge-density waves and superconductivity, presently a field of lively debate (Loret et al. 2019, Lian et al. 2019, Yang et al. 2018). Studies show also a close relation to the quantum Hall effect (Tang et al. 2019, Qin et al. 2020).

A phase modulation, when quantized, can yield low-energy collective excitation spectra with quanta $\hbar\omega$ from

$$\phi = \phi(\mathbf{r}, t) \propto \sin(\mathbf{qr} - \omega t). \tag{15}$$

These are called *phasons* and are observed in $LaGe_2$ and also in twisted transition-metal dichalcogenide bilayers (Maity et al. 2020). They could have a measurable effect on transport of lattice energy (Manley et al. 2018) and, at low temperature, on electron scattering (Huberman and Overhauser 1982).

2 Nonresonant Free-Carrier Absorption

Far away from the resonance transition at $\omega = \omega_p$, we observe the *nonresonant* part of the absorption with a much reduced amplitude, tailing from the resonance peak through the extrinsic optical absorption range and extending toward the band edge.

2.1 Dispersion Relation for Free Carriers

The optical absorption and reflection induced by free carriers outside of the resonance absorption provide valuable information about carrier relaxation and, for holes, about the band structure near $\mathbf{k} = 0$. The relationship between free carriers and an external field is the same as developed in Sect. 1.1 (Lax 1963), except that here the restoring forces are negligible and electrons respond only to external electromagnetic forces and damping:

$$m_n\left(\frac{d^2\Delta x}{dt^2}-\gamma\frac{d\Delta x}{dt}\right)=-eE_x\exp(-i\omega t). \tag{16}$$

This equation of motion was introduced for the classical treatment of free-carrier dispersion (Drude model) by Drude (1900); it provides an independent opportunity to measure the electron scattering by optical absorption and will be discussed in more detail in the following section. The damping parameter γ can be related to the energy-relaxation time $\tau_{\rm el}$ for a free electron by $\gamma = 1/\tau_{\rm el}$. Equation 16 has the solution

$$\Delta x=-\frac{eE}{m_n}\;\frac{1}{\omega^2-i\omega/\tau_{\rm el}}. \tag{17}$$

The electric polarization P arising from bound *and* free electrons is then

$$P=\varepsilon_0\big(\varepsilon_{\rm opt}-1\big)E+en\,\Delta x. \tag{18}$$

The complex dielectric function is related to P by $P=\varepsilon_0(\widetilde{\varepsilon}-1)E$. Using this, we obtain from Eq. 18

$$\widetilde{\varepsilon}(\omega)=\varepsilon_{\rm opt}\left(1-\frac{\omega_{\rm p}^2}{\omega^2+i\omega/\tau_{\rm el}}\right), \tag{19}$$

where $\omega_{\rm p}$ is given by Eq. 4. This dielectric function describes the response of free electrons on an external electromagnetic field. Equation 19 can be separated into real and imaginary parts as

$$\varepsilon'=n_{\rm r}^2-\kappa^2=\varepsilon_{\rm opt}\left(1-\frac{\omega_{\rm p}^2}{\omega^2}\;\frac{\omega^2\tau_{\rm el}^2}{1+\omega^2\tau_{\rm el}^2}\right) \tag{20}$$

and

$$\varepsilon''=2n_{\rm r}\kappa=\varepsilon_{\rm opt}\frac{\omega_{\rm p}^2}{\omega^2}\;\frac{\omega\tau_{\rm el}}{1+\omega^2\tau_{\rm el}^2}. \tag{21}$$

Equations 20 and 21 have a form similar to the resonance case of Eqs. 44 and 45 in ▶ Chap. 11, "Photon–Phonon Interaction," except that the resonance frequency here is zero. Free carriers give a *negative* contribution to ε' and, hence, to the index of refraction $n_{\rm r}$.

There is another approach to the free-carrier dielectric function and optical absorption. Is is used here for educational purposes in order to arrive directly at the complex conductivity (see Sect. 1.1.4 in ▶ Chap. 10, "Interaction of Light with Solids"):

$$\frac{d\Delta x}{dt} = v_x = -\frac{eE}{m_n}\frac{\tau_{\mathrm{el}}}{1 - i\omega\tau_{\mathrm{el}}}. \tag{22}$$

This velocity can be used to define an electric current density[3]

$$\tilde{j}_x = -e\,n\,\tilde{v}_x = \tilde{\sigma}E, \tag{23}$$

with a complex conductivity

$$\tilde{\sigma} = \sigma' + i\sigma'' = \sigma_0\left(\frac{1}{1+\omega^2\tau_{\mathrm{el}}^2} + i\frac{\omega\tau_{\mathrm{el}}}{1+\omega^2\tau_{\mathrm{el}}^2}\right), \tag{24}$$

with $\sigma_0 = e^2 n\tau_{\mathrm{el}}/m_n$ as the dc conductivity. Introducing Eq. 24 into $\mathbf{j}(\omega)$ given in Eq. 26 of ▶ Chap. 10, "Interaction of Light with Solids," we obtain with $\sigma_0 = \omega_{\mathrm{p}}^2\tau_{\mathrm{el}}\varepsilon_0\varepsilon_{\mathrm{opt}}$ again (Eqs. 20 and 21).

We can experimentally obtain ω_{p} and $\varepsilon_{\mathrm{opt}}$ by measuring the reflectance. In the frequency range where $n_{\mathrm{r}}^2 \gg \kappa^2$ we obtain from Eq. 54 in ▶ Chap. 10, "Interaction of Light with Solids,"

$$R_0 \cong \frac{(n_{\mathrm{r}}-1)^2}{(n_{\mathrm{r}}+1)^2} \text{ or } n_{\mathrm{r}} \cong \frac{1+\sqrt{R_0}}{1-\sqrt{R_0}}. \tag{25}$$

The index of refraction as a function of ω for the condition $n_{\mathrm{r}}^2 \gg \kappa^2$ refers to $\omega\tau_{\mathrm{el}} \gg 1$, yielding from Eq. 20

$$n_{\mathrm{r}}^2 \cong \varepsilon_{\mathrm{opt}}\left(1 - \frac{\omega_{\mathrm{p}}^2}{\omega^2}\right) = \varepsilon_{\mathrm{opt}} - \frac{\lambda^2}{4\pi^2}\frac{e^2}{c^2\varepsilon_0^2 m_n}. \tag{26}$$

In the last expression, we have used $\omega = 2\pi c/\lambda$. The frequencies $\omega_{\min}$ and $\omega_{\max}$ at which the reflection minimum and maximum occur are those for which $R_0 \ll 1$ (or $n_{\mathrm{r}} \cong 1$) and $R_0 \cong 1$ (or $n_{\mathrm{r}} \cong 0$), respectively; they are related to ω_{p} and $\varepsilon_{\mathrm{opt}}$ using Eq. 26 by

$$\omega_{\min} \cong \omega_{\mathrm{p}}\sqrt{\frac{\varepsilon_{\mathrm{opt}}}{\varepsilon_{\mathrm{opt}}-1}} \text{ and } \omega_{\max} \cong \omega_{\mathrm{p}}. \tag{27}$$

With known $\varepsilon_{\mathrm{opt}}$, we have a useful method for determining ω_{p}, and with it the effective carrier mass in ω_{p} (Eq. 4), provided the carrier density is known by

[3]This is an ac current at the frequency ω which, for higher ω (i.e., for $\omega \geq \tau_{\mathrm{el}}^{-1}$), is substantially different from the dc current, as indicated in Sect. 1.1.4. of ▶ Chap. 10, "Interaction of Light with Solids."

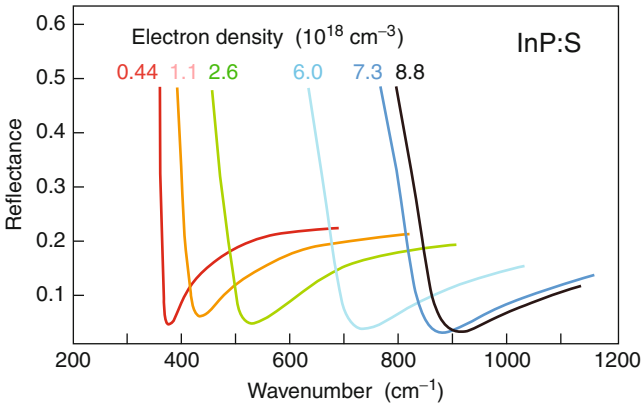


Fig. 4 Infrared reflectance of *n*-type InP with various doping densities (After Kim and Bonner 1983)

independent means. We can vice versa extract the carrier density from reflection spectra if the effective mass is known. Typical IR reflection spectra observed for *n*-type InP are shown in Fig. 4, with reflection minima shifting to higher energies with increasing carrier density according to Eqs. 4 and 27.

To conclude, when the concentration of free carriers is sufficiently high, the reflection in the IR range between the wavelength corresponding to the bandgap and the Reststrahlen wavelength is strongly influenced by free carriers. From an analysis of the reflection spectra, we can obtain the plasma frequency. For a review of optical characterization of free carriers in semiconductors see Palik and Holm (1979). For the dynamics of free carriers see Elsaesser and Woerner (1999).

2.2 Free-Electron Absorption

The optical absorption coefficient due to free electrons is obtained from Eq. 24 in ▶ Chap. 10, "Interaction of Light with Solids," and Eq. 21:

$$\alpha_o = \frac{2\omega\kappa}{c} = \frac{\omega}{c}\,\frac{\varepsilon''}{n_r} = \frac{\omega_p^2}{n_r\ c}\ \ \frac{\tau_{el}}{1+\omega^2\tau_{el}^2}. \tag{28}$$

The strength of the absorption is $\propto \omega_p^2$. For high enough frequencies ($\omega\tau_{el} \gg 1$), Eq. 28 becomes the classical $\alpha_o \propto \lambda^2$ relation

$$\alpha_o = \frac{\omega_p^2}{\omega^2}\ \ \frac{1}{n_r\, c\tau_{el}} = \frac{\omega_p^2\lambda^2}{4\pi^2\, n_r\, c^3\tau_{el}}. \tag{29}$$

$$= 3 \times 10^{-4} \frac{10}{\varepsilon_{\text{opt}}} \frac{m_0}{m_n} \frac{n}{10^{16}(\text{cm}^{-3})} \frac{\lambda^2(\mu\text{m})^2}{n_\text{r}} \frac{10^{-13}\text{s}}{\tau_{\text{el}}}.$$

Such a behavior is often observed for the free-electron absorption in the extrinsic range beyond the band edge. The free-electron absorption is a continuous absorption tailing off from the lattice resonances toward the band edge (Fig. 5). This absorption is rather small near the absorption edge for semiconductors with normal electron densities and becomes observable in the IR only at higher levels of doping ($n > 10^{17}$ cm^{-3} – see Fig. 5a) when the probing frequency is closer to the plasma frequency – see Sect. 1.1.

In the following section, we will show that this λ^2 relation is modified by specific scattering mechanisms of the carriers when this scattering depends on the energy of the carrier within the band; here $\tau_{\text{el}} = \tau_{\text{el}}(\lambda)$, and α_o becomes a more complicated function of λ.

2.2.1 Effect of Scattering Mechanisms on Free-Electron Absorption

Examining the electronic band structure $E(\mathbf{k})$ of typical conduction bands, we recognize that direct optical transitions of electrons in the conduction bands, which are essentially vertical in $E(\mathbf{k})$, are rather rare events. They are restricted to transitions at specific values of $\mathbf{k}$, which find an allowed $E(\mathbf{k})$ for the given photon energy (e.g., $h\nu_{\text{dir}}(3)$ in Fig. 6) *and* a sufficiently large occupation probability of the ground state from which such a transition (3) starts. Only from the minimum of the

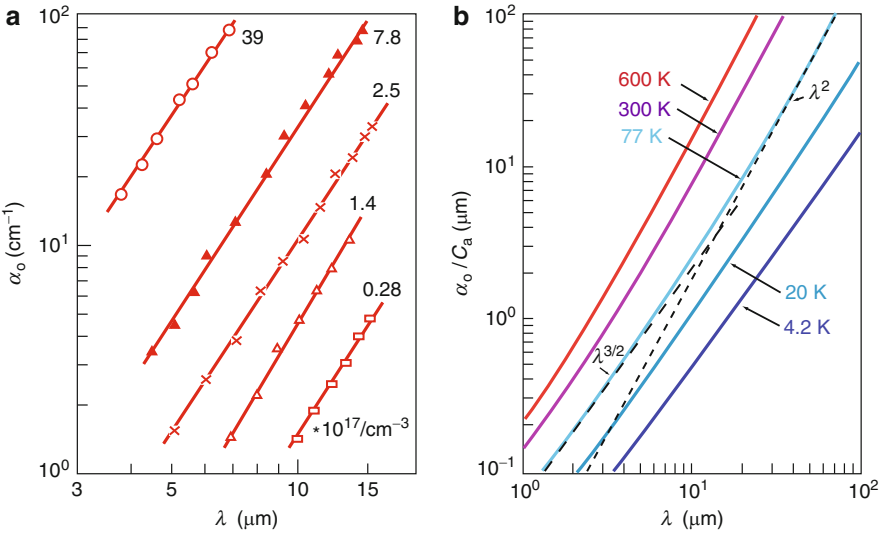


Fig. 5 (**a**) Optical absorption of free electrons with predominant acoustic phonon scattering beyond the band edge of *n*-type InAs at room temperature for various electron densities (After Dixon 1960). (**b**) Optical absorption of free electrons as in (**a**), at various temperatures, with C_a (μm^2) a proportionality factor (After Seeger 1973)

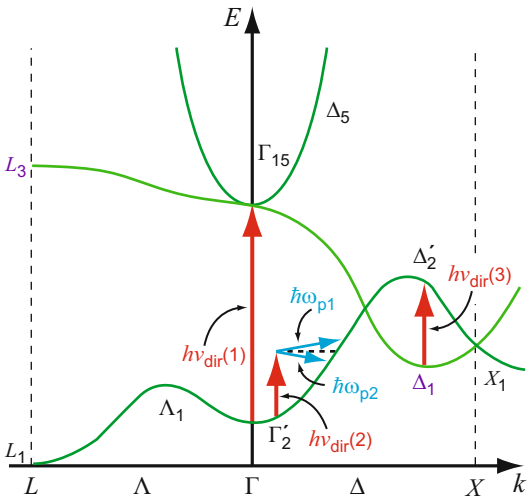


Fig. 6 Direct ($h\nu_{\mathrm{dir}}$) and indirect transitions ($h\nu_{\mathrm{indir}} = h\nu_{\mathrm{dir}} \pm \hbar\omega_{\mathrm{p}}$) for free electrons within conduction bands. The indirect transition requires, in addition to the absorption of a photon $h\nu_{\mathrm{dir}}$, the absorption ($\hbar\omega_{\mathrm{p1}}$) or emission ($\hbar\omega_{\mathrm{p2}}$) of a phonon of the proper energy and momentum

conduction band to a band above it are such transitions plentiful (at the Γ-point with $h\nu_{\mathrm{dir}}$ (1) in Fig. 6). For most other photon energies, only indirect transitions are possible – see Sect. 2 in ▶ Chap. 13, "Band-to-Band Transitions." They require the supply or emission of a phonon in addition to the absorption of the photon, as shown in Fig. 6 as transition $h\nu_{\mathrm{indir}}(2) = h\nu_{\mathrm{dir}}(2) \pm \hbar\omega_{\mathrm{p}}$.

The transition probability for such indirect transitions can be calculated from a second-order perturbation theory (see Sakurai and Napolitano 2011) including the Hamiltonian for a photon and the Hamiltonian for a phonon transition. This is necessitated by the fact that three particles are involved in the transition: a photon, an electron, and a phonon or a lattice defect. The interaction is much like that of the scattering of an electron with a phonon: the electron is first elevated by the photon to a virtual state within the conduction band.

The mathematical formulation of the theory for the scattering with different types of phonons is similar to the analysis of the electron scattering for carrier transport (▶ Chap. 23, "Carrier Scattering at Low Electric Fields") or for higher energies within the band (▶ Chap. 24, "Carrier Scattering at High Electric Fields" – see also Seeger 2004). In a first approximation, the optical absorption coefficient can be expressed by the classical free-electron absorption α_{o} Eq. 29 and by replacing the electron relaxation time τ_{el} with the one relating to the predominant scattering mechanism. This permits the replacement of the yet unknown τ_{el} in Eq. 29 with an appropriate expression (see Seeger 2004).

For instance, we obtain for predominant *acoustic deformation potential scattering* with a phonon energy being much smaller than the photon energy $\hbar\omega_q \ll h\nu$:

$$\alpha_{\mathrm{o,ac},E} \cong \alpha_{\mathrm{o,ac}} \frac{h\nu}{4\ kT} \sqrt{\frac{T}{T_{\mathrm{el}}}} \sinh\left(\frac{h\nu}{2\ kT_{\mathrm{el}}}\right) K_2\left(\frac{h\nu}{2\ kT_{\mathrm{el}}}\right), \tag{30}$$

where T_{el} is the electron temperature[4] and $h\nu$ is the energy of the absorbed photon; K_2 is a modified Bessel function (Poole 1998).

When only small deviations from thermal equilibrium are considered ($T_{\mathrm{el}} \cong T$) and the photon energy is small compared to the thermal energy ($h\nu \ll kT$), then the factor following $\alpha_{\mathrm{o,ac}}$ approaches 1 and $\alpha_{\mathrm{o,ac},E} \cong \alpha_{\mathrm{o,ac}}$ with

$$\alpha_{\mathrm{o,ac}} = \alpha_{\mathrm{o}} \frac{2^{7/2}(m_n kT)^{3/2} \Xi_{\mathrm{c}}^2}{3\pi^{3/2}\hbar^4 C_l} \tag{31}$$

and α_{o} given by Eq. 29. Here Ξ_{c} is the deformation potential for the conduction band, and C_l is the appropriate elastic stiffness constant. Also, $\alpha_{\mathrm{o,ac}}$ shows the classical λ^2 behavior (contained in α_{o}) as indicated in Fig. 5a.

For photons with higher energies (for $h\nu \gg kT$), however, one obtains

$$\alpha_{\mathrm{o,ac},E} \cong \frac{\alpha_{\mathrm{o,ac}}}{2} \sqrt{\frac{\pi}{2}\frac{h\nu}{kT}} \propto \lambda^{3/2}, \tag{32}$$

which gives a somewhat lower slope (when the emission of many phonons becomes possible), which is indicated in Fig. 5b.

There are substantial differences for the scattering of electrons which are excited into higher band states. From here, scattering is greatly enhanced by rapid generation of LO phonons (Sect. 3 in ▶ Chap. 24, "Carrier Scattering at High Electric Fields"). This scattering can be described as the scattering of a *heated electron gas* with T_{el} much larger than T.

The absorption coefficients for other phonon-scattering mechanisms show a wavelength dependence similar to the acoustic one for long-wavelength excitation. However, at low temperatures, it becomes nonmonotonic as soon as the electron energy exceeds that of the rapidly generated optical phonons (Fig. 7a). At low temperature LO phonons are frozen-out; a low scattering and therefore a low absorption results. When the electrons gain energy above $\hbar\omega_{\mathrm{LO}}$ with decreasing λ of the exciting light, the relaxation time decreases dramatically, and therefore $\alpha_{\mathrm{o,opt},E}$ increases. At high temperatures this resonance is hidden since there are enough electrons thermally excited near $\hbar\omega_{\mathrm{LO}}$.

The maximum is due to resonance scattering when the electron energy within the band equals that of the LO phonon. This nonmonotonic behavior is observed for *optical deformation potential scattering* as well as for *polar optical*

[4]An elevated (above the lattice temperature) electron temperature is used here to indicate the occupancy of states higher in the conduction band by electrons.

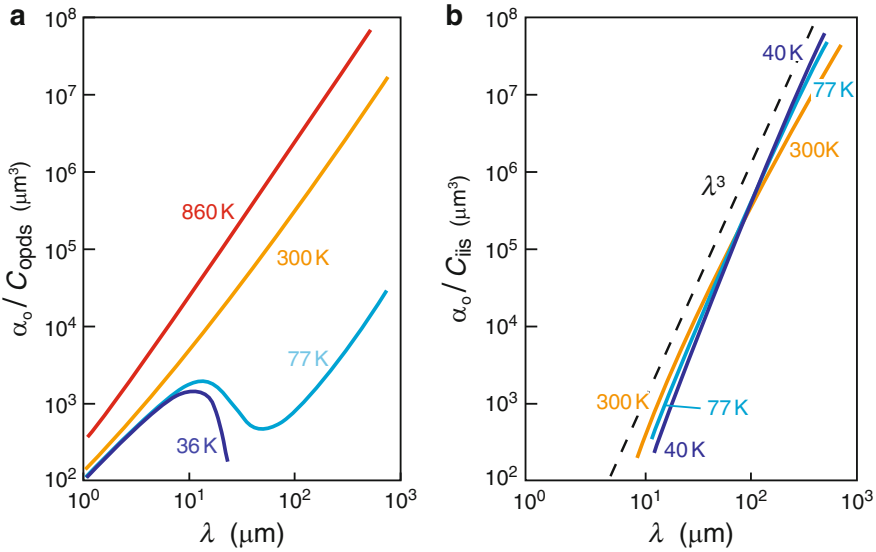


Fig. 7 (**a**) Optical absorption of free electrons computed for predominant optical deformation potential scattering with $\Theta = 720$ K and C_{opds} a proportionality factor of dimension cm^{-2} at various temperatures. (**b**) Optical absorption of free electrons computed for ionized impurity scattering with C_{iis} a proportionality factor of dimension cm^{-2} at various temperatures (After Seeger 1973)

scattering, for which the wavelength dependence increases from λ^2 to $\lambda^{2.5}$ for higher energies.

For lower temperatures ($kT < \hbar\omega_{LO}/2$) near thermal equilibrium ($T_{el} = T$) one obtains for the ratio of optical and acoustic deformation potential scatterings:

$$\frac{\alpha_{o,\,opt}}{\alpha_{o,\,ac}} = \frac{4}{\sqrt{\pi}}\left(\frac{D_o v_l}{\Xi_c \omega_0}\right)^2 \sqrt{\frac{T}{\Theta}}, \tag{33}$$

where D_o and Ξ_c are the optical and acoustic deformation potential constants, and v_l is the longitudinal sound velocity. Here Θ is the Debye temperature.

Ionized impurity scattering (Wolfe 1954) causes a stronger dependence on the wavelength, which for higher energies approaches λ^3:

$$\alpha_{o,\,ion} \propto \frac{N_{ion} Z_{ion} e^4}{3\sqrt{2}\pi^{3/2}(\varepsilon\varepsilon_0)^2 (mkT_{el})^{1/2} h\nu} \sinh\left(\frac{h\nu}{2\ kT_{el}}\right) K_0\left(\frac{h\nu}{2\ kT_{el}}\right) \propto \lambda^3, \tag{34}$$

where K_0 is a modified Bessel function (Poole 1998). The computed absorption behavior for scattered ionized impurities is shown in Fig. 7b.

With different scattering mechanisms, the optical absorption coefficients for each one of them is added (the inverse of the relaxation times are added – see Eq. 29) in order to obtain the total optical absorption:

$$\alpha_{o,\,\text{total}} = \sum_i \alpha_{o,\,i}. \tag{35}$$

When studying a family of absorption curves at different temperatures, we can obtain some information about the predominant scattering mechanism, especially about scattering at higher electron energies for semiconductors with sufficient electron densities.

2.2.2 Free-Hole Absorption

The absorptions of free electrons and holes follow similar principles. However, although most electron transitions within the conduction band are indirect, requiring the emission or absorption of phonons (Fan 1967), the most prevalent long-wavelength transitions for holes are direct, as shown in Fig. 8. Therefore, under otherwise similar conditions, the absorption coefficient for holes is much larger than that for electrons.

Closely spaced valence bands (Fig. 8a) provide possibilities for a wide spectrum of direct transitions at very long wavelengths. In addition, the change from heavy to light holes, detected by simultaneous conductivity measurements, assists in identifying the specific feature in the absorption spectrum.

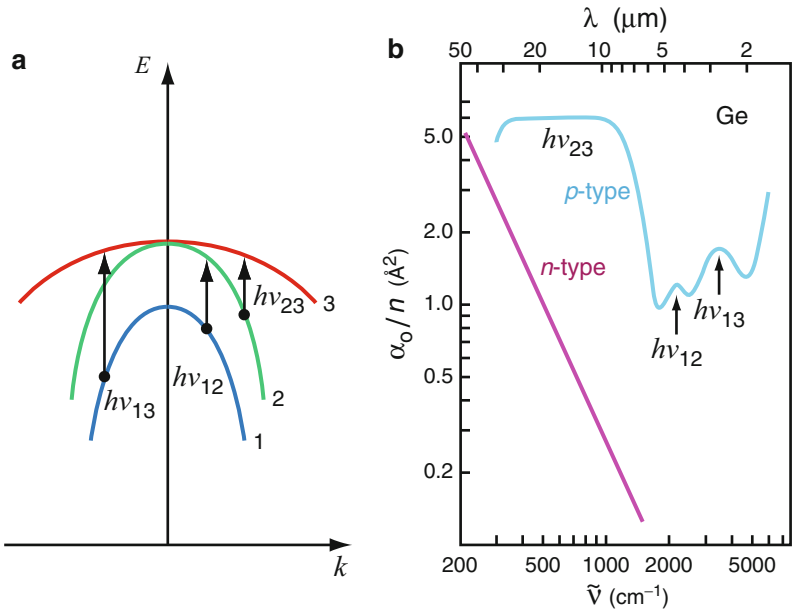


Fig. 8 (**a**) Schematic of intervalence-band transitions. (**b**) Optical absorption cross-section as a function of wavelength (wavenumbers) for *n*- and *p*-type Ge (After Kaiser et al. 1953)

Steps in the optical absorption spectrum are observed when the energy of the holes is sufficient to permit optical phonon relaxation. Such a step was calculated for a single threshold in the electron energy for $\hbar\omega_{\mathrm{LO}}$ emission (see section above and Fig. 7a).

3 Carrier Dispersion in Electric and Magnetic Fields

The dispersion equation with an additional magnetic field is obtained from Maxwell's equations, exactly as in Sect. 1.1.3 from Eq. 17 in ▶ Chap. 10, "Interaction of Light with Solids," except that the magnetic field causes a change in wave propagation or polarization and therefore requires a vector relation for the amplitude function $\mathbf{E}(x)$, resulting in

$$\frac{\partial^2}{\partial \mathbf{r}^2}\mathbf{E} + \frac{\omega^2}{c^2}\left(\underline{\boldsymbol{\varepsilon}} + \frac{i}{\varepsilon_0\,\omega}\underline{\boldsymbol{\sigma}}(\mathbf{B})\right)\ \mathbf{E} = 0. \tag{36}$$

where $\underline{\boldsymbol{\varepsilon}}$ and $\underline{\boldsymbol{\sigma}}$ are the *dielectric* and the *conductivity matrices*, respectively. The dielectric matrix is not influenced by the magnetic field. For isotropic crystals, we can simply use the dielectric constant, requiring, however, matrix notation to permit matrix calculus. For a magnetic induction in the z direction, the $\underline{\boldsymbol{\sigma}}$ matrix simplifies to

$$\underline{\boldsymbol{\sigma}} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \\ \sigma_{yx} & \sigma_{yy} & \\ & & \sigma_{zz} \end{pmatrix}. \tag{37}$$

The components of this matrix are obtained from the equation of motion, which is identical to Eq. 16, except for the addition of the Lorentz force $e(\mathbf{v}_{\mathrm{drift}} \times \mathbf{B})$:

$$m_n\left(\frac{d\mathbf{v}_{\mathrm{drift}}}{dt} + \frac{\mathbf{v}_{\mathrm{drift}}}{\tau_{\mathrm{mag}}}\right) = e(\mathbf{E} + \mathbf{v}_{\mathrm{drift}} \times \mathbf{B}). \tag{38}$$

We assume an isotropic effective mass m_n and an isotropic relaxation time τ_{mag} independent on the velocity. With $\mathbf{E} = \mathbf{E}(x)\exp(-i\omega t)$ and $en\mathbf{v}_{\mathrm{drift}} = \underline{\boldsymbol{\sigma}}(B)\mathbf{E}$, we obtain

$$\sigma_{xx} = \sigma_{yy} = \sigma_0 \frac{1 - i\omega\tau_{\mathrm{mag}}}{\left(1 - i\omega\tau_{\mathrm{mag}}\right)^2 + \left(\omega_{\mathrm{c}}\,\tau_{\mathrm{mag}}\right)^2} \tag{39}$$

$$\sigma_{xy} = -\sigma_{yx} = \sigma_0 \frac{\omega_{\mathrm{c}}\,\tau_{\mathrm{mag}}}{\left(1 - i\omega\tau_{\mathrm{mag}}\right)^2 + \left(\omega_{\mathrm{c}}\,\tau_{\mathrm{mag}}\right)^2} \tag{40}$$

$$\sigma_{zz} = \sigma_0 \frac{1}{1 - i\omega\tau_{\text{mag}}}, \tag{41}$$

with $\omega_c = eB/m_n$ as the cyclotron frequency and $\sigma_0 = e^2 n\tau_{\text{mag}}/m_n = en\mu$ as the dc conductivity. Given the magnetic induction $\mathbf{B} = (0, 0, B_z)$, we can distinguish from the general dispersion equation[5]

$$k^2\mathbf{E} = \frac{\omega^2}{c^2}\underline{\boldsymbol{\varepsilon}}\mathbf{E} + i\frac{\omega}{\varepsilon_0 c^2}\underline{\boldsymbol{\sigma}}(B)\mathbf{E}, \tag{42}$$

a number of cases depending on the relative orientation of the propagation (**k**) and polarization (**E**) of the light interacting with the semiconductor. These cases are identified in Table 2.

For the *longitudinal case* one has

$$k_z^2 E_x = \frac{\omega^2}{c^2}\varepsilon_{\text{opt}} E_x + i\frac{\omega}{\varepsilon_0 c^2}\left(\sigma_{xx}E_x + \sigma_{xy}E_y\right). \tag{43}$$

$$k_z^2 E_y = \frac{\omega^2}{c^2}\varepsilon_{\text{opt}} E_y + i\frac{\omega}{\varepsilon_0 c^2}\left(-\sigma_{xy}E_x + \sigma_{xx}E_y\right). \tag{44}$$

For *circular polarizarized light* ($E_x \pm i\,E_y$) we obtain by using $k_z^2(\omega/c)^2(n + i\kappa)^2$ and adding Eq. 43 $\pm\, i \times$ Eq. 44

$$(n_\pm + i\kappa_\pm)^2 = \varepsilon_{\text{opt}} + i\frac{\sigma_\mp}{\varepsilon_0\,\omega} \quad \text{with} \quad \sigma_\mp = \sigma_{xx} \mp i\sigma_{xy}. \tag{45}$$

After separation of real and imaginary parts we arrive at

Table 2 Various cases for dispersion relations discussed in the text, depending on the relative orientation of **B**, **k**, and **E**

Magnetic induction	$\mathbf{B} = (0, 0, B_z)$	
Wave propagation	Longitudinal Faraday configuration $\mathbf{k} = (0, 0, k_z)$	Transverse Voigt configuration $\mathbf{k} = (0, k_y, 0)$
Electric vector linear Linear Circular	$\mathbf{E} = (0, E_y, 0)$ $\mathbf{E} = (E_x, 0, 0)$ $\mathbf{E} = (E_x \pm i\,E_y)$	$\mathbf{E} = (0, 0, E_z)$ $\mathbf{E} = (E_x, 0, 0)$

[5]Obtained from Eq. 36 with $\mathbf{E} = \mathbf{E}_0 \exp[i(\mathbf{k}\mathbf{r} - \omega t)]$.

$$\varepsilon'_{\pm} = \left(n_{\pm}^2 - \kappa_{\pm}^2\right) = \varepsilon_{\text{opt}}\left(1 - \frac{\omega_{\text{p}}^2\,\tau_{\text{mag}}^2}{\omega}\,\frac{\omega \pm \omega_{\text{c}}}{\left[(\omega \pm \omega_{\text{c}})\ \tau_{\text{mag}}\right]^2 + 1}\right) \tag{46}$$

$$\varepsilon''_{\pm} = 2n_{\pm}\kappa_{\pm} = \varepsilon_{\text{opt}}\,\frac{1}{\omega}\,\frac{\omega_{\text{p}}^2\,\tau_{\text{mag}}}{\left[(\omega \pm \omega_{\text{c}})\ \tau_{\text{mag}}\right]^2 + 1}. \tag{47}$$

with the + or – sign for right- or left-polarized light, respectively, and ω_{p} the plasma frequency (Eq. 4).

For the *transverse case* there are three dispersion equations:

$$k_y^2 E_x = \frac{\omega^2}{c^2}\,\varepsilon_{\text{opt}} E_x + i\,\frac{\omega}{\varepsilon_0\,c^2}\left(\sigma_{xx}E_x + \sigma_{xy}E_y\right), \tag{48}$$

$$0 = \frac{\omega^2}{c^2}\,\varepsilon_{\text{opt}} E_y + i\,\frac{\omega}{\varepsilon_0\,c^2}\left(-\sigma_{xy}E_x + \sigma_{xx}E_y\right), \tag{49}$$

$$k_y^2 E_z = \frac{\omega^2}{c^2}\,\varepsilon_{\text{opt}} E_z + i\,\frac{\omega}{\varepsilon_0\,c^2}\,\sigma_{zz}E_z, \tag{50}$$

which yield for *parallel polarization* (**E**||**B**)

$$\widetilde{\varepsilon}_{||}\left(n_{||} + i\kappa_{||}\right)^2 = \varepsilon_{\text{opt}} + i\,\frac{\sigma_{zz}}{\varepsilon_0\,\omega}, \tag{51}$$

with

$$\varepsilon' = \varepsilon_{\text{opt}}\left(1 - \frac{\left(\omega_{\text{p}}\,\tau_{\text{mag}}\right)^2}{\left(\omega\,\tau_{\text{mag}}\right)^2 + 1}\right) \text{and}\, \varepsilon'' = \varepsilon_{\text{opt}}\,\frac{1}{\omega}\,\frac{\omega_{\text{p}}^2\,\tau_{\text{mag}}}{\left(\omega\,\tau_{\text{mag}}\right)^2 + 1}. \tag{52}$$

For the *perpendicular polarization* (**E**⊥**B**) we have

$$\widetilde{\varepsilon}_{\perp}(n_{\perp} + i\kappa_{\perp})^2 = \varepsilon_{\text{opt}} + i\,\frac{1}{\varepsilon_0\,\omega}\left(\sigma_{xx} + i\,\frac{\sigma_{xy}^2}{\varepsilon_{\text{opt}}\varepsilon_0\omega + i\sigma_{xx}}\right) \tag{53}$$

with

$$\varepsilon' = \varepsilon_{\text{opt}}\left(1 - \frac{\left(\omega_{\text{p}}\,\tau_{\text{mag}}\right)^2\ \beta}{\left(\omega\,\tau_{\text{mag}}\right)^2\beta^2 + \alpha^2}\right) \text{and}\ \varepsilon'' = \varepsilon_{\text{opt}}\,\frac{1}{\omega}\,\frac{\omega_{\text{p}}^2\,\tau_{\text{mag}}\,\alpha}{\left(\omega\,\tau_{\text{mag}}\right)^2\beta^2 + \alpha^2} \tag{54}$$

and with the auxiliary functions

$$\alpha = 1 + \frac{\omega^2 \tau_{\mathrm{mag}}^2 \omega_{\mathrm{c}}^2}{\omega^2 + \left[\left(\omega^2 - \omega_{\mathrm{p}}^2\right)\ \tau_{\mathrm{mag}}\right]^2} \text{ and } \beta = 1 - \frac{\left(\omega^2 - \omega_{\mathrm{p}}^2\right) \tau_{\mathrm{mag}}^2 \omega_{\mathrm{c}}^2}{\omega^2 + \left[\left(\omega^2 - \omega_{\mathrm{p}}^2\right)\ \tau_{\mathrm{mag}}\right]^2}. \quad (55)$$

There are two characteristic frequencies entering the dispersion equation: the plasma (ω_{p}) and the cyclotron (ω_{c}) frequencies, which determine possible resonances. The damping is determined by the appropriate relaxation time and gives the width of the resonance peak. These equations describe all possible interactions of electromagnetic radiation with a semiconductor while exposed to a dc magnetic induction. Examples for such interactions are cyclotron resonance, magnetoplasma reflection, and the Faraday and Voigt effects. They will be discussed in the following sections. For more detail, see Madelung (1978), Roth (1982), Basu (1997), and Sugano and Kojima (2000).

3.1 Magnetoplasma Reflection

A relatively simple means to determine one or both of the characteristic frequencies ω_{c} and ω_{p} – and thereby the effective mass, carrier type, and carrier density – is to measure the spectral distribution of the reflectivity with $\mathbf{k} \| \mathbf{B}$ or $\mathbf{k} \perp \mathbf{B}$, i.e., for Faraday or Voigt configuration, respectively. The effect in reflectivity can be derived by substituting the magneto-optical constants $n_{\pm}$, $\kappa_{\pm}$, $n_{\|}$, $n_{\perp}$, $\kappa_{\|}$, and $\kappa_{\perp}$ in the Fresnel formula for the reflection coefficient (Eq. 54 in ▶ Chap. 10, "Interaction of Light with Solids"):

$$R_0 = R_{\perp} = R_{\|} = \frac{\left(n_{\mathrm{r}} - 1\right)^2 + \kappa^2}{\left(n_{\mathrm{r}} + 1\right)^2 + \kappa^2}. \quad (56)$$

Faraday Configuration Arranging the propagation of the interacting light parallel to the magnetic induction and using circular polarized light yields a reflectivity

$$R_{\pm} = \frac{\left(n_{\pm} - 1\right)^2 + \kappa_{\pm}^2}{\left(n_{\pm} + 1\right)^2 + \kappa_{\pm}^2}. \quad (57)$$

For $n_{+}^2 \gg \kappa_{-}^2$, the reflectivity shows a maximum (see Eq. 46) when the index of refraction tends to vanish ($n_{\pm} \cong 0$) – that is, when

$$\omega \cong \omega_{\mathrm{p}} \pm \frac{1}{2}\omega_{\mathrm{c}} + \frac{\sqrt{\varepsilon_{\mathrm{opt}}}}{8}\ \frac{\omega_{\mathrm{c}}^2}{\omega_{\mathrm{p}}} \quad (58)$$

for left (−) or right (+) circular polarized light. Consequently, a shift in the plasma edge by $\pm\, \omega_{\mathrm{c}}/2$ is seen (when $\omega_{\mathrm{p}} \gg \omega_{\mathrm{c}}$) when a magnetic induction is applied. This

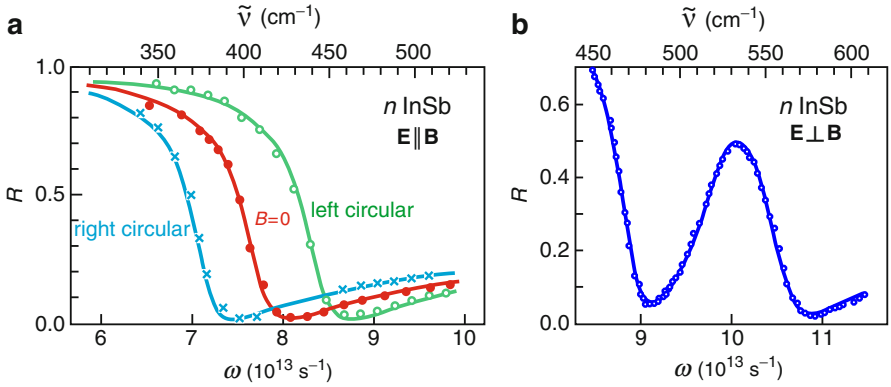


Fig. 9 (**a**) Shift of the plasma absorption edge in *n*-type InSb at room temperature for $n = 10^{18}$ cm^{-3}, $\tau_{el} = 2.8 \cdot 10^{-13}$s, and $m_n/m_0 = 0.035$ at $\mathbf{B} = 0$ and ± 2.54 T (Faraday configuration); the curves are calculated (After Palik et al. 1962). (**b**) Transverse magnetoplasma reflection (Voigt configuration) in *n*-type InSb at room temperature for $n = 1.8 \cdot 10^{18}$ cm^{-3}, $\tau_{el} = 3.6 \cdot 10^{-13}$s, and $B = 3.52$ T (After Wright and Lax 1961)

is shown in Fig. 9a for InSb and can be used to determine ω_c and thereby the effective mass.

Voigt Configuration With plane-polarized light at normal incidence to the crystal surface and the magnetic vector parallel to the surface, we obtain for $\mathbf{E} \perp \mathbf{B}$ as a condition for minimum reflection ($R_\perp \cong 0$ because of $n_\perp \to 1$) from Eq. 54, neglecting the damping term:

$$\omega^4 \left(\frac{\varepsilon_{opt} - 1}{\varepsilon_{opt}} \right) - \omega^2 \left[\left(\frac{\varepsilon_{opt} - 1}{\varepsilon_{opt}} \right) \left(\omega_p^2 + \omega_c^2 \right) + \omega_p^2 \right] + \omega^4 = 0 \qquad (59)$$

which has two minima at

$$\omega_1^2 \cong \omega_p^2 \left(\frac{\varepsilon_{opt}}{\varepsilon_{opt} - 1} \right) + \varepsilon_{opt}\, \omega_c^2 \quad \text{and} \quad \omega_2^2 \cong \omega_p^2 - \omega_c^2 \left(\varepsilon_{opt} - 1 \right). \qquad (60)$$

They are shown in Fig. 9b for *n*-type InSb from which the cyclotron frequency, and thereby the effective mass, can be obtained. Once the cyclotron frequency is known, the plasma frequency can also be obtained and thereby the carrier density. For more information on magneto-optical and magnetoplasma effects, see Palik and Wright (1967), Maan (1993), and Seeger (2004).

3.2 Cyclotron-Resonance Absorption and Faraday Effect

3.2.1 Cyclotron-Resonance Absorption

Cyclotron-resonance absorption occurs when the frequency of the impinging electromagnetic radiation (usually in the microwave range) equals the cyclotron frequency (eB/m_n). It is measured with $\mathbf{k}\|\mathbf{B}$ and linear polarized $\mathbf{E} = (E_x, 0, 0)$ radiation, which renders (Eq. 43)

$$k_z^2 = \frac{\omega^2}{c^2}\varepsilon_{\text{opt}} + i\frac{\omega}{\varepsilon_0 c^2}\sigma_{xx}. \tag{61}$$

With $k_z^2 = (\omega/c)^2 \tilde{n}_\text{r}^2 = (\omega/c)^2(n_\text{r} + i\kappa)^2$ we obtain with (Eq. 39)

$$2n_\text{r}\kappa = \frac{1}{\varepsilon_0\omega}\sigma_{xx} = \frac{\sigma_0}{\varepsilon_0\omega}\frac{1 + \left(\omega_\text{c}^2 + \omega^2\right)\tau_\text{mag}^2}{\left[1 + \left(\omega_\text{c}^2 - \omega^2\right)\ \tau_\text{mag}^2\right]^2 + 4\omega^2\tau_\text{mag}^2}. \tag{62}$$

Cyclotron resonance effectively determines ω_c, which is a direct measure of the effective mass. By rotating the semiconductor with respect to $\mathbf{B}$, one obtains m_n as a function of the crystallographic orientation. This was discussed in more general terms in Sect. 1.2.5 of ▶ Chap. 8, "Bands and Bandgaps in Solids."

In addition to the conventional method of observing cyclotron resonance by microwave absorption, one can detect the resonances optically by exciting carriers near the band edges. With a magnetic field applied, electrons are forced into orbits with the cyclotron frequency ω_c and the band states split into Landau levels with spacing $\hbar\omega_\text{c}$, see Sect. 2 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients." The ensuing mixed collective mode frequency in Voigt configuration is then composed of three frequencies,

$$\begin{aligned}\omega_\pm = &\frac{1}{2}\left(\omega_\text{p}^2 + \omega_\text{c}^2 + \omega_\text{LO}^2\right)\\ &\pm\frac{1}{2}\sqrt{\left(\omega_\text{p}^2 + \omega_\text{c}^2 + \omega_\text{LO}^2\right)^2 - 4\left(\omega_\text{c}^2\omega_\text{LO}^2 + \omega_\text{p}^2\omega_\text{LO}^2\right)}.\end{aligned} \tag{63}$$

Consequently, the resonances become magnetic field dependent. These resonances are given by $\omega_\text{c}^2 + \omega_\text{p}^2 = \omega_\text{LO}^2$ (Palik and Furdyna 1970). In Faraday configuration, strong absorption occurs at $\omega = \omega_\text{c}$ or $\omega = \omega_\text{TO}$, the former is independent of lattice coupling. In addition to the resonant Raman scattering at ω_c, two-photon resonances occur when $\omega_1 - \omega_2 = 2\omega_\text{c}$ (Patel and Slusher 1968).

In *superlattices*, the confinement of cycling electrons within each well can be easily detected by the optical method described before (Cavenett and Pakulis 1985): the resonances become angle-dependent and are pronounced only within the plane of the superlattice, where electrons can follow the Lorentz force.

3.2.2 Faraday Effect

The dispersion relation for circular polarized light (Eqs. 45 to 47) with **k**||**B** shows that the propagation velocity for circular polarized light $c/n_{\pm}$ is different for right- or left-hand polarization. Therefore, a linear polarized light beam, composed of an equal fraction of left- and right-polarized components, experiences a turning of its polarization plane with progressive traveling through a semiconductor sample with a thickness d. The corresponding *Faraday angle* is defined by

$$\theta_F = \frac{\omega d}{2c}(n_+ - n_-) \cong \frac{\omega d}{4c\sqrt{\varepsilon_{opt}}}\left(n_+^2 - n_-^2\right), \tag{64}$$

since $\left(n_+^2 - n_-^2\right) = (n_+ + n_-)(n_+ - n_-) \cong 2\sqrt{\varepsilon_{opt}}(n_+ - n_-)$. The angle can be obtained from Eq. 46 for $\kappa_+^2 \ll n_-^2$ and $\left(\omega_c \tau_{mag}, \omega \tau_{mag}\right) \gg 1$,

$$\theta_F = \frac{360^\circ}{2\pi} \quad \frac{n e^3 B}{m_n^2 \varepsilon_0 \sqrt{\varepsilon_{opt}}\, 2c\left(\omega^2 - \omega_c^2\right)} d \quad \text{for } \omega \gg \omega_c. \tag{65}$$

The quantity $\theta_F/(B\ d)$ is known as the *Verdet coefficient* and is proportional to $n\lambda^2$. The proportionality of the Verdet coefficient with λ^2 is shown in Fig. 10 for n-InSb with the electron density and temperature as the family parameters. It shows an increase of the Verdet coefficient with carrier density in agreement with Eq. 65.

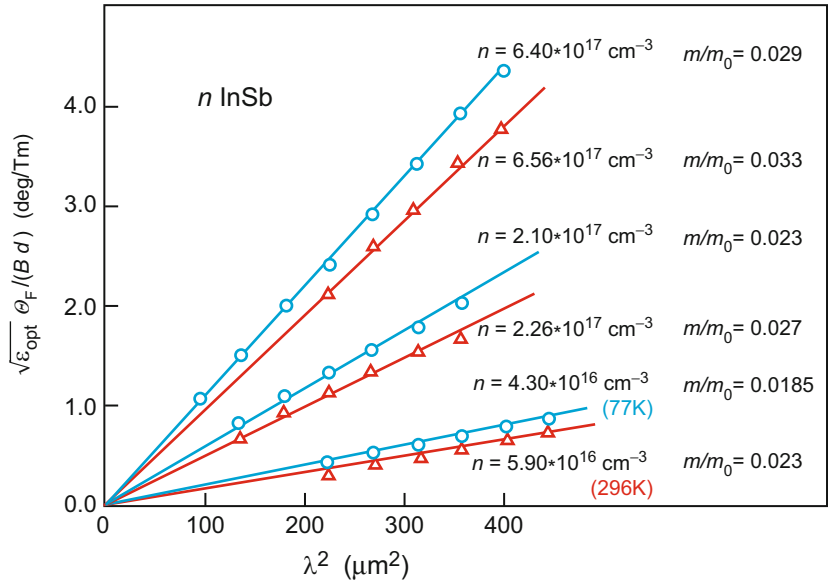


Fig. 10 Faraday rotation in n-type InSb with electron density and temperature as parameters; *blue* and *red* curves are for $T = 77$ K and 290 K, respectively. The corresponding average effective mass ratios for the six curves are shown at the *right side* of the figure (After Pidgeon 1962)

It also shows an increase with decreasing temperature due to a temperature-dependent effective mass (see below). With known carrier density, the Verdet coefficient yields the value of the effective mass. The changes in m_n, as a function of T and n, are due to changes in the electron distribution within the band, indicating that the effective mass increases for electrons higher in the conduction band – see Sect. 1.2.8 in ▶ Chap. 8, "Bands and Bandgaps in Solids."

Faraday rotation for electrons is opposite in sign to the rotation for holes. In a multiband semiconductor, both carriers must be considered. For instance, with heavy and light holes, the quantity n/m_n^2 in Eq. 65 must be replaced by $n_{\mathrm{lh}}/m_{\mathrm{lh}}^2 + n_{\mathrm{hh}}/m_{\mathrm{hh}}^2$ for holes.

4 Plasmon Dispersion in 2D Semiconductors

Collective oscillations of carriers (electrons or holes) are altered if they are free to move in two spatial dimensions but have their motion constrained in the third dimension. Two-dimensional plasmons were observed first for electrons on the surface of liquid helium (Grimes and Adams 1976). Most studied two-dimensional free-carrier systems in semiconductors are the two-dimensional electron gas (2DEG) created in semiconductor inversion or accumulation layers of modulation-doped field-effect transistors (Fig. 11a) and multiple layers of 2DEGs, so-called layered electron gas (LEG), realized in modulation-doped multilayer stacks of semiconductor heterojunctions (Fig. 11b). In these structures the free electrons are spatially separated from the ionized donor impurities and scattering by charged impurities is only weak, yielding a high carrier mobility (▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors").

The two-dimensional plasmon resonance of the collective free-carrier motion was calculated for single quantum wells (Stern 1967) and inversion systems (Chaplik 1972), yielding for long wavelengths ($q_{\|} \to 0$)

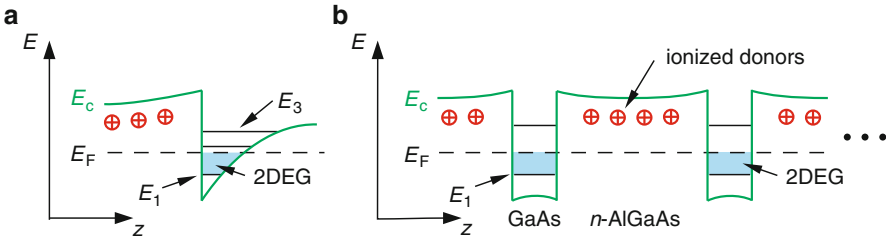


Fig. 11 Two-dimensional electron gas (2DEG) in modulation-doped GaAs/n-$Al_xGa_{1-x}As$ structures created (**a**) at the heterojunction and (**b**) in a multi-quantum-well superlattice. E_c and E_F denote the conduction-band edge and the Fermi energy, respectively. Free electrons are confined in the triangular-shaped potential at the interface forming an inversion layer (**a**) or in the undoped GaAs quantum wells (**b**)

$$\omega_{\mathrm{p}}^{2\mathrm{D}} = \sqrt{\frac{n_{2\mathrm{D}}\, e^2}{2\varepsilon_{\mathrm{eff}}\, \varepsilon_0\, m_n}}\sqrt{q_{\|}}. \tag{66}$$

with the areal carrier density n_{2D} (in units of cm^{-2}) and the in-plane wavevector of the plasmon $q_{\|}$. ε_{eff} is an effective dielectric function, which accounts for the geometry of the system.[6] Equation 66 shows a square-root dependence on the wavevector for 2D plasmons, in contrast to dispersionless 3D plasmons (Eq. 4).

In a *layered electron gas* with an *infinite* number N of equally spaced 2DEGs according to Fig. 11b the plasma frequency gets (Jain and Allen 1985; Oleg et al. 1982)

$$\omega_{\mathrm{p}}^{\mathrm{LEG}} = \left(\frac{n_{2\mathrm{D}}\, e^2}{2\varepsilon_{\mathrm{eff}}\, \varepsilon_0\, m_n} q_{\|} \frac{\sinh q_{\|} d}{\cosh q_{\|} d - \cos q_{\perp} d}\right)^{1/2} \tag{67}$$

with the normal component $q_{\perp}$ of the wavevector. Equation 67 contains a transition between regimes of different dimensionality. For large separation d between the 2DEG planes ($q_{\|}\, d \gg 1$) the equation reduces to that of a single 2D plama expressed by Eq. 66. For long in-plane wavelengths ($q_{\|}\, d \ll 1$) and in-phase oscillation of all planes ($q_{\perp} = 0$) the plasma frequency is similar to that of a 3D system with an electron density n_{2D}/d (Das Sarma and Quinn 1982).

The in-plane wavevector $q_{\|}$ can be varied in inelastic light scattering by changing the angle Θ of the incident light outside the sample with respect to the normal of the sample surface. Choosing the angle between incident and detected backscattered light to 90°, the components of the scattering wavevector are

$$q_{\|} = \frac{2\pi}{\lambda}(\sin\Theta - \cos\Theta) \quad \text{and} \quad q_{\perp} = \frac{4\pi}{\lambda} n_{\mathrm{r}}\left(1 - \frac{1}{4n_{\mathrm{r}}^2}\right). \tag{68}$$

The resulting dispersion of the 2D plasmon resonance in a layered 2D electron gas (Fig. 11b) is shown in Fig. 12. The dispersion curves indicate the dependence on the sample structure ($q_{\|}d$) and doping (n_{2D}).

In a layered electron gas with a *finite* number N of 2DEG layers, the plasmon resonance is fanned out into a band due to the reduced symmetry (the mirror symmetry at *each* 2DEG is lost) and the Couplomb interaction between the 2DEG layers. A superlattice with N layers has N eigenmodes for the correlated motion of electrons in the different layers. The normal components of the wavevectors for these eigenmodes are given by

[6] In single and multiple quantum wells ε_{eff} is the dielectric function of the barriers. In a metal-oxide-semiconductor junction with oxide thickness d_{ox} and perfectly screening gate, $\varepsilon_{eff} = ½(\varepsilon_{semiconductor} + \varepsilon_{ox} \coth(q_{\|}\, d_{ox}))$, see Chaplik 1972.

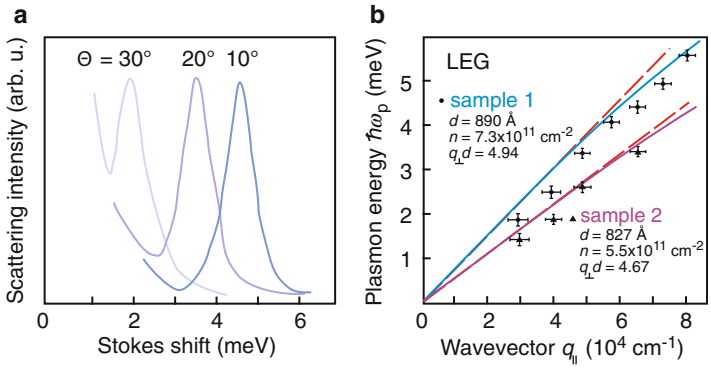


Fig. 12 (**a**) Plasmon Raman line of a layered electron gas (LEG) for different incident angles Θ with respect to the sample normal, yielding different in-plane wavevectors $q_{\|}$. (**b**) Dispersion relation of the 2D plasma frequency for two LEG samples with different electron densities n^{2D} and values of $q_{\perp}\, d$; upper data points refer to the spectra of panel (**a**). *Solid* and *dashed lines* refer to a calculation according to Eq. 67 and a linear dispersion, respectively (After Oleg et al. 1982)

$$q_{\perp i} = i\frac{2\pi}{N\,d}, \quad i = 1 \ldots N, \tag{69}$$

where d is the superlattice period. $q_{\perp i}\, d$ represents the phase relation beween the layers for the plasmon mode i; for small N the quantity $q_{\perp i}$ loses the physical meaning of a wavenumber.

The calculated dispersion relation for a layered electron gas with $N = 6$ 2DEG layers is shown in Fig. 13a; the parameters are those of sample 1 in Fig. 12. Each single 2DEG layer has a dispersion $\omega_p \propto \sqrt{q}$, leading to $\omega_p \to 0$ as $q \to 0$ for each branch. In the highest-energy branch electrons in each plane oscillate in phase; for $q_{\|}\, L \ll 1$ the entire layered electron gas of thickness $L = (N\text{-}1)d$ then acts as a single 2DEG layer with an electron density of $N \times n$. The two modes above the 3D plasmon band are surface modes (Giuliani and Quinn 1983). For large values $q_{\|}\, L$ the plasmons on the two outer surfaces do not couple and these modes are degenerate; for small $q_{\|}\, L$ the surface modes split into a symmetric and an antisymmetric mode. We note that the surface modes shown in Fig. 13a lie outside the 3D plasmon band. Such a feature originates from a dielectric mismatch between the top and bottom side of the outermost 2DEG layers, if the dielectric constant outsite the superlattice is less than inside. The resulting image charges modify the Coulomb interaction and increase the plasmon energy.

The number of modes increases in a finite layered electron gas as the number of layers N increases as shown in Fig. 13b. The figure also illustrates that the band of the 3D plasmon then becomes gradually more densely populated and eventually continuous for $N \to \infty$; the boundaries of the 3D plasmon are obtained by putting $\cos(q_{\perp} d) = \pm 1$ in Eq. 67.

The discrete plasmon modes were also observed in experiment (Pinczuk et al. 1986; Fasol et al. 1986). More recently also collective plasmon modes in

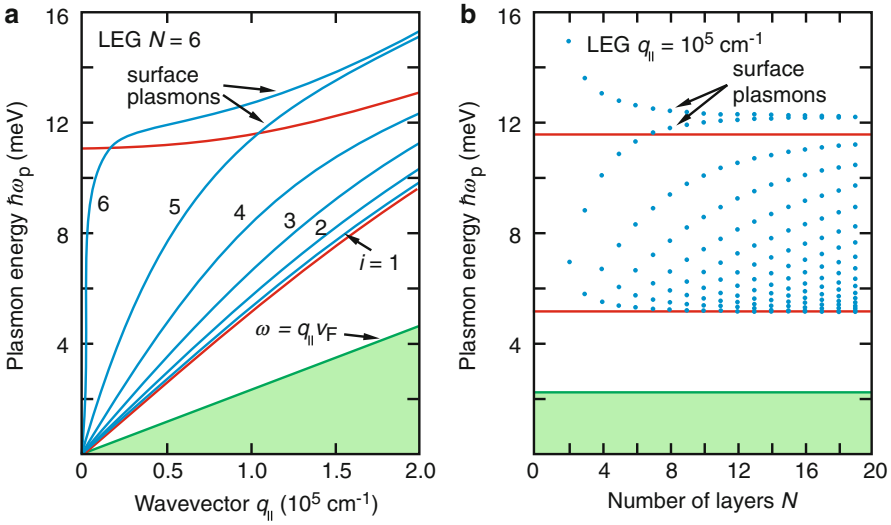


Fig. 13 (**a**) Calculated plasmon energies in a layered electron gas with six 2DEG layers. The *red lines* represent the boundaries of the 3D plasmon band; the *green shaded* area signifies the single-particle continuum. (**b**) Plasmon energies for various numbers N of 2DEG layers in a layered electron gas, calculated for a fixed in-plane plasmon wavevector $q_{\parallel} = 10^5$ cm^{-1} (After Jain and Allen 1985)

systems of mixed dimensionality such as coupled 3D-2D or 3D-1D or 2D-1D systems were studied (Hwang et al. 2018, Wang et al. 2021). A measurement of the plasmon branches in a finite layered electron gas by inelastic light scattering with a GaAs/(AlGa)As sample corresponding to the schematic of Fig. 11b is shown in Fig. 14. Solid and dashed lines represent calculated plasmon energies for antisymmetric and symmetric modes, respectively.

5 Summary

Plasmon oscillations of the free-electron ensemble with respect to the ion cores lead to a resonance absorption in the IR range described by two branches; the plasma frequency depends on the square root of the free-carrier density and represents the low- and high-energy asymptote of the ω_- and ω_+ branches, respectively. An additional resonance absorption occurs at high (~10 eV) photon energy for valence-electron plasmons. The plasma resonance can be used to determine optically the density, effective mass, and relaxation time of free carriers. The effective mass and its anisotropy can be measured by simple changes of the relative alignment of crystal, electrical, and magnetic fields. The other parameters can often be

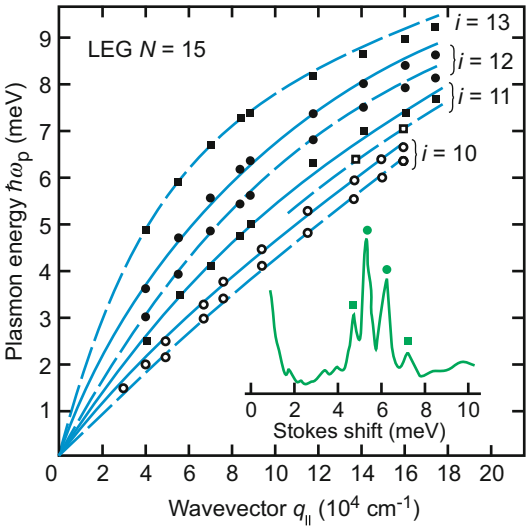


Fig. 14 Experimental plasmon dispersions of a finite layered electron gas with 15 2DEG layers. The *inset* shows the discrete plasmon modes observed in inelastic light scattering for $q_{\parallel} = 8.5 \times 10^4$ cm^{-1}. *Squares* and *circles* indicate weak and strong lines, *filled* and *open* symbols were measured with incident wavelengths of 676 nm and 763 nm, respectively (After Pinczuk et al. 1986)

obtained more easily by electrical measurements; however, optical measurements can be performed when electrical contacts cannot be applied, or electrical measurements are otherwise impeded. If the free carriers are confined to two dimensions the plasmon frequency depends on the in-plane wavevector. In multiple two-dimensional layers of free carriers discrete plasmon branches exist.

Away from the resonance free-carrier absorption and reflection with or without magnetic fields provide also important information about carrier parameters – most significantly about the effective carrier mass, its anisotropy, and its dependence on other variables, such as carrier density and temperature. The knowledge of the effective carrier mass is essential for the analysis of a large variety of semiconducting and electro-optical device properties.

References

Basu PK (1997) Theory of optical processes in semiconductors. Oxford University Press, New York

Cavenett BC, Pakulis EJ (1985) Optically detected cyclotron resonance in a GaAs/$Ga_{0.67}Al_{0.33}As$ superlattice. Phys Rev B 32:8449

Chan S-K, Heine V (1973) Spin density wave and soft phonon mode from nesting Fermi surfaces. J Phys F: Met Phys 3:795

Chaplik AV (1972) Possible crystallization of charge carriers in low-density inversion layers. Sov Phys - JETP 35:395

Das Sarma S, Quinn JJ (1982) Collective excitations in semiconductor superlattices. Phys Rev B 25:7603

Dixon JR (1960) Proceedings of 5th international conference physics semicond, Prague, p 366

Drude P (1900) Zur Elektronentheorie der Metalle. Ann Phys (Leipzig) 1:566, and 3: 369 (On the electron theory of metals, in German)

Egerton RF (2009) Electron energy-loss spectroscopy in the TEM. Rep Prog Phys 72:016502

Eichberger M, Schäfer H, Krumova M, Beyer M, Demsar J, Berger H, Moriena G, Sciaini G, Dwayne Miller RJ (2010) Snapshots of cooperative atomic motions in the optical suppression of charge density waves. Nature 468:799

Elsaesser T, Woerner M (1999) Femtosecond infrared spectroscopy of semiconductors and semiconductor nanostructures. Phys Rep 321:253

Fan HY (1967) Effects of free carriers on the optical properties. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 3. Academic Press, New York, p 405

Fasol G, Mestres N, Hughes HP, Fischer A, Ploog K (1986) Raman scattering by coupled-layer plasmons and in-plane two-dimensional single-particle excitations in multi-quantum-well structures. Phys Rev Lett 56:2517

Foteinopoulou S, Devarapu GCR, Subramania GS, Krishna S, Wasserman D (2019) Phonon-polaritonics: enabling powerful capabilities for infrared photonics. Nanophotonics 8:2129

Giuliani GF, Quinn JJ (1983) Charge-density excitations at the surface of a semiconductor superlattice: a new type of surface polariton. Phys Rev Lett 51:919

Grimes CC, Adams G (1976) Observation of two-dimensional plasmons and electron-ripplon scattering in a sheet of electrons on liquid helium. Phys Rev Lett 36:145

Harper PG, Hodby JW, Stradling RA (1973) Electrons and optic phonons in solids-the effects of longitudinal optical lattice vibrations on the electronic excitations of solids. Rep Prog Phys 36:1

Huberman M, Overhauser AW (1982) Open-orbit magnetoresistance spectra of potassium. Phys Rev B 25:2211

Hwang EH, Hu Ben Yu-Kuang, Das Sarma S (2018) Dimensionally mixed coupled collective modes. Phys Rev B 98:161304

Hybertsen MS, Louie SG (1986) Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. Phys Rev B 34:5390

Irmer G, Röder C, Himcinschi C, Kortus J (2013) Phonon polaritons in uniaxial crystals: A Raman scattering study of polaritons in α-GaN. Phys Rev B 88:104303

Jain JK, Allen PB (1985) Plasmons in layered films. Phys Rev Lett 54:1985

Kaiser W, Collins RJ, Fan HY (1953) Infrared absorption in *p*-type germanium. Phys Rev 91:1380

Kim OK, Bonner WA (1983) Infrared reflectance and absorption of *n*-type InP. J Electron Mater 12:827

Kong X, Albert S, Bengoechea-Encabo A, Sanchez-Garcia MA, Calleja E, Trampert A (2012) Plasmon excitation in electron energy-loss spectroscopy for determination of indium concentration in (In, Ga)N/GaN nanowires. Nanotechnology 23:48

Larson P, Dvorak M, Wu Z (2013) Role of the plasmon-pole model in the *GW* approximation. Phys Rev B 88:125205

Lax B (1963) Proceedings of international school physics "Enrico Fermi", Course XXII (Smith RA ed), p 240. Academic Press, New York

Leon DA, Cardoso C, Chiarotti T, Varsano D, Molinari E, Ferretti A (2021) Frequency dependence in *GW* made simple using a multipole approximation. Phys Rev B 104:115157

Lian C-S, Heil C, Liu X, Si C, Giustino F, Duan W (2019) Coexistence of superconductivity with enhanced charge density wave order in the two-dimensional limit of $TaSe_2$. J Phys Chem Lett 10:4076

Loret B, Auvray N, Gallais Y, Cazayous M, Forget A, Colson D, Julien M-H, Paul I, Civelli M, Sakuto A (2019) Intimate link between charge density wave, pseudogap and superconducting energy scales in cuprates. Nat Phys 15:771

Maan JC (1993) Intra- and interband magneto-optical properties of bulk semiconductors and heterostructures. In: Martinez G (ed) Optical properties of semiconductors. Springer Netherlands, Dordrecht

Madelung O (1978) Introduction to solid state theory. Springer, Berlin/New York

Maity I, Naik MH, Maiti PK, Krishnamurthy HR, Jain M (2020) Phonons in twisted transition-metal dichalcogenide bilayers: Ultrasoft phasons and a transition from a superlubric to a pinned phase. Phys Rev Research 2:013335

Manley ME, Stonaha PJ, Abernathy DL, Chi S, Sahul R, Hermann RP, Budai JD (2018) Supersonic propagation of lattice energy by phasons in fresnoite. Nat Commun 9:1823
Mooradian A, McWhorter AL (1967) Polarization and intensity of Raman scattering from plasmons and phonons in gallium arsenide. Phys Rev Lett 19:849
Mooradian A, Wright GB (1966) Observation of the interaction of plasmons with longitudinal optical phonons in GaAs. Phys Rev Lett 16:999
Oleg D, Pinczuk A, Gossard AC, Wiegmann W (1982) Plasma dispersion in a layered electron gas: a determination in GaAs-(AlGa)As heterostructures. Phys Rev B 25:7867
Overhauser AW (1978) Charge-density waves and isotropic metals. Adv Phys 27:343
Palik ED, Furdyna JK (1970) Infrared and microwave magnetoplasma effects in semiconductors. Rep Prog Phys 33:1193
Palik ED, Holm RT (1979) Optical characterization of semiconductors. In: Zemel JN (ed) Nondestructive evaluation of semiconductor materials and devices. Plenum Press, New York, pp 315–345
Palik ED, Wright GB (1967) Free-carrier magnetooptical effects. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 3. Academic Press, New York, p 421
Palik ED, Teitler S, Henvis BW, Wallis RF (1962) Magneto-optical studies of semiconductors using polarized radiation. In: Stickland AC (ed) Proceedings of 6th international conference physics semicond, Exeter, pp 288–294
Patel CKN, Slusher RE (1968) Raman scattering by polaritons in presence of electron plasma in GaAs. Phys Rev Lett 22:282
Pidgeon CR (1962) PhD thesis, University Reading
Pinczuk A, Lamont MG, Gossard AC (1986) Discrete plasmons in finite semiconductors multilayers. Phys Rev Lett 56:2092
Pines D (1999) Elementary excitations in solids. Perseus Books, Reading
Platzman PM, Wolff PA (1973) Waves and interactions in solid state plasmas. In: Ehrenreich H, Seitz F, Turnbull D (eds). Solid state physics, vol Supplement 13, Academic Press, New York
Poole CP Jr (1998) The physics handbook. Wiley, New York
Qin F, Li S, Du ZZ, Wang CM, Zhang W, Yu D, Lu H-Z, Xie XC (2020) Theory for the charge-density-wave mechanism of 3D quantum Hall effect. Phys Rev Lett 125:206601
Roth LM (1982) Dynamics of electrons in electric and magnetic fields. In: Moss TS, Paul W (eds) Handbook of semiconductors, vol 1. North Holland Publishing, Amsterdam, pp 451–485
Rossnagel K (2011) On the origin of charge-density waves in select layered transition-metal dichalcogenides. J Phys Condens Matter 23:213001
Sakurai JJ, Napolitano J (2011) Modern quantum mechanics, 2nd edn. Addison-Wesley, Boston
Seeger K (1973) Semiconductor physics. Springer, New York
Seeger K (2004) Semiconductor physics, 9th edn. Springer, New York
Stern F (1967) Polarizability of a two-dimensional electron gas. Phys Rev Lett 18:546
Storeck G, Horstmann JG, Diekmann T, Vogelgesang S, von Witte G, Yalunin SV, Rossnagel K, Ropers C (2020) Structural dynamics of incommensurate charge-density waves tracked by ultrafast low-energy electron diffraction. Struct Dyn 7:034304
Sugano S, Kojima N (eds) (2000) Magneto-optics. Springer, Berlin
Tang F, Ren Y, Wang P, Zhong R, Schneeloch J, Yang SA, Yang K, Lee PA, Gu G, Qiao Z, Zhang L (2019) Three-dimensional quantum Hall effect and metal–insulator transition in $ZrTe_5$. Nature 569:537
Tiginyanu IM, Irmer G, Monecke J, Vogt A, Hartnagel HL (1997) Porosity-induced modification of the phonon spectrum of *n*-GaAs. Semicond Sci Technol 12:491
Wang S, Yoo SJ, Zhao S, Zhao W, Kahn S, Cui D, Wu F, Jiang L, Utama MIB, Li H, Li S, Zibrov A, Regan E, Wang D, Zhang Z, Watanabe K, Taniguchi T, Zhou C, Wang F (2021) Gate-tunable plasmons in mixed-dimensional van der Waals heterostructures. Nat Commun 12:5039
Wilson JA, DiSalvo FJ, Mahajan S (1975) Charge-density waves and superlattices in the metallic layered transition metal dichalcogenides. Adv Phys 24:17

Wolfe R (1954) On the theory of optical absorption in metals and semiconductors. Proc Phys Soc London Sect A 67:74

Wright GB, Lax B (1961) Magnetoreflection experiments in intermetallics. J Appl Phys 32:2113

Yang Y, Fang S, Fatemi V, Ruhman J, Navarro-Moratalla E, Watanabe K, Taniguchi T, Kaxiras E, Jarillo-Herrero P (2018) Enhanced superconductivity upon weakening of charge density wave transport in 2*H*-TaS_2 in the two-dimensional limit. Phys Rev B 98:035203

Band-to-Band Transitions

13

Karl W. Böer and Udo W. Pohl

Contents

1 Optical Absorption Spectrum 500
1.1 The Joint Density of States 500
1.2 Absorption Coefficient and Dielectric Function 504
1.3 The Fundamental Absorption Edge 509
2 Direct and Indirect Transitions 511
2.1 Indirect Transitions 511
2.2 Allowed and Forbidden Transitions 513
2.3 Band-to-Band Magnetoabsorption 516
3 Transitions in Quantum Wells 518
3.1 Energy Levels in Multiple Quantum Wells 518
3.2 Absorption in Quantum Wells 520
4 Optical Bandgap of Amorphous Semiconductors 521
4.1 Intrinsic Absorption 521
4.2 Extrinsic Absorption in Glasses 523
5 Summary 525
References 525

Abstract

Optically induced band-to-band transitions are resonance transitions and related to the band structure by the momentum matrix-element and the joint density of states. For transitions near the band edge, the theory of optical transitions between the valence and conduction bands can be simplified with an effective-mass approximation, assuming parabolic band shapes and arriving at quantitative expressions for the absorption as a function of the photon energy. Depending on

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_13

the conduction-band behavior, strong direct or weak indirect transitions occur at the band edge. In addition, a contribution of forbidden transitions modifies the absorption further away from the band edge. Deviations from the ideal, periodic crystal lattice provide tailing states extending beyond the band edge, usually as an Urbach tail which decreases exponentially with distance from the band edge. In quantum wells the two-dimensional joint density of states leads to a steplike increase of the absorption for increasing photon energy.

Keywords

Absorption coefficient · Band structure · Band tailing · Direct transitions · Forbidden transitions · Indirect transitions · Joint density of states · Momentum-matrix element · Optical transitions · Parabolic bands · Urbach tail · Van Hove singularities

1 Optical Absorption Spectrum

The description of optical band-to-band transitions requires a quantum mechanical analysis (Bassani 1966; Bassani and Pastori Parravicini 1975). For a simplified treatment of transitions near the fundamental bandgap, an effective-mass model can be used. This chapter focuses on the absorption process; the subsequent radiative or nonradiative recombination processes are treated in ▶ Chap. 30, "Carrier Recombination and Noise." The absorption spectrum is directly related to the imaginary part of the dielectric function introduced in ▶ Chap. 10, "Interaction of Light with Solids." One distinguishes *direct* band-to-band transitions, which are essentially vertical transitions in $E(\mathbf{k})$, and *indirect* transitions, which involve phonons, permitting major changes of $\mathbf{k}$ during the transitions. We will first discuss the direct transitions.

1.1 The Joint Density of States

Light of sufficiently short wavelength with

$$\mathbf{E} = A_0 \mathbf{e} \exp[i(\mathbf{kr} - \omega t)] \tag{1}$$

initiates electronic transitions from one band to another. We have used A_0 here as the amplitude to avoid confusion with the energy of an optical transition, E. $\mathbf{e}$ is the unit vector of the electric polarization and $\mathbf{k}$ is the wavevector of the light, traveling in $\mathbf{r}$ direction.

The number of optically induced transitions at the same $\mathbf{k}$ (i.e., neglecting the wavevector of the photon) between band μ and band ν is proportional to the square of the momentum matrix elements given by

$$\mathbf{e}\,\mathbf{M}_{\mu\nu}(\mathbf{k}) = \mathbf{e}\int_V \psi_\mu^*(\mathbf{k},\mathbf{r})\ \left(-i\hbar\frac{\partial}{\partial\mathbf{r}}\right)\ \psi_\nu(\mathbf{k},\mathbf{r})\ d\mathbf{r}, \tag{2}$$

with the integral extending over the crystal volume V and $\mathbf{e}$ the unit polarization vector. The proportionality factor for a specific transition is

$$P_{\mu\nu}(\mathbf{k}) = \frac{\pi}{\varepsilon_0 h}\left(\frac{eA_0}{m_0 c}\right)^2 \delta\big(E_\mu(\mathbf{k}) - E_\nu(\mathbf{k}) - h\nu\big), \tag{3}$$

where δ is the Dirac delta function. The factor $1/h$ indicates the quantum nature of the transition, $e/(m_0 c)$ stems from the interaction Hamiltonian between light and electrons, and A_0 from the amplitude of the light (i.e., the value of the Poynting vector A_0^2). The delta function switches on this contribution when a transition occurs from one state to another, i.e., when $E_\mu(\mathbf{k}) - E_\nu(\mathbf{k}) = h\nu$. No broadening of any of these transitions is assumed. Close proximity to adjacent transitions and Kramers–Kronig interaction make such a broadening consideration unnecessary, except for very pronounced features.

After integration over all states within the first Brillouin zone and all bands between which the given photon $h\nu$ can initiate transitions, we obtain for the number of such transitions per unit volume and time *Fermi's golden rule*:

$$\begin{aligned} W(\nu) &= \sum_{\mu,\nu}\int_{\mathrm{BZ}} \frac{2}{(2\pi)^3} P_{\mu\nu}\left|\mathbf{e}\cdot\mathbf{M}_{\mu\nu}(\mathbf{k})\right|^2 d\mathbf{k} \\ &= \frac{2}{(2\pi)^3}\frac{\pi}{\varepsilon_0 h}\left(\frac{eA_0}{m_0 c}\right)^2 \sum_{\mu,\nu}\int_{\mathrm{BZ}} \left|\mathbf{e}\cdot\mathbf{M}_{\mu\nu}(\mathbf{k})\right|^2 \delta\big(E_\mu(\mathbf{k}) - E_\nu(\mathbf{k}) - h\nu\big)\, d\mathbf{k}, \end{aligned} \tag{4}$$

where the integration is over the entire Brillouin zone (BZ). The factor $V_{\text{unit cell}}/(2\pi)^3$ normalizes the $\mathbf{k}$ vector density within the Brillouin zone; $V_{\text{unit cell}}$ cancels from the integration of Eq. 2, the factor 2 stems from the spin degeneracy. Equation 4 follows directly from a perturbation theory. The matrix elements vary little within the Brillouin zone; therefore, we can pull these out in front of the integral. This leaves only the delta function inside the integral. The integral identifies the sum over all possible transitions which can be initiated by photons with a certain energy $h\nu$, and it is commonly referred to as the *joint density of states* between these two bands. It is given by

$$J_{\mu\nu}(\nu) = \frac{2}{(2\pi)^3}\int_{\mathrm{BZ}} \delta\big(E_\mu(\mathbf{k}) - E_\nu(\mathbf{k}) - h\nu\big)\, d\mathbf{k}. \tag{5}$$

The argument of the delta function is a function of $\mathbf{k}$. Since the photon momentum is negligibly small, $\mathbf{k}_{\text{initial}} = \mathbf{k}_{\text{final}}$ applies for the transition (assumed

to be direct), yielding for the delta function[1] $\delta\left(E_\mu - E_\nu - h\nu\right) = \delta(k_{\text{inital}} - k_{\text{final}})/\left|\partial/\partial\mathbf{k}\left(E_\mu(\mathbf{k}) - E_\nu(\mathbf{k})\right)\right|_{E_\mu - E_\nu = h\nu}$. The joint density of states can then be written as an integral over a surface of constant energy instead of an integral in $\mathbf{k}$ space:

$$J_{\mu\nu}(\nu) = \frac{2}{(2\pi)^3}\int \frac{dS}{\left|\partial/\partial\mathbf{k}\left(E_\mu(\mathbf{k}) - E_\nu(\mathbf{k})\right)\right|_{E_\mu - E_\nu = h\nu}}, \tag{7}$$

where dS is an element on the equal-energy surface in $\mathbf{k}$ space, with the surface depicted by

$$E_\mu(\mathbf{k}) - E_\nu(\mathbf{k}) = h\nu; \tag{8}$$

here μ stands for any one of the conduction bands and ν for any one of the valence bands.

For each of the transitions, $\mathbf{k}$ is constant; it is a *direct transition*. When the slopes of $E(\mathbf{k})$ of both bands are different, Eq. 4 is fulfilled only for an infinitesimal surface area. In contrast, a larger area about which such transitions can take place appears near points at which both bands have the *same* slope, i.e., near *critical points* (van Hove 1953 and Phillips 1956). Here

$$\partial E_\mu(\mathbf{k})/\partial\mathbf{k} = \partial E_\nu(\mathbf{k})/\partial\mathbf{k} = 0 \text{ or } \partial E_\mu(\mathbf{k})/\partial\mathbf{k} - \partial E_\nu(\mathbf{k})/\partial\mathbf{k} = 0, \tag{9}$$

and the expression under the integral of Eq. 7 has a singularity. By integration this results in a kink in the joint density of states. The singularity related to Eq. 7 is referred to as a *van Hove singularity*.

The types of critical points are referred to as $M_0 \ldots M_3$, $P_0 \ldots P_2$, and Q_0 or Q_1 in three-, two-, and one-dimensional $E(\mathbf{k})$ representations, respectively; two- and one-dimensional representations are applicable to superlattices or quantum wire configurations (Sect. 3 of ▶ Chap. 8, "Bands and Bandgaps in Solids"). They describe the relative curvature of the lower and upper bands with respect to each other. In a first-order expansion at a critical point $\mathbf{k}_{\text{cp}}$ in $\mathbf{k}$ space, we can express this by

$$E_\mu(\mathbf{k}) - E_\nu(\mathbf{k}) = E_{\mu\nu,0}\left(\mathbf{k}_{\text{cp}}\right) + \sum_{i=1}^{3} a_i \left(\mathbf{k} - \mathbf{k}_{\text{cp}}\right)^2, \tag{10}$$

where the components of $\mathbf{k}$ and $\mathbf{k}_{\text{cp}}$ are along the principal axes. Equation 10 may be interpreted as a flat lower band ($E_{\mu\nu,0}$) and the appropriately changed curvature of the upper band describing the energy difference between the original valence and conduction bands. Here positive quantities a_i are proportionality factors measuring

[1]The rewriting of Eq. 5 can be understood from the behavior of the Dirac delta function

$$\int g(x)\ \ \delta[f(x)]\,dx = g(x_0)\ \left|\frac{d\ f(x)}{dx}\right|^{-1}_{x=x_0}. \tag{6}$$

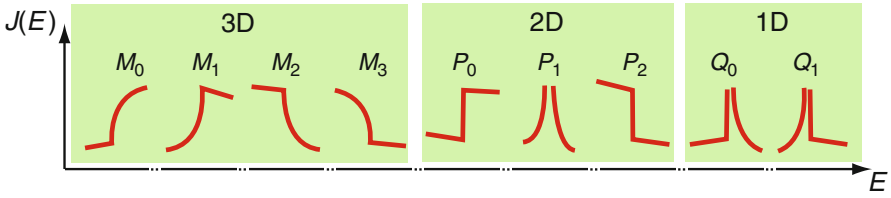


Fig. 1 Joint density-of-state representation near critical points (schematic) in three (M_i), two (P_i), and one (Q_i) dimensions of $E(\mathbf{k})$ with subscripts identifying the number of negative signs of the factors a_i in Eq. 10

the *relative* band curvatures, and the combination of their signs identifies the type of critical point,[2] with a schematic representation shown in Fig. 1.

The integrand in Eq. 7 can be written $\pi\, dk_3/a_1$, yielding for the joint density of states the expression $J_{\mu\nu}(\nu) = \frac{\text{const}}{8\pi^2 a_1}\int dk_3$ (Cohen and Chelikowsky 1988). $J_{\mu\nu}$ is thus proportional to the extent of the surface in k_3 direction. At an M_0 critical point, the surface extends to $k_3 = \pm\left((\omega - \omega_{\text{cp}})/a_3\right)^{1/2}$, leading to $J_{\mu\nu}(\nu) \propto \left(\nu - \nu_{\text{cp}}\right)^{1/2}$ rising from $h\nu_{\text{cp}}$. The structure of other critical points follows similar geometrical arguments (Cohen and Chelikowsky 1988). In optical spectra of α_{o}, R, or ε'', the critical points lead to sharp van Hove singularities, albeit the matrix elements (Eq. 2) effect some deviation from the ideal structure.

Matrix elements are mainly influenced by *selection rules*, which are similar to those in atomic spectroscopy. The symmetry properties of ψ_μ and ψ_ν in Eq. 2 – for example, whether even or odd under reflection or inversion and whether they are the same or different from each other – determine whether the matrix element has a finite value or vanishes; judgment can be rendered on group theoretical arguments (see Bassani and Pastori Parravicini 1975). Comparing the results of such an analysis with experimental observation of strong or weak optical absorption provides quite convincing arguments by assigning certain features of the absorption spectrum to the appropriate critical points of the $E(\mathbf{k})$ behavior.

The photon which is absorbed in the band-to-band transition can be released in a radiative recombination process. Such a luminescence – referred to as spontaneous emission – is important for optoelectronic devices. Luminescence is time-reversed to optical absorption; therefore, the same matrix elements and selection rules apply for both processes. The luminescence intensity is obtained from a product of the matrix element with the joint density of states and the distributions of the excited carriers. Radiative recombination is discussed in ▶ Chap. 30, "Carrier Recombination and Noise."

[2] 0, 1, 2, or 3 negative factors a_i in Eq. 10 yield for M_0 to M_3 a maximum, saddle point, saddle point, or minimum, respectively, in three dimensions; 0, 1, or 2 negative factors for P_0 to P_2 a maximum, saddle point, or minimum, respectively, in two dimensions; and 0 or 1 negative factors for Q_0 or Q_1 a maximum or minimum, respectively, in one dimension.

1.2 Absorption Coefficient and Dielectric Function

The optical absorption coefficient α_o (Sect. 1.1.3 of ▶ Chap. 10, "Interaction of Light with Solids") is proportional to the number of optical transitions per unit volume and unit time. The coefficient can be calculated from simple optical principles (Bassani and Pastori Parravicini 1975)[3]: it is given by the absorbed energy per unit time and volume, $h\nu \cdot W(\nu)$, divided by the energy flux $2\pi A_0^2 \nu^2 n_r/c$.[4] With Eq. 4 we now obtain

$$\alpha_o(\nu) = \frac{2}{(2\pi)^3} \frac{e^2}{2\varepsilon_0 n_r c m_0^2 \nu} \sum_{\mu,\nu} \int_{BZ} \left|\mathbf{e} \cdot \mathbf{M}_{\mu\nu}(\mathbf{k})\right|^2 \delta\left(E_\mu(\mathbf{k}) - E_\nu(\mathbf{k}) - h\nu\right) d\mathbf{k}. \quad (11)$$

For a further analysis of the integral, we need more knowledge about $E(\mathbf{k})$. This will be done near the fundamental absorption edge in Sect. 1.3 where some simplifying assumptions for the valence and conduction bands are introduced. For a short review, see Madelung (1981).

With the availability of synchrotron radiation sources, light sources of sufficient power and stability have become available to perform absorption and reflection spectroscopy for higher bands in the vacuum UV and soft x-ray range. Since the optical absorption here is very strong, requiring extremely thin crystals for absorption spectroscopy which are difficult to obtain with sufficient crystal perfection, it is preferable to obtain such spectra in reflection.

A more sensitive set of methods relates to modulation spectroscopy (Sect. 1.2.3 of ▶ Chap. 10, "Interaction of Light with Solids") where a crystal variable – e.g., the electric field, temperature, or light intensity – is modulated and the changing reflection signal is picked up by look-in technology. To compare the optical properties with the band structure, one must now express the transitions in terms of the complex dielectric constant $\tilde{\varepsilon} = \varepsilon' + i\varepsilon''$ $(= \varepsilon_1 + i\varepsilon_2)$. The optical absorption constant α_o is then obtained from $\varepsilon'' = \alpha_o n_r c/(2\pi\nu)$ (see Eqs. 22 and 24 of ▶ Chap. 10, "Interaction of Light with Solids"), with

$$\varepsilon'' = \frac{2}{(2\pi)^3} \frac{e^2}{4\pi\varepsilon_0 m_0^2 \nu^2} \sum_{\mu,\nu} \int_{BZ} \left|\mathbf{e} \cdot \mathbf{M}_{\mu\nu}(\mathbf{k})\right|^2 \delta\left(E_\mu(\mathbf{k}) - E_\nu(\mathbf{k}) - h\nu\right) d\mathbf{k}. \quad (12)$$

This representation has the advantage of not depending on the index of refraction, which causes a distortion of the calculated $\varepsilon''(\nu)$ distribution. With the first Kramers–Kronig relation (Eq. 20 of ▶ Chap. 11, "Photon–Phonon Interaction") and $\omega_a = \left(E_\mu - E_\nu\right)/\hbar$, we obtain for ε' from Eq. 12 after conversion of the δ function (Bassani and Pastori Parravicini 1975)

[3]Note that former literature used Gaussian cgs units.

[4]The energy flux is equal to the optical energy density given by the square of the wave amplitude per wavelength interval, $2\pi A_0^2 \nu^2 n_r^2/c^2$, times the light velocity within the semiconductor c/n_r.

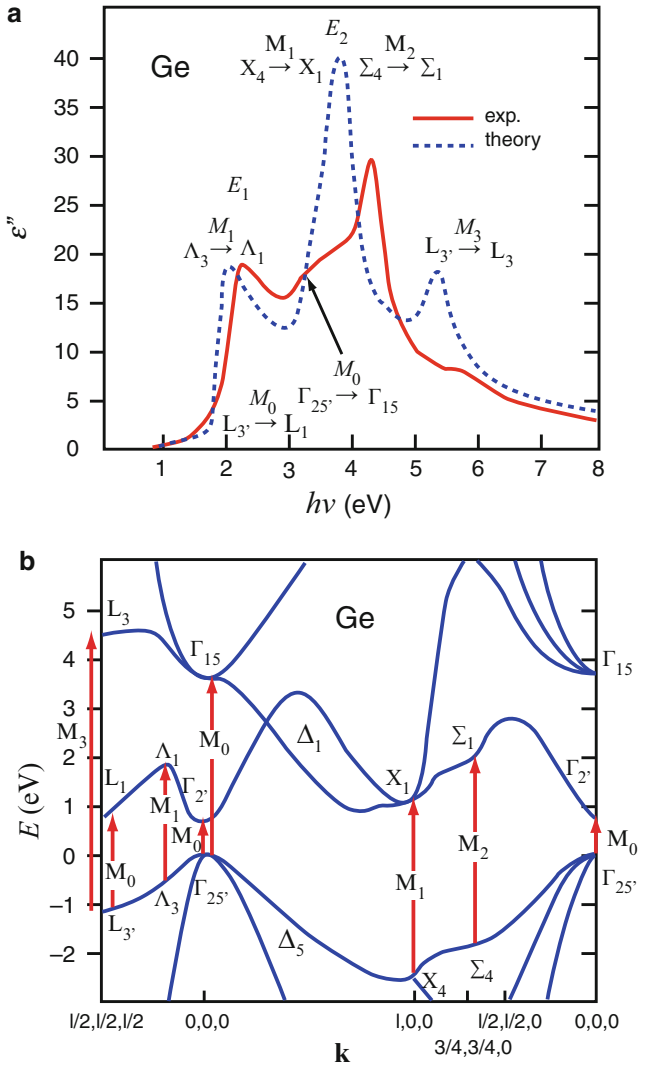


Fig. 2 (**a**) Optical absorption spectrum with critical points and (**b**) energy bands with corresponding transitions for Ge (After Brust et al. 1962). The *dashed curve* is the joint density of states, weighted by oscillator strength f, computed by the empirical pseudofunctional method

$$\varepsilon' = 1 + \frac{2e^2}{\varepsilon_0 m_0^2} \sum_{\mu,\nu} \int_{\mathrm{BZ}} \frac{2}{(2\pi)^3} \frac{\left|\mathbf{e} \cdot \mathbf{M}_{\mu\nu}(\mathbf{k})\right|^2}{\left[E_\mu(\mathbf{k}) - E_\nu(\mathbf{k})\right]/\hbar} \frac{d\mathbf{k}}{\left[E_\mu(\mathbf{k}) - E_\nu(\mathbf{k})\right]^2/\hbar^2 - \omega^2}. \tag{13}$$

As an example, the $\varepsilon''(\nu)$ spectrum is given in Fig. 2 for Ge. It shows several shoulders and spikes that can be connected to critical points in the corresponding

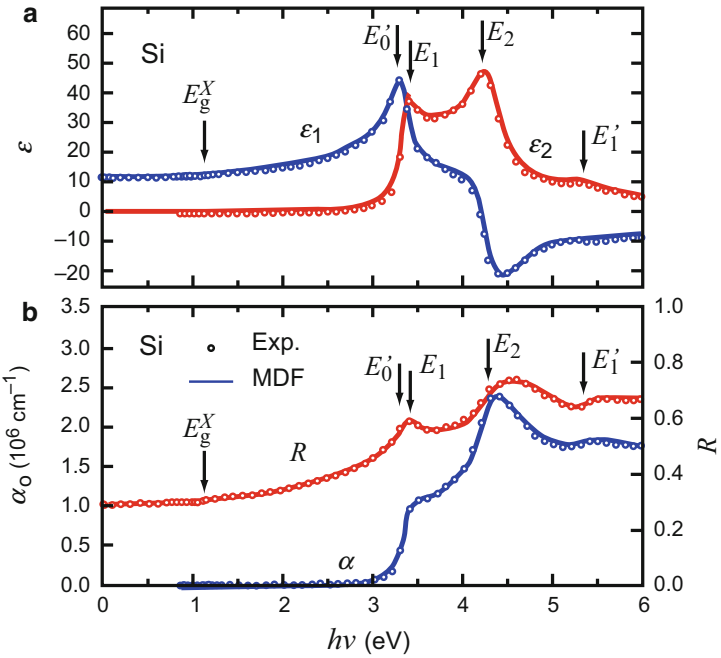


Fig. 3 (**a**) Real and imaginary parts of the dielectric function as well as (**b**) absorption coefficient and normal-incidence reflectivity of Si at 300 K as a function of the photon energy (After Adagi 2005). *Circles* and *solid lines* are, respectively, experimental data and calculations based on a model dielectric function (*MDF*) with a given number of interband critical points

theoretical curve, such as the L, Γ, and X transitions (see Sect. 4.2 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"). The theoretical spectrum was obtained from an empirical pseudopotential calculation (Phillips 1966) by numerically sampling $E_\mu(\mathbf{k}) - E_\nu(\mathbf{k})$ in the entire Brillouin zone at a large number of $\mathbf{k}$ values. By ordering and adding the transitions that occur at the same value of $h\nu$, possibly involving different band combinations, and repeating for other values of $h\nu$, we obtain the joint density of states. These yield the given $\varepsilon''(\nu)$ as shown in Fig. 2. The biggest peaks relate to ranges in $E(\mathbf{k})$ where both bands involved in the transition run parallel to each other for an extended $\mathbf{k}$ range.

Locations for intense transitions in the Brillouin zone of several III–V semiconductors are reported by Sobolev and Perevoshchikov (2016). For Si the spectral distribution between 0 and 6 eV of ε' and ε'' (also written ε_1 and ε_2) and their relation to the experimentally obtained optical constants (see Eqs. 22 and 24 of ▶ Chap. 10, "Interaction of Light with Solids") is shown in Fig. 3.

Advances in ab initio calculations – mostly within density-functional theory and local-density approximation (DFT-LDA, Sect. 2.2.2 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals") – made it possible to compare the optical absorption spectrum calculated from first principles with experiment. Commonly LDA eigenvalues and eigenfunctions are used to calculate ε''. Several corrections

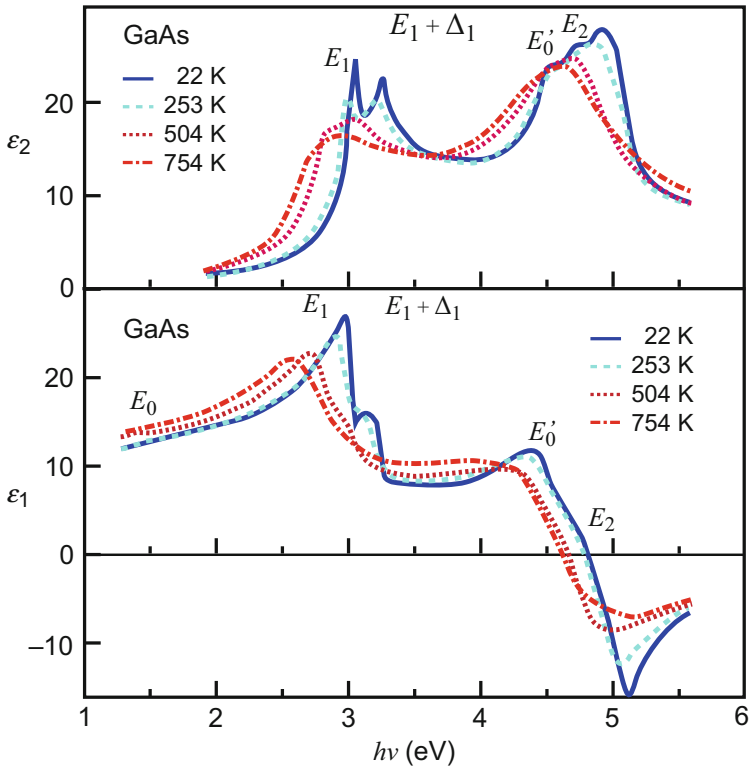


Fig. 4 Real and imaginary parts of the dielectric constant of GaAs as a function of the photon energy with the temperature as the family parameter (After Lautenschlager et al. 1987b)

are, however, to be made to obtain quantitative agreement with experiment. These are self-energy corrections (Sect. 2.2.3 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"; Hybertsen and Louie 1985; Aulbur et al. 2000) for correcting the underestimation of the energy gap in LDA and excitonic effects associated with Coulomb interaction between electron and hole (Albrecht et al. 1998).

The influence of temperature on the principal bandgap was discussed in Sect. 2.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids." An example for changes within higher bands is shown in Fig. 4 for the real and imaginary parts of the dielectric constant for GaAs (for Si see Lautenschlager et al. 1987a; Shkrebtii et al. 2010).

Absorption from Core Levels When the excitation takes place from a sharp core level rather than from a wide valence band, the structure of the band to which this transition occurs, the conduction band, can be obtained directly. When comparing excitation spectra from core levels with those from the valence band, however, a difference occurs because of the different relaxation of the excited state which is stronger when the hole is localized at the core than when it is more widely spread out in the valence band (Zunger 1983).

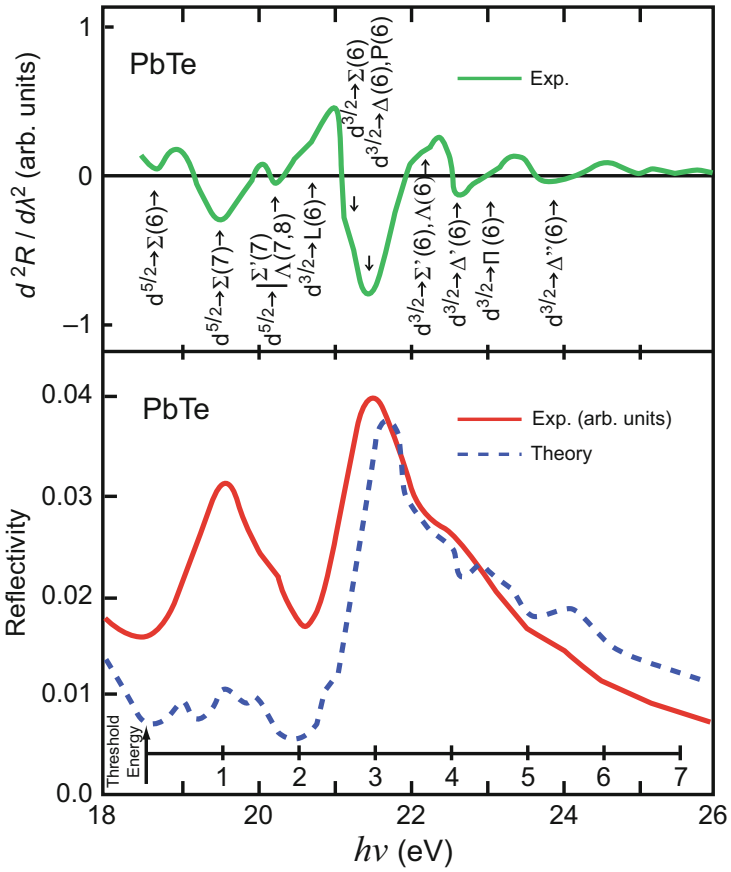


Fig. 5 Reflectivity of PbTe in the conduction-band range with excitation from the core levels $d^{3/2}$ and $d^{5/2}$ (distance from the conduction-band edge ~18.6 eV; see shifted scale.) The *upper* part of the figure shows the second derivative of the reflectivity and exposes more structure of the reflection spectrum. The transitions to the corresponding critical points are identified (After Martinez et al. 1975)

The joint density of states involving excitation from a core level is identical with the density of states in the conduction band. A typical example for a core-to-conduction-band spectrum of the reflectivity is given in Fig. 5. The experimental curves usually show less detail than those obtained from band-structure calculations (Martinez et al. 1975). This may be caused by electron–hole interaction and local field effects which are insufficiently accounted for in the single-electron approximation used for band calculation (Hanke and Sham 1974). These effects can be included following methods similar to those applied in valence–conduction-band transitions (Aulbur et al. 2000; Albrecht et al. 1998).

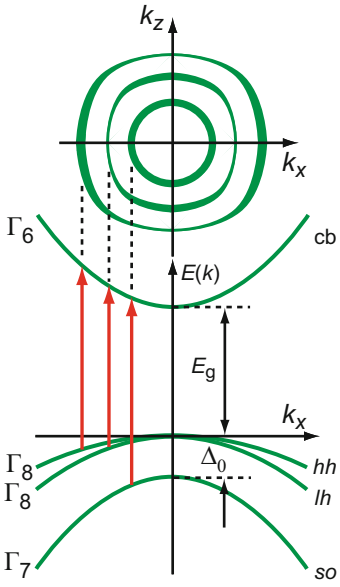


Fig. 6 Optical vertical excitation transitions for monochromatic photons at the band edge of GaAs near $k = 0$ from the heavy hole (*hh*), light hole (*lh*), and spin–orbit split-off (*so*) valence bands into the conduction band. The *upper* diagram indicates the resulting electron distribution in **k** space. The flattened outer rings indicate the strongly warped *hh* and the lesser warped *lh* valence bands (Sect. 1.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids"); the split-off valence band is not warped (After Lyon 1986)

1.3 The Fundamental Absorption Edge

Attention will now be focused on the energy range near the threshold for valence–conduction-band transitions, the *fundamental absorption edge*. Usually, several excitation processes are possible between different subbands for a given optical excitation energy. For a photon energy slightly exceeding the bandgap, transitions from different valence bands into different $E(k)$ values of the conduction band are possible near $k = 0$, as shown in Fig. 6. The distribution of these states is represented by nearly spherical shells (upper part of Fig. 6). Deviations from spherical shells occur because of the warping of the valence bands. Thermalization will average the excited electron distribution to approach a Boltzmann distribution (▶ Chap. 21, "Equilibrium Statistics of Carriers").

The transition between the top of one valence band to the bottom of the lowest conduction band when both lie at $k = 0$ is discussed first. This transition is responsible for the *direct optical absorption edge*. The joint density of states is estimated from a parabolic approximation of both bands:

$$h\nu = E_c(\mathbf{k}) - E_v(\mathbf{k}) = E_g + \frac{\hbar^2 k^2}{2m_n} + \frac{\hbar^2 k^2}{2m_p} = E_g + \frac{\hbar^2 k^2}{2\mu} \tag{14}$$

with E_g as the bandgap energy and μ as a reduced carrier mass

$$\frac{1}{\mu} = \frac{1}{m_n} + \frac{1}{m_p}. \tag{15}$$

Recognizing that Eq. 5 yields

$$J_{\mathrm{cv}} = \frac{2}{(2\pi)^3} \frac{d}{d(h\nu)} \left(\frac{4\pi}{3} k^3\right)$$

and using Eq. 14, we obtain for the joint density of states

$$J_{\mathrm{cv}} = \frac{1}{2\pi^2} \left(\frac{2\mu}{\hbar^2}\right)^{3/2} \sqrt{h\nu - E_{\mathrm{g}}}. \tag{16}$$

After introducing J_{cv} into Eq. 11, we obtain for the optical absorption coefficient near the band edge (Moss et al. 1973):

$$\alpha_{\mathrm{o,cv}} = \frac{e^2 (2\mu)^{3/2} |\mathbf{e} \cdot \mathbf{M}_{\mathrm{cv}}|^2}{\varepsilon_0 n_{\mathrm{r}}\, c m_0^2 h^3 \nu} \sqrt{h\nu - E_{\mathrm{g}}}. \tag{17}$$

The momentum matrix element can be approximated for transitions from the three valence bands in zincblende-type semiconductors across the bandgap (near E_{g}) from the Kane estimate (Kane 1957) when using

$$|\mathbf{e} \cdot \mathbf{M}_{\mathrm{cv}}|^2 \cong \left(\frac{m_0}{\hbar}\right)^2 P^2, \tag{18}$$

where P is the interband (momentum matrix) parameter, which can be obtained from the $\mathbf{k} \cdot \mathbf{p}$ perturbation theory (Sect. 2.1.1 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals" and Sect. 1.2.8 of ▶ Chap. 8, "Bands and Bandgaps in Solids"):

$$P^2 = \left(\frac{1}{m_n^0} - \frac{1}{m_0}\right) \frac{3\hbar^2}{2} \frac{E_{\mathrm{g}} + \Delta_0}{3E_{\mathrm{g}} + 2\Delta_0} E_{\mathrm{g}} \tag{19}$$

(compare with Eq. 33 of ▶ Chap. 8, "Bands and Bandgaps in Solids"). Here Δ_0 is the spin–orbit splitting energy and m_n^0 the effective mass at the bottom of the conduction band. With known m_n, E_{g}, and Δ_0, we obtain a numerical value for P, and therefore for the matrix element,[5]

[5]The momentum matrix-element with dimension $\mathrm{W^2 s^4 cm^{-2}}$ should not be confused with the often used oscillator strength

$$f_{\mathrm{cv}} = \frac{2}{3m_0 h\nu} |\mathbf{e} \cdot \mathbf{M}_{\mathrm{cv}}|^2, \tag{20}$$

which is dimensionless and of the order of one, while the matrix element is not. The factor 1/3 in Eq. 19 is due to averaging, with $|M_x|^2 = |M_y|^2 = |M_z|^2 = 1/3|\mathbf{M}|^2$; factor 2 accounts for the spin.

$$|\mathbf{e}\cdot\mathbf{M}_{\mathrm{cv}}|^2=\frac{3m_0}{2m_n^0}\ \left(m_0-m_n^0\right)\frac{E_\mathrm{g}+\Delta_0}{3E_\mathrm{g}+2\Delta_0}E_\mathrm{g}. \tag{21}$$

When this is introduced into Eq. 17, we obtain for $h\nu\cong E_\mathrm{g}$ and $m_n^0\cong m_n$:

$$\alpha_{\mathrm{o,cv}}\cong\frac{3\,e^2\sqrt{2m_0}}{\varepsilon_0\,n_\mathrm{r}\,h^2c}\left(\frac{m_n\ \ m_p}{m_0\left(m_n+m_p\right)}\right)^{3/2}\frac{m_0-m_n}{m_n}f\left(E_\mathrm{g}\right) \tag{22}$$

with

$$f\left(E_\mathrm{g}\right)=\frac{E_\mathrm{g}+\Delta_0}{3E_\mathrm{g}+2\Delta_0}\sqrt{h\nu-E_\mathrm{g}}, \tag{23}$$

which increases proportional to the square root of the energy difference from the band edge:

$$\alpha_{\mathrm{o,cv}}=\alpha_{\mathrm{o,dir}}\propto\sqrt{h\nu-E_\mathrm{g}}. \tag{24}$$

2 Direct and Indirect Transitions

The optical transitions discussed in the previous section are *direct transitions*; they involve a direct optical excitation from the valence to the conduction band, using only photons. This is represented by a vertical transition in the reduced $E(\mathbf{k})$ diagram, since the momentum of a photon, which is added to or subtracted from the electron momentum when a photon is absorbed or emitted, is negligibly small: the ratio of photon to electron momentum can be estimated from $p_{\mathrm{ph}}/p_{\mathrm{el}}=(h\nu/c)/(\hbar k)$ which, at the surface of the Brillouin zone, is $\cong(h/\lambda)/(\hbar\pi/a)=2a/\lambda\cong 10^{-3}$ for visible light.

2.1 Indirect Transitions

Nonvertical transitions are possible when, in addition to the photon, *phonons* are absorbed or emitted during the transition. These transitions are *indirect transitions*. The phonons can provide a large change in the electron momentum $\mathbf{k}$ (the phonon momentum is identified by $\mathbf{q}$):

$$\mathbf{k}'=\mathbf{k}\pm\mathbf{q}. \tag{25}$$

Because of the necessity of finding a suitable phonon for the optical transition to obey energy and momentum conservation for the transition, the probability of an indirect transition is less than that of a direct transition by several orders of magnitude. An estimate of the matrix elements for phonon-assisted electron

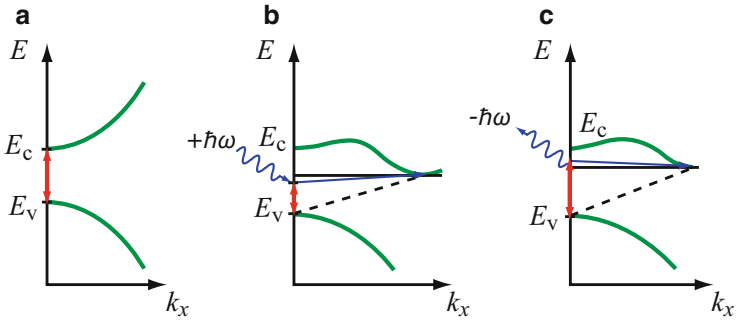


Fig. 7 (**a**) Direct and (**b**) indirect transitions with absorption and (**c**) with emission of a phonon

transitions follows the corresponding selection rules including the symmetry of the phonon (Lax and Hopfield 1961).

Indirect transitions identify the bandgap for materials where the lowest energy minimum of the conduction band is not at the same **k** as the highest maximum of the valence band. For instance, the highest maximum of the valence band is at $\mathbf{k} = 0$, but the lowest minimum of the conduction band is at $\mathbf{k} \neq 0$ (Fig. 7). These materials are *indirect bandgap semiconductors*; Si, Ge, GaP, AlAs, and AlSb are examples. Most other semiconductors are *direct gap semiconductors*.

When, together with the photon, a phonon is absorbed, we obtain for the absorption constant

$$\alpha_o^{abs} = \frac{2\pi e^2 f_{BE}|M|^2}{\varepsilon_0 n_r c m_0^2 2\pi\nu} \iint_{BZ} \frac{2d\mathbf{k}_1 2d\mathbf{k}_2}{(2\pi)^3 (2\pi)^3} \delta(E_c(\mathbf{k}_2) - E_v(\mathbf{k}_1) - h\nu + \hbar\omega), \qquad (26)$$

where f_{BE} is the Bose–Einstein distribution function for the phonon and M is the matrix element for the simultaneous absorption of a phonon and a photon (Bassani and Pastori Parravicini 1975). Using a parabolic band approximation and integrating over the delta function yields

$$\alpha_o^{abs} = \frac{2\pi e^2 f_{BE}|M|^2}{\varepsilon_0 n_r c m_0^2 2\pi\nu} \frac{1}{8(2\pi)^3} \left(\frac{2m_p}{\hbar^2}\right)^{3/2} \left(\frac{2m_n}{\hbar^2}\right)^{3/2} (h\nu - E_g + \hbar\omega)^2. \qquad (27)$$

A similar value for the absorption constant is obtained when a phonon is emitted, except $+\hbar\omega$ is replaced by $-\hbar\omega$ and f_{BE} is replaced by $1 - f_{BE}$. The total optical absorption of the indirect transition results as the sum

$$\alpha_{o,\,ind} = \alpha_o^{abs} + \alpha_o^{emi}. \qquad (28)$$

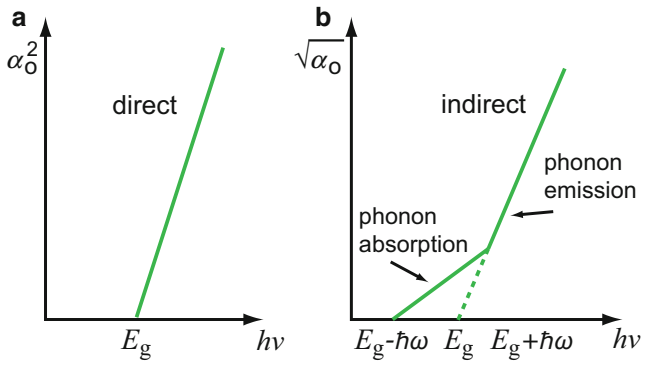


Fig. 8 Simple theoretical behavior of the absorption coefficient for (**a**) direct and (**b**) indirect transitions

It has a quadratic dependence on the photon energy (for more details, see Moss et al. 1973):

$$\alpha_{o,\,ind} \propto f_{BE}\left(h\nu + \hbar\omega - E_g\right)^2 + \left(1 - f_{BE}\right)\left(h\nu - \hbar\omega - E_g\right)^2, \tag{29}$$

where $\hbar\omega$ is the energy of a phonon of proper momentum and energy. The two branches of indirect transitions (Eq. 29 and Fig. 8) show a different temperature dependence. The phonon absorption branch vanishes at low temperatures, when the phonons are not thermally excited (frozen-out).

Weak indirect transitions, resulting in a factor $\sim 10^3$ lower absorption, can be observed only in a wavelength range where they do not compete with direct transitions. Indirect transitions are followed at higher energies by direct transitions with a secondary band edge. Figure 9 shows an example for Ge where the indirect band edge is preceding the direct edge.

2.2 Allowed and Forbidden Transitions

The matrix element (Eq. 2) contains two terms; the second one was neglected in the previous discussions. The one described before relates to an *allowed transition* and has a value close to 1 for its oscillator strength; here the selection rules are fulfilled. The other term relates to a *forbidden transition*; it represents a transition in which the Bloch function of the electron wave is orthogonal to the electric polarization of the light. The forbidden transition becomes finite but small when the electron momentum changes slightly during a transition from the ground to the excited state, which is caused by the finite, small momentum of the photon. It becomes important for the agreement between theory and experiment for the absorption spectrum further away from the band edge (Johnson 1967). This is indicated in Fig. 10b for the example of InSb.

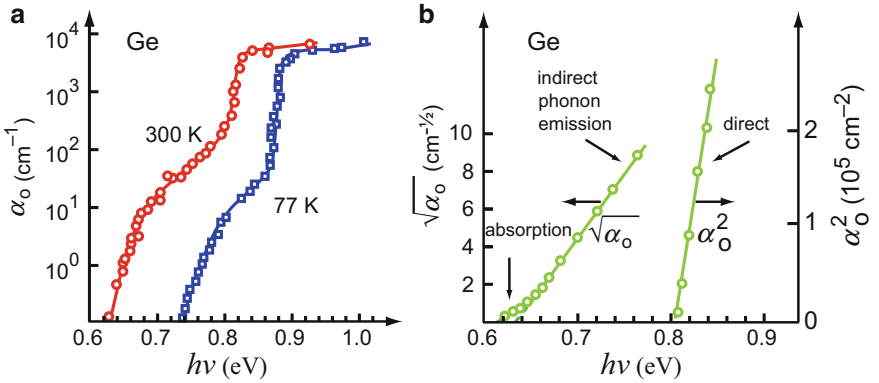


Fig. 9 (**a**) Absorption coefficient near the band edge for Ge (After Dash and Newman 1955). (**b**) Analysis of the *lower* ($10^{-1} \ldots 10^{1}$ cm^{-1}) and *upper* ($10^{1} \ldots 5 \times 10^{2}$ cm^{-1}) part of the absorption range at 300 K, giving evidence for indirect and direct absorption edges following each other at 0.66 and 0.81 eV, respectively (After Bube 1974)

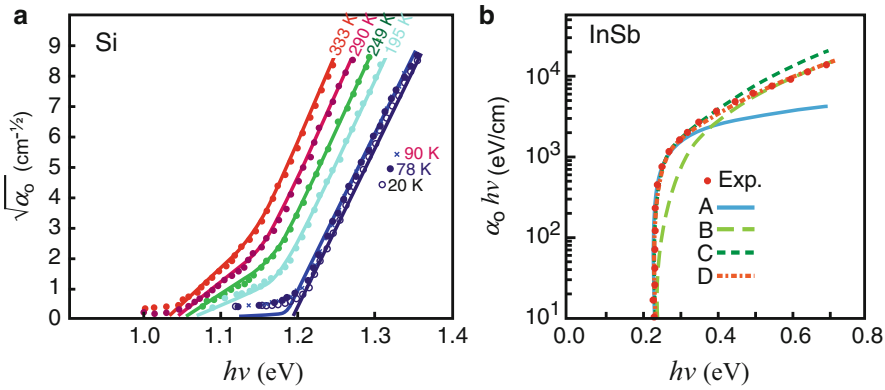


Fig. 10 (**a**) Absorption coefficient of Si with temperature as the family parameter indicating decreasing contribution of the phonon absorption branch with decreasing T (After MacFarlane and Roberts 1955). (**b**) Absorption coefficient near the band edge of InSb; (*A*) direct allowed transitions; (*B*) direct forbidden transitions; (*C*) as curve (*A*), however, corrected for nonparabolic bands; (*D*) as curve (*C*), however, corrected for the k dependence of matrix elements (After Johnson 1967)

The matrix element for such a forbidden transition may be calculated from the $\mathbf{k} \cdot \mathbf{p}$ perturbation theory with

$$\mathbf{M}_{cv}(\mathbf{k}) = (\mathbf{k} - \mathbf{k}_0)\,|\partial/\partial\mathbf{k}(\mathbf{M}_{cv}(\mathbf{k}))|_{\mathbf{k}=\mathbf{k}_0}. \tag{30}$$

This transition-matrix element increases for larger values of $\mathbf{k}$ and therefore provides a larger contribution to the optical absorption (see Fig. 10b).

Forbidden transitions are always additional components for both direct and indirect transitions. The absorption coefficient for these transitions varies as a function of photon energy as (Bardeen et al. 1956)

$$\alpha_{\mathrm{o,\,dir,\,forb}} = A_{\mathrm{df}}\left(h\nu - E_{\mathrm{g}}\right)^{3/2}, \tag{31}$$

$$\alpha_{\mathrm{o,\,ind,\,forb}} = A_{\mathrm{if}}\left(h\nu \pm \hbar\omega - E_{\mathrm{g}}\right)^{3}. \tag{32}$$

The proportionality coefficient for forbidden direct transitions is given for parabolic bands as

$$A_{\mathrm{df}} = \frac{2\pi^2 e^2 (2\mu)^{3/2}}{3\varepsilon_0 n_{\mathrm{r}} c m_0^2 h^5 \nu} \left|\partial/\partial\mathbf{k}(\mathbf{M}_{\mathrm{cv}})\right|^2_{\mathbf{k}=\mathbf{k}_0} \tag{33}$$

(Moss et al. 1973).

Transitions from Different Valence Bands Transitions from the two upper valence bands, which are split at $\mathbf{k} = 0$ in *anisotropic lattices*, can be separated by using polarized light. For instance, the band-edge transitions from the two valence bands, which are separated by crystal-field splitting in CdS (Sect. 1.2.3 of ▶ Chap. 8, "Bands and Bandgaps in Solids"), can be identified. When polarized light with its electric vector perpendicular to the **c** axis is used, transitions from one band are allowed, whereas transitions from the other band are only allowed when the electric vector is parallel to the **c** axis (Fig. 11). Transitions from the spin–orbit split-off band overlap and are more difficult to separate because they cannot be turned off individually.

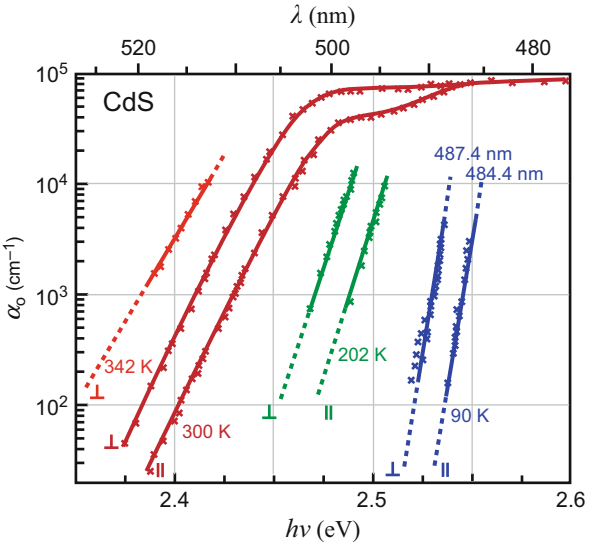


Fig. 11 Absorption edge of CdS measured in polarized light at various temperatures (After Dutton 1958)

In anisotropic semiconductors, certain band degeneracies in valence and conduction bands are removed and the optical absorption depends on the relative orientation of the crystal with respect to the light beam and its polarization. Only such transitions which have a component of the dipole orientation in the direction of the electrical vector of the light can be excited. Consequently, the absorption spectrum becomes angle dependent: it is anisotropic. An example for such an anisotropy related to valence-band splitting was given above.

2.3 Band-to-Band Magnetoabsorption

In a magnetic field, band degeneracies are lifted. Band states are split into different *Landau levels*. Such a splitting is proportional to the magnetic induction B. It is described in more detail in Sect. 2 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients." Direct band-to-band transitions at $\mathbf{k} \cong 0$ in a magnetic field are then given by transitions between the different Landau levels of the valence and conduction bands.

For a simple two-band model, the splitting of the bands is shown in Fig. 12a. The allowed transitions near $\mathbf{k} = 0$ are – for the field B_z along the z axis – given by

$$\begin{aligned}\Delta E &= E_g + \left(n_q + \tfrac{1}{2}\right)\ (\hbar\ \omega_e + \hbar\ \omega_h) + \Delta m_j\,(g_e + g_h)\,\mu_B\,B_z + \frac{\hbar^2 k^2}{2\mu} \\ &= E_g + \left(n_q + \tfrac{1}{2}\right)\ \frac{\hbar\ eB_z}{\mu} + \Delta m_j\,g_{tot}\,\mu_B\,B_z + \frac{\hbar^2 k^2}{2\mu}\quad,\quad n_q = 0, 1, 2, 3, \ldots,\end{aligned} \tag{34}$$

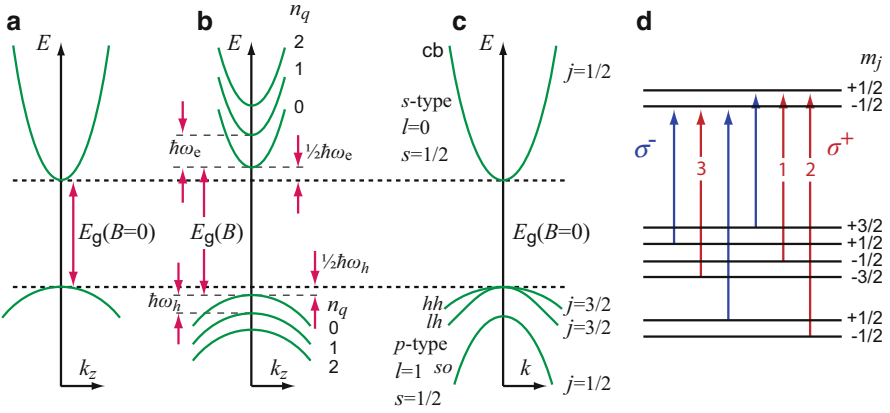


Fig. 12 (**a**) Band structure near $k = 0$ for single bands (**a**) without and (**b**) with a magnetic induction B; B splits the bands into equidistant Landau levels. (**c**) Band structure for zincblende semiconductors. (**d**) Allowed transitions for spin-split Landau levels of one pair of valence and conduction bands with equal quantum number n_q. Positive g factors are assumed for the splitting; numbers at the transitions indicate the absorption-intensity ratio

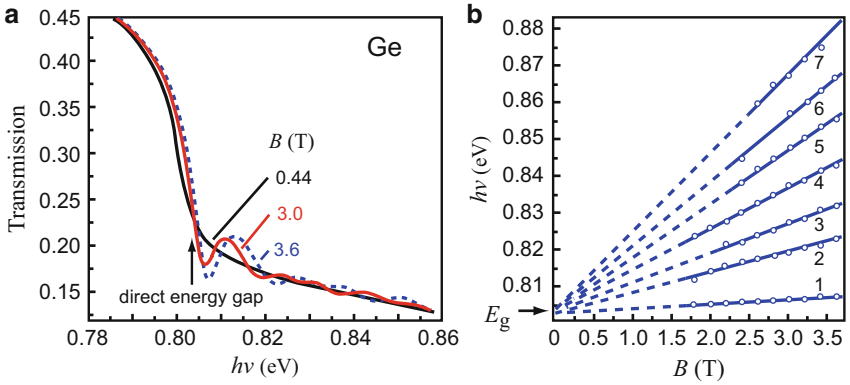


Fig. 13 (**a**) Transmission spectrum of Ge with minima at absorbing transitions. (**b**) Energy of principal minima in panel (**a**) as a function of the magnetic induction (After Zwerdling et al. 1957)

where n_q is the quantum number denoting the Landau levels (equal in the valence and conduction band for allowed transitions), $\hbar\omega_{e,h} = eB/m_{e,h}$ is the cyclotron-resonance energy for electrons or holes, and $\Delta m_j = \Delta(m_l + s_z)$ is the difference in the respective azimuthal plus spin quantum numbers of the two bands; μ_B is the Bohr magneton, and μ is the reduced effective mass $\left(m_n^{-1} + m_p^{-1}\right)^{-1}$. Here g is the respective Landé g factor (see also Sect. 2 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients"), which is derived from the g tensor (Yafet 1963); it depends on the angle θ between the magnetic induction and the principal axis of the band ellipsoid:

$$g = \sqrt{g_\parallel^2 \cos^2\theta + g_\perp^2 \sin^2\theta}, \text{ where } g_\parallel = g_{zz} \text{ and } g_\perp = g_{xx} = g_{yy}.$$

The transitions are subject to *selection rules* which depend on the polarization of the radiation. These selection rules are

$$\begin{aligned} &\text{For } \mathbf{E}\|\mathbf{B}: \quad \Delta k_z = 0, \quad \Delta n_q = 0, \quad \Delta m_j = 0 \\ &\text{For } \mathbf{E}\perp\mathbf{B}: \quad \Delta k_z = 0, \quad \Delta n_q = 0, \quad \Delta m_j = \pm 1, \end{aligned} \tag{35}$$

for right and left circular polarized light with $\Delta m_j = +1$ and -1 indicated as σ^+ and σ^-, respectively. The intensity ratio of the different transitions is indicative of the original degree of spin orientation. For instance, for equidistributed spin of conduction electrons, the intensity ratio of absorption for right polarized light ($-\frac{3}{2} \rightarrow -\frac{1}{2}$ to $-\frac{1}{2} \rightarrow +\frac{1}{2}$) is 3:1; see Sect. 3 of ▶ Chap. 9, "Magnetic Semiconductors."

In typical semiconductors, the valence band is p-like and the conduction band is s-like (Fig. 12c). There are hence four possibilities at $k = 0$ for the angular momentum quantum number $m_j = m_l + s_z$ of the valence band: $m_j = -3/2$, $-1/2$, 1/2, and 3/2. Those of the conduction band are $-1/2$ and +1/2. For $\mathbf{E}\perp\mathbf{B}$ and $s_z = -1/2$, we have transitions from $m_j = +1/2$ and $-3/2$; for $s_z = 1/2$, we have transitions from $m_j = +3/2$ and $-1/2$. The permitted transitions are shown in

Fig. 12d; such a scheme applies for each pair of Landau levels with equal quantum numbers n_q for valence and conduction band.

The general behavior can be seen in Fig. 13a with Landau levels of Ge at room temperature identified. They shift linearly with the magnetic induction (Fig. 13b) and permit a measurement of the reduced effective mass μ^{-1} from the slopes, and the bandgap from an extrapolation[6] to $B \to 0$. At low temperatures, the transitions from the light and heavy-hole bands and from the split-off band are resolved (Zawadzki and Lax 1966; Roth et al. 1959). Oscillatory magnetoabsorption was observed for many semiconductors, e.g., for GaAs (Vrehen 1968), InN (Millot et al. 2011), InP (Rochon and Fortin 1975), ZnSe (Aliev et al. 1996), GaSe (Watanabe et al. 2003), or InSe (Millot et al. 2011). An analysis of such spectra also provides information about band anisotropies (Sari 1972). For a review, see Mavroides (1972). Interband transitions between Landau levels were also observed in emission spectra (Gubarev et al. 1993).

3 Transitions in Quantum Wells

One of the most direct confirmations of the simple quantum-mechanical model is obtained from the measurement of the optical spectra in low-dimensional structures (quantum wells, quantum wires, or quantum dots). Since the limited interaction volume of a single structure leads to only weak absorption, usually either absorption spectra of multiple similar structures[7] or photoluminescence excitation spectra are measured.

3.1 Energy Levels in Multiple Quantum Wells

Optically induced absorption transitions in multiple quantum wells give an instructive picture of the energy-level structure and reflect selection rules.[8] A single quantum well contains a set of energy levels that can approximately be calculated from a square-well potential with infinite barrier height, yielding

$$E_n = \frac{\hbar^2 \pi^2}{2 m_n l_{\mathrm{QW}}^2} n^2, \; n = 1, \; 2, \; 3, \; \ldots \; , \tag{36}$$

see Eq. 54 of ▶ Chap. 8, "Bands and Bandgaps in Solids."

[6]For semiconductors without inversion symmetry (e.g., GaAs), the valence-band extrema do not occur exactly at $k = 0$, and the extrapolations of the measured lines in Fig. 13b do not precisely meet in one point at $B = 0$; a detailed analysis is given by Zwerdling et al. (1957).

[7]Such structures are multiple quantum wells or arrays of quantum wires with a sufficient thickness of the separating barrier material or ensembles of quantum dots.

[8]For a study of quantum wires, see, e.g., Ihara et al. (2007).

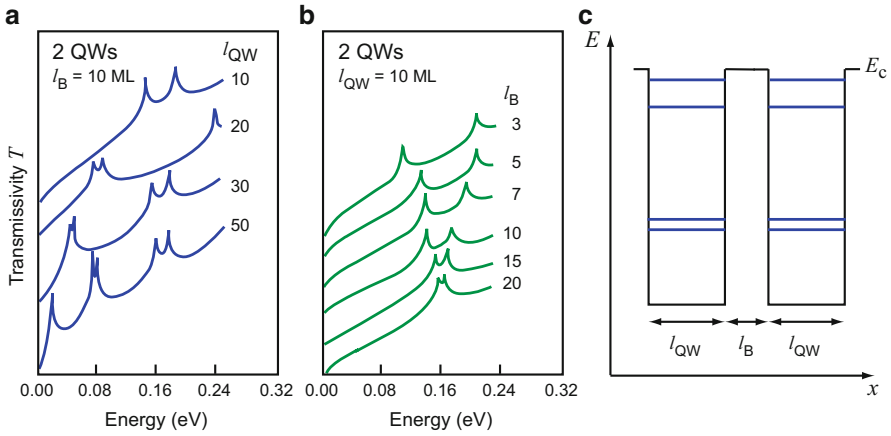


Fig. 14 Optical transmission spectrum of two GaAs quantum wells coupled by one $Al_{0.35}Ga_{0.65}As$ barrier layer with (**a**) variation of the well thickness l_{QW} and (**b**) variation of the barrier thickness l_B expressed in numbers of monolayers (ML; 10 ML ≅ 28 Å) and showing the split into two levels for each quantum state. (**c**) Schematics of the well and barrier geometry for a 30/10/30 ML structure, the third curve in (**a**) (After Torabi et al. 1987)

When multiple wells in a periodic superlattice are separated by barriers with a small thickness, their states interact and the energy levels start to split. Eventually, the interaction of many wells leads to *mini-bands* (see Sect. 3.1.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids"). A development of mini-bands is shown in Fig. 14. Two quantum wells are formed by quantum well (QW) layers of GaAs and are separated by one barrier (B) layer of $Al_{0.35}Ga_{0.65}As$. In a sequence of experiments, the well or barrier width was varied. For each well the transitions show a shift toward higher energies according to Eq. 36 and wider spacing of the lines with decreased *well width* l_{QW}. In addition, the lines split and the splitting increases with reduced *barrier width* l_B, as shown in Fig. 14b. With decreasing barrier width, these lines also become broader (Schulman and McGill 1981); here, the interaction between the wells becomes more probable (see Sect. 3.1.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids"). The observation agrees with the calculation of a simple one-dimensional square well/barrier potential.

When more wells in a periodic superlattice structure can interact, the lines spread to mini-bands: due to the interaction of a sufficient number of layers, the discrete features shown in Fig. 14 are broadened into continuous mini-bands, as indicated in a sequence of curves in Fig. 15 and discussed in the following section. The mean free path of electrons, however, gives their coherence length across which superlattice periodicity is recognized.

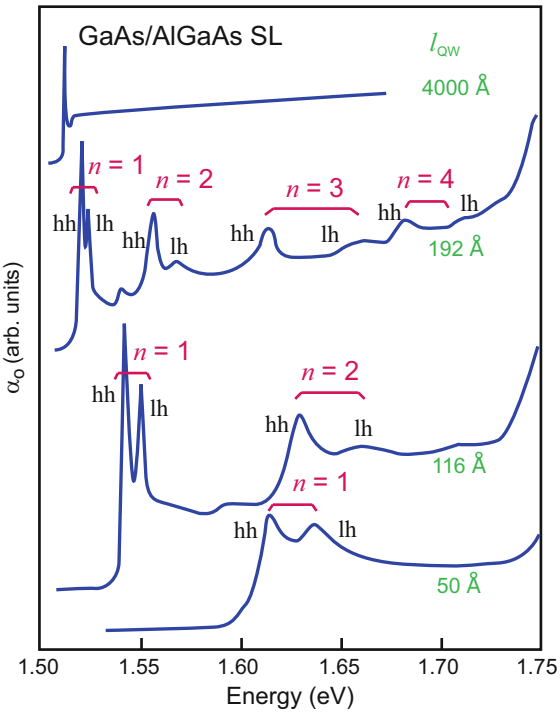


Fig. 15 Optical absorption spectrum of GaAs/GaAlAs superlattices with different lattice constants. The doublets are due to transitions from light and heavy hole mini-bands. The $l_{QW} = 4000$ Å sample shows bulk properties with the exciton peak at 1.513 eV – see Fig. 8 of ▶ Chap. 14, "Excitons" (After Dingle et al. 1974)

3.2 Absorption in Quantum Wells

When the well width is very large, the optical absorption spectrum for superlattices is identical to that of the bulk (well) material (Fig. 15, uppermost curve). With decreasing well width, the onset of optical absorption shifts to higher energies, and relatively sharp lines appear within the band of the well material, as shown in the second and third curves. The lines correspond to the eigenstates of conduction electrons in this well, which for a simple square well are given in Eq. 36, indicating a $1/l_{QW}^2$ dependence (Dingle 1975; Miller et al. 1980; Pinczuk and Worlock 1982). The bound eigenstates penetrate only about 25 Å into the AlGaAs barriers, which are 250 Å thick and hence separate the wells. The doublets shown in this figure are due to transitions from the light- and heavy-hole mini-bands which were omitted in Fig. 16 for clarity. The lines become broader and mini-bands develop as the wells and the barriers become thinner (lowest curve in Fig. 15).

The envelope functions of the different eigenstates of valence and conduction electrons within the well are shown in Fig. 16a for a deep, isolated quantum well. For the optical transitions, the figure indicates maximum overlap between eigenfunctions of the same subband index n, while transitions between subbands of different indices have rather small transition probabilities. This is the basis for the corresponding selection rules.

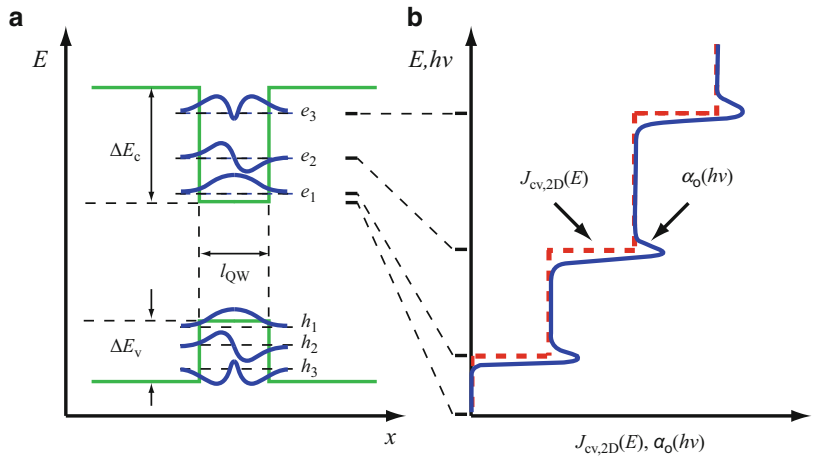


Fig. 16 (**a**): Envelope functions of three eigenstates in a single quantum well. (**b**) Corresponding joint density of states from valence- to conduction-band transitions (*dashed curve*) and absorption coefficient including a simple excitonic feature preceding each step (*solid curve*)

The two-dimensional joint density-of-state function (Fig. 16b) is steplike for quantum wells (or superlattices with wide barriers). These steps become smooth when mini-bands develop and broaden with decreasing barrier width – see Fig. 38 of ▶ Chap. 8, "Bands and Bandgaps in Solids." Due to excitonic contributions to absorption (see Sect. 2.1 of ▶ Chap. 14, "Excitons"), peaks occur slightly below each subband threshold, which causes a modified, increased absorption at each lower edge shown by the solid curve of Fig. 16b. This is in general agreement with the experiment – see Fig. 15 and also Sect. 2.1 of ▶ Chap. 14, "Excitons."

The confinement of the electron eigenfunctions in narrow mini-bands yields high oscillator strength for optical transitions. The optical absorption into the mini-bands is a direct one, even though the well material may have an indirect bandgap.

4 Optical Bandgap of Amorphous Semiconductors

4.1 Intrinsic Absorption

The optical bandgap of amorphous semiconductors is less well defined than in a crystalline material (Sect. 4.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids"), due to substantial tailing of defect states into the bandgap (Mott and Davis 1979; Sa-yakanit and Glyde 1987; see also Sect. 1 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors"). Usually, an *effective bandgap* is taken from the optical absorption spectrum plotted in an $(\alpha_o\, hv)^{1/2}$ versus $(hv - E_g)$ presentation and extrapolated to $\alpha_o = 0$ or, for practical purposes in thin films, presented as E_{04} – the energy is used at which α_o reaches a value of 10^4 cm^{-1}. Plotting the absorption

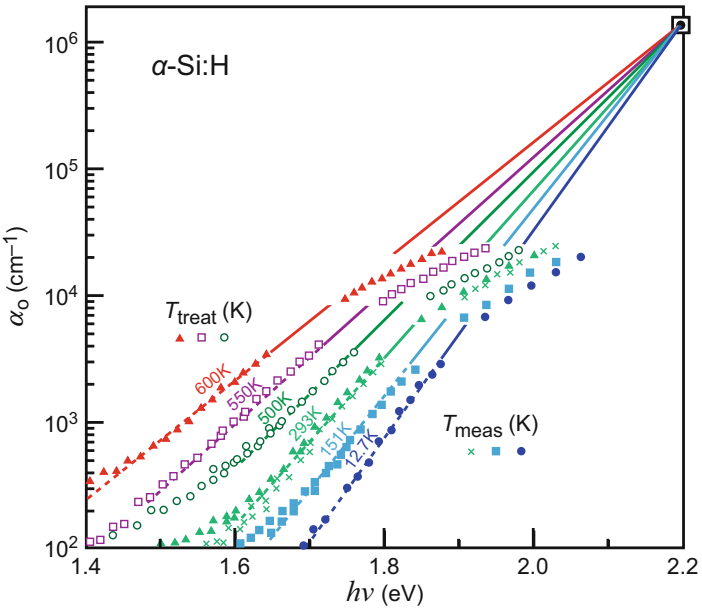


Fig. 17 Optical absorption of α-Si:H as a function of the photon energy with measurement temperature T_{meas} or treatment temperature T_{treat} indicated, pointing to an additive effect of temperature and structural disorder (After Cody et al. 1981)

coefficient versus bandgap energies in a semilogarithmic graph, we obtain a straight line and deduce from its slope the *Urbach parameter* E_0:

$$\alpha_o(\nu) = \alpha_{o,0}\exp\left(\frac{h\nu - E_g}{E_0}\right), \tag{37}$$

which gives the steepness of the level distribution near the band edge.[9] More recently also a model of an *average* dipole matrix element for tailing states was applied to obtain the optical gap and tailing widths of conduction and valence band from optical absorption; the study indicates a linear dependence of tailing on disorder in α-Si:H (Li et al. 2018).

A particularly striking example of band tailing of α-Si:H is given in Fig. 17 which demonstrates the relation between an increased severity of disorder and the slope of the band tailing (Cody et al. 1981; see also Canul et al. 2020). Such

[9]An exponential Urbach tail (with Urbach energy in the meV range) was also observed in absorption spectra of high-quality GaAs/AlGaAs quantum wells (Bhattacharya et al. 2015). The broadening is assigned to disorder originating from the electric field of zero-point oscillations of LO phonons in polar semiconductors, yielding a fundamental limit to the Urbach slope. In ternary InGaN/GaN quantum wells, Urbach tails – also known from bulk alloys – could likewise clearly be assigned to compositional disorder (Piccardo et al. 2017).

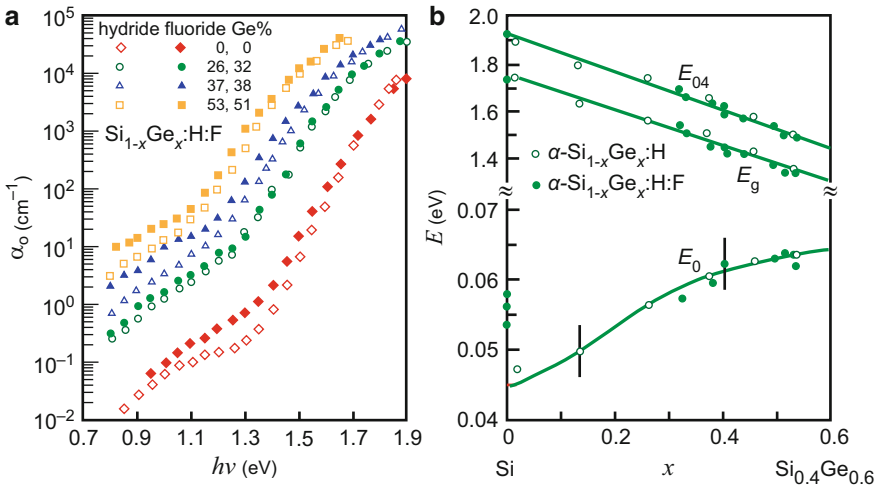


Fig. 18 (**a**) Optical absorption spectra of α-$Si_{1-x}Ge_x$:H and α-$Si_{1-x}Ge_x$:F. (**b**) Variation of the bandgap energies with composition (After Mackenzie et al. 1988)

increased severity is caused by a heat treatment of α-Si:H at temperatures above 700 K, when hydrogen is released: thereby dangling bonds are created which act as major defects (see Sect. 2.2.3 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors"). The tailing, described by Eq. 37, increases from E_0 = 50–100 meV with increasing treatment temperature. Recent modeling indicates that the bandgap energy is not fixed but depends somewhat on hydrogen saturation (Legesse et al. 2014).

Another example for hydrogenated (or fluorinated) amorphous Si and $Si_{1-x}Ge_x$ alloys is shown in Fig. 18. The α-Si:H shows a relatively steep absorption edge, indicating direct absorption; the slope decreases with alloying.

The hydrogenated α-$Si_{1-x}Ge_x$ alloys show little bowing in spite of the large lattice mismatch between Si and Ge. This is characteristic for one-atomic semiconductors which show less restraint against lattice deformation than two-atomic or higher-atomic lattices.

4.2 Extrinsic Absorption in Glasses

The absorption of amorphous materials below the band edge decreases rapidly, following the Urbach tail, and reaches values that can be much lower than those for high-purity crystals. It is difficult to make crystals with very low defect densities, while in amorphous structures, most defects, i.e., wrong bonds or impurities, are incorporated into the host material and yield a much lower optical absorption far from the absorption edge. This makes glasses good materials for optical fibers in communication systems.

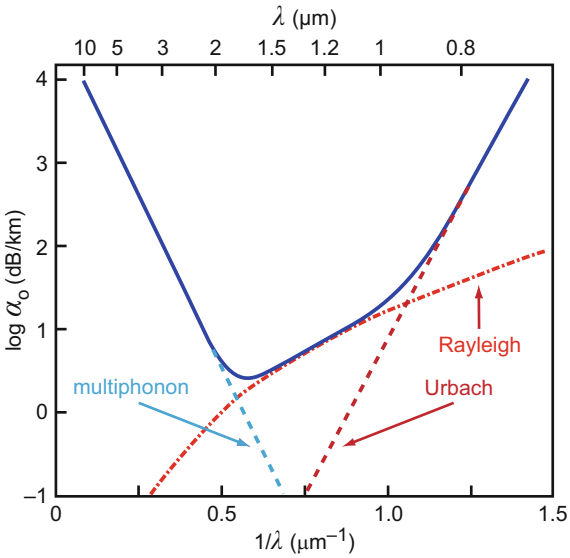


Fig. 19 Attenuation of light in intrinsic glasses as a function of the wavelength for a hypothetical glass (After Lines 1986)

Fused silica (SiO_2) is used at a length of up to 30 km between repeater stations with an absorption of ~0.16 dB/km,[10] corresponding to an absorption constant of $\sim 3.7 \times 10^{-7}$ cm^{-1}, compared to the lowest absorption of high-purity GaAs of 2×10^{-3} cm^{-1} (Lines 1986). Optical absorption and pulse spreading because of finite dispersion ($\partial n_r/\partial\lambda$) are the limiting factors for the length of an optical transmission line (Lines 1986). Absorption losses are usually caused by larger inclusions, imperfections, and impurities and require purification in the range 10^{13} cm^{-3}, i.e., in the parts per billion (ppb) range. The remaining losses in SiO_2 fibers are due to water contamination, which is difficult to remove.

When fibers are prepared well enough so that only intrinsic phenomena limit the absorption, then three effects have to be considered:

1. The Urbach tail, responsible for electronic transitions, and extending in energy from the bandgap downward
2. Multiphonon absorption, extending from the *Reststrahl* frequency upward
3. Rayleigh scattering (see Sect. 3.3.1 of ▶ Chap. 11, "Photon–Phonon Interaction"), due to density fluctuation in the midfrequency range, as shown in Fig. 19

The lowest intrinsic absorption values predicted for SiO_2 are 0.1 dB/km at $\lambda = 1.55$ μm with ~90% of the attenuation stemming from Rayleigh scattering, ~10% from multiphonon absorption, and negligible contribution from the Urbach tail.

[10] 1 dB (*decibel*) is equal to the logarithm (base ten) of the ratio of two power levels having the value 0.1, that is, $\log_{10}(p_1/p_0) = 0.1$ or $p_1/p_0 = 1.259$. For the example given above of 0.16 dB/km, one measures a reduction of the light intensity by 3.8% per km. The intensity is reduced to 33% after 30 km.

5 Summary

The theory of optical transitions between the valence and conduction bands was sketched in rather general terms. Fermi's golden rule yields the transition probability from the momentum matrix-element and the joint density of states. The joint density of states has characteristic kinks or maxima at critical points; they occur at energies where the **k**-depending slopes of the contributing bands are equal. For transitions near the band edge, the theory can be simplified by assuming parabolic band shapes. This effective-mass approach yields quantitative expressions for the dielectric function and the absorption as a function of the photon energy. In quantum wells, the two-dimensional joint density of states leads to a steplike increase of the absorption for increasing photon energy, modified by additional excitonic absorption at the step edges.

Depending on the conduction-band behavior, one has strong direct or weak indirect transitions at the band edge; the latter require the participation of a phonon to fulfill energy and momentum conservation. Contributions of forbidden transitions additionally modify the absorption further away from the band edge. Depending on the slope of the optical absorption in the intrinsic range near the band edge, direct or indirect transitions can be identified unambiguously, and, with the help of an external magnetic field, the effective mass can be determined from the period of magneto-oscillations of the near-edge absorption due to Landau splitting of the band states.

Deviations from the ideal, periodic crystal lattice provide tailing states extending beyond the band edge, usually as an Urbach tail which decreases exponentially with distance from the band edge. The resulting tailing of the absorption coefficient below the bandgap energy is particularly pronounced for amorphous semiconductors.

References

Adagi S (2005) Properties of group-IV, III-V and II-VI semiconductors. Wiley, Chichester

Albrecht S, Reining L, Del Sole R, Onida G (1998) Ab initio calculation of excitonic effects in the optical spectra of semiconductors. Phys Rev Lett 80:4510

Aliev GN, Datsiev RM, Ivanov SV, Kop'ev PS, Seisyan RP, Sorokin SV (1996) Optics and magneto-optics of ZnSe heteroepitaxial layers. J Crystal Growth 159:523

Aulbur WG, Jonsson L, Wilkins JW (2000) Quasiparticle calculations in solids. Solid State Phys 54:1

Bardeen J, Blatt FJ, Hall LH (1956) Indirect transition from the valence to the conduction band. In: Breeckenridge R, Russel B, Halm T (eds) Proceedings of photoconductivity conference. Wiley, New York, p 146

Bassani GF (1966) Methods of band calculations applicable to III–V compounds. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 1. Academic Press, New York, p 21

Bassani GF, Pastori Parravicini G (1975) Electronic states and optical transitions in solids. Pergamon Press, Oxford

Bhattacharya R, Mondal R, Khatua P, Rudra A, Kapon E, Malzer S, Döhler G, Pal B, Bansal B (2015) Measurements of the electric field of zero-point optical phonons in GaAs quantum wells support the Urbach rule for zero-temperature lifetime broadening. Phys Rev Lett 114:047402

Brust D, Phillips JC, Bassani F (1962) Critical points and ultraviolet reflectivity of semiconductors. Phys Rev Lett 9:94

Bube RH (1974) Electronic properties of crystalline solids. Academic Press, New York
Canul A, Thapa D, Huso J, Bergman L, Williams RV, Machleidt R (2020) Mixed-strategy approach to band-edge analysis and modeling in semiconductors. Phys Rev B 101:195308
Cody GD, Tiedje T, Abeles B, Brooks B, Goldstein Y (1981) Disorder and the optical absorption edge of hydrogenated amorphous silicon. Phys Rev Lett 47:1480
Cohen ML, Chelikowsky JR (1988) Electronic structure and optical properties of semiconductors. Springer, Berlin
Dash WC, Newman R (1955) Intrinsic optical absorption in single-crystal germanium and silicon at 77°K and 300°K. Phys Rev 99:1151
Dingle R (1975) Confined carrier quantum states in ultrathin semiconductor hetero-structures. In: Queisser HJ (ed) Festkörperprobleme, vol 15, Advances in solid state physics. Pergamon/ Vieweg, Braunschweig, p 21
Dingle R, Wiegmann W, Henry CH (1974) Quantum states of confined carriers in very thin $Al_xGa_{1-x}As$-GaAs-$Al_xGa_{1-x}As$ heterostructures. Phys Rev Lett 33:827
Dutton D (1958) Fundamental absorption edge in cadmium sulfide. Phys Rev 112:785
Gubarev SI, Ruf T, Cardona M, Ploog K (1993) Magnetoluminescence of GaAs in the quasiclassical limit. Phys Rev B 48:1647
Hanke W, Sham LJ (1974) Dielectric response in the Wannier representation: application to the optical spectrum of diamond. Phys Rev Lett 33:582
Hybertsen MS, Louie SG (1985) First-principles theory of quasiparticles: calculation of band gaps in semiconductors and insulators. Phys Rev Lett 55:1418
Ihara T, Hayamizu Y, Yoshita M, Akiyama H, Pfeiffer LN, West KW (2007) One-dimensional band-edge absorption in a doped quantum wire. Phys Rev Lett 99:126803
Johnson EA (1967) Absorption near the fundamental edge. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 3. Academic Press, London, pp 153–258
Kane EO (1957) Band structure of indium antimonide. J Phys Chem Solids 1:249
Lautenschlager P, Garriga M, Vina L, Cardona M (1987a) Temperature dependence of the dielectric function and interband critical points in silicon. Phys Rev B 36:4821
Lautenschlager P, Garriga M, Logothetidis S, Cardona M (1987b) Interband critical points of GaAs and their temperature dependence. Phys Rev B 35:9174
Lax M, Hopfield JJ (1961) Selection rules connecting different points in the Brillouin zone. Phys Rev 124:115
Legesse M, Nolan M, Fagas G (2014) A first principles analysis of the effect of hydrogen concentration in hydrogenated amorphous silicon on the formation of strained Si-Si bonds and the optical and mobility gaps. J Appl Phys 115:203711
Li Z, Lin SH, Qiu GM, Wang JY, Yu YP (2018) A method for determining band parameters from the optical absorption edge of amorphous semiconductor: Application to α-Si:H. J Appl Phys 124:025702
Lines ME (1986) Ultralow-loss glasses. Annu Rev Mater Sci 16:113
Lyon SA (1986) Spectroscopy of hot carriers in semiconductors. J Lumin 35:121
Macfarlane GG, Roberts V (1955) Infrared absorption of silicon near the lattice edge. Phys Rev 98:1865
Mackenzie KD, Burnett JH, Eggert JR, Li YM, Paul W (1988) Comparison of the structural, electrical, and optical properties of amorphous silicon-germanium alloys produced from hydrides and fluorides. Phys Rev B 38:6120
Madelung O (1981) Introduction to solid state theory. Springer, Berlin/New York
Martinez G, Schlüter M, Cohen ML (1975) Electronic structure of PbSe and PbTe II – optical properties. Phys Rev B 11:660
Mavroides JG (1972) Magneto-optical properties. In: Abeles F (ed) Optical properties of solids. North Holland, Amsterdam
Miller RC, Kleinman DA, Nordland WA Jr, Gossard AC (1980) Luminescence studies of optically pumped quantum wells in GaAs-AlxGa1-xAs multilayer structures. Phys Rev B 22:863

Millot M, Ubrig N, Poumirol J-P, Gherasoiu I, Walukiewicz W, George S, Portugall O, Léotin J, Goiran M, Broto J-M (2011) Determination of effective mass in InN by high-field oscillatory magnetoabsorption spectroscopy. Phys Rev B 83:125204

Moss TS, Burrell GJ, Ellis B (1973) Semiconductor optoelectronics. Wiley, New York

Mott NF, Davis EA (1979) Electronic processes in non-crystalline materials. Claredon Press, Oxford, UK

Piccardo M, Li C-K, Wu Y-R, Speck JS, Bonef B, Farrell RM, Filoche M, Martinelli L, Peretti J, Weisbuch C (2017) Localization landscape theory of disorder in semiconductors. II. Urbach tails of disordered quantum well layers. Phys Rev B 95:144205

Phillips JC (1956) Critical points and lattice vibration spectra. Phys Rev 104:1263

Phillips JC (1966) The fundamental optical spectra of solids. Solid State Phys 18:55

Pinczuk A, Worlock JM (1982) Light scattering by two-dimensional electron systems in semiconductors. Surf Sci 113:69

Rochon P, Fortin E (1975) Photovoltaic effect and interband magneto-optical transitions in InP. Phys Rev B 12:5803

Roth LM, Lax B, Zwerdling S (1959) Theory of optical magneto-absorption effects in semiconductors. Phys Rev 114:90

Sari SO (1972) Excitonic effects in Landau transitions at the E_1 edges of InSb and GaSb. Phys Rev B 6:2304

Sa-yakanit V, Glyde HR (1987) Urbach tails and disorder. Comments Condens Matter Phys 13:35

Schulman JN, McGill TC (1981) Complex band structure and superlattice electronic states. Phys Rev B 23:4149

Shkrebtii AI, Ibrahim ZA, Teatro T, Richter W, Lee MJG, Henderson L (2010) Theory of the temperature dependent dielectric function of semiconductors: from bulk to surfaces. Application to GaAs and Si. Phys Status Solidi B 247:1881

Sobolev VV, Perevoshchikov DA (2016) Localization of the interband transitions in the bulk of the Brillouin zone of III–V compound crystals. Semiconductors 50:572

Torabi A, Brennan KF, Summers CJ (1987) Photoluminescence studies of coupled quantum well structures in the AlGaAs/GaAs system. Proc SPIE 0792:152

van Hove L (1953) The occurrence of singularities in the elastic frequency distribution of a crystal. Phys Rev 89:1189

Vrehen QHF (1968) Interband magneto-optical absorption in gallium arsenide. J Phys Chem Solids 29:129

Watanabe K, Uchida K, Miura N (2003) Magneto-optical effects observed for GaSe in megagauss magnetic fields. Phys Rev B 68:155312

Yafet Y (1963) *g* factors and spin-lattice relaxation of conduction electrons. In: Seitz F, Turnbull D (eds) Solid state physics, vol 14. Academic Press, New York, p 1

Zawadzki W, Lax B (1966) Two-band model for Bloch electrons in crossed electric and magnetic fields. Phys Rev Lett 16:1001

Zhen Li, Shun Hui Lin, Gui Ming Qiu, Jiang Yong Wang, and Yun Peng Yu (2018) A method for determining band parameters from the optical absorption edge of amorphous semiconductor: Application to a-Si:H. J Appl Phys 124:025702

Zunger A (1983) One-electron broken-symmetry approach to the core-hole spectra of semiconductors. Phys Rev Lett 50:1215

Zwerdling S, Lax B, Roth LM (1957) Oscillatory magneto-absorption in semiconductors. Phys Rev 108:1402

Excitons

14

Karl W. Böer and Udo W. Pohl

Contents

1 Optical Transitions of Free Excitons 530
1.1 Frenkel Excitons 532
1.2 Wannier–Mott Excitons 535
1.3 Exciton–Polaritons 547
1.4 Trions and Biexcitons in Bulk Crystals 553
2 Excitons in Low-Dimensional Semiconductors 555
2.1 Excitons in Quantum Wells 555
2.2 Biexcitons and Trions in Quantum Wells 558
2.3 Excitons in Quantum Wires and Quantum Dots 561
3 Confined Excitons in Quantum Optics 568
3.1 Representation of a Qubit 568
3.2 Qubit Photons from Semiconductor Quantum Dots 577
4 Summary 583
References 584

Abstract

Optical band-to-band absorption can produce an electron and a hole in close proximity which attract each other via Coulomb interaction and can form a hydrogen-like bond state, the exciton. The spectrum of free Wannier–Mott excitons in bulk crystals is described by a Rydberg series with an effective Rydberg constant given by the reduced effective mass and the dielectric constant. A small dielectric constant and large effective mass yield a localized Frenkel exciton resembling an excited atomic state. Excitons increase the

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_14

absorption slightly below the band edge significantly. The interaction of photons and excitons creates a mixed state, the exciton–polariton, with photon-like and exciton-like dispersion branches. An exciton can bind another exciton or carriers to form molecules or higher associates of excitons. Free charged excitons (trions) and biexcitons have a small binding energy with respect to the exciton state. The binding energy of all excitonic quasiparticles is significantly enhanced in low-dimensional semiconductors. Basic features of confined excitons with strongest transitions between electron and hole states of equal principal quantum numbers remain similar. The three-dimensional confinement of quantum dots allows for forming stable antibinding exciton associates. Excitonic states of quantum dots are a prominent physical basis for the realization of photonic qubits. The exciton emission is deployed for producing single photons on demand with qubits encoded into the polarization or another degree of freedom. Pairs of entangled photons are obtained from the biexciton-exciton cascade. Photons used as flying qubits enable inherently secure data transmission; semiconductor-based quantum optics is a highly active field of research and development.

Keywords

Band-to-band absorption · Biexciton · Binding energy · Bloch sphere · Bound exciton · Confined exciton · Coulomb interaction · Dark exciton · Dielectric constant · Effective mass · Electron-hole exchange interaction · Entanglement · Exciton complex · Exciton spectra · Excitonic quasiparticles · Exciton–polariton · Free exciton · Frenkel exciton · Indirect-gap exciton · Quantum information processing · Quasi-hydrogen states · Qubit · Rydberg constant · Rydberg series · Trion · Wannier-Mott exciton

1 Optical Transitions of Free Excitons

The transitions discussed in the previous chapters considered electron–hole pair generation by the absorption of a photon with bandgap energy E_g and neglected the Coulomb attraction between the created electron and hole. With photons of $h\nu \cong E_g$, both electron and hole do not have enough kinetic energy at low temperatures to separate. They form a bound state. This state can be modeled by an electron and a hole, circling each other much like the electron and proton in a hydrogen atom, except that they have almost the same mass;[1] hence, in a semiclassical model, their center of rotation lies closer to the middle on their interconnecting

[1]This causes the breakdown of the adiabatic approximation. The error in this approximation is on the order of the fourth root of the mass ratio. For hydrogen this is $(m_n/M_H)^{1/4} \cong 10\%$ and is usually acceptable. For excitons, however, the error is on the order of 1 and is no longer acceptable. This is relevant for the estimation of exciton molecule formation discussed in Sect. 1.4.

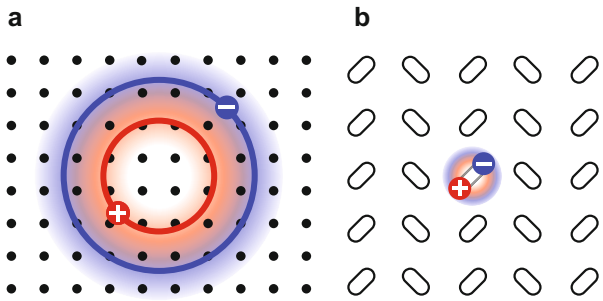


Fig. 1 (**a**) Exciton with a large radius extending over many lattice constants and a center of mass slightly shifted toward the heavier hole (Wannier–Mott exciton). (**b**) Exciton with a small radius localized at a molecule in an organic crystal or an atom in an organic crystal (Frenkel exciton)

axis (Fig. 1). This bound state is called an *exciton*. These excitons have a significant effect on the optical absorption close to the absorption edge. There are various kinds of excitons. *Free excitons*, the main topic of Sect. 1, are free to move in the crystal; further classification distinguishes the degree of localization and free excitons in direct or indirect semiconductors. *Confined excitons* experience spatial restrictions by heterojunctions in low-dimensional semiconductors and are discussed in Sect. 2. *Bound excitons* are excitons trapped in the potential of a defect, e.g., an impurity atom, when the interaction to the defect becomes larger than their thermal energy; their analysis requires a knowledge of the specific defect and is considered in Sect. 2 of ▶ Chap. 18, "Shallow-Level Centers." Furthermore, an exciton may be (weakly) bound to a free carrier or another exciton, forming a charged exciton (a *trion*) or an *exciton complex*. These states are discussed in Sects. 1.4, 2.2, and 2.3.

The Hamiltonian to describe a free exciton can be approximated as

$$H = H_0 + U,\ U = -\frac{e^2}{4\pi\varepsilon\varepsilon_0|\mathbf{r}_e - \mathbf{r}_h|}, \tag{1}$$

where H_0 is the kinetic energy of the electron–hole pair in the center-of-mass frame (we neglect the center-of-mass kinetic energy) and U describes the Coulomb attraction between the electron and hole, screened by the dielectric constant ε. The Schrödinger equation $H\psi = E\psi$ has as eigenvalues a series of quasi-hydrogen states

$$E_{\mathrm{exc},n} = -R_\infty \frac{\mu}{m_0}\frac{1}{\varepsilon^2}\frac{1}{n^2} \text{ with } R_\infty = \frac{m_0 e^4}{8\varepsilon_0^2 h^2} = 13.6 \text{ eV}, \tag{2}$$

where R_∞ is the Rydberg energy, μ is the reduced exciton mass given by $\mu^{-1} = m_n^{-1} + m_p^{-1}$ for simple parabolic bands, and n as the principal quantum number. The exciton radius is a quasi-hydrogen radius

$$a_{\mathrm{exc},n} = a_\mathrm{H}\frac{\varepsilon}{\mu/m_0}n^2 \text{ with } a_\mathrm{H} = \frac{h^2\varepsilon_0}{\pi m_0 e^2} = 0.529 \text{ Å}, \tag{3}$$

where a_H is the Bohr radius of the hydrogen atom. For a review, see Bassani and Pastori-Parravicini (1975), Haken (1976), or Singh (1984).

Depending on the reduced exciton mass and dielectric constant, one distinguishes between *Wannier–Mott excitons*, which extend over many lattice constants and are free to move through the lattice, and *Frenkel excitons*, which have a radius comparable to the interatomic distance. Frenkel excitons become localized and resemble an atomic excited state (for more detail, see Singh 1984). For the large Wannier–Mott excitons, the screening of the Coulomb potential is appropriately described by the static dielectric constant $\varepsilon_{\text{stat}}$ which is used in Eq. 2.

When the lattice interaction is strong, the electron–hole interaction can be described by an *effective* dielectric constant $\varepsilon^* = \left(\varepsilon_{\text{opt}}^{-1} + \varepsilon_{\text{stat}}^{-1}\right)^{-1}$, which provides less shielding. A further reduction of the correlation energy by modifying the effective dielectric constant was introduced by Haken (1963 – see Eq. 97 of ▶ Chap. 10, "Interaction of Light with Solids"):

$$U = -\frac{e^2}{4\pi\varepsilon_0|\mathbf{r}_e - \mathbf{r}_h|}\left\{\frac{1}{\varepsilon_{\text{stat}}} + \frac{1}{2}\left(\frac{1}{\varepsilon_{\text{opt}}} - \frac{1}{\varepsilon_{\text{stat}}}\right)\left[\exp\left(\frac{r_e}{r_{\text{pe}}}\right) + \exp\left(\frac{r_h}{r_{\text{ph}}}\right)\right]\right\}. \quad (4)$$

Here, r_{pe} and r_{ph} are the electron and hole *polaron* radii,[2] respectively. The radius of the exciton consequently shrinks, and the use of the effective mass becomes questionable. Here, a tight-binding approximation becomes more appropriate to estimate the eigenstates of the exciton, which is now better described as a Frenkel exciton.

For both types of excitons, one obtains eigenstates below the bandgap energy by an amount given by the binding energy of the exciton. In estimating the binding energy, the band structure of valence and conduction bands must be considered, entering into the effective mass and dielectric function. Such structures relate to light- and heavy-hole bands, energy and position in **k** of the involved minima of the conduction band, and other features determining band anisotropies. This will be explained in more detail in the following sections. We will first discuss some of the general features of Frenkel and Wannier–Mott excitons.

1.1 Frenkel Excitons

Frenkel excitons (Frenkel 1931 – see also Landau 1933) are observed in ionic crystals with relatively small dielectric constants, large effective masses, and strong coupling with lattice, as well as in organic molecular crystals (see below). These excitons show relatively large binding energies, usually in excess of 0.5 eV, and are also referred to as tight-binding excitons. They cannot be described in a simple hydrogenic model.

[2]When a quasi-free charge carrier (electron or hole) moves through a crystal with strong lattice polarization, it is surrounded by a polarization cloud. Carrier plus polarization form a *polaron*, a quasiparticle with an increased effective mass (see Sect. 1.2 of ▶ Chap. 22, "Carrier-Transport Equations").

1.1.1 Excitons in Alkali Halides

In alkali halides Frenkel excitons with lowest energy are localized at the negatively charged halogen ion, which have lower excitation levels than the positive ions. Figure 2 shows the absorption spectrum of KCl with two relatively narrow Frenkel exciton absorption lines. They relate to the two valence bands at the Γ point that are shifted by spin–orbit splitting. The doublet can be interpreted as excitation of the Cl^- ion representing the valence bands in KCl. This absorption produces tightly bound excitons, but does not produce free electrons or holes; that is, it does not produce photoconductivity as does an absorption at higher energy. The excited state of the Cl^- is considered the Frenkel exciton (see Kittel 1966). It may move from one Cl^- to the next Cl^- ion by quantum-mechanical exchange. The long-range Coulomb attraction between electron and hole permits additional excited states which have a hydrogen-like character, although with higher binding energy (~1 eV) than in typical semiconductors because of a large effective mass and relatively small dielectric constant. An extension of a short-range (tight-binding) potential with a Coulomb tail is observed in a large variety of lattice defects (see ▶ Chap. 19, "Deep-Level Centers") and provides characteristics mixed between a deep-level and a shallow-level series. It creates a mixture of properties with Frenkel and Wannier–Mott contributions (see Tomiki 1969).

Lowest excitation energies $E_{exc,0}$ of Frenkel excitons in alkali halides are listed in Table 1 (Song and Williams 1993). The large difference to the bandgap energy E_g yields large binding energies $E_B = E_g - E_{exc,0}$, which significantly exceed those of Wannier excitons.

The strength of the exciton absorption substantially exceeds that of the band edge which coincides with the series limit ($n = \infty$). The features to the right of this limit in Fig. 2, labeled with roman numerals, result from excitation into the higher conduction bands.

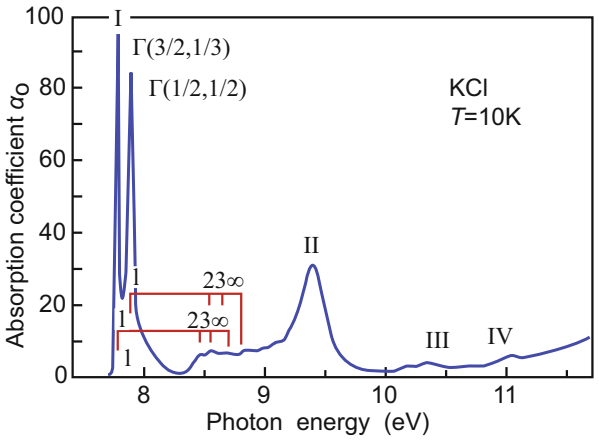


Fig. 2 Absorption spectrum of KCl at 10 K with two narrow exciton peaks identified as transitions at the Γ point. The hydrogen-like series are due to the Coulomb tail of the potential (After Tomiki 1969)

Table 1 Experimental lowest-excitation energies $E_{exc,0}$ of Frenkel excitons for alkali-halide crystals; E_g and E_B denote, respectively, the bandgap energy and the resulting binding energy

Crystal	E_g (eV)	$E_{exc,0}$ (eV)	E_B (eV)
LiF	13.7	12.8	0.9
NaF	11.5	10.7	0.8
NaCl	8.8	7.9	0.9
NaBr	7.1	6.7	0.4
KF	10.8	9.9	0.9
KCl	8.7	7.8	0.9
KBr	7.4	6.7	0.7

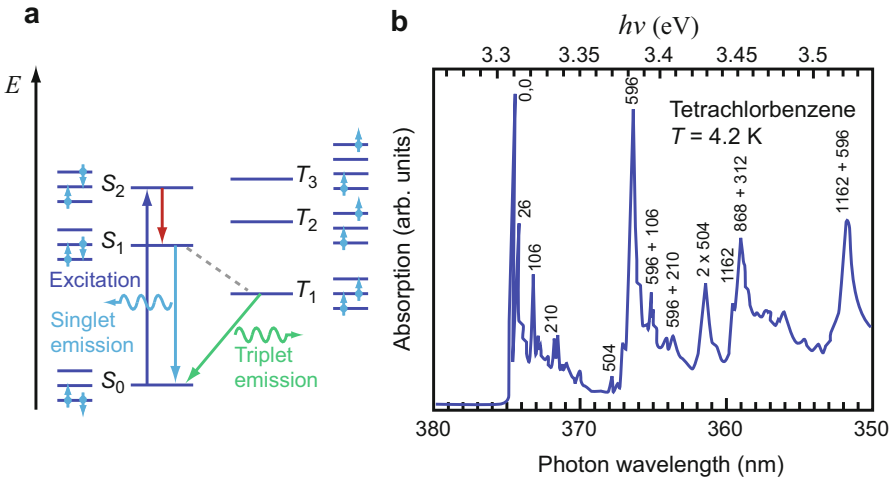


Fig. 3 (**a**) Singlet/triplet exciton schematics with ground state S_0, excited singlet states S_1 and S_2, and triplet states T_1 to T_3; occupied spin states are indicated beside the electronic levels. (**b**) Optical absorption spectrum of singlet–triplet excitons in a tetrachlorobenzene crystal platelet, measured at 4.2 K with unpolarized light; line labels denote energy differences in units of cm^{-1} with respect to the zero-phonon line labeled 0,0 (After George and Morris 1970)

1.1.2 Excitons in Organic Crystals

Frenkel excitons are observed particularly in organic molecular crystals, such as in anthracene, naphthalene, benzene, etc., where the binding forces within the molecule (covalent) are large compared to the binding forces between the molecules (van der Waals interaction, see Sect. 1.5 of ▶ Chap. 3, "The Structure of Semiconductors"). Here, localized excited states within the molecules are favored.

If more than one electron is involved in the excited state, we can distinguish singlet and triplet excited states, while the ground state is always a singlet state (see Fig. 3a). In recombination, the singlet–singlet transition is allowed (it is a luminescent transition), while the triplet–singlet transition is spin forbidden. Consequently, the triplet state has a long lifetime, depending on possible triplet/singlet mixing, and a much weaker

luminescence. Studies indicate that triplet excitons are generated in molecular crystals by fission of singlet excitons in excited states (Müller et al. 2006; Najafov et al. 2010).

Such singlet and triplet excitons are common in organic semiconductors and have been discussed extensively (see Pope and Swenberg 1982, Köhler and Bässler 2015). Their importance has also been recognized in inorganic semiconductors in the neighborhood of crystal defects (Cavenett 1984; Davies et al. 1984) and also in layered semiconductors (GaS and GaSe; Cavenett 1980). The luminescence of singlet excitons is employed in organic LED (OLED) devices used to create displays for, e.g., mobile phones or TV screens or for solid-state lighting; for more information, see Shinar (2004) and Kamtekar et al. (2010).

An example of a molecular singlet and triplet Frenkel exciton is shown in Fig. 3. Panel (a) gives the level scheme of organic molecules with an even number of π electrons (i.e., no ions or radicals); the absorption spectrum for tetrachlorobenzene is shown in panel (b). The transition near 375 nm labeled 0,0 refers to the $T_1 \leftarrow S_0$ absorption; transitions at shorter wavelength originate from additional excitations of molecule and lattice vibrations with energies indicated at the lines. When the unit cell contains more than one identical atom or molecule, an additional small splitting of the excited eigenstates occurs. This is referred to as *Davydov splitting* and is observed in molecule crystals but not in isolated molecules.

For the exciton transport, we have to distinguish between triplet and singlet excitons – for reviews, see Knox (1984), Mikhnenko et al. (2015). For the latter, a dipole–dipole interaction via radiation, i.e., luminescence and reabsorption, contributes to the exciton transport. Such a mechanism is negligible for triplet excitons, which have a longer lifetime and therefore much lower luminescence. For exciton diffusion, see Sect. 2 of ▶ Chap. 22, "Carrier-Transport Equations."

1.2 Wannier–Mott Excitons

Wannier–Mott excitons are found in most of the typical semiconductors and extend over many lattice constants (see Wannier 1937 and Mott 1938). In the center-of-mass frame, their eigenfunctions, which solve the Schrödinger equation with the Hamiltonian Eq. 1, can be written as the sum of two terms: a translational part $\phi(\mathbf{R})$ describing the motion of the entire exciton as a particle with mass $M = m_n + m_p$ and a rotational part $\phi_n(\mathbf{r})$ related to the rotation of electron and hole about their center of mass:

$$\psi = \phi(\mathbf{R}) + \phi_n(\mathbf{r}). \tag{5}$$

The center-of-mass coordinate $\mathbf{R}$ and the electron–hole separation $\mathbf{r}$ are given by

$$\mathbf{R} = \frac{m_n \mathbf{r}_e + m_p \mathbf{r}_h}{m_n + m_p} \text{ and } \mathbf{r} = \mathbf{r}_e - \mathbf{r}_h. \tag{6}$$

The Schrödinger equation of the translation is

$$-\frac{\hbar^2}{2M}\frac{\partial^2}{\partial \mathbf{R}^2}\phi(\mathbf{R}) = E_K\ \phi(\mathbf{R}), \tag{7}$$

with the eigenfunctions and eigenvalues determined by the wavevector $\mathbf{K} = \mathbf{k}_e + \mathbf{k}_h$ of the entire exciton:

$$\phi(\mathbf{R}) = e^{i\mathbf{K}\cdot\mathbf{R}},\ E_K = \frac{\hbar^2|\mathbf{K}|^2}{2M}. \tag{8}$$

The rotation is described by

$$\left(-\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial \mathbf{r}^2} - \frac{e^2}{4\pi\varepsilon\varepsilon_0|\mathbf{r}|}\right)\ \phi_n(\mathbf{r}) = E_n\ \phi_n(\mathbf{r}), \tag{9}$$

solved by the quasi-hydrogen eigenvalues given in Eq. 2.

For isotropic parabolic bands, the minimum optical energy needed to excite an electron and to create an exciton is slightly smaller than the bandgap energy:

$$E_{\mathrm{g,exc}} = E_\mathrm{g} - R_\infty \frac{\mu}{m_0}\frac{1}{\varepsilon_{\mathrm{stat}}^2}\frac{1}{n^2} + \frac{\hbar^2\ K^2}{2\left(m_n + m_p\right)} \text{ with } n = 1,\ 2,\ \ldots; \tag{10}$$

the second term is the quasi-hydrogen binding energy (Eq. 2) with $\varepsilon = \varepsilon_{\mathrm{stat}}$, and the third term is the kinetic energy due to the center-of-mass motion of the exciton. This term leads to a broadening of optical transitions compared to those of bound and confined excitons. The bandgap energy can be determined from two transitions of the series; e.g., from the $1S$ and $2S$ transition energies $E_\mathrm{g} = E_{\mathrm{g,exc}} + R_\infty^{\mathrm{eff}} = \frac{4}{3}(E(2S) - E(1S)) + R_\infty^{\mathrm{eff}}$ is concluded, with R_∞^{eff} the effective Rydberg energy (or binding energy $E_{\mathrm{exc},n=1}$) of the exciton. The values of the binding energy $E_{\mathrm{exc},n=1}$ (Eq. 2) and the quasi-hydrogen Bohr radius $a_{\mathrm{exc},n=1}$ (Eq. 3) show a clear trend in the dependence on bandgap energy: Fig. 4 shows the increase of the binding energy and the decrease of the exciton Bohr radius for increasing low-temperature bandgap. For a listing of exciton energies, see Table 2. The Wannier–Mott exciton is mobile and able to diffuse (see ▶ Chap. 22, "Carrier-Transport Equations"). Since it has no net charge, it is not influenced in its motion by an electric field[3] and does not contribute directly to the electric current.

The ionization energy[4] of these excitons in typical semiconductors is on the order of 10 meV (Table 2 and Thomas and Timofeev 1980); hence, the thermal energy kT at room temperature (26 meV) is sufficient to dissociate most of them.

The principal quantum number n defines S states ($l = 0$) which contribute to electric dipole transitions in direct-gap semiconductors with allowed transitions, while P states ($l = 1$) contribute to dipole-forbidden transitions (see triplet excitons

[3]It is, however, influenced by the *gradient* of an electric field or by strain; see, e.g., Tamor and Wolfe 1980.

[4]The ionization energy is also referred to as binding energy or Rydberg energy.

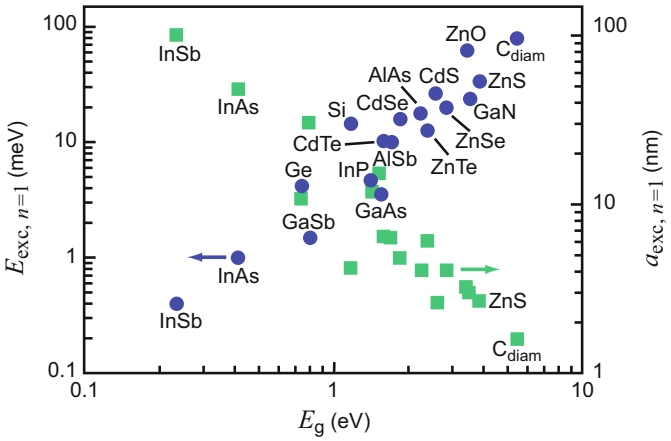


Fig. 4 Exciton binding energy (*blue symbols*) and ground-state Bohr radius for semiconductors (*green symbols*) at low temperature. Ordinate and abscissa values are from Table 2 and from Table 8 of ▶ Chap. 8, "Bands and Bandgaps in Solids," respectively

in previous section). With introduction of symmetry-breaking effects, such as external fields, external stresses, or those in the neighborhood of crystal defects (Gislason et al. 1982), and moving slightly away from $\mathbf{K} = 0$, the other quantum numbers, l and m, need to be considered. This results in a more complex line spectrum. Exciton transitions with $n \rightarrow \infty$ merge into the edge of the band continuum (Fig. 5).

At low temperatures, excitons have a major influence on the optical absorption spectrum. This can be seen from the matrix elements M_{cv} for transitions from near the top of the valence band to the vicinity of the bottom of the conduction band. When considering exciton formation, the band-to-band transition matrix elements (e.g., Eq. 21 of ▶ Chap. 13, "Band-to-Band Transitions") are modified by multiplication with the eigenfunction of the exciton $\phi_n(\mathbf{r})$ as discussed in the following.

1.2.1 Direct-Gap Excitons

For an excitonic single-photon absorption at the Γ point in a direct-bandgap material, the matrix element is given by

$$M_{cv}^{(exc)} = M_{cv}\,\phi_n(0), \tag{11}$$

with M_{cv} given by Eq. 21 of ▶ Chap. 13, "Band-to-Band Transitions." The strength of the absorption is proportional to the square of the matrix element, which yields

$$\alpha_{o,\,vc,\,exc} = \alpha_{o,\,vc}|\phi_n(0)|^2, \tag{12}$$

where $\alpha_{o,vc}$ is the valence-to-conduction band optical absorption coefficient neglecting excitons. In the case of isotropic parabolic bands, the eigenfunction of the n^{th} exciton state is related to the ground state with $n = 1$ by

Table 2 Experimental parameters of 1*S* Wannier–Mott excitons for some semiconductors at low temperature; *d* and *i* denote direct and indirect excitons, respectively, and *A, B, C* excitons with holes from the respective valence bands in wurtzite crystals; *T* signifies the measurement temperature; $T = 0$ refers to extrapolated data

Crystal	*d*/*i*	Ground-state energy $E^{1S}_{g,exc}$ (eV)	*T*(K)	Binding energy $E_{exc,n=1}$ (meV)	Bohr radius $a_{exc,n=1}$ (Å)	Reduced mass μ (m_0)
C_{diam}	*i*	5.409	1.6	80	16	0.191
Si	*i*	1.1545	1.6	14.7	42	0.145
Ge	*i*	0.7405	2.1	4.2 (±3/2) 3.1 (±1/2)	108 143	0.079 0.059
AlN	*d*	6.025 (*A*) 6.243 (*B*) 6.257 (*C*)	1.7 1.7 1.7	∼55 (*A*) ∼55 (*B*) ∼55 (*C*)		
AlAs	*i*	2.228	2	18	41	0.134
AlSb	*i*	1.677	4.2	10	64	0.092
α-GaN	*d*	3.476 (*A*) 3.482 (*B*) 3.499 (*C*)	4.2 4.2 4.2	24.0 (*A*) 22.8 (*B*) 24.5 (*C*)	31 (*A*) 33 (*B*) 31 (*C*)	0.164 (*A*) 0.156 (*B*) 0.168 (*C*)
β-GaN	*d*	3.2725		26	32	0.156
GaAs	*d*	1.5152	2	3.6	155	0.044
GaSb	*d*	0.8099	2	1.5	306	0.027
InP	*d*	1.4814	2	4.8	120	0.055
InAs	*d*	0.4157	4.8	1.0	494	0.016
InSb	*d*	0.2347	1.7	0.4	1017	0.009
ZnO	*d*	3.3756 (*A*) 3.3811 (*B*) 3.4327 (*C*)	1.5 1.5 1.5	63 (*A*) 50 (*B*) 49 (*C*)	32 (*A*) 35 (*B*) 37 (*C*)	0.062 (*A*) 0.057 (*B*) 0.054 (*C*)
α-ZnS	*d*	3.8069	0	34.2	27	0.106
ZnSe	*d*	2.8052	0	19.9	41	0.119
ZnTe	*d*	2.3816	0	12.7	62	0.078
α-CdS	*d*	2.5529 (*A*) 2.5676 (*B*)	1.6 1.8	27 (*A*) 31 (*B*) 30 (*C*)	26 25 25	0.197 (*A*) 0.204 (*B*) 0.211 (*C*)
CdSe	*d*	1.8265 (*A*) 1.8503 (*B*)	0 0	15.7 (*A*) 16.7 (*B*)	48 45	0.106 (*A*) 0.112 (*B*)
CdTe	*d*	1.595	4.8	10.5	65	0.085
Cu_2O	*d*	2.1472 ($n = 2$)	1.2	98 ($n \geq 2$)	11	0.407

$$|\phi_n(0)|^2 = n^{-3}|\phi_1(0)|^2. \tag{13}$$

For exciton states[5] below the bandgap, it follows

[5] $\phi_n(0) \neq 0$ applies only for *S* states.

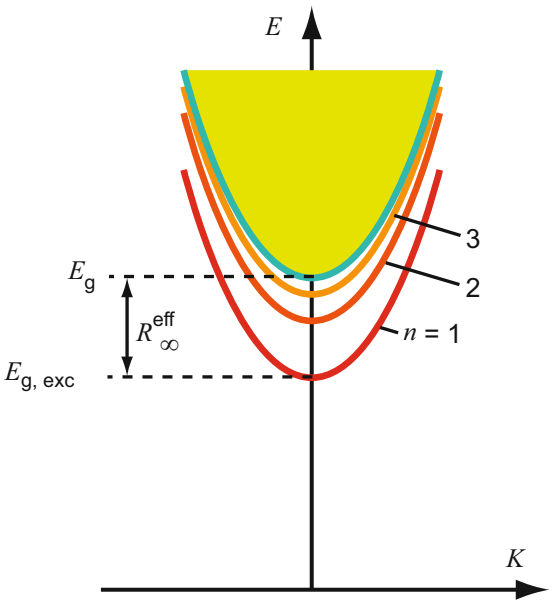


Fig. 5 Band model with exciton levels that result in a hydrogen-like line spectrum for direct-gap semiconductors. The shaded area above E_g indicates unbound continuum states

$$|\phi_n(0)|^2_{h\nu<E_g} = \begin{cases} \dfrac{1}{a_{qH}^3 n^3} & \text{allowed} \\ \dfrac{n^2-1}{\pi a_{qH}^5 n^5} & \text{forbidden} \end{cases} \tag{14}$$

for the indicated type of transitions (Sects. 1.3 and 2.2 of ▶ Chap. 13, "Band-to-Band Transitions") and with the quasi-hydrogen radius $a_{qH} = a_H \varepsilon_{stat} m_0/\mu = a_{exc,n=1}$ for $\varepsilon = \varepsilon_{stat}$. For higher excited states, the line intensity decreases proportional to n^{-3} or $n^{-5}(n^2-1)$ for allowed or forbidden transitions, respectively. The optical absorption per center is spread over a large volume element of radius $a_{qH} \cdot n^2$; therefore, the corresponding matrix element is reduced accordingly. The spacing of the absorption lines is given by

$$E_{g,exc}^{(dir)} = E_g - E_{exc,n}; \tag{15}$$

E_g is also the limit of the line series (Fig. 5). Hence, we expect one strong line for the ground state in absorption, followed by much weaker lines for the excited states which converge at the absorption edge (see Fig. 7).

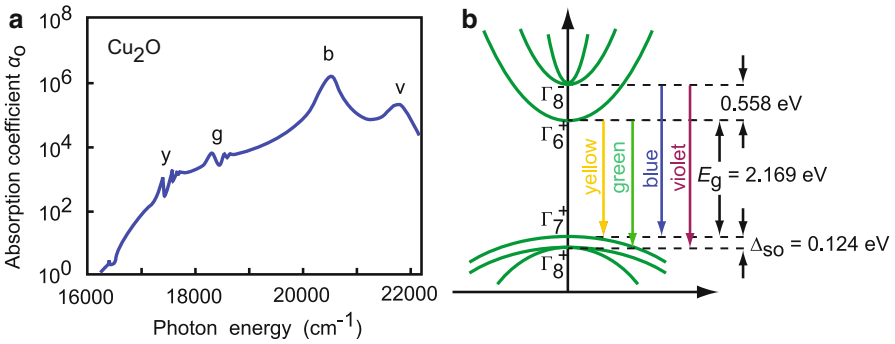


Fig. 6 (**a**) Absorption spectrum and (**b**) band structure near the Γ point of Cu_2O with strong dipole-allowed (*blue* and *violet*) and weak forbidden (*yellow* and *green*) transitions. The absorption below the *yellow* series corresponds to indirect transitions to the 1*S* exciton via phonon coupling

Electric dipole-forbidden 1*S* transitions[6] are observed in only a few semiconductors. Most extensively investigated is Cu_2O with *d*-like valence bands, which has two series of hydrogen-like levels – the yellow (superscript *y*) and the green (superscript *g*) series from the Γ_7 and Γ_8 valence bands, respectively. The observed level spectra are given in Fig. 6, with

$$E_n^{(y)} = \left(2.1661 - 0.0971/n^2\right)\ \text{eV and } E_n^{(g)} = \left(2.2975 - 0.1565/n^2\right)\text{eV} \qquad (16)$$

for the *P* levels of these two series. According to the second case of Eq. 14, the series starts with $n = 2$ (Fig. 7b) and is observed up to $n = 25$ in pure Cu_2O crystals[7] (Kazimierczuk et al. 2014). In addition, there are two dipole-allowed excitons in the blue and violet range of the spectrum from the two valence bands into the higher conduction band Γ_{12} (Compaan 1975). Figure 7b shows that with sufficient perturbation by an electric field, the electric dipole selection-rule is broken, and the *S* transitions are also observed in the yellow (and also in the green) series. Another material showing forbidden exciton spectra is SnO_2.

In contrast to the strongly absorbing Frenkel exciton with a highly localized electron–hole wavefunction in the ground state, the intensity of the Wannier–Mott exciton lines is reduced by $(a/a_{qH})^3$: the larger the quasi-hydrogen radius a_{qH} is

[6]Strictly, such transitions cannot occur at $\mathbf{k} = 0$; however, a slight shift because of the finite momentum of the photon permits the optical transition to occur because of a weak electric quadrupole coupling (Elliott 1961). Such transitions can also be observed under a high electric field using modulation spectroscopy (Washington et al. 1977). Dipole-forbidden transitions are easily detected with Raman scattering (Sect. 1.3) or two-photon absorption (for Cu_2O, see Uihlein et al. 1981), which follow different selection rules.

[7]With a correspondingly large exciton Bohr radius of 1.04 μm for $n = 25$, compared to ∼1 nm for $n = 1$.

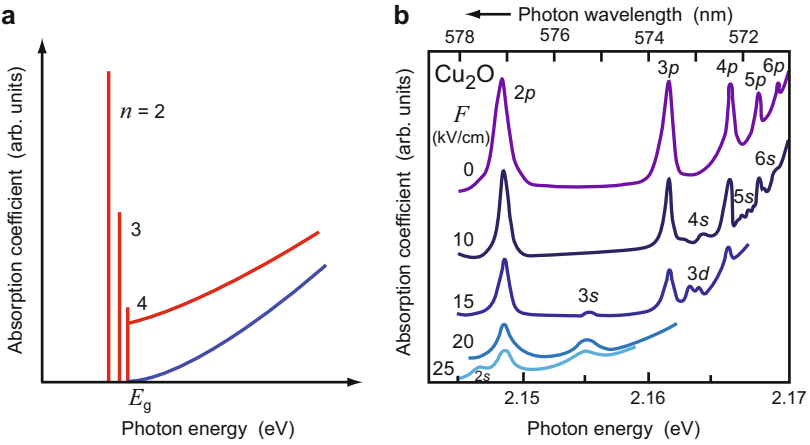


Fig. 7 (**a**) Direct-bandgap dipole-forbidden transitions including (*red curves*) and excluding exciton excitation (*blue curve*). (**b**) Transitions in Cu_2O at 4 K (After Grosmann 1963)

compared to that of the corresponding atomic eigenfunction a, the weaker is the corresponding absorption line (see Fig. 7b and Kazimierczuk et al. 2014).

For $h\nu > E_g$ near the band edge, the absorption is substantially increased due to the effect of the Coulomb interaction. With exciton contribution, a semiconductor with a *direct* bandgap has an increased absorption for transitions between spherical parabolic bands, given by (for details, see Bassani and Pastori-Parravicini 1975)

$$\alpha^{(\mathrm{dir})}_{\mathrm{o,vc,exc}} = \begin{cases} \dfrac{2\pi\nu}{n_r c}\gamma_e \dfrac{\exp\gamma_e}{\sinh\gamma_e} & \text{allowed} \\ \dfrac{2\pi\nu}{n_r c}\gamma_e \dfrac{\left[1+(\gamma_e/\pi)^2\right]\exp\gamma_e}{\sinh\gamma_e} & \text{forbidden} \end{cases}, \text{with } \gamma_e = \pi\sqrt{\frac{E_{\mathrm{exc},n=1}}{h\nu - E_g}}. \quad (17)$$

The Coulomb interaction of the electron and hole influences the relative motion, and the optical absorption in the entire band-edge range is thereby enhanced, as shown in Fig. 8. In semiconductors with large $\varepsilon_{\mathrm{stat}}$ and small reduced mass μ, only the first exciton peak is usually observed: in GaAs the relative distance $E_g - E_{\mathrm{exc}}(\mathbf{K} = 0)$ is only 3.4 meV. Higher absorption lines, which are too closely spaced, are reduced in amplitude and merge with the absorption edge in most direct-gap III–V compounds. Figure 8 shows the pronounced excitonic absorption below the band-edge energy E_g and the onset of continuum absorption above E_g. For an advanced discussion, see Beinikhes and Kogan (1985).

A line spectrum including higher excited states can be observed more easily when it lies adjacent to the reduced absorption of forbidden transitions. It is also easier to observe in materials with a somewhat higher effective mass and lower dielectric

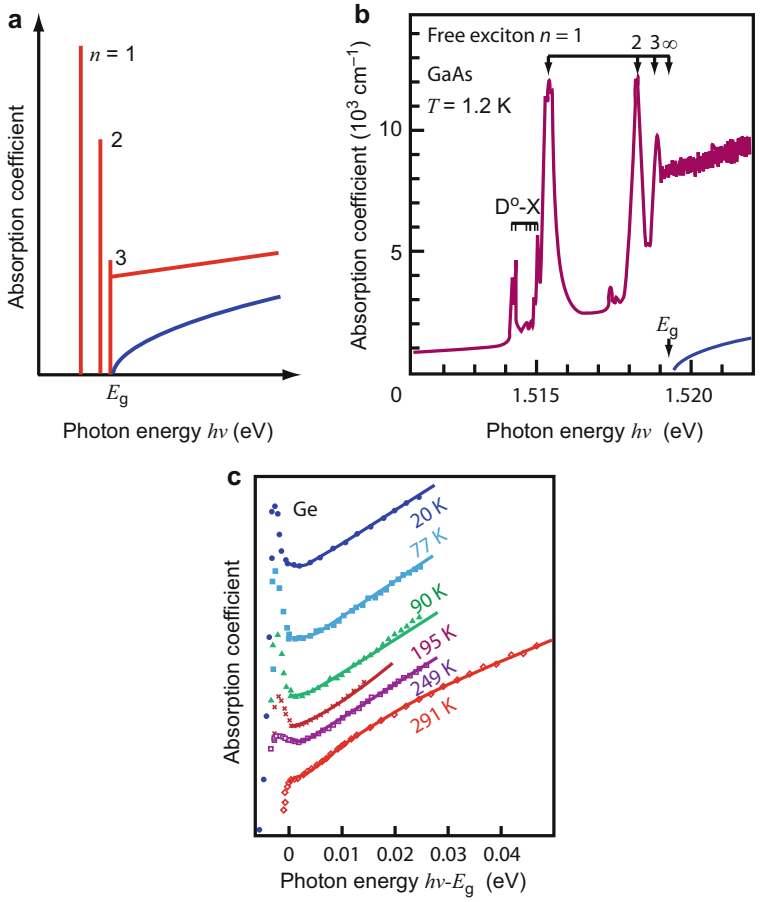


Fig. 8 Direct-bandgap dipole-allowed transitions including exciton excitation (i.e., effects of Coulomb interaction, *red curves* in (**a**)) and excluding this (*blue curves* in **a** and **b**). Examples are (**b**) for transitions in GaAs (After Weisbuch et al. 2000) and (**c**) for Ge (After McLean 1963); (**c**) shows the decrease of absorption at higher temperatures where excitons can no longer exist; measured curves are shifted due to the temperature dependence of the bandgap energy $E_g(T)$

constant to obtain a wide enough spacing of these lines. Well-resolved line spectra of higher excited states can be observed when they do not compete with other transitions and are not inhomogeneously broadened by varying lattice environments.

1.2.2 Complexity of Exciton Spectra

The simple hydrogen-like model described above must be modified in real semiconductors because of several contributions (Flohrer et al. 1979):

- The anisotropy of effective masses and dielectric constants
- The electron–hole exchange interaction
- The exciton–phonon interaction

- The action of local mechanical stress or electrical fields
- The interaction with magnetic fields

These contributions act on quantum numbers l, m, and s not included in the discussed model; they introduce anisotropies and lift the degeneracy of conduction or valence bands as considered in the following.

The *anisotropy of the effective masses* produces excitons, elongated in the direction in which the mass is smallest. A compression in the direction of the largest effective mass reduces the quasi-hydrogen Bohr radius a_{qH} in this direction by a factor of less than 2 and increases E_{qH} up to a factor of 4 (Shinada and Sugano 1966). The *reduced exciton mass* μ, entering the expression for the exciton energy in Eq. 10, is given by

$$\frac{1}{\mu} = \frac{2}{3}\frac{1}{\mu_{\perp}} + \frac{1}{3}\frac{\varepsilon_{\perp}}{\varepsilon_{\parallel}}\frac{1}{\mu_{\parallel}}. \tag{18}$$

The meaning of $\perp$ and $\parallel$ depends on the crystal structure. For instance, in *wurtzite-type semiconductors*, $\parallel$ means parallel to the **c** direction. For calculation of the reduced effective mass, we distinguish six effective masses: $m_n^{\perp}, m_n^{\parallel}, m_{pA}^{\perp}, m_{pA}^{\parallel}, m_{pB}^{\perp}, m_{pB}^{\parallel}$ with the A and B valence bands split by the crystal field with Γ_9 and Γ_7 symmetry, respectively, neglecting the spin–orbit split-off band, the C band.

The *lifting of band degeneracies* occurs generally when bands are split by, e.g., intrinsic anisotropy in a crystal or, following the application of mechanical stress, an electric or a magnetic field. The exciton line spectrum can be distinguished with respect to transitions from different valence bands, which result in different exciton line series with different spacings because of a different reduced mass. Taking splitting and warping of the valence bands (Sects. 1.2.2 and 1.2.3 of ▶ Chap. 8, "Bands and Bandgaps in Solids") into consideration, one obtains a splitting of the P-like exciton states (Baldereschi and Lipari 1973). This is shown in Fig. 9 as a function of the reduced effective mass. The reduced mass μ can in turn be expressed as a function of the Luttinger valence-band parameters:

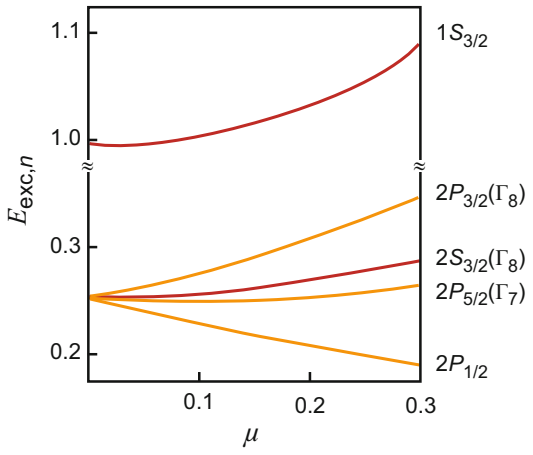


Fig. 9 Exciton binding energy for *S* and *P states* as a function of the reduced mass (After Baldereschi and Lipari 1973)

$$\mu = \frac{6\gamma_3 + 4\gamma_2}{5} \frac{m_n m_0}{m_0 + m_n \gamma_1}. \tag{19}$$

For a review, see Rössler (1979) and Hönerlage et al. (1985).

The *electron–hole exchange interaction* due to the coupling of the electron and hole spins contains a short-range (J_{SR}) and a long-range term (J_{LR}) (also referred to as analytic and nonanalytic contributions), which depend on the wavevector of the exciton motion: $J_{\text{longitudinal}} = J_{SR} + \frac{2}{3}J_{LR}$ and $J_{\text{transversal}} = J_{SR} - \frac{1}{3}J_{LR}$. Thereby the fourfold degenerate $A(n = 1)$ exciton is split into two Γ_6 and Γ_5 exciton states, each twofold degenerate (compare Fig. 10b); the fourfold $B(n = 1)$ exciton splits into the Γ_5 exciton, which is twofold degenerate, and into the nondegenerate Γ_1 and Γ_2 exciton states. The $A(\Gamma_6)$ and $B(\Gamma_2)$ states are not affected by the exchange interaction. For details, see Denisov and Makarov (1973) and Flohrer

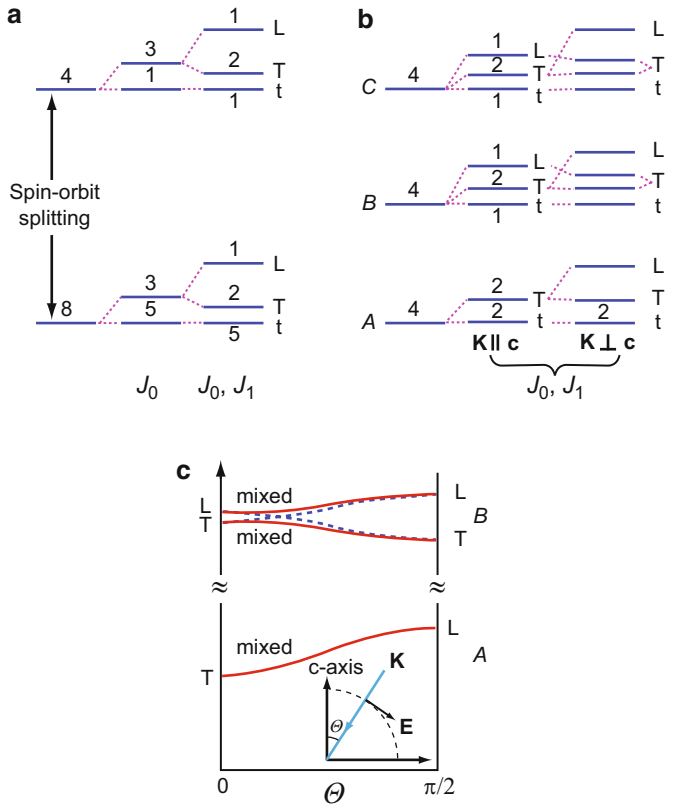


Fig. 10 Splitting of exciton levels for (**a**) cubic and (**b**) uniaxial crystals from the spin-triplet states (t) into dipole-allowed spin singlets and to the longitudinal (L) and transverse (T) dipole-allowed states due to exchange effects; numbers denote degeneracies. Levels in (**b**) are given for exciton wavevectors **K** parallel and perpendicular to the crystal **c** axis. (**c**) Mixed longitudinal and transverse modes of the A and B excitons as a function of the angle θ of **K** with respect to **c** (After Cho 1979)

et al. (1979) who model explicitly wurtzite CdS. The splitting for ZnO is discussed by Lambrecht et al. (2002) and Wagner et al. (2013), for GaN by Rodina et al. (2001), and the large splitting in AlN by Ishii et al. (2020).

The *exciton–phonon interaction* depends strongly on the carrier–lattice coupling, which is weak for predominantly covalent semiconductors, intermediate for molecular crystals, and strong for ionic crystals (alkali halides). Excitons can interact with phonons in a number of different ways. One distinguishes exciton interaction (for reviews, see Vogl 1976 and Yu 1979) by:

- Nonpolar optical phonons via the deformation potential (a short-range interaction; Loudon 1963)
- Longitudinal optical phonons via the induced longitudinal electrical field (Fröhlich interaction; Fröhlich 1954)
- Acoustic phonons via the deformation potential for a large wavevector, since for $\mathbf{q} \cong 0$ the exciton experiences a nearly uniform strain resulting in a near-dc field, resulting in no interaction with the electrically neutral exciton (Kittel 1963)
- Piezoelectric acoustic phonons via the longitudinal component similar to the Fröhlich interaction (Mahan and Hopfield 1964)

Another interaction involves three-particle scattering among photons, phonons, and excitons, such as the relevant Brillouin and Raman scattering (see Yu 1979 and Reynolds and Collins 1981).

In solids with a strong coupling of electrons (and holes) with the lattice, the exciton–phonon interaction can become large enough to cause *self-trapping* of an exciton (Kabler 1964). This can be observed in predominantly ionic crystals with a large bandgap energy and occurs because of a large energy gain due to distortion of the lattice by the exciton. It results in a very large increase of the effective mass of the exciton, which is usually a Frenkel exciton. It is distinguished from a polaron by the short-range interaction of the exciton dipole, compared to the far reaching Coulomb interaction of the electron/polaron. The self-trapped exciton may be a significant contributor to photochemical reactions. It is best studied in alkali halides; for a review, see Toyozawa (1980), Song and Williams (1993), and Li et al. (2019).

In anisotropic crystals with an external perturbation, we must consider the relative direction of the exciting optical polarization **e**, the exciton wavevector **K**, and the crystallographic axis **c**. One distinguishes σ and π modes when **e** is $\perp$ or $\parallel$ to the plane of incident light, respectively (*transverse and longitudinal excitons*). When the incident angle $\theta \neq 90^\circ$, mixed modes appear (Fig. 10c). The resulting exciton lines in cubic and hexagonal systems for $\mathbf{k} \perp \mathbf{c}$ and $\mathbf{k} \parallel \mathbf{c}$ are given in Fig. 10a, b.

Many of the band degeneracies are lifted when the crystal is exposed to internal or external perturbation. Internally, this can be done by alloying as reported, e.g., for $CuCl_{1-x}Br_x$ bulk crystals (Kato et al. 1970) or for $In_xGa_{1-x}As/GaAs$ quantum dots (Zieliński 2021). Externally, degeneracies are lifted by external fields such as mechanical stress and electrical or magnetic fields (reviewed by Cho 1979).

1.2.3 Indirect-Gap Excitons

Indirect-gap excitons are associated with optical transitions to a satellite minimum at $\mathbf{K} \neq 0$. The energy of an indirect exciton is

$$E_{\mathrm{g,\,exc}}^{(\mathrm{indir})} = E_{\mathrm{g}}^{(\mathrm{indir})} - E_{\mathrm{exc},\,n} + \frac{\hbar^2}{2\mu}(\mathbf{K} - \mathbf{q}_0)^2, \tag{20}$$

with μ as the reduced mass between the valence bands and the satellite minima of the conduction band. In order to compensate for the electron momentum $\mathbf{k}_0$ at the satellite minimum, the transition requires an absorption or emission of a phonon of appropriate energy and momentum. The phonon momentum $\mathbf{q}_0$ is equal to $\mathbf{k}_0$ (see Fig. 11).

During the indirect excitation process, an electron and a hole are produced with a large difference in wavevectors. Such an excitation can proceed to higher energies within the exciton dispersion using a slightly higher photon energy. The excess in the center-of-mass momentum is balanced by only a small change in phonon momentum. Therefore, we observe an onset for each branch of appropriate phonon processes for indirect-bandgap transitions, following selection rules, rather than a line spectrum observed for direct-bandgap material. Here, relaxing phonon processes are hardly observed since they have a much smaller probability.

One obtains for the absorption coefficient caused by these indirect transitions

$$\alpha_{\mathrm{o,\,cv,\,exc}}^{(\mathrm{indir})} = \alpha_{\mathrm{o}}^{*}\left[f_{\mathrm{BE}}\sqrt{h\nu - E_{\mathrm{g}}^{(\mathrm{indir})} + E_{\mathrm{g,exc},n} + \hbar\omega_q} + (f_{\mathrm{BE}} + 1)\sqrt{h\nu - E_{\mathrm{g}}^{(\mathrm{indir})} + E_{\mathrm{g,exc},n} - \hbar\omega_q}\right], \tag{21}$$

where α_{o}^{*} is a proportionality factor. This factor includes the absorption enhancement relating to the square of the exciton envelope function, and f_{BE} is the

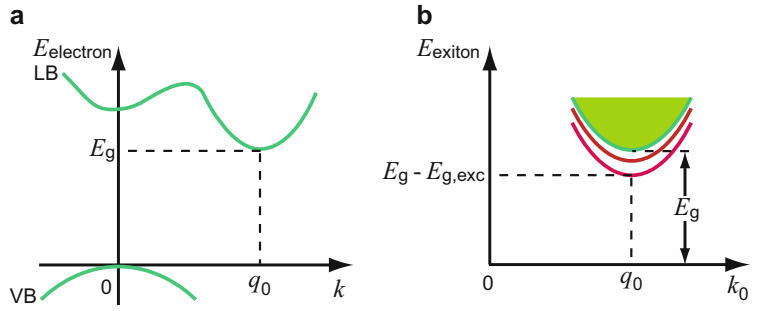


Fig. 11 Dispersion relation for indirect-bandgap excitons: (**a**) Typical band dispersion $E(k)$ of an indirect semiconductor. (**b**) Satellite minimum with ground and first excited state parabolas of indirect excitons

Bose–Einstein distribution function for the phonons. The two terms describe transitions with absorption and emission of a phonon, respectively. The allowed spectrum consequently has two edges for the ground state of the exciton, plus or minus the appropriate phonon energy. It is shown in Fig. 12b for the indirect gap of Ge. Excited exciton states disappear because of the strong ($1/n^3$) dependence of the oscillator strength. The branch caused by phonon absorption also disappears at lower temperatures as less phonons are available.

For indirect-bandgap semiconductors like Ge, also transitions at $\mathbf{K} = 0$, i.e., direct-gap exciton transitions, can be observed; they proceed into a higher band above the bandgap. As an example, the exciton line near the direct transition at the Γ point of Ge at 0.883 eV for 77 K is shown in Fig. 12b, c.

A more structured series of square-root shaped steps at the indirect excitonic transition is shown for GaP in Fig. 13. Here, different types of involved phonons are observed more pronounced. The corresponding energies of the absorbed or emitted phonons are 12.8, 31.3, and 46.5 meV for TA, LA, and LO phonons in GaP, respectively.

1.3 Exciton–Polaritons

In some of the previous sections, the interaction of light with electrons was described by near band-to-band transitions close to $\mathbf{K} = 0$, creating direct-bandgap excitons. This interaction requires a more precise analysis of the dispersion relation. The exciton and photon dispersion curves are similar to those for phonons and photons (Fig. 14). Both show the typical splitting due to the von Neumann noncrossing principle. In this range, the distinction between a photon and the exciton can no longer be made. The interacting particle is a polariton or, more specifically, an *exciton–polariton*.

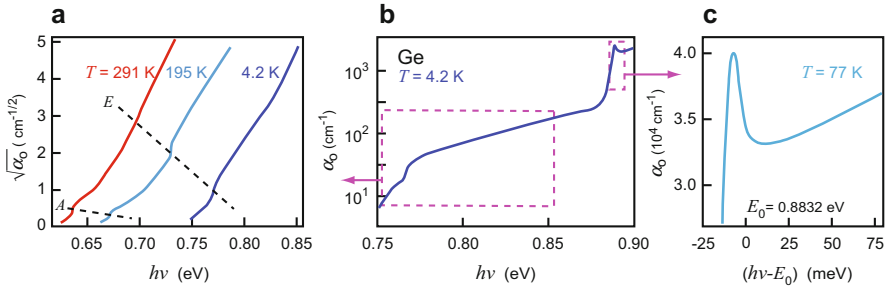


Fig. 12 (**b**) Weak indirect and strong direct excitonic transitions in Ge. (**a**) Detail of the indirect exciton transition preceding the band edge ($\Gamma_{8vb} \rightarrow \Gamma_{6cb}$) of Ge (After MacFarlane et al. 1957). *Dashed lines* indicate threshold for absorption (*A*) or emission (*E*) of a phonon. (**c**) Direct exciton at $k = 0$ and band-to-band transition from Γ_{8vb} into Γ_{7cb} of Ge at 77 K. Note different ordinate scalings

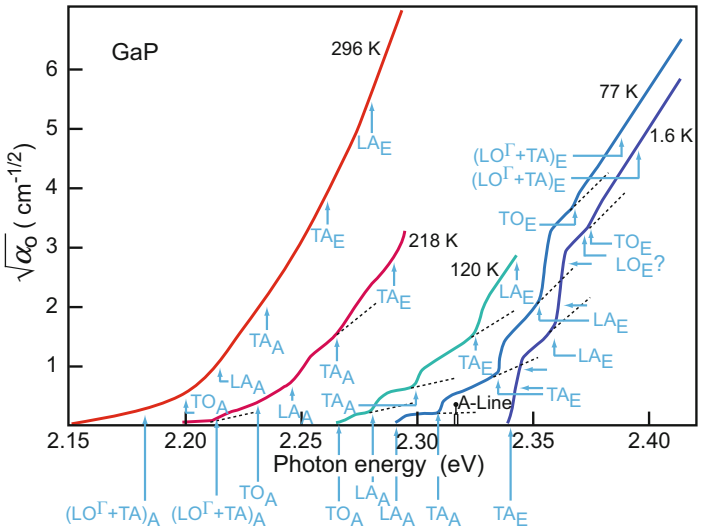


Fig. 13 Indirect exciton transition in GaP with threshold energies for indirect, phonon-assisted transitions at various temperatures. The additional absorption edges connected to different types of phonons or phonon pairs are identified accordingly (After Dean and Thomas 1966)

Entering the resonance frequency for excitons into the polariton dispersion equation (Eq. 68 of ▶ Chap. 11, "Photon–Phonon Interaction") and including spatial dispersion ($\hbar^2K^2\omega_t^2/\mu^2$), we obtain the *exciton–polariton equation*

$$\nu^4 - \nu^2\left(\nu_l^2 + \frac{c^2k^2}{\varepsilon^*}\right) + \frac{c^2k^2}{\varepsilon^*}\nu_t^2 = 0 \tag{22}$$

with a kinetic energy term

$$h\nu_{(l,t)} = h\nu_{\mathrm{exc}(l,t)} + \frac{\hbar^2K^2}{2\left(m_n + m_p\right)}. \tag{23}$$

The ability of the exciton to move through the lattice represents propagating modes of excitation within the semiconductor. They are identified by the term $\propto K^2$ and have a group velocity ($\propto \partial E/\partial K$) on the order of 10^7 cm/s. Consequently, the dielectric constant is wavevector dependent and can be written as

$$\varepsilon^* = \varepsilon_{\mathrm{opt}}\left(1 + \frac{\nu_l^2 - \nu_t^2}{\nu_t^2 - \nu^2 + \beta K^2 - i\gamma\nu}\right), \text{ where } \beta = \frac{h\nu_t}{m_n + m_p}. \tag{24}$$

There is now an additional spatial dispersion term βK^2 in the denominator in contrast to an otherwise similar equation for the phonon–polaritons, which does not have spatial dispersion.

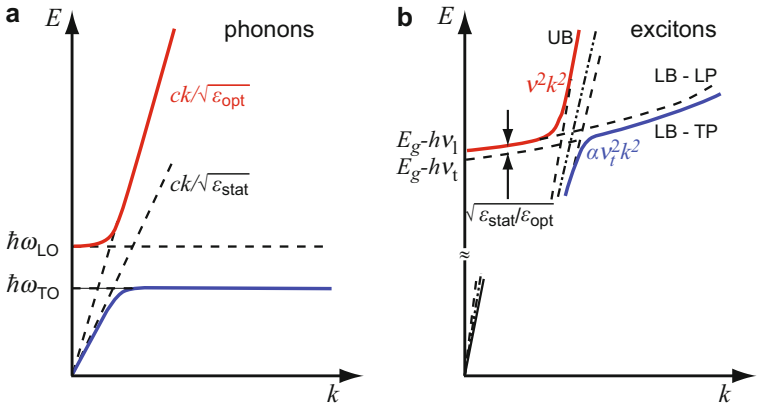


Fig. 14 Comparison between (**a**) a phonon–polariton and (**b**) an exciton–polariton (schematic and not to scale) with upper (*UB*) and lower-branch (*LB*) longitudinal (*LP*) and transverse polaritons (*TP*)

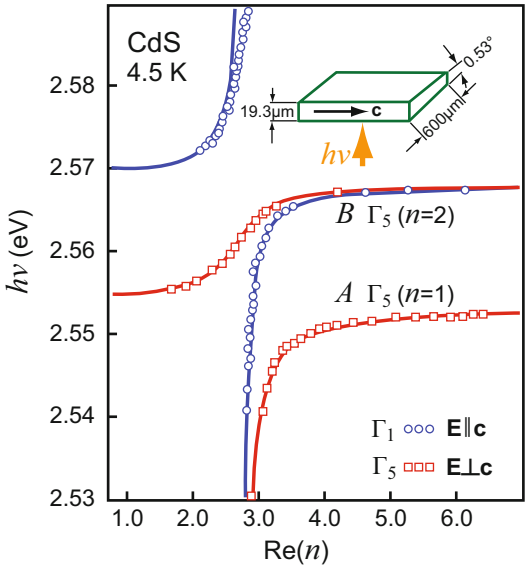


Fig. 15 Energy of *A* and *B* exciton–polaritons as a function of the real part of the index of refraction for **E** ∥ and ⊥ to **c** in a wedge-shaped CdS crystal compared to theoretical curves (After Broser et al. 1981)

The polariton dispersion-equation has several branches. These branches depend on crystal anisotropy and the relative orientation of the polarization of light, exciton **K** vector, and crystallographic axes. As many as four lower and two upper branches are predicted and observed by single- and two-photon excitation processes. An example is shown for CdS in Fig. 15, with upper and lower branches for longitudinal and transverse excitons pointed out; the given abscissa values are related to the wavevector by $\mathrm{Re}(k) = (\omega/c)\mathrm{Re}(\tilde{n})$. See also Girlanda et al. (1994); for a review, see Hönerlage et al. (1985).

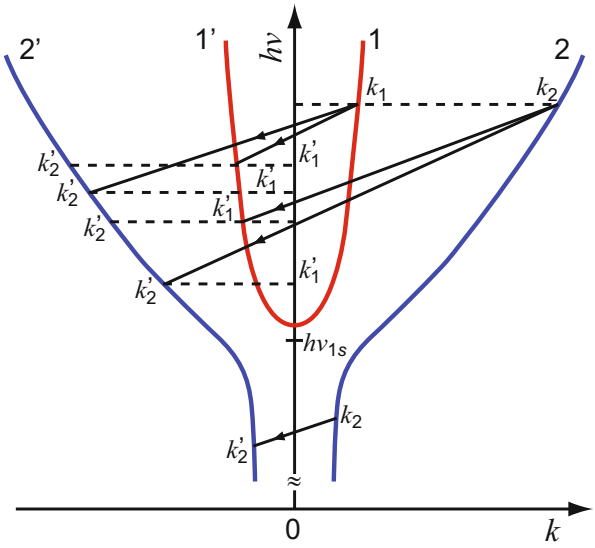


Fig. 16 Dispersion curves of free exciton–polaritons with inner (1, 1′) and outer branches (2, 2′), indicating the Stokes processes of Brillouin scattering between different branches

If the impinging light has an energy below, but close to, a free exciton line in direct-bandgap semiconductors, a longitudinal acoustic phonon can supply the missing energy, and resonant Brillouin scattering with exciton–polaritons occurs. When the energy supplied by the photon lies below the 1*S* exciton, one observes a single backward scattered Stokes line at $h\nu_{1s}$ (Fig. 16) with a transition $k_2 \rightarrow k_2'$. When the frequency of the photon lies above $h\nu_{1S}$, four Stokes-shifted lines are expected, for which energy and momentum conservation is fulfilled. The same number of *blue*-shifted anti-Stokes lines is additionally observed (not shown in the figure for clarity). The exciton–polariton state is created with phonons near the center of the Brillouin zone, which interact strongly with optical radiation.

In contrast to the phonon–polariton scattering discussed in Sect. 3.1 of ▶ Chap. 11, "Photon–Phonon Interaction," the lower branch of the exciton–polariton is not flat but parabolic in its upper part, since the exciton is mobile and can acquire kinetic energy which causes an $E \propto k^2$ behavior. In addition, there are several lower branches according to the different excited states of the exciton; for a review, see Yu (1979). A measured spectrum,[8] showing several of these branches from the ground and excited states of the exciton, is shown in Fig. 17. Some of these relate to interaction with LA phonons, others with TA phonons, as indicated in the figure.

[8]The analysis of the measured reflection spectrum as a function of the wavelength and incident angle is rather involved. A relatively simple method for measuring the central part of the exciton–polariton spectrum in transmission through a prismatic crystal was used by Broser et al. (1981) (see Fig. 15).

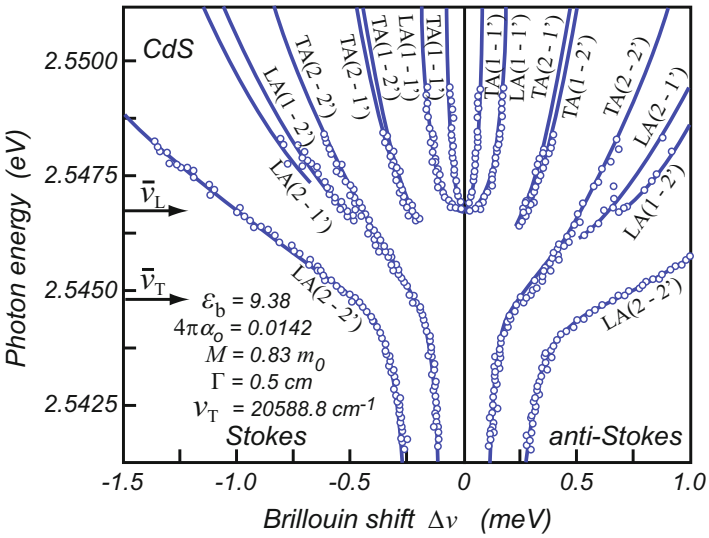


Fig. 17 Brillouin shift of the scattered laser light in CdS as a function of the laser energy. Theoretical transitions refer to the indexing shown in Fig. 16. Parameters: dielectric constant $\varepsilon_b = 9.38$, oscillator strength $4\pi f = 0.0142$, excitonic mass 0.83 m_0, phenomenological damping constant $\gamma = 0.5$ cm, transverse exciton energy 2.5448 eV, and longitudinal exciton energy 2.5466 eV. The numerals 1 and 2 refer to inner and outer branch polaritons (After Wicksted et al. 1984)

Additional information about the exciton–polariton dispersion relation is obtained from *hyper-Raman scattering*. More intense optical excitation is required to induce a two-photon excitation of a virtual biexciton[9] (see Fig. 18). Hyper-Raman scattering, when using two photons of frequencies ν_1 and ν_2, each having an energy slightly below the bandgap energy and propagating with wavevectors $\mathbf{k}_1$ and $\mathbf{k}_2$ inside the semiconductor, creates a new intermediate state with

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{K} \text{ and } h\nu_1 + h\nu_2 = E(\mathbf{K}). \tag{25}$$

If the energy $h\nu_1 + h\nu_2$ is close to the resonant state (the biexciton), one observes a strongly enhanced transition (see Fröhlich 1981). The virtual biexciton decays into two quasiparticles, one of which is observed, while the other remains in the crystal to conserve energy and momentum. There are three possibilities:

$h\nu_1 + h\nu_2 = E_{LP}(\mathbf{K}_{LP}) + E_{LP}(\mathbf{K}_{LP})$ (two lower-branch polaritons)
$h\nu_1 + h\nu_2 = E_{LP}(\mathbf{K}_{LP}) + E_{UP}(\mathbf{K}_{UP})$ (one lower, one upper polariton)
$h\nu_1 + h\nu_2 = E_{LP}(\mathbf{K}_{LP}) + E_{L,exc}(\mathbf{K}_{L,exc})$ (one lower polariton, one longitudinal exciton)

[9] A state close to an actual biexciton state (Sect. 1.4) which immediately decays into other states.

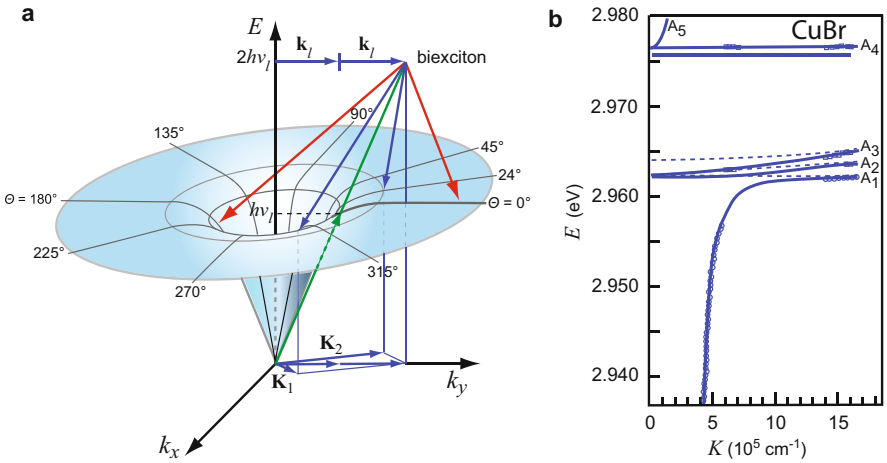


Fig. 18 (**a**) Hyper-Raman scattering illustrated by one lower polariton branch drawn in two dimensions (k_x, k_y). Exciting laser photon, $h\nu_l(k_l)$; virtual biexciton, $2h\nu_l(2k_l)$; indicated recombination, backscattering (*red arrows*) one polariton-like and one exciton-like state, *arrow* at 180° paired to *arrow* at 0° for leftover polariton; forward scattered pair, *blue arrows* as alternatives. (**b**) Dispersion curves of five branches of polaritons in CuBr for $\mathbf{e}\|[001]$ and $\mathbf{K}\|[110]$ (After Hönerlage et al. 1985)

For all of these alternatives, the condition

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{K}_i + \mathbf{K}_j \tag{26}$$

is fulfilled.

Figure 18a depicts the creation of a biexciton from two photons (two green arrows up) with energies $h\nu_1 = h\nu_2 = h\nu_l$ and wavevector $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{k}_l$ along the k_y axis; both photons are provided by the same laser with energy $h\nu_l$, and the biexciton has $2h\nu_l$ energy and $2\mathbf{k}_l$ wavevector. This biexciton decays into two lower polaritons (pair of red arrows down) which are offset by an angle to fulfill momentum conservation and land at different points ($h\nu$) on the polariton surface; one of these polaritons is observed, and θ is the scattering angle between incident photon and observed polariton. The point of landing depends on the energy of the initiating photon pair and the angle of observation of the emitted (scattered) photon [backscattering, small θ (<30°), red arrows; forward scattering (observation through the sample), large θ, blue arrows]. This makes hyper-Raman scattering a three-photon process; it is determined by the third-order term in the susceptibility. The entire polariton spectrum is obtained by changing the energy of the exciting light (the resulting virtual biexciton), the angle and energy of the emitted photon, and calculating energy and momentum of the leftover polariton. Several polariton branches have consequently been observed (Fig. 18b).

Hyper-Raman scattering follows selection rules other than those for normal Raman scattering or IR absorption. It thereby yields additional information about the lattice vibrational spectrum, e.g., about dipole modes in centrosymmetrical lattices which are forbidden in normal Raman scattering. It also permits excitation deep inside a crystal. For reviews, see Denisov et al. (1987) and Hönerlage et al. (1985); see also García-Cristóbal et al. (1998).

Besides the bulk polaritons discussed so far, there are also *surface-polariton modes*; they progress in a thin layer near the semiconductor surface and can be used to reveal properties of the crystal near the surface. Grazing incident light or reflection measurement is used for their detection (Hopfield and Thomas 1963); for Raman measurements, see Davydov et al. (1997). The region near the surface cannot be penetrated by bulk-excitons to a thickness of either the space-charge region at the surface or the bulk-exciton diameter, whichever is larger (Altarelli et al. 1979). These surface exciton–polaritons are reviewed by Fischer and Lagois (1979), Agranovich and Mills (1982).

1.4 Trions and Biexcitons in Bulk Crystals

Free Trions An exciton can be weakly bound to an electron or another exciton to form a trion or biexciton, respectively (Lampert 1958; Moskalenko 1958; a proceeding on trions is found in Phys. Stat. Sol. B 227 (2), 2001). The trion is composed of either two electrons and a hole (this one is similar to an H^- ion) or two holes and an electron. Therefore, it is negatively or positively charged. Predicted by Gerlach (1974) and observed first by Thomas and Rice (1977), their binding energy in Ge is on the order of 0.2 meV; this trion has an effective mass of about 20% more than the sum of the free electron and hole masses, a radius about 50% larger than that of an exciton, and an ionization energy about 10% of the ionization energy of excitons.

Reports on *free* trions in semiconductors are scarce. Calculations yield states which are stable against dissociation into an exciton and a free carrier in both three-dimensional (3D) and lower-dimensional semiconductors. The binding energy (with respect to the exciton energy) depends on the ratio of electron and hole effective masses and is generally small for bulk semiconductors (Combescot 2019), typically one tenth for 3D trions compared to 2D quantum-well trions (Fig. 19); often experimental assignments are not unequivocal.

Characteristic features distinguish trions from excitons. Their formation is influenced by the position of the Fermi energy, as explicitly demonstrated for 2D trions (Sect. 2.2). In optical transitions, the initial electron or hole momentum is transferred to the final trion state, which leads to low-energy tails at the trion and exciton lines; in addition, exciton–electron scattering gives rise to a high-energy tail (trion continuum) at the exciton transition (Esser et al. 2001). In a magnetic field, trions show circularly polarized transitions for singlet and triplet states. Since trions have a significantly increased binding energy in low-dimensional semiconductors, they are well studied there and discussed in more detail in Sect. 2.

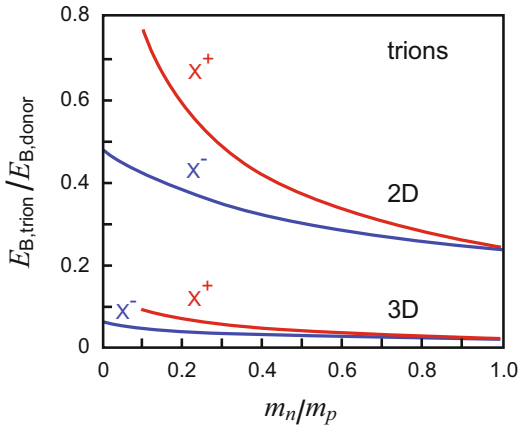


Fig. 19 Calculated binding energies of 3D and 2D negative (X^-) and positive trions (X^+) as a function of the electron and hole effective mass ratio (After Stébé and Ainane 1989). Energies are given with respect to the binding energy of a donor in a 3D crystal (The generally used symbol for trions is X^+; in initial work also the symbol X_2^+ was used, in analogy to a positively charged H_2^+ molecule)

Free Biexcitons Biexcitons are exciton molecules X_2 similar to a hydrogen molecule (when $m_p \gg m_n$) or a positronium molecule (when $m_p \cong m_n$). Biexcitons form a stable bound state of two excitons for all ratios of electron to hole effective masses and in all dimensions from 3D to 0D. The binding energy of a biexciton is higher than that of a trion and is typically of the order of 10–20% of that for an exciton. For isotropic bands, it decreases monotonically with increasing ratio of m_n/m_p from $E_B(X_2)/R_\infty^{\text{eff}}(X) \cong 0.3$ for $m_n/m_p \to 0$ to 0.03 for $m_n/m_p = 1$ (Akimoto and Hanamura 1972; Brinkman et al. 1973). This is similar to the decrease in relative binding energy from a hydrogen molecule $E_B(H_2)/R_\infty(H) = 4.7/13.6 = 0.35$ to the relative binding energy of a positronium (Ps) molecule $E_B(Ps_2)/E_B(Ps) = 0.13/6.8 = 0.02$. Such biexcitons have been observed in Si (Hanamura and Haug 1977; Thewalt and Rostworowski 1978) and can be observed readily in II–VI semiconductors (Zn and Cd chalcogens) or CuCl and CuBr (see Haken and Nikitine 1975 and Ueta and Nishina 1976). Biexciton and exciton spectra of a bulk-like thick ZnS layer are given in Fig. 20. Tensile strain splits the valence band of the zincblende ZnS into the heavy- and light-hole bands, each producing an exciton emission. The exciton emission increases approximately linearly with excitation intensity I_{ex}, while the biexciton emission increases roughly quadratically[10] ($I_{X_{LH}} \propto I_{ex}^{1.1}, I_{XX} \propto I_{ex}^{1.9}$): the probability to form a biexciton molecule XX is proportional to the density of *each* of the two participating excitons X; the dominating line in Fig. 20 is the light-hole biexciton X_{lh}.

A high density of excitons hence favors the formation of biexcitons. However, the low binding energy requires low temperatures which, in turn, favors further condensation into an electron–hole liquid (Sect. 3.2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"). This condensation can be suppressed by applying a uniaxial

[10]Deviations from a pure quadratic dependence are due to the short radiative lifetime for the involved species in direct-bandgap semiconductors, preventing a thermal equilibrium of the population.

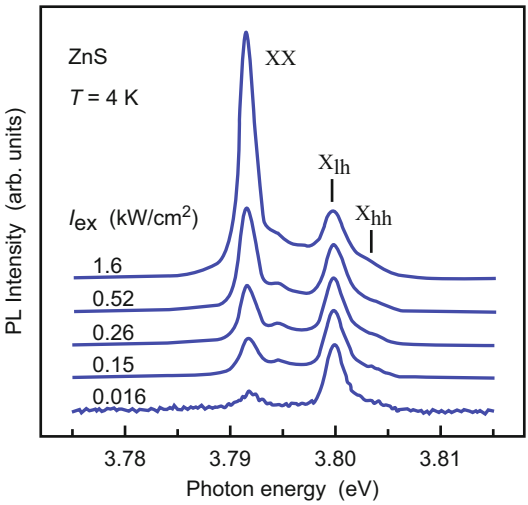


Fig. 20 Photoluminescence spectra of biexcitons (XX) and light- (X_{lh}) and heavy-hole exciton (X_{hh}) in a 6 μm-thick tensely strained ZnS layer on GaAs substrate. Spectra are normalized with respect to the peak intensity of the X_{lh} line (After Yamada et al. 2000)

stress (Gourley and Wolfe 1978) or biaxial strain as in epitaxial layers (Fig. 20); larger biexciton signals are then observed. Further discussion of excitonic molecules is provided by Kulakovskii et al. (1985).

The line shape of biexcitons, observed by luminescence or scattering experiments, is typically asymmetric (similar to that of trions) because of the recoil energy when they recombine: the line has a larger low-energy tail. During this process, there is always a remaining partner that can take up part of the energy as kinetic energy. For a review, see Hanamura (1976) and Hönerlage et al. (1985). At higher densities, excitons or excitonic molecules can no longer exist, but form an electron–hole plasma, as described in Sect. 3.2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers."

2 Excitons in Low-Dimensional Semiconductors

In low-dimensional semiconductors, the density of states is modified (Sect. 3.2 and Fig. 37 of ▶ Chap. 8, "Bands and Bandgaps in Solids"), yielding an increase at the band-edge energy for gradually decreased dimensionality. In addition, the dielectric constant and effective mass are anisotropic and result in ellipsoidal excitonic eigenfunctions. The exciton binding-energy is increased by confining barriers, affecting energies of the Rydberg series, the Bohr radius, and the oscillator strength. We first focus on two-dimensional excitons.

2.1 Excitons in Quantum Wells

Carriers in a quantum well are free to move in two spatial directions (x,y), while confining barriers lead to quantized states in the third dimension (z) (see Sect. 3.1.1

of ▶ Chap. 8, "Bands and Bandgaps in Solids"). The confinement applies also for excitons, and the two-dimensional exciton energy gets

$$E_{g,\mathrm{exc},2D} = E_g + E_z - R_\infty \frac{\mu}{m_0}\frac{1}{\varepsilon_{\mathrm{stat}}^2}\frac{1}{\left(n-\frac{1}{2}\right)^2} + \frac{\hbar^2\left(K_x^2+K_y^2\right)}{2\left(m_n+m_p\right)} \text{ with} \quad (27)$$

$$n = 1, 2, \ldots .$$

Here, E_z is the energy of the quantization in z direction given by the quantum number $n_{QW} = 1$ in Eq. 54 of ▶ Chap. 8, "Bands and Bandgaps in Solids," for infinite barriers and shown in Fig. 31 of ▶ Chap. 8, "Bands and Bandgaps in Solids," for finite barriers. The third term is the effective Rydberg energy R_∞^{eff} (or binding energy $E_{\mathrm{exc},n=1}$; see Eq. 10); if the exciton wavefunction does not penetrate significantly into the barriers, the material parameters of R_∞^{eff} remain unchanged; consequently, the binding energy of the two-dimensional 1*S* exciton is increased by a factor of 4 compared to the 3D 1*S* exciton (Shinada and Sugano 1966). In real quantum wells with finite barriers, the factor is smaller and depends on the well width (see Fig. 21). The confinement results in elliptical orbits with a highly compressed coordinate in the direction normal to the quantum-well plane. Here, the orbiting electron and hole approach each other closely, which causes the increase in their binding energy. Initial work on quantum wells was done by Dingle et al. (1974). For a review, see Ploog and Döhler (1983) and Miller and Kleinman (1985).

Since the binding energy depends on the well width, variations of this width by a roughness of the interfaces between well and barriers (also on a scale of atomic

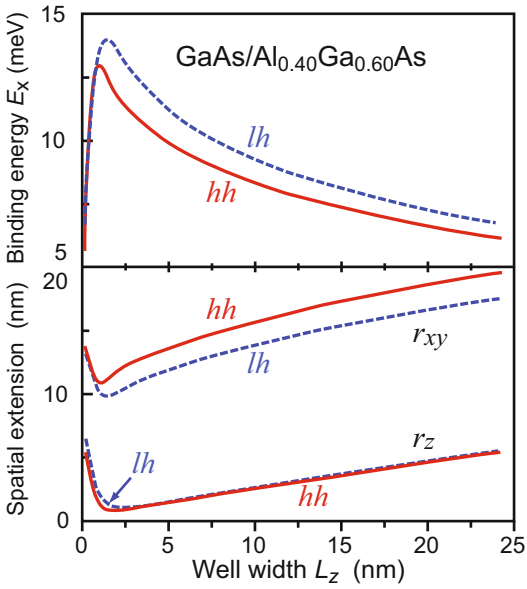


Fig. 21 Calculated binding energy and spatial extension of heavy- (*hh*) and light-hole excitons (*lh*) in a GaAs/$Al_{0.4}Ga_{0.6}As$ quantum well (After Grundmann and Bimberg 1988)

monolayers) also affect the binding energy. This applies not only to excitons but also to charged excitons (trions) and higher exciton complexes (Filinov et al. 2005). The binding energy depends also on the effective mass (see Eq. 27); since heavy and light holes have different effective masses, their splitting in exciton spectra can be used to measure the strain in quantum wells (Kudlek et al. 1992).

The increased binding energy and decreased spatial extension in two dimensions are accompanied by an increased oscillator strength f, originating from a larger overlap of the electron and hole wavefunctions. In three dimensions, $f(n) \propto n^{-3}$; in two dimensions n is replaced by $n - \frac{1}{2}$, yielding a substantial increase in transition probability. Exciton absorption in quantum wells causes the near-band edge features shown in Fig. 15 of ▶ Chap. 13, "Band-to-Band Transitions." These Wannier–Mott excitons are observed in bulk material only at low temperatures, but remain visible to much higher temperatures in quantum wells. The substantially increased lifetime of excitons at higher temperatures is due to the increased exciton binding energy, which is caused by the two-dimensional confinement.

In the energy diagram (Fig. 22a), the exciton energies are shown as lines below the electron levels e_i. The bonding energy of the 1S exciton is indicated as $E_B = e_1 - E_{exc,n=1}$. The exciton energy, however, lies above the gap energy of the well material (labeled by QW in Fig. 22a). We distinguish light- (lh) and heavy-hole (hh) excitons and excitons relating to the first or higher electron level. An example is given in Fig. 22b for a single GaAs/AlGaAs quantum well: excitons are shown combining up to the third conduction-band level with up to the fourth valence-band level.

The linewidth of the exciton absorption is given by its relaxation time (discussed in Sect. 2.2 of ▶ Chap. 32, "Dynamic Processes"). Additional broadening is caused by

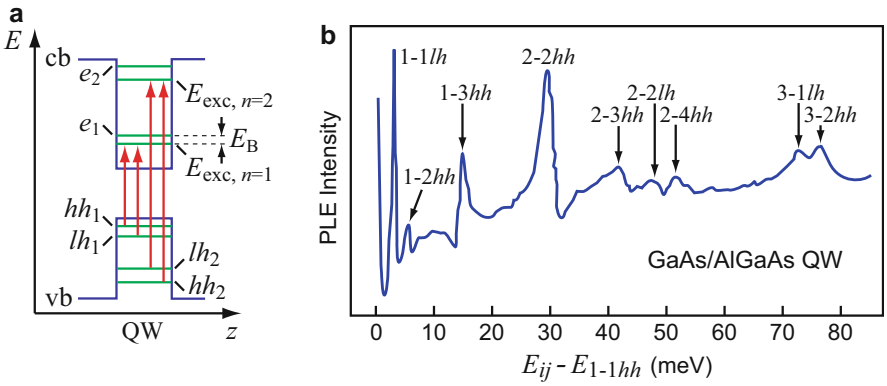


Fig. 22 (**a**) Exciton states for light (lh) and heavy holes (hh) with 1S exciton binding energy E_B indicated. *Red arrows* mark the strongest transitions. (**b**) Photoluminescence excitation (*PLE*) spectrum of hh valence- or lh valence-band excitons in a $L = 22$ nm-wide GaAs quantum well with AlGaAs barriers, measured at $T = 5$ K (After Koteles et al. 1987). The first and second indices identify the quantum level in the well of the conduction band and valence band, respectively; energies are given with respect to the e_1-hh_1 exciton emission

disorder in the alloy of the quantum well or the barriers and, particularly when severely confined, also by the quality of the well interfaces (Christen and Bimberg 1990). Roughness in these interfaces causes broadening by well-width fluctuation (Bajaj and Reynolds 1987). At very low temperatures, well-size fluctuations are resolved as different spikes separated by <1 meV, as shown by Yu et al. (1987).

Electric and uniaxial stress fields cause characteristic changes in the exciton spectrum of quantum wells. The electric field perpendicular to the layers causes a Stark shift toward longer wavelength. New peaks become visible, caused by transitions which were forbidden without perturbation, e.g., such from the mth level of the valence band to the nth level of the conduction band with $m \neq n$ (also seen for the strained QW in Fig. 22b). Such transitions become permitted because of a field-induced deformation of electron and hole eigenfunctions which now overlap. It is shown in Fig. 23a for 1-2 *hh* and 1-3 *hh* excitons which are not observed in these samples at zero bias. The changes in peak position with the electric field illustrate the anticrossing of two levels, demonstrating the von Neumann noncrossing rule when the states interact with each other (see Fig. 23b, c).

Similarly, shifts and anticrossing of levels are observed when uniaxial stress is applied, as shown by Koteles et al. (1987) for the transitions given in Fig. 22b. For earlier works, see Miller et al. (1985).

Exiton-polaritons like those occurring in the bulk (Sect. 1.3) are also observed in two-dimensional semiconductors. They were widely studied in quantum wells embedded in microcavities to produce a high density of bosonic quasiparticles consisting of a superposition of an exciton and a cavity photon. Since a (nonequilibrium) ensemble of such polaritons macroscopically may occupy the same quantum state, they can form a condensate much like an (equilibrium) Bose–Einstein condensate. Reviews on this topic are given by Deng et al. (2010) and Byrnes et al. (2014).

It should be noted that also triplet excitons with parallel spin of electron and hole are observed in quantum wells (Blackwood et al. 1994). Due to their total angular momentum $|M| = 2$, they couple only weakly to photon fields by residual bright-state mixing and are referred to as *forbidden* or *dark excitons*. Although even observable in bulk samples (Ekardt et al. 1979), their oscillator strength increases in structures of reduced dimensionality. Dark excitons are hence discussed for zero-dimensional structures, see Sects. 2.3.2 and 3.2.1.

2.2 Biexcitons and Trions in Quantum Wells

Biexcitons and trions experience in low-dimensional structures a substantial increase of binding energy due to the confining potential similar to confined excitons.

Two-Dimensional Trions There are many reports on trions in quantum wells; for reviews, see Bar-Joseph (2005) and Shields et al. (1995a). A slight, often unintentional n-type or p-type doping provides excess carriers and increases the probability for $X^- = (e,e,h)$ or $X^+ = (e,h,h)$ formation, respectively, after creation

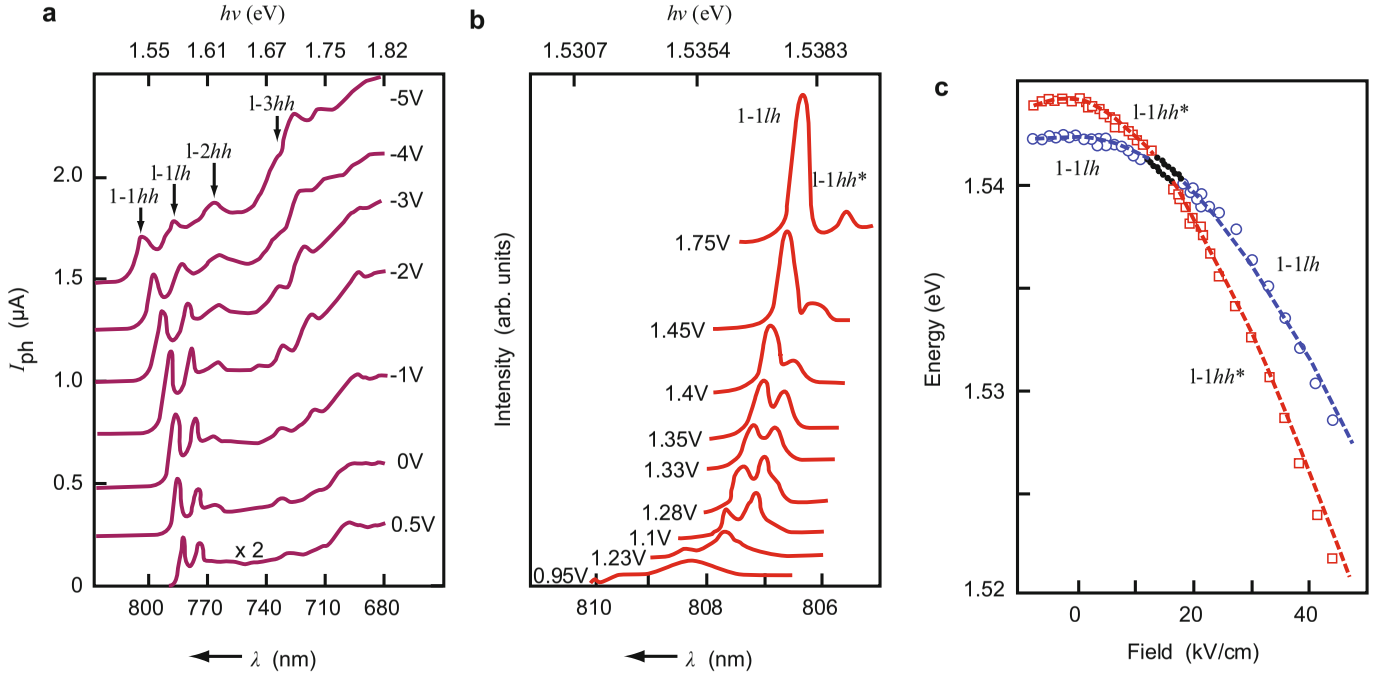


Fig. 23 Shift of exciton peaks with external electric field in GaAs/$Al_{0.3}Ga_{0.7}As$ quantum wells. (**a**) Well width 8 nm, $T = 10$ K, applied voltage across 210 nm. (**b**, **c**) Anticrossing of the 2*S* state observed in the 1-1*hh* transition with the 1*S* state observed in the 1-1*lh* transition for 16 nm well width at 4.3 K. Data in the anticrossing region are shown as *solid circles*; the *dashed curves* are calculated Stark shifts neglecting exciton coupling. Family parameter is the applied voltage (After Collins et al. 1987)

of excitons. Using a modulation-doped quantum-well structure (similar to Fig. 11b of ▶ Chap. 12, "Photon–Free-Electron Interaction"), the two-dimensional remote electron density can be controlled, yielding an increased negative trion emission at higher electron density (see Fig. 24). Vice versa the reflectivity of the trion resonance was used to measure the carrier density (Astakhov et al. 2002a).

The separation between the X and X^- emissions in Fig. 24 is the binding energy of the trion, i.e., the energy of the second electron bound to the exciton. It varies with well width and reaches about 2 meV for X^- in 10 nm-wide GaAs QWs (Bar-Joseph 2005). Since trions can also be bound to donors which provide the remote electron density, the experimentally determined X^- binding energy may be overestimated; significantly smaller values (factor ½) were evaluated from an extrapolation to large donor distances for comparable GaAs QW samples (Solovyev and Kukushkin 2009). II–VI semiconductors have larger exciton and consequently also larger X^- binding energies. For ZnSe, up to 8.9 meV for thin QWs (2.9 nm) was determined for X^- and slightly smaller values for X^+ (Astakhov et al. 2002b).

The binding energy for the second hole in positively charged (X^+) trions should theoretically slightly exceed that of the second electron value for X^- by 17% in the 2D limit due to the larger hole effective mass (Stébé and Ainane 1989). Experimentally roughly similar values are observed due to the uncertainties noted above and the quite small difference (Shields et al. 1995b).

Two-Dimensional Biexcitons A biexciton binding energy of ~22% of the respective exciton binding energy was calculated for quantum wells (Singh et al. 1996), in agreement with experimental values (Birkedal et al. 1996). This implies a largely similar dependence of the binding energy on QW width for both biexcitons

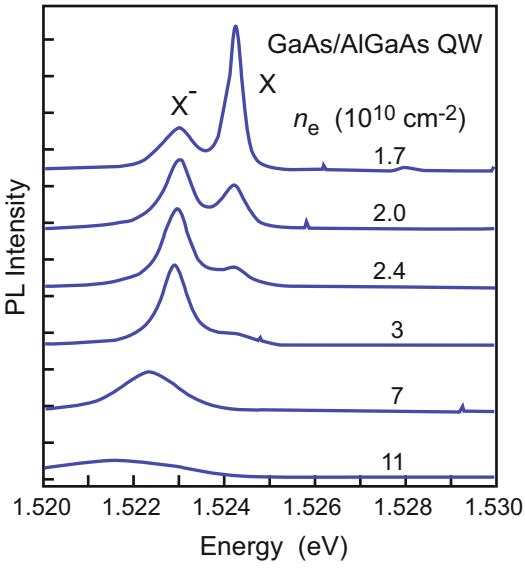


Fig. 24 Negative trion photoluminescence in a modulation-doped GaAs/AlGaAs quantum well with an electron density controlled by a negative gate voltage. At high electron sheet-density, the heavy-hole exciton emission X disappears, and the trion emission X^- dominates (After Bar-Joseph 2005)

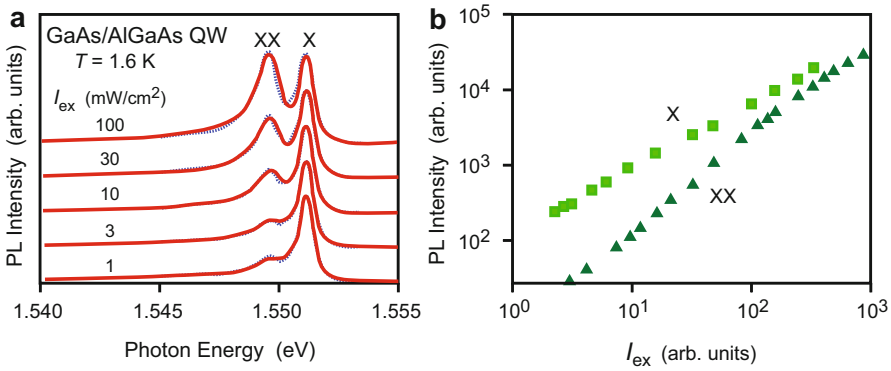


Fig. 25 (**a**) Biexciton emission (XX) in a GaAs/$Al_{0.33}Ga_{0.67}As$ quantum well at different excitation levels. The spectra (*blue*) are normalized with respect to the exciton peak (X); continuous curves (*red*) are model calculations. (**b**) Dependence of exciton and biexciton luminescence intensity on excitation density (After Phillips et al. 1992)

(XX) and excitons (X), a finding also observed in quantum dots (Zieliński et al. 2015). Values are typically slightly larger than for trions.

The superlinear increase of the biexciton emission-line with increase of excitation density I_{ex} is shown for a GaAs/AlGaAs quantum well in Fig. 25. Similar to the bulk case (Fig. 20), the measured dependence of biexciton to exciton intensity $I_{XX} \propto I_X^{1.6}$ has an exponent below 2 due to the short radiative lifetime; the displayed XX and X lines are from heavy-hole excitons.

2.3 Excitons in Quantum Wires and Quantum Dots

Starting from a two-dimensional quantum well, further reduction of dimensionality toward a one-dimensional quantum wire or a zero-dimensional quantum dot requires some patterning to define an additional lateral confinement. The small dimensions needed to obtain quantum-size effects can usually not be accomplished by patterning a quantum-well structure using lithography techniques: the electronic properties of such structures are governed by interface defects. A variety of techniques was developed instead to realize 1D and 0D structures with high optical quality (see Sect. 2.2 of ▶ Chap. 3, "The Structure of Semiconductors"). Most of these techniques lead to complicate confinement potentials, and often an additional quantum well is coupled to the quantum wires or quantum dots.

2.3.1 Excitons in Quantum Wires

Work on epitaxial quantum wires (QWRs) was mostly performed using a V-shaped or T-shaped geometry (see Sect. 2.2.2 and Fig. 20 of ▶ Chap. 3, "The Structure of Semiconductors"). Such structures based on GaAs are particularly well studied; the effective-mass approximation describes quantum effects quite good, and the

fabrication technology is well developed. The lateral confinement of these wire geometries is small, typically 30–40 meV, leading to small subband energy spacings of only 10 meV. For a review, see Akiyama (1998) and Wang and Voliotis (2006). Another approach for fabricating QWRs is whisker growth forming nanowires (Fig. 22 of ▶ Chap. 3, "The Structure of Semiconductors") or colloidal synthesis of nanorods; the vapor-liquid-solid (VLS) growth of whiskers is widely used for self-organized fabrication of uniform nanowires of compound semiconductors. In these structures, the confinement and dielectric contrast to the environment is much larger than in epitaxial structures. A review on VLS growth of semiconductor nanowires is given in Dick (2008) and Pohl (2020).

The exciton binding energy in a 1D quantum wire is expected to be stronger than in 2D or 3D. Analytical solutions of the Schrödinger equation for eigenenergies of states bound in a bare Coulomb potential for d dimensions ($d = 1,2,3,\ldots$) yield in each dimension a Rydberg series (Ogawa and Takagahara 1991):

$$E_n^{(d)} = -\mathrm{R}_\infty^{\mathrm{eff}} \left(n + \frac{d-1}{2} \right)^{-2}, n = 0,\ 1,\ 2,\ \ldots . \tag{28}$$

For 3D and 2D, we recognize the Rydberg series of hydrogen in Eq. 2 and of 2D excitons in Eq. 27. For 1D a singularity occurs for the lowest state $n = 0$; the $1/r$ singularity of the Coulomb potential is removed upon integration in 2D and 3D, but it remains as a logarithmic singularity in 1D. This suggests the attractive force between electron and hole being stronger in 1D than in 2D or 3D. In a descriptive picture a particle may move *around* the origin of a Coulomb potential in 2D or 3D, while it moves *through* the origin in 1D. Experimentally, an enhancement of the exciton binding energy to 27 ± 3 meV (six to seven times larger than the bulk value) was observed in T-shaped quantum wires (Someya et al. 1996); binding energies exceeding 100 meV were found in semiconductor–insulator quantum wires (nanowires), where the effect is further enhanced by the strong dielectric contrast (Muljarov et al. 2000; Giblin et al. 2011).

Luminescence properties of one-dimensional excitons confined in a V-shaped quantum wire are shown in Fig. 26. The GaAs QWR is formed at the bottom of an AlGaAs V groove and has a crescent-like shape (Fig. 21 of ▶ Chap. 3, "The Structure of Semiconductors"); at the groove sidewalls, additional quantum wells are formed during deposition of the upper AlGaAs barrier material, giving rise to the dominating luminescence in panel *a*. Placing the spectrometer on the QWR emission and varying the energy of the excitation photons yields the PLE spectrum shown in panel *b*. Labels mark exciton transitions, which similar to quantum wells are strong when quantum numbers of electrons and hole states are equal. An assignment of the transitions is not trivial due to the complicate geometry of the confinement potential; the transition energies indicated in Fig. 26b refer to model calculations, using a QWR shape extracted from electron micrographs. No distinction between heavy and light holes is made: the calculations reveal a strong mixing of the light- and heavy-hole character in the valence-band states (Vouilloz et al. 1998).

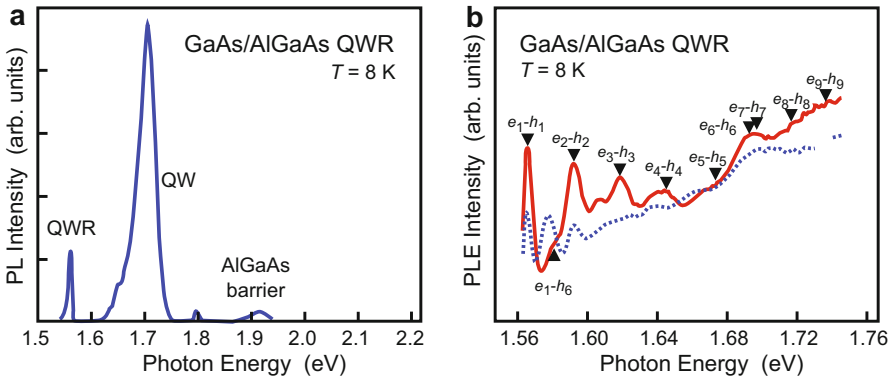


Fig. 26 (**a**) Photoluminescence (*PL*) of a 2.5 nm-thick V-shaped GaAs/$Al_{0.3}Ga_{0.7}As$ quantum wire (*QWR*); the strong QW luminescence originates from sidewall quantum-wells. (**b**) PL excitation spectra polarized parallel (*solid line*) or perpendicular (*dotted line*) the [110]-oriented wire. *Arrows* mark calculated positions of excitonic e_n-h_m interband transitions (After Vouilloz et al. 1997)

2.3.2 Excitons in Quantum Dots

The confinement in all three spatial directions leads to fully quantized electron and hole states of quantum-dot excitons without a kinetic contribution. Exciton transitions are consequently sharp like the discrete spectral lines of atoms.[11] The three-dimensional confinement may prevent dissociation of exciton complexes which are not stable in presence of a translational degree of freedom. A quantum dot with type I band alignment to the surrounding barrier material (Sect. 3 of ▶ Chap. 8, "Bands and Bandgaps in Solids") provides a confinement for negative and positive carriers, creating a multilevel system depending on the number of trapped carriers.

Energy Levels of Excitons

We first discuss a single trapped electron and hole forming a confined exciton.[12] The total angular momentum M of heavy-hole excitons (X) is composed of the electron spin ($s = \pm 1/2$) and the heavy-hole angular momentum ($j = \pm 3/2$), producing four

[11] Still a significant broadening of exciton transitions (of single quantum dots) well above the natural linewidth is observed due to the interaction of the quantum dot with its environment. The interaction with acoustic phonons (deformation potential coupling) and optical phonons (Fröhlich coupling) leads to broad transitions at increased temperature (Rudin et al. 1990); in addition, randomly fluctuating electrical fields of charged defects in the vicinity of the dots lead to a spectral jitter of the transitions on a very short time scale (spectral diffusion) even at low temperature (Türck et al. 2000).

[12] We consider only the heavy-hole exciton due to its lower energy in the generally compressively strained epitaxial quantum dots and their larger oscillator strength.

degenerate exciton states referred to as bright ($M = \pm 1$) and dark ($M = \pm 2$) states; the labels indicate whether they couple to the photon field or not.[13]

The electron-hole exchange interaction may be divided into two ranges. The short-range part leads to a splitting Δ_0 of the exciton multiplett into bright and dark subspaces, independent of the symmetry of the confinement potential. In addition, it mixes the dark states, lifting their degeneracy and forming a dark doublet ($|2\rangle \pm |-2\rangle$) with splitting Δ_2, see Fig. 27. The long-range part also contributes to the bright-dark splitting Δ_0; in addition, it results in a splitting Δ_1 of the bright-exciton states in quantum dots with a low symmetry. The bright-dark splitting was found to increase with decreasing size of a quantum dot due to the strong influence of quantum confinement on Coulomb-interaction energies; the variation of the exchange energy Δ_0 of InGaAs/GaAs dots could be well described by a $1/d_{\mathrm{QD}}^3$ dependence, where d_{QD} is the diameter of the quantum dot (Bayer et al. 2002).

Bright Excitons

The splitting Δ_1 of the bright states is usually called fine-structure splitting. This (often small) splitting is generally observed in the luminescence of quantum dots grown with zincblende semiconductors. Major reasons for the low symmetry of the quantum dot are a structural elongation in one lateral direction (Seguin et al. 2005) and strain-induced piezoelectric fields (Schliwa et al. 2007), but also the symmetry of the interface to the surrounding barrier material provides a small contribution (Bester and Zunger 2005).

If more carriers than a single electron-hole pair are trapped in a quantum dot, the luminescence of an individual dot shows additional radiative exciton relaxations. The excitonic emission spectrum of a single quantum dot shown in Fig. 28 is dominated by the recombination of the neutral exciton labeled X; in addition emissions from positively and negatively charged excitons (X^-, X^+) and those of neutral (XX) and positive biexcitons (XX^+) are observed. We note that the emission energy of several lines is *greater* than that of the exciton; this means that their binding energy with respect to the exciton energy is *negative*: they are in an antibinding state. The energy to keep the particles in a combined state is provided by the 3D confinement.

The binding energy of the exciton complexes depends on the size of the quantum dots. The binding energies of the biexciton (XX) and the positive trion (X^+) increase as the exciton energy decreases, i.e., for larger quantum dots (Fig. 29b); in this trend, the biexciton changes from antibinding to a binding state, as also indicated in Fig. 29a. The negative trion (X^-) is largely unaffected by the QD size and has always positive binding energy (i.e., the emission is shifted to the red). These features originate from the Coulomb interaction and correlation

[13]Pure states with $|M| = 2$ are generally not observed in luminescence experiments (in absence of magnetic fields and confinements with reduced symmetry which can mix them with $|M| = 1$ states), since a $j = \pm 1$ photon cannot induce a radiative transition to the $M = 0$ ground state.

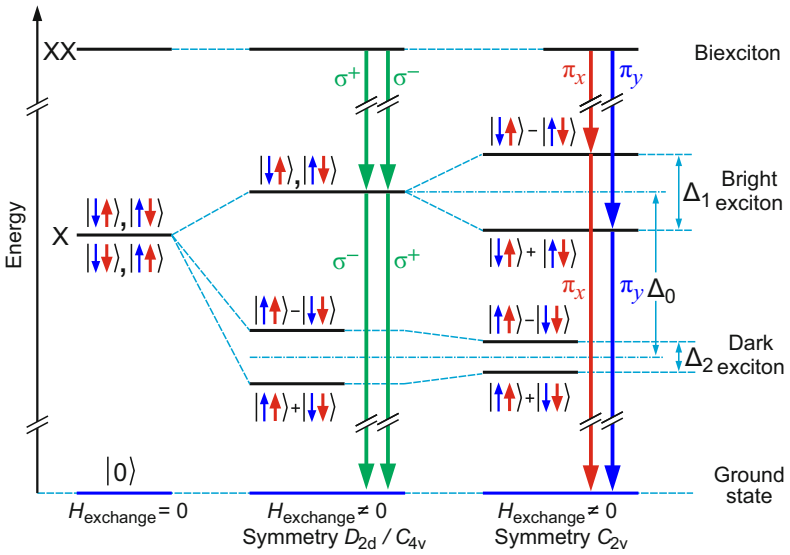


Fig. 27 Energy scheme illustrating the effect of the exchange interaction $H_{exchange}$ on the exciton bright states. The brackets $|\,\rangle$ indicate the spin orientation of the electron (blue arrow) and the hole (thicker red arrow); σ^+,σ^- denotes circularly polarized light, and π_x, π_y linearly polarized light. D_{2d}, C_{4v}, and C_{2v} indicate the symmetry of the confinement potential

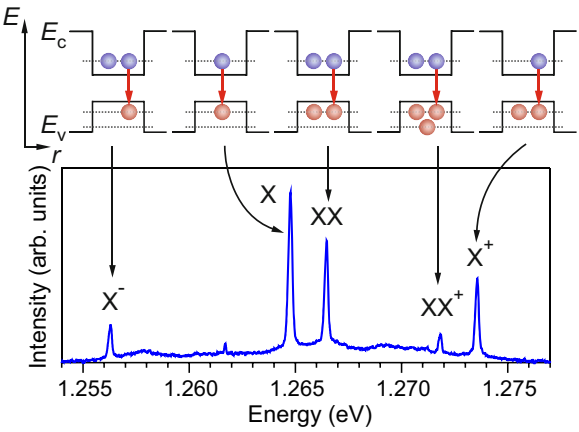


Fig. 28 *Bottom*: Luminescence of a single InAs/GaAs quantum dot showing sharp recombination transitions of neutral and charged excitations and biexcitons (After Rodt et al. 2005). *Top*: Schematic band structure of carriers confined in the quantum dot with recombining electron (*blue*) and hole (*red*) indicated

of the confined particles. The negative trion consists of two electrons and one hole confined in the QD; its binding energy is governed by the difference between the two Coulomb terms $C(e,h)$ and $C(e,e)$. The binding energy of the positive trion X^+ correspondingly depends on $C(e,h)$ and $C(h,h)$. Due to the larger effective mass of holes and the small size of the QDs, the wavefunction of the hole is stronger localized than that of the electron. Consequently $|C(e,e)| < |C(e,h)| < |C(h,h)|$ and the negative trion has a positive binding energy, while the positive trion has a

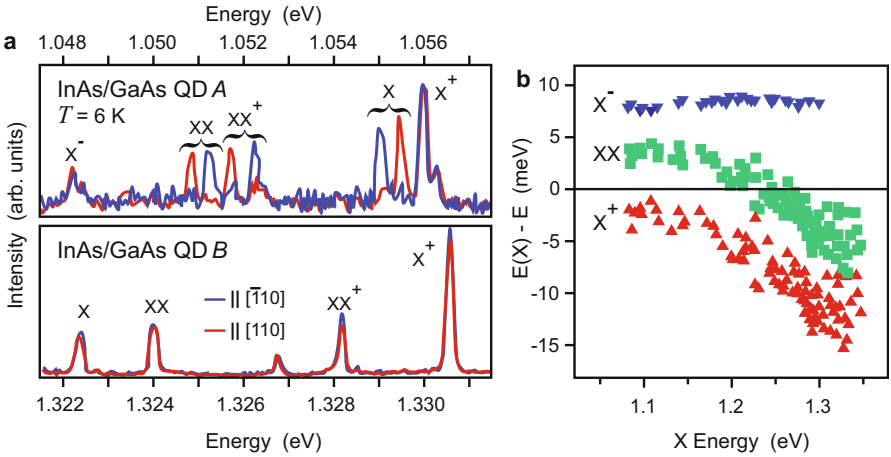


Fig. 29 (**a**) Exciton and biexciton luminescence of a large (*top*) and a small InAs/GaAs quantum dot (*bottom*) for linear polarization along the in-plane $[\bar{1}10]$ or [110] directions. (**b**) Energy differences between the recombination energies of confined exciton complexes (X^-, XX, X^+) and the neutral exciton (X) as a function of the exciton emission-energy (After Pohl 2008)

negative binding energy. Calculations show that the *trend* is due to the number of bound hole states in the QD (Rodt et al. 2005); larger dots (smaller exciton energy) have more bound states and consequently more configuration interaction and increased binding energy.

The exciton emission is split into a doublet as shown in Fig. 29a for different polarization directions; this so-called fine-structure splitting appears in reversed order also in the biexciton emission, and for varied QD sizes it has a trend comparable to that of the biexciton binding energy (Seguin et al. 2005). The mirrored appearance in the exciton and biexciton emission is due to the splitting of the exciton state, which is the final state of the radiative recombination of the biexciton XX to an exciton X (the biexciton state and the exciton ground state are unsplit). The bright-exciton emission from a single quantum dot has recently received much advertence for realizing photonic qubits as discussed in Sect. 3.2.

Dark Excitons

In a "dark" exciton the electron and the heavy-hole spins are oriented *parallel* to each other, yielding an angular momentum $|M| = 2$. Since pure occupied states with $|M| = 2$ do not couple to a photon field there must occur some mixing with the bright-exciton $|M| = 1$ states to allow for an observation in radiative recombination to the $M = 0$ ground state. Some residual mixing occurs already in bulk semiconductors due to the long-range part of the exchange interaction (Ekardt et al. 1979). A significant mixing can be induced either by applying an external magnetic field (Bayer et al. 2002, Germanis et al. 2018) or by lowering the symmetry of an

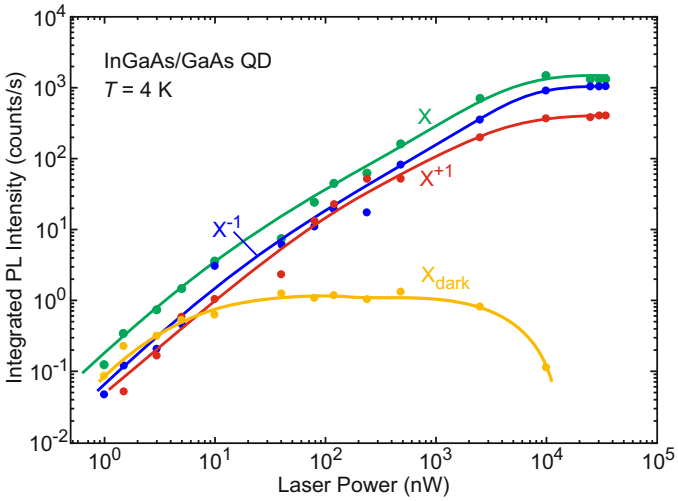


Fig. 30 Luminescence intensity of the dark exciton (*yellow dots and curve*), the bright exciton (*green*), and the trions (*blue, red*) of an InGaAs/GaAs quantum dot as a function of nonresonant laser excitation at 2.725 eV (After Schwartz et al. 2015)

ideally symmetric quantum dot (Zieliński et al. 2015). Without external magnetic field the oscillator strength of the dark exciton is generally some orders of magnitude smaller than that of the bright exciton. Due to the slow radiative decay of the dark exciton (τ_{rad} on the order of 1 μs), the rate of carrier accumulation in the dot exceeds the radiative rate of the dark states by far under usually applied excitation conditions; the additional carriers destroy the dark exciton, and its luminescence cannot be observed (Schwartz et al. 2015). However, when at very low excitation rates the carrier accumulation rate is comparable to the dark-exciton radiative rate, its emission intensity may become comparable to that of the bright-exciton emissions, see Fig. 30.

Since dark-exciton states with weak bright-state mixing couple only weakly to a photon field, also their photo excitation is weak. They can, however, quite efficiently be occupied from a radiative decay of metastable biexcitonic triplet states with parallel heavy-hole spin (Poem et al. 2010). Emission from two of these triplet states[14] are denoted XX_T in Fig. 31. The long lifetime of the dark exciton may be interesting for photonic qubit implementation as outlined in Sect. 3.2.1.

[14]The three triplet states XX_{T+3}, XX_{T0}, and XX_{T-3}, with parallel electron and hole spin are distinguished by the projection of their the total angular momentum (spin + orbital) on the symmetry axis of the quantum dot given by the growth direction.

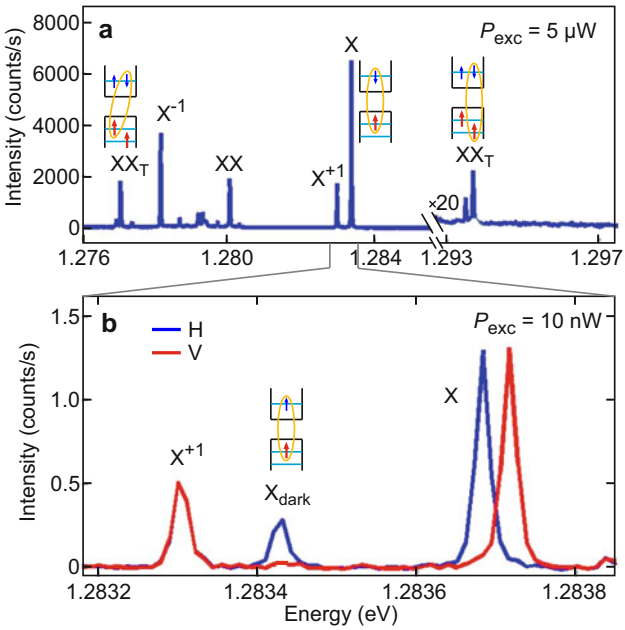


Fig. 31 Photoluminescence spectra of a single quantum dot excited **(a)** by high-intensity nonresonant cw light and **(b)** polarization-sensitive at low excitation intensity. The *insets* at relevant lines illustrate the carrier occupation in the dot; *blue and red arrows* denote electron and heavy-hole spins, respectively. The recombining *e-h* pair is indicated by a *yellow oval* (After Schwartz et al. 2015)

3 Confined Excitons in Quantum Optics

Electronic states of confined excitons are a prominent physical basis for the realization of photonic qubits. Such qubits are needed for the implementation of quantum-information processing and quantum-communication channels and are an active field of research and development. Single photons generated on demand within a semiconductor quantum dot represent a possible platform for qubits. Entangled pairs of two photons produced in quantum dots can be applied as qubits, which are particularly well suited for long haul quantum communication; these qubits enable, e.g., inherently secure data transmission (quantum cryptography). In the following, we discuss the implementation of qubits, properties of qubit photons, and their generation in semiconductors.

3.1 Representation of a Qubit

3.1.1 The Qubit

In our computers, the basic unit of information is encoded in a binary digit, a *bit*, which may have only two distinct values: 0 or 1; there is no value defined in

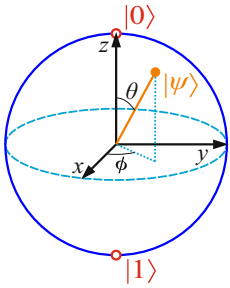


Fig. 32 Representation of a qubit using a Bloch sphere. The *dashed circle* at the equatorial plane $z = 0$ represents states with an equal share of $|0\rangle$ or $|1\rangle$ states

between. The physical realization of such a bit is the level of a small DC voltage of an electronic circuit, being either "low" or "high." Another common implementation is the orientation of the magnetization in the surface of a hard disk.

A *qubit*, or quantum bit, is the basic unit of *quantum* information. Like the bit, its information is usually taken to have the value "0" or "1," but it may also have any value in between. The qubit is represented by a complex vector; it is described by a linear combination of the basis states $|0\rangle$ and $|1\rangle$, yielding the coherent superposition

$$|\psi\rangle = \alpha|0\rangle + \beta\Big|1\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}. \tag{29}$$

The complex numbers α and β are referred to as probability amplitudes. Their absolute squares represent probabilities of the qubit being (partly) in state $|0\rangle$ and (the other part) in state $|1\rangle$; since the total probability has to be unity, they fulfill the condition $|\alpha|^2 + |\beta|^2 = 1$. In addition to their amplitude, the quantities α and β also have a relative phase, giving rise to quantum interference phenomena.

The possible quantum states for a single qubit are usually illustrated using a *Bloch sphere*,[15] which shows the phase and angles of the probability amplitudes. Figure 32 depicts $|\psi\rangle$ as a point on the Bloch sphere in terms of polar coordinates $(r = 1, \theta, \phi)$. Introducing a complex relative phase $e^{i\phi}$, we actually only need to consider the upper hemisphere ($\theta \leq \pi/2$); points at the lower hemisphere differ only by a phase factor $e^{i\pi} = -1$. The probability amplitudes then read

$$\alpha = \cos(\theta/2), \beta = e^{i\phi}\sin(\theta/2), \tag{30}$$

where $e^{i\phi} = \cos\phi + i\sin\phi$ is the relative phase and $0 \leq \theta \leq \pi, 0 \leq \phi \leq 2\pi$. While a classical bit is represented on a Bloch sphere solely at two locations (topmost and lowermost points), a pure qubit state can be represented by any point on the surface of the sphere. The point on the positive x-axis at the dashed

[15]The Bloch sphere is a generalization of the representation of a complex number $z = x + iy$ with $|z|^2 = z^*z = 1$ as a point in the complex plane on the unit circle: $z^*z = (x - iy)(x + iy) = x^2 + y^2$.

circle in Fig. 32, e.g., is given by $\left(|0\rangle + |1\rangle\right)/\sqrt{2}$, that on the positive y-axis at the dashed circle by $\left(|0\rangle + i|1\rangle\right)/\sqrt{2}$.

The *coherent* superposition of the probability amplitudes according to Eq. 29 leads to a state represented by a point on the surface of the Bloch sphere; this is called a *pure* qubit state. The qubit may be put in a *mixed* state by interaction with other qubits or by decoherence. Such states are represented by points *inside* the Bloch sphere. Coherence is generally lost by an interaction of the qubit state with its environment; it is hence also lost by *measuring* the qubit: by gaining information about the state, the magnitudes of α and β are altered. Prior to the measurement, the qubit can be in a coherent quantum superposition of the two states "0" or "1," but upon measurement the qubit yields one bit of the mutually exclusive outcomes "0" or "1"; the result depends on the qubit's quantum state immediately prior to measurement. Operations on single qubits used in quantum information processing can be well described using the Bloch sphere.

Physical Implementations of Qubits

Any quantum-mechanical two-level system can be used as a qubit. There are many physical implementations to create such a system. Often, they employ the spin (up or down with respect to a given reference), either that of trapped ions, or of atomic nuclei, or of electrons. Other schemes utilize photons and encode the information into the polarization. The basis may also be defined by the *number* of quanta, which are given by electrons, photons, flux quanta, or other quantized quantities. Some examples are given in Table 3.

A general issue of qubit systems is the limited duration of their intrinsic coherence. Any system must be adequately isolated from its environment to prevent unintended decoherence, while still providing sufficient coupling to the control signal. In addition, qubits within a register comprising multiple qubits must have individual energy levels to be separately addressed. A further issue is the error rate, which must be sufficiently low to allow for error-correcting operations.

Table 3 Some implementations of qubits. The choice of the basis is arbitrary, the meaning of $|0\rangle$ and $|1\rangle$ may be swapped or differently defined; for a photon, e.g., a diagonal or circularly polarized basis may be used as well

Qubit carrier	Qubit property	Qubit basis	
		$\|0\rangle$	$\|1\rangle$
Photon	Polarization	Horizontal	Vertical
	Photon number	No photon	1 photon
Electron	Spin	Up	Down
	Charge	No electron	1 electron
Josephson junction	Current of flux quantum	Clockwise	Anticlockwise
	Cooper pair charge	n pairs	$(n+1)$ pairs
Quantum dot	Spin	Up	Down
	State	Ground state	Excited state

Qubits in Quantum Information Processing

In a quantum computer, the information to be processed is encoded in the basic units of qubits, which are represented by vectors. Processing of these vectors is essentially matrix vector-multiplication, performed by sequences of quantum gates. Such gates execute reversible, norm-preserving linear transformations on the qubits, which can be considered rotations of the vectors on the Bloch sphere. There are seven general criteria any system for quantum-information processing must meet (DiVincenzo 2000): (1) Implementation of registers with many well characterized qubits, (2) the ability to initialize them to a known state, (3) availability of universal gate operations, (4) longer decoherence than total gate operation times, (5) a proper readout (measurement) capability, (6) implementation of quantum interfaces for qubit interconversion (storage, communication between registers), and (7) availability of quantum interfaces to flying qubits (photons) for long-distance optical communication. We recognize the typical features of a classical computer.

Today an impressive progress is being made in parallel developments on various platforms, which all meet at least some of the requirements. Particularly advanced systems are based on either Josephson charge qubits or photonic qubits and are on the verge to solve real problems, outperforming conventional computers in certain tasks of specific computational complexity (Wendin 2017). Practical future systems will probably be hybrid systems comprising different types of specialized quantum components: qubits, resonators, interfaces, memories (Kurizki et al. 2015). The quantum computer will likely serve as an accelerator for tackling special quantum problems, embedded in an environment of classical high-performance computing.

Quantum Processing Based on Josephson Junctions There are many types of qubit implements using superconducting Josephson junctions: Flux qubits, phase qubits, charge qubits, and others, some of them derived from the level spectrum of the Cooper pair box – a superconducting island connected via Josephson junctions to a grounded electron reservoir (Bouchiat et al. 1998). The transmon version of this device is a charge-noise resilient qubit (Paik et al. 2011), presently applied in advanced processing demonstration systems (Arute et al. 2019; García-Pérez et al. 2020). Qubits are very sensitive to a variety of noise sources; low logical error rates hence proved experimentally crucial. State-of-the-art quantum platforms typically have physical error rates near 10^{-3}. A correctable code can be obtained making use of redundancy via code words with more bits;[16] exponential error suppression for growing number of qubits has recently been demonstrated, an important step toward a scalable fault-tolerant quantum computer (Chen et al. 2021).

[16]A simple classical example is a code word with three bits: 000 and 111; a (single) bit flip can be corrected by a majority vote: 010→000.

Quantum Processing Based on Photonic Qubits Photons used as "flying" qubits offer the advantage of being decoherence-free quantum systems for which measurements and single-qubit operations can be easily performed (e.g., entanglement by a simple 50%/50% beam splitter). The qubit is usually encoded in the polarization, phase, frequency, or time bin of an individual photon (Gisin et al. 2002); basically, several degrees of freedom of the same photon can be used simultaneously. However, since photons interact only weakly, deterministic two-qubit gates are difficult to implement. Only five of the seven requirements for a quantum computer are presently essentially satisfied with photon qubits (Slussarenko and Pryde 2019).

A scalable scheme for photonic quantum computing that required merely linear optics components, single photon detection (measurements), and classical forward-feed were introduced by Knill et al. (2001). While a photonic gate-based quantum computer remains challenging today, a cluster-based quantum computation as proposed by Nielsen (2004) appears presently viable. Photon qubits are well-suited to such a one-way quantum computing procedure, and cluster states[17] combined with *non*deterministic gates can simulate an entangling gate. Moreover, such a scheme can be made tolerant to the main source of noise – the photon loss in an optical system. Since total error correction for compensating photon loss is hardly viable today, present photonic systems are "noisy intermediate-scale quantum devices."

A reconfigurable two-qubit photonic processor for efficiently calculating the expectation value of a Hamiltonian was developed by Peruzzo et al. (2014); in this approach, a single long coherent evolution was replaced by a number of shorter coherent calculations proportional to the number of terms in the Hamiltonian. The quantum variational eigensolver was applied to the ground-state calculation of a helium-hydride cation $He–H^+$ as a subroutine for exponential speedup of the mainly classical calculation. Storage of the global quantum state and quantum energy estimation requires in quantum hardware exponentially fewer resources than the exact diagonalization by classical hardware. Using a modified version of this quantum variational eigensolver and a total number of 22 *transmon* qubits, Robert et al. (2021) recently simulated the folding of the 10-amino-acid protein angiotensin.

3.1.2 Requirements for Qubit Photons

Different Sources of Radiation

Photon sources have quite different properties. They may be classified as source of thermal, coherent, or nonclassical radiation. The characteristics of a light source can be assessed by measuring the coherence properties in a correlation of emitted photons (Hanbury Brown and Twiss 1956). The resulting second-order autocorrelation function $g^{(2)}(\tau)$ reads for an electromagnetic wave with intensity I (Lounis and Orrit 2005)

[17]Cluster states are formed by an entangled network of qubits (Raussendorf and Briegel 2001; Nielsen 2004).

$$g^{(2)}(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2}. \tag{31}$$

Here the $\langle\ \rangle$ brackets signify the average emission rate (photons/second) or, in a quantum-mechanical picture of photon statistics, the expectation value of the photon occupation number. In the single-photon picture, the quantity $g^{(2)}(\tau)$ can be considered the probability for the emission of a photon at time $t + \tau$, after the preceding emission of a photon at time t. The special case $\tau \to 0$ describes the simultaneous emission of more than one photon. For $\tau \to 0$, Eq. 31 can be reduced to

$$g^{(2)}(0) = 1 + \frac{\Delta\mu^2 - \mu}{\mu^2}, \tag{32}$$

where μ is the mean photon number and $\Delta\mu^2$is the variance. $g^{(2)}(0)$ yields characteristic differences for various kinds of radiation sources.

Thermal Radiation The emission of a thermal light source (e.g., an incandescent lamp) is an incoherent superposition of the emission of a large number of independent emitters. The probability distribution of photons $P_{\text{thermal}}(n, \mu)$ follows the Bose-Einstein distribution of black body radiation (Lounis and Orrit 2005), yielding

$$P_{\text{thermal}}(n,\mu) = \mu^n / \left[(\mu+1)^{n+1}\right], \tag{33}$$

where n is the number of photons. Fluctuations in the number of photons n is larger than the mean photon number μ; the variance is given by $\Delta\mu^2 = \mu^2 + \mu$. We note that for thermal radiation, the state with no photon (i.e., $n = 0$) always has highest occupation probability (see Fig. 33a).

Inserting the variance into the second-order autocorrelation Eq. 32, we obtain for thermal light the value $g^{(2)}_{\text{thermal}}(0) = 2$, while away from $\tau = 0$ the mean value is $g^{(2)}_{\text{thermal}}(\tau) = 1$, cf. Fig. 33b. The increased probability to emit more than one photon at a time is referred to as *bunching*; classical (thermal) photon sources emit photons in irregularly occurring bunches.

Coherent Radiation The emission of a coherent light source (e.g., a stable laser) has a much lower fluctuation of n. The probability distribution of photons $P_{\text{coherent}}(n, \mu)$ is given by the Poisson distribution

$$P_{\text{coherent}}(n,\mu) = e^{-\mu}\mu^n / n!. \tag{34}$$

The variance is much smaller than for thermal radiation and equals the mean value, $\Delta\mu^2 = \mu$. The corresponding Gaussian noise is referred to as shot noise or photon

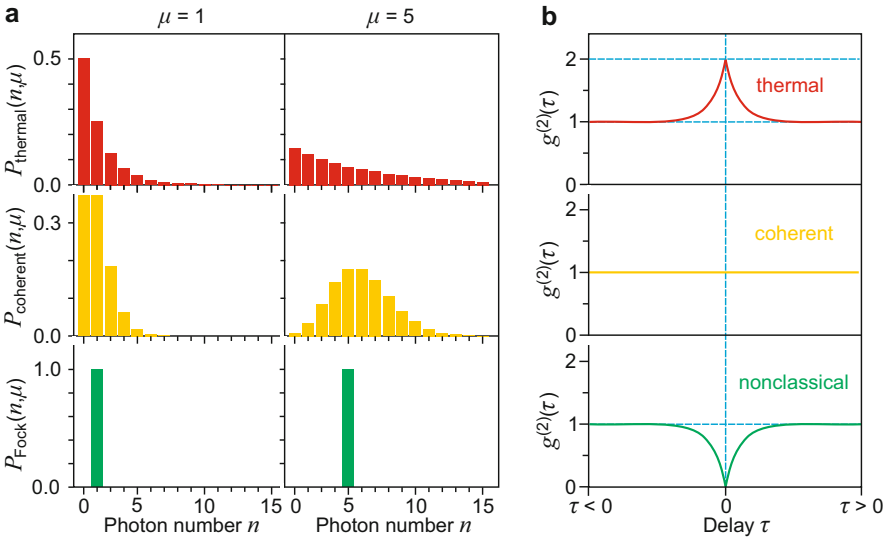


Fig. 33 (**a**) Probability $P(n, \mu)$ for the emission of photons, when a mean photon number μ (either 1 or 5) is given, illustrated for three different kinds of light sources; n is the number of photons with the sum of their probability yielding 1. $P(n, \mu)$ is shown for a thermal source (*top*), a coherent source (*central row*), and a nonclassical source (*bottom*); note different scaling. (**b**) Schematic of the second-order autocorrelation function $g^{(2)}(\tau)$ of the different sources

noise. Coherent radiation describes the emission of a macroscopic laser; the shot noise is the lower bound for the noise of the laser.

The variance leads with Eq. 32 to a second-order autocorrelation $g^{(2)}_{\text{coherent}}(0) = 1$ for a coherent radiation source. There is a maximum probability to find μ photons in the mode, with no bunching effect for $\tau \to 0$ as illustrated in Fig. 33b.

Nonclassical Radiation Nonclassical light, also called squeezed light, is characterized by Fock states $|n_i\rangle$. A Fock state represents a well-defined photon number n; the probability distribution is then given by

$$P_{\text{Fock}}(n) = \delta_{n,n_i}. \tag{35}$$

The mean value μ equals the expectation value n_i, and the variance is zero. This behavior is distinctly different from that of the other sources. Since there is no variance and $\mu = n_i$, we obtain from Eq. 32

$$g^{(2)}(0) = 1 - \frac{1}{n_i}. \tag{36}$$

An ideal source, which emits always a single photon at a time ($n_i = 1$), is hence characterized by $g^{(2)}(0) = 0$. There is an absence of coincidences of two photons in

one-photon states; this effect of anticorrelation is referred to as antibunching – see Fig. 33b. A (non-ideal) source emitting exactly two photons simultaneously ($n_i = 2$) yields $g^{(2)}(0) = 0.5$. This value is considered the upper limit of an experimental single-photon source.

The probability $P(n, \mu)$ for the emission of photons from the three different radiation sources and their second-order autocorrelation are compared in Fig. 33.

Single and Entangled Photons

The properties of photons suitable for encoding qubits depend on the application. Simple communication channels using "flying" qubits merely need single photons with high purity,[18] produced at a high rate and on demand. More advanced quantum communication requires *entangled* photon pairs or even *indistinguishable* photons; the latter are inevitably required for quantum computing. Certainly single-photon Fock states are ideal for all these tasks, but they are difficult to fabricate.

Single Photons Practical implementations for single-photon generation usually rely on faint laser pulses or spontaneous parametric downconversion (Gisin et al. 2002). Such photons can be applied for quantum cryptography by quantum key distribution using, e.g., the BB84 protocol presented by Bennett and Brassard (1984).

A strongly attenuated coherent laser pulse (small μ) produces approximately single-photon number-states, but still obeys Poisson statistics (Fig. 33a). There is hence a finite probability that a nonempty weak coherent pulse contains more than one photon, $P_{\text{multi}} \approx \mu/2$; this quantity can be made very small, but then most laser pulses are empty and the transferred bit rate gets small.

In spontaneous parametric downconversion, the nonlinear interaction of a pump-laser pulse with a non-centrosymmetric crystal (e.g., KDP) occasionally splits one pump photon into two lower-frequency photons (Lounis and Orrit 2005); one photon can be used as a heralding trigger to activate detection of the other one. The efficiency of a suitable setup may exceed that of attenuated laser emission, but the photon states still obey Poisson statistics.

An ideal single-photon source emits one and only one photon upon a trigger. A single two-level system is a good approach, because it cannot emit two photons at a time. Promising implementations are the nitrogen-vacancy center in diamond[19] and electron-hole pairs in a semiconductor quantum dot; the latter are compatible with semiconductor technology and discussed in Sect. 3.2.1.

[18] Meaning $g^{(2)}(0) \ll 0.5$, ideally 0; there must actually be only a single photon (and not more) present at a time.

[19] A substitutional nitrogen atom with a vacancy at an adjacent lattice position (Kurtsiefer et al. 2000).

Entangled Photon Pairs *Entanglement* is a fundamental manifestation of quantum superposition and nonclassical correlations for pure states (Wang et al. 2016). For illustration we consider the equal superposition of two qubits, which both have their probability amplitudes $\alpha = 0$ and $\beta = 1$; this yields[20] a composite system with the entangled state $|\Phi^+\rangle = 1/\sqrt{2}\left(|00\rangle + |11\rangle\right)$. In this two-qubit Bell state, it is no longer possible to assign a single state vector to any of the two subsystems, only to the total system. A measurement of $|\Phi^+\rangle$ results in equal probabilities of measuring either the product state $|00\rangle$ or $|11\rangle$, i.e., the result is $\left(1/\sqrt{2}\right)^2 = 1/2$. It is not possible to tell whether the first qubit has value "0" or "1." When these subsystem qubits are spatially separated and one of this pair is measured, its value is known; then immediately also the (identical) value of the second qubit is instantaneously known, even without measuring and whatever the distance separating the pair (Giustina et al. 2015).

Entanglement of two photons may refer to different degrees of freedom, e.g., polarization, energy, or time bin. If this property is employed for encoding in communication networks, their quantum correlation can be used for quantum key distribution in simple (Sasaki et al. 2011) and larger networks based on trusted nodes (Lo et al. 2014). The no-cloning theorem of quantum mechanics essentially ensures secure communication even in the presence of eavesdropping (Wootters and Zurek 1982).

Indistinguishable Photons Advanced applications such as quantum repeaters or quantum computing employ entanglement distribution via Bell-state measurements; this requires quantum-light sources emitting *indistinguishable* photons. If two independent photons with this property impinge simultaneously on a 50%/50% beam splitter, they interfere in such a way that they inevitably *both* exit from one of the two output ports (Hong et al. 1987): they coalescence into a two-photon state. Pure spontaneous emission by an ideal two-level system leads to perfectly indistinguishable photons. Practical implementations suffer, however, from dephasing (Bylander et al. 2003) and spectral diffusion (Türck et al. 2000) by, e.g., phonon coupling and charge fluctuations in the environment.

Production of indistinguishable photons from spatially separated single-photon sources requires precise energy matching. In measurements given in Fig. 34, thermal shifting of one quantum-dot emission was applied to match the emission energy of another quantum dot in a different sample (Thoma et al. 2017). Both dots were quasi-resonantly excited into an excited state (p shell) by a pulsed laser, and their spatially superimposed photons were directed on a 50%/50% beam splitter with two output ports for detecting the two-photon interference. The individual dots act as single photon emitters as shown by the Hanbury Brown and Twiss autocorrelation function in Fig. 34b. The (not perfect) coalescence into a two-photon state

[20]There are four (maximally entangled) two-qubit states designated as *Bell states*. In addition to $|\Phi^+\rangle$ noted in the text these are $|\Phi^-\rangle$ (with "+" replaced by "-" in the sum) and $|\Psi^{+/-}\rangle$ with α and β being 0 and 1 in the first and 1 and 0 in the second qubit and the sum built with either "+" or "−."

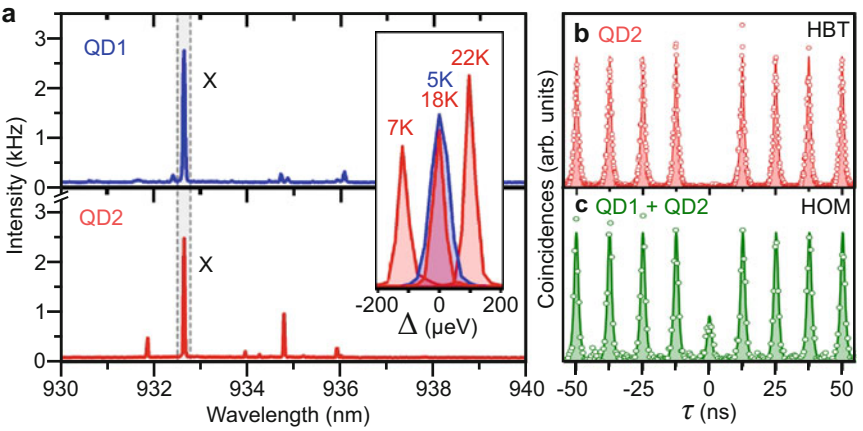


Fig. 34 (**a**) Photoluminescence spectra of two InGaAs/GaAs quantum dots (*QD1*, *QD2*) located in different samples. Inset: Thermal tuning of QD2 to match the emission energy of QD1. (**b**) Hanbury Brown and Twiss autocorrelation measurement of QD2 yielding $g^{(2)}_{\mathrm{HBT}}(0) = 0.01 \pm 0.01$. (**c**) Hom-Ou-Mandel histogram $g^{(2)}_{\mathrm{HOM}}(\tau)$ of the two remote QD single-photon sources yielding photon indistinguishability with a visibility of $V_{1+2} = 29\%$. (After Thoma et al. 2017)

is seen in the Hom-Ou-Mandel coincidence histogram of the mutual interference of the two remote single-photon sources shown in Fig. 34c. For perfect two-photon Fock states $|2\rangle$, the histogram $g^{(2)}_{\mathrm{HOM}}(\tau)$ shows a vanishing number of coincidences at the two interferometer outputs for zero time delay $\tau \rightarrow 0$; see also Fig. 37 for indistinguishable photons from a single source.

3.2 Qubit Photons from Semiconductor Quantum Dots

Several approaches are being developed to implement a solid-state base for quantum-information technology. An early scheme proposed the use of nuclear spins of ^{31}P donors implanted into spin-zero ^{28}Si crystals (Kane 1998); decoherence times of 30 s (Muhonen et al. 2014) and robust two-qubit gates demonstrate substantial progress in the field (Kalra et al. 2014; McCallum et al. 2021). Another advanced spin system are qubits of nitrogen-vacancy centers in diamond (Doherty et al. 2013), which were, e.g., employed for simulating the bond-dissociation curve of the helium hydride cation He-H$^+$ (Wang et al. 2015). Much work focuses on semiconductor quantum dots, which can provide both single and pairs of entangled photon qubits on demand.

3.2.1 Single-Photon Emission from Bright and Dark Excitons

Bright Excitons The bright-exciton states illustrated in Fig. 27 may serve as excited state of a two-level system for producing single photons on demand after

optical or electrical excitation. This applies as well for other emissions from the quantum dot shown in Fig. 28 and discussed in Sect. 2.3.2. In contrast to the exciton states, those of the trions (X^- and X^+) are not split. In the negative trion, the two electrons occupy a singlet state with zero angular momentum; there is hence no exchange interaction with the angular momentum of the additional hole. An analogous case applies for X^+.

Many approaches for implementing quantum dots as single-photon sources are being pursued (Arakawa and Holmes 2020). Highest single-photon purity is obtained using resonant optical excitation (Ollivier et al. 2020; see also Fig. 34b). Electrically driven deterministic single-photon sources based on epitaxial quantum dots are of particular interest though, albeit all efficient sources still require cryogenic cooling to suppress phonon coupling and to freeze out nonradiative pathways.[21]

The emission of an individual site-controlled InGaAs/GaAs quantum dot, implemented into a *pin* diode structure, is shown in Fig. 35 (Pohl et al. 2020). Besides the exciton and biexciton doublets, we observe the narrow line of the negative trion X^-. The photons of this emission were analyzed in a cw Hanbury Brown and Twiss correlation measurement, yielding an almost pure single-photon characteristic with $g^{(2)}(0)$ near zero.

There are presently two major issues for fabricating a single-photon source based on quantum-dot emission. (1) The efficiency of photon extraction from the device is generally poor due to internal reflection originating from the refractive index step between host material and environment. A variety of cavities and photonic structures were developed to overcome these losses and enhance the

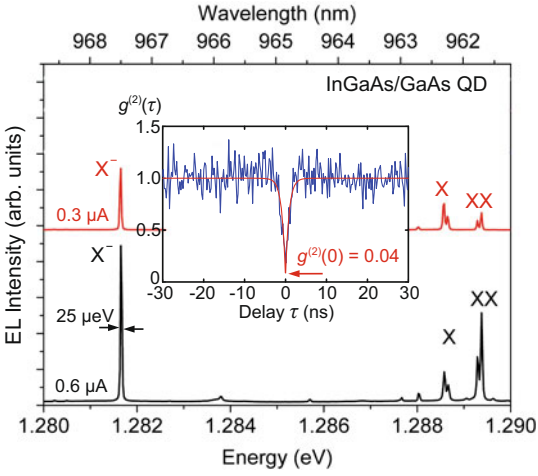


Fig. 35 Low-temperature electroluminescence spectra of a single selected InGaAs/GaAs quantum dot in a *pin* diode. Inset: The photon autocorrelation measurement of the X^- line recorded at an injection current of 0.32 μA shows clear antibunching. (After Unrau et al. 2012)

[21]Optically excited single-photon emission with $g^{(2)}(0) = 0.33$ at room temperature was achieved using a core-shell GaN/AlGaN nanowire quantum-dot (Holmes et al. 2014).

emission directionality (Rodt et al. 2020 and references in Arakawa and Holmes 2020). (2) To achieve indistinguishability of photons (e.g., for entanglement swapping), single-photon sources with precisely matched emission energy are required within the emission's homogeneous linewidth of about 1–2 μeV. Spectral fine-tuning is accomplished by temperature tuning due to the bandgap shift (Farrow et al. 2008; see also Fig. 34a), electric fields via the quantum-confined Stark effect (Ellis et al. 2018), and strain-tuning (Ding et al. 2010). Still the purity, collection efficiency, and two-photon interference visibility will have to be improved for deploying the sources in practical systems.

Employing the *spin* of a confined bright exciton as an exciton spin-qubit requires full coherent control, i.e., to rotate it to a defined position at will. This can be achieved by one single optical pulse of defined polarization and duration (Stievater et al. 2001). The qubit must, however, be used within the short (sub-ns) lifetime, which is limited by radiative recombination of the electron-hole pair. This restriction makes the long-living dark excitons interesting.

Dark Excitons Dark excitons couple only weakly to a photon field,[22] see Sect. 2.3.2. They may still be beneficial for storing or generating a photon qubit, because a dark-exciton qubit lives three orders of magnitude longer than the bright-exciton qubit and maintains its coherence for about 100 ns (Schwartz et al. 2015). Coherent control of the dark-exciton spin was demonstrated by applying a resonant excitation of an excited dark-exciton state (Schwartz et al. 2015b). This spin-parallel state provides some optical activity by a small mixing with the spin-antiparallel bright-exciton states. The steps for defining and later probing this state is illustrated in Fig. 36a. A resonant polarized laser pulse occupies the excited dark-exciton state $|X^*_{dark}\rangle$ ("writing"), which rapidly relaxes to the ground state $|X_{dark}\rangle$ nonradiatively by a spin-preserving phonon emission. Spin state and population of the dark exciton are then probed ("reading") by excitation with a second laser into the triplet biexciton state $|XX_T\rangle$; the magnitude of the absorption into this state depends on the dark-exciton population and polarization at the photon absorption time. The emitted biexciton photon monitors the temporal dependence of the absorption.

The dark-exciton state is a superposition of spin-parallel states with some addition of spin-antiparallel (bright) states. The polarization of the creating pulse defines the phase and amplitude between the σ^- circular polarization-induced transition (left side in Fig. 36) and the σ^+ polarization (right side). After relaxation to the ground state, the coherent dark-exciton state evolves in time. Since this state is a coherent superposition of both transitions, it precesses between the $|+2\rangle$ and $|-2\rangle$ spin states. The monitoring $|XX_T\rangle$ emission shows clear Rabi oscillations of this cyclic precession,[23] differing from that of the simultaneously measured bright exciton (Fig. 36b).

[22]The luminescence of the dark exciton is shown in Fig. 31b.

[23]With a π pulse, the system changes, e.g., from "qubit state" $|0\rangle$ to $|1\rangle$; with 2π, it returns to its initial state.

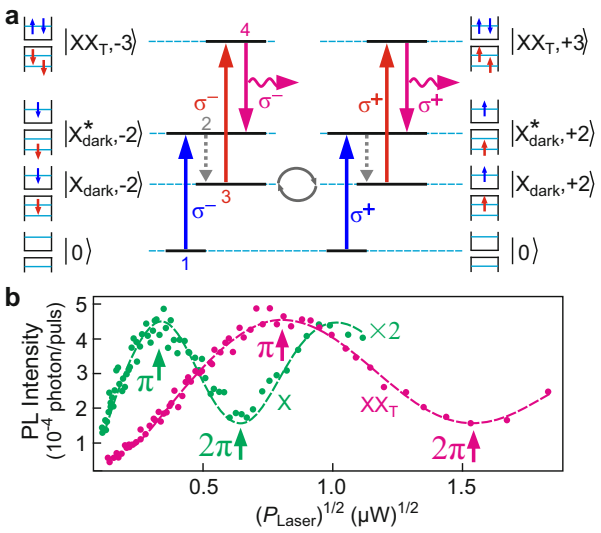


Fig. 36 (**a**) Left: Steps for creating a dark exciton $|X_{dark}, -2\rangle$ via an excited state $|X^*_{dark}, -2\rangle$: *1* excitation (*blue arrow*), *2* nonradiative relaxation (*gray dotted arrow*), 3 probing (*red arrow*), *4* radiative detection. $|J\rangle$ describes the total angular momentum; the *gray cyclical arrow* indicates precession between the $|+2\rangle$ and $|-2\rangle$ spin states. (**b**) Luminescence intensity depending on the square root of the laser power; the laser is tuned to excited resonances of the dark exciton (*magenta*) and bright exciton (*green*). *Dashed lines* are model fits to the experimental data; intensities corresponding to π and 2π area pulses are marked. (After Schwartz et al. 2015b)

The scheme presented above requires a setup with two dedicated lasers. It should be noted that more recently also the creation of a dark exciton by a *single* laser pulse was proposed, employing an optical excitation with a longitudinal field component (Holtkemper et al. 2021).

3.2.2 Generation of Entangled Photon Pairs

The biexciton-exciton cascade XX → X → 0 depicted in Fig. 27 provides two energy-indistinguishable paths for zero fine-structure splitting of the intermediate X level.[24] The two photons emitted from such a cascade then form a pair with entangled polarization. This was verified using a respective InGaAs/GaAs quantum dot optically excited into the biexciton level (Müller et al. 2014). Cross-correlation measurements proved that, e.g., after a right-circular photon (emitted from the XX level) generally a left-circular photon (emitted from X) followed, i.e., an antibunching occurred for parallel polarizations. By contrast, in the linear and diagonal bases, the photons show strong bunching when the polarizations of the biexcitonic and excitonic photons are parallel, and antibunching when they are orthogonal. Still the two photons have different energy, separated by the exciton binding-energy.[25]

[24]The circularly polarized transitions are a superposition of two cross-polarized linear transitions. The fine-structure splitting is negligible if it is smaller than the radiadive linewidth.

[25]Even if the biexciton binding-energy gets zero for larger quantum dots (Fig. 29b), a nonzero fine-structure splitting of such dots leads usually to photon pairs of different energies.

Usually the exciton fine-structure splitting of the exciton is finite, and for alloyed self-assembled InGaAs quantum dots, a residual splitting is *generally* expected (Singh and Bester 2010). Much effort was devoted to reduce the splitting, e.g., by externally applied piezoelectric-induced strain (Zhang et al. 2015) or electric field tuning (Bennett et al. 2010). It should be noted that even for a finite fine-structure splitting, entanglement can be obtained due to Heisenberg's relation: a high temporal resolution in determining the dynamics of the XX-X decay is connected to a larger uncertainty in energy. If the latter is larger than the splitting, the "which-path information" is lost (Akopian et al. 2006).

Characteristics of a bright state-of-the-art source of strongly entangled photon pairs from a XX → X → 0 cascade are schown in Fig. 37. A "droplet" GaAs/AlGaAs quantum dot was integrated into a resonant cavity composed of a lateral concentric AlGaAs Bragg grating and a broadband gold mirror beneath; thereby a high photon-extraction efficiency of 85% for both X and XX emissions was achieved (Liu et al. 2019). Due to the Purcell effect (see Sect. 2.2.2 of ▶ Chap. 10, "Interaction of Light with Solids"), the lifetime of the exciton is shortened from 210 ps to 60 ps, corresponding to a lifetime-limited linewidth of ~11 μeV. Hereby a high time-averaged fidelity is obtained despite a measured fine-structure splitting of 4.8 μeV of this dot. The degree of correlation in a particular polarization basis is given by (Hudson et al. 2007)

$$C = \frac{g^{(2)}_{\mathrm{XX,X}}(0) - g^{(2)}_{\mathrm{XX},\overline{\mathrm{X}}}(0)}{g^{(2)}_{\mathrm{XX,X}}(0) + g^{(2)}_{\mathrm{XX},\overline{\mathrm{X}}}(0)}, \tag{37}$$

where $g^{(2)}_{\mathrm{XX,X}}(0)$ and $g^{(2)}_{\mathrm{XX},\overline{\mathrm{X}}}(0)$ are the second-order correlations for the co-polarized and cross-polarized photon pairs in the considered basis at zero delay time, respectively. The correlations in the polarization bases determined from the measured coincidence histograms in Fig. 37a, b are $C_{\mathrm{linear}} = 0.92$, $C_{\mathrm{circular}} = -0.80$ (Liu et al. 2019); additionally, $C_{\mathrm{diagonal}} = 0.81$ was extracted. From these values, the entanglement fidelity f for the polarization between the emitted photon pairs can be calculated by (Hudson et al. 2007)

$$f = \frac{1 + C_{\mathrm{linear}} + C_{\mathrm{diagonal}} - C_{\mathrm{circular}}}{4}, \tag{38}$$

yielding a value of 0.88 for this photon-pair source.

Hom-Ou-Mandel two-photon interferences to measure the indistinguishability of the photons are shown in Fig. 37c–f (Liu et al. 2019). The quantum dot is optically excited under a two-photon resonant scheme by two consecutive pulses, and the emitted single photons are coupled to an unbalanced Mach–Zehnder interferometer equipped with a delay to compensate for the pulse delay; in one

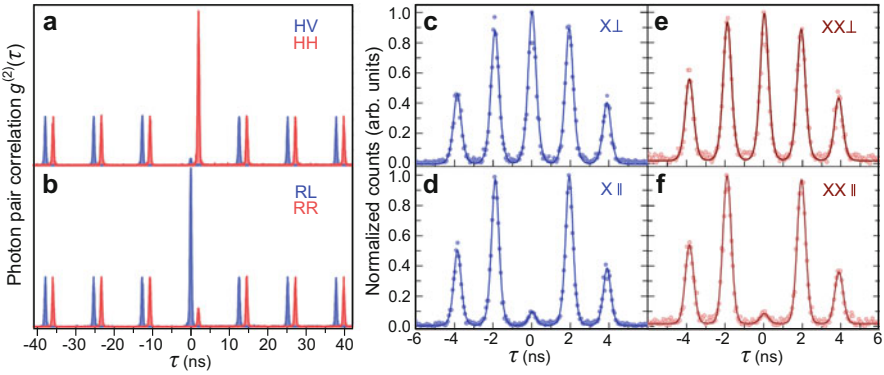


Fig. 37 (**a**, **b**) Polarization entanglement measured for two photons from the XX → X → 0 cascade of a quantum dot. In the polarization-dependent cross-correlation histogram for (**a**) linear and (**b**) circular basis, the *red data* of *HH* (both photons horizontally) and *RR* (both photons right circularly polarized) are shifted for clarity. (**c–f**) Hom-Ou-Mandel interference measurement for determining the indistinguishability of the two photons. Two-photon interference for cross-polarized (**c**) and co-polarized (**d**) X photons, and for cross-polarized (**e**) and co-polarized (**f**) XX photons. *Solid curves* are fit to the data. (After Liu et al. 2019)

arm of the interferometer, a half-wave plate may prepare cross-polarized or copolarized photons to make them either distinguishable or indistinguishable in polarization. Finally, the photons are interfered at a beam splitter in co- and cross-polarized configurations. In the copolarized configuration (Fig. 37d, f), the coincidence counts at zero delay are greatly suppressed, indicating two-photon interference of $|2\rangle$ Fock states at the beam-splitter output ports; in cross-polarized configuration (Fig. 37c, e), however, the central peaks have similar counts as the side peaks, indicating no indistinguishability. From the central peak areas at $\tau = 0$, visibilities of two-photon interference of 0.901 for X and 0.903 for XX are extracted.

An interesting case occurs for a XX → X → 0 cascade with an equal value of the biexciton binding energy and the exciton fine-structure splitting, $\left|E_{\mathrm{b}}^{\mathrm{XX}}\right| = E_{\mathrm{FSS}}$ $(\equiv \Delta_1)$. The *horizontally* polarized path (π_x in Fig. 27) then emits two equally polarized photons of equal energy, $E_{\mathrm{H}}^{\mathrm{XX}} = E_{\mathrm{H}}^{\mathrm{X}}$, which are additionally highly correlated in time. The autocorrelation $g_{\mathrm{auto}}^{(2)}(\tau)$ of such *twin photons* emitted from a suitable quantum dot shows a pronounced symmetrical bunching at $\tau = 0$. In contrast, the cross-correlation $g_{\mathrm{cross}}^{(2)}(\tau)$ of the two *vertically* polarized photons with start trigger from the XX emission and stop from the (lower energy) X emission exhibits a clear antibunching for $\tau < 0$, immediately followed by a pronounced bunching for $\tau > 0$ (Heindel et al. 2017).

4 Summary

Excitons are quasiparticles combining an electron and a hole within their mutual Coulomb field; they resemble hydrogen atoms, except that the positively charged partner is not a heavy proton but a hole of about the same mass as an electron. Excitons show a hydrogen-like excitation spectrum; however, because of the perturbation from the lattice, other than *S* states are also observed in optical transitions, while in a free hydrogen atom, all optical transitions are equivalent to transitions between *S* states. One distinguishes large (Wannier–Mott) and small (Frenkel) excitons. Large excitons, with a typical diameter above three lattice constants, have a small ionization energy, typically on the order of 20 meV, while small excitons have ionization energies on the order of 1 eV. Wannier–Mott excitons are found in typical, mostly covalent semiconductors with large dielectric constant and small effective mass, while Frenkel excitons are found in ionic or organic molecule crystals. Many kinds of excitons are observed depending on the type of hole within the light, heavy, or spin–orbit split-off band and on the type of electron: from a band at Γ (a direct exciton) or from a satellite valley (an indirect exciton). The interaction of an exciton with a charge carrier may form a charged exciton, a trion; similarly neutral or charged biexcitons can be formed. The binding energy of such free exciton molecules or higher associates is typically well below 20% of the exciton binding energy.

In low-dimensional semiconductors the binding energy and oscillator strength of all excitonic associates are significantly enhanced. Strongest transitions of confined excitons occur between electron and hole states of equal principal quantum numbers, similar to bulk excitons. In quantum dots, the three-dimensional confinement allows for forming stable antibinding exciton associates.

Excitons have a profound effect on the optical absorption spectrum by permitting absorption slightly below the corresponding band edge and by substantially increasing the absorption near the edge, but inside the intrinsic band-to-band range. Their analysis permits significant insight into the bonding type of the semiconductor, the dielectric function, and effective mass, all of which influence the spectrum. In low-dimensional semiconductors, they can be used as an optical probe of the perfection of the interlayer boundaries on an atomic scale and can also be used to analyze in detail the influence of a variety of field perturbations, e.g., internal strain. As such, they offer an important potential for analytical purposes.

Excitonic states of quantum dots are used as physical basis for the realization of photonic qubits, with their information encoded in the polarization, phase, frequency, or time bin of an individual photon. Exciton and biexciton emission of a quantum dot is employed to produce single photons and pairs of entangled photons on demand. Tuning of quantum-dot emission by strain, temperature, or other means to match the emission energy of a distant quantum dot can produce indistinguishable photons deployed, e.g., for entanglement swapping in quantum repeaters. The quantum correlation of the generated photons can be applied for quantum-key distribution to establish inherently secure data communication. Spin-forbidden optical transitions from the dark exciton, which is composed of electron and hole

with parallel spins, become accessible by residual mixing with bright-exciton states; due to their long lifetime and coherence dark excitons may become interesting for qubit storage. Photonic quantum technology based on semiconductors is presently a rapidly developing field.

References

Agranovich VM, Mills DL (1982) Surface polaritons. North-Holland, Publ. Co., Amsterdam
Akimoto O, Hanamura E (1972) Excitonic molecule. I. Calculation of the binding energy. J Phys Soc Jpn 33:1537
Akiyama H (1998) One-dimensional excitons in GaAs quantum wires. J Phys Condens Matter 10:3095
Akopian N, Lindner NH, Poem E, Berlatzky Y, Avron J, Gershoni D, Gerardot BD, Petroff PM (2006) Entangled photon pairs from semiconductor quantum dots. Phys Rev Lett 96:130501
Altarelli M, Bachelet G, Del Sole R (1979) Theory of exciton effects in semiconductor surface spectroscopy. J Vac Sci Technol 16:1370
Arakawa Y, Holmes MJ (2020) Progress in quantum-dot single photon sources for quantum information technologies: A broad spectrum overview. Appl Phys Rev 7:021309
Arute F et al (2019) Quantum supremacy using a programmable superconducting processor. Nature 574:505
Astakhov GV, Kochereshko VP, Yakovlev DR, Ossau W, Nürnberger J, Faschinger W, Landwehr G, Wojtowicz T, Karczewski G, Kossut J (2002a) Optical method for the determination of carrier density in modulation-doped quantum wells. Phys Rev B 65:115310
Astakhov GV, Yakovlev DR, Kochereshko VP, Ossau W, Faschinger W, Puls J, Henneberger F, Crooker SA, McCulloch Q, Wolverson D, Gippius NA, Waag A (2002b) Binding energy of charged excitons in ZnSe-based quantum wells. Phys Rev B 65:165335
Bajaj KK, Reynolds DC (1987) An overview of optical characterization of semiconductor structures and alloys. Proc SPIE 0794:2
Baldereschi A, Lipari NO (1973) Spherical model of shallow acceptor states in semiconductors. Phys Rev B 8:2697
Bar-Joseph I (2005) Trions in GaAs quantum wells. Semicond Sci Technol 20:R29
Bassani F, Pastori-Parravicini G (1975) Electronic states and optical transitions in solids. Pergamon Press, Oxford
Bayer M, Ortner G, Stern O, Kuther A, Gorbunov AA, Forchel A, Hawrylak P, Fafard S, Hinzer K, Reinecke TL, Walck SN, Reithmaier JP, Klopf F, Schäfer F (2002) Fine structure of neutral and charged excitons in self-assembled In(Ga)As/(Al)GaAs quantum dots. Phys Rev B 65:195315
Beinikhes IL, Kogan ShM (1985) Influence of valence band degeneracy on the fundamental optical absorption in direct-gap semiconductors in the region of exciton effects. Sov Phys JETP 62:415
Bennett CH, Brassard G (1984) Proc IEEE Int Conf Computers, Systems and Signal Processing, Bangalore, pp 175–179
Bennett AJ, Pooley MA, Stevenson RM, Ward MB, Patel RB, Boyer de la Giroday A, Sköld N, Farrer I, Nicoll CA, Ritchie DA, Shields AJ (2010) Electric-field-induced coherent coupling of the exciton states in a single quantum dot. Nat Phys 6:947
Bester G, Zunger A (2005) Cylindrically shaped zinc-blende semiconductor quantum dots do not have cylindrical symmetry: Atomistic symmetry, atomic relaxation, and piezoelectric effects. Phys Rev B 71:045318
Birkedal D, Singh J, Lyssenko VG, Erland J, Hvam JM (1996) Binding of quasi-two-dimensional biexcitons. Phys Rev Lett 76:672

Blackwood E, Snelling MJ, Barley RT, Andrews SR,Foxon CTB (1994) Exchange interaction of excitons in GaAs heterostructures. Phys Rev B 50:14246
Bouchiat V, Vion D, Joyez P, Esteve D, Devoret MH (1998) Quantum coherence with a single Cooper pair. Physica Scripta T76:165
Brinkman WF, Rice TM, Bell B (1973) The excitonic molecule. Phys Rev B 8:1570
Broser I, Broser R, Beckmann E, Birkicht E (1981) Thin prism refraction: a new direct method of polariton spectroscopy. Solid State Commun 39:1209
Buckley S, Rivoire K, Vuckovic J (2012) Engineered quantum dot single-photon sources. Rep Prog Phys 75:26503
Bylander J, Robert-Philip I, Abram I (2003) Interference and correlation of two independent photons. Eur Phys J D 22:295
Byrnes T, Kim NY, Yamamoto Y (2014) Exciton–polariton condensates. Nat Phys 10:803
Cavenett BC (1980) Optical detection of exciton resonances in semiconductors. J Phys Soc Jpn 49(Suppl A):611
Cavenett BC (1984) Triplet exciton recombination in amorphous and crystalline semiconductors. J Lumin 31/32:369
Chen Z et al. (2021) Google Quantum AI Exponential suppression of bit or phase errors with cyclic error correction. Nature 595:383
Cho K (1979) Internal structure of excitons. In: Cho K (ed) Excitons. Springer, Berlin, p 15
Christen J, Bimberg D (1990) 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
Collins RT, Vina L, Wang WI, Mailhiot C, Smith DL (1987) Electronic properties of quantum wells in perturbing fields. Proc SPIE 0792:2
Combescot R (2019) Trion ground-state energy: Simple results. Phys Rev B 100:245201
Compaan A (1975) Surface damage effects on allowed and forbidden phonon Raman scattering in cuprous oxide. Solid State Commun 16:293
Davies JJ, Cox RT, Nicholls JE (1984) Optically detected magnetic resonance of the triplet state of copper-center – donor pairs in CdS. Phys Rev B 30:4516
Davydov VYu, Subashiev AV, Cheng TS, Foxon CT, Goncharuk IN, Smirnov AN, Zolotareva RV, Lundin WV (1997) Surface polariton Raman spectroscopy in cubic GaN epitaxial layers. Mater Sci Forum 264:1371
Dean PJ, Thomas DG (1966) Intrinsic absorption-edge spectrum of gallium phosphide. Phys Rev 150:690
Deng H, Haug H, Yamamoto Y (2010) Exciton-polariton Bose-Einstein condensation. Rev Mod Phys 82:1489
Denisov MM, Makarov VP (1973) Longitudinal and transverse excitons in semiconductors. Phys Status Solidi B 56:9
Denisov VN, Mavrin BN, Podobedov VB (1987) Hyper-Raman scattering by vibrational excitations in crystals, glasses and liquids. Phys Rep 151:1
Dick KA (2008) A review of nanowire growth promoted by alloys and non-alloying elements with emphasis on Au-assisted III-V nanowires. Prog Crystal Growth Character Mater 54:138
Ding F, Singh R, Plumhof JD, Zander T, Krápek V, Chen YH, Benyoucef M, Zwiller V, Dörr K, Bester G, Rastelli A, Schmidt OG (2010) Tuning the exciton binding energies in single self-assembled InGaAs/GaAs quantum dots by piezoelectric-induced biaxial stress. Phys Rev Lett 104:067405
Dingle R, Wiegmann W, Henry CH (1974) Quantum states of confined carriers in very thin $Al_xGa_{1-x}As$-GaAs-$Al_xGa_{1-x}As$ heterostructures. Phys Rev Lett 33:827
DiVincenzo DP (2000) The physical implementation of quantum computation. Fortschr Phys 48:771
Doherty MW, Manson NB, Delaney P, Jelezko F, Wrachtrup J, Hollenberg LCL (2013) The nitrogen-vacancy colour centre in diamond. Phys Rep 528:1

Ekardt W, Lösch K, Bimberg D (1979) Determination of the analytical and the nonanalytical part of the exchange interaction of InP and GaAs from polariton spectra in intermediate magnetic fields. Phys Rev B 20:3303

Elliott RJ (1961) Symmetry of excitons in Cu_2O. Phys Rev 124:340

Ellis DJP, Bennett AJ, Dangel C, Lee JP, Griffiths JP, Mitchell A, Paraiso T-K, Spencer P, Ritchie DA, Shields AJ (2018) Independent indistinguishable quantum light sources on a reconfigurable photonic integrated circuit. Appl Phys Lett 112:211104

Esser A, Zimmermann R, Runge E (2001) Theory of trion spectra in semiconductor nanostructures. Phys Status Solidi B 227:317

Farrow T, See P, Bennett AJ, Ward MB, Atkinson P, Cooper K, Ellis DJP, Unitt DC, Ritchie DA, Shields AJ (2008) Single-photon emitting diode based on a quantum dot in a micro-pillar. Nanotechnology 19:345401

Filinov AV, Riva C, Peeters FM, Lozovik YuE, Bonitz M (2005) Influence of well-width fluctuations on the binding energy of excitons, charged excitons, and biexcitons in GaAs-based quantum wells. Phys Rev B 70:035323

Fischer B, Lagois J (1979) Surface exciton polaritons. In: Cho K (ed) Excitons. Springer, Berlin, p 183

Flohrer J, Jahne E, Porsch M (1979) Energy levels of A and B excitons in wurtzite-type semiconductors with account of electron-hole exchange interaction effects. Phys Status Solidi B 91:467

Frenkel JI (1931) On the transformation of light into heat in solids II. Phys Rev 37:1276

Fröhlich H (1954) Electrons in lattice fields. Adv Phys 3:325

Fröhlich D (1981) Aspects of nonlinear spectroscopy. In: Treusch J (ed) Festkörperprobleme. Advances in solid state physics, vol 21. Vieweg, Braunschweig, p 363

García-Cristóbal A, Cantarero A, Trallero-Giner C, Cardona M (1998) Resonant hyper-Raman scattering in semiconductors. Phys Rev B 58:10443

García-Pérez G et al. (2020) IBM Q Experience as a versatile experimental testbed for simulating open quantum systems. npj Quantum Inf 6:1

George GA, Morris GC (1970) The absorption, fluorescence and phosphorescence of single crystals of 1,2,4,5-tetrachlorobenzene and 1,4-dichlorobenzene at low temperatures. Mol Cryst Liq Cryst 11:61

Gerlach B (1974) Bound states in electron-exciton collisions. Phys Status Solidi B 63:459

Germanis S, Atkinson P, Hostein R, Gourdon C, Voliotis V, Lemaître A, Bernard M, Margaillan F, Majrab S, Eble B (2018) Dark-bright exciton coupling in asymmetric quantum dots, Phys Rev B 98:155303

Giblin J, Vietmeyer F, McDonald MP, Kuno M (2011) Single nanowire extinction spectroscopy. Nano Lett 11:3307

Girlanda R, Savasta S, Quattropani A (1994) Theory of exciton-polaritons in semiconductors with nearly degenerate exciton levels. Solid State Commun 90:267

Gisin N, Ribordy G, Tittel W, Zbinden H (2002) Quantum cryptography. Rev Mod Phys 74:145

Gislason HP, Monemar B, Dean PJ, Herbert DC, Depinna S, Cavenett BC, Killoran N (1982) Photoluminescence studies of the 1.911-eV Cu-related complex in GaP. Phys Rev B 26:827

Giustina M, Versteegh MAM, Wengerowsky S, Handsteiner J, Hochrainer A, Phelan K, Steinlechner F, Kofler J, Larsson J-Å, Abellán C, Amaya W, Pruneri V, Mitchell MW, Beyer J, Gerrits T, Lita AE, Shalm LK, Nam SW, Scheidl T, Ursin R, Wittmann B, Zeilinger A (2015) Significant-loophole-free test of Bell's theorem with entangled photons. Phys Rev Lett 115:250401

Gourley PL, Wolfe JP (1978) Spatial condensation of strain-confined excitons and excitonic molecules into an electron-hole liquid in silicon. Phys Rev Lett 40:526. And: Properties of the electron-hole liquid in Si: zero stress to the high-stress limit. Phys Rev B 24:5970 (1981)

Grosmann M (1963) The effect of perturbations on the excitonic spectrum of cuprous oxide. In: Kuper CG, Whitfield GD (eds) Polarons and excitons. Oliver and Boyd, London, p 373

Grundmann M, Bimberg D (1988) Anisotropy effects on excitonic properties in realistic quantum wells. Phys Rev B 38:13486
Haken H (1963) Theory of excitons II. In: Kuper CG, Whitfield GD (eds) Polarons and excitons. Oliver and Boyd, Edinburgh, p 295
Haken H (1976) Quantum field theory of solids. North Holland Publishing, Amsterdam
Haken H, Nikitine S (eds) (1975) Excitons at high densities. Springer tracts in modern physics. Springer, New York
Hanamura E (1976) Excitonic molecules. In: Seraphin BO (ed) Optical properties of solids. North Holland Publishing, Amsterdam, pp 81–142
Hanamura E, Haug H (1977) Condensation effects of excitons. Phys Rep 33:209
Hanbury Brown R, Twiss RQ (1956) Correlation between photons in two coherent beams of light. Nature 177:27
Heindel T, Thoma A, von Helversen M, Schmidt M, Schlehahn A, Gschrey M, Schnauber P, Schulze J -H, Strittmatter A, Beyer J, Rodt S, Carmele A, Knorr A, Reitzenstein S (2017) A bright triggered twin-photon source in the solid state. Nat Commun 8:14870
Holmes MJ, Choi K, Kako S, Arita M, Arakawa Y (2014) Room-temperature triggered single photon emission from a III-nitride site-controlled nanowire quantum dot. Nano Lett 14:982
Holtkemper M, Quinteiro GF, Reiter DE, Kuhn T (2021) Dark exciton preparation in a quantum dot by a longitudinal light field tuned to higher exciton states. Phys Rev Research 3:013024
Hönerlage B, Lévy R, Grun JB, Klingshirn C, Bohnert K (1985) The dispersion of excitons, polaritons and biexcitons in direct-gap semiconductors. Phys Rep 124:161
Hong CK, Ou ZY, Mandel L (1987) Measurement of subpicosecond time intervals between two photons by interference. Phys Rev Lett 59:2044
Hopfield JJ, Thomas DG (1963) Theoretical and experimental effects of spatial dispersion on the optical properties of crystals. Phys Rev 132:563
Hudson AJ, Stevenson RM, Bennett AJ, Young RJ, Nicoll CA, Atkinson P, Cooper K, Ritchie DA, Shields AJ (2007) Coherence of an entangled exciton–photon state. Phys Rev Lett 99:266802
Ishii R, Funato M, Kawakami Y (2020) Long-range electron-hole exchange interaction in aluminum nitride. Phys Rev B 102:155202
Kabler MN (1964) Low-temperature recombination luminescence in alkali halide crystals. Phys Rev 136:A1296
Kalra R, Laucht A, Hill CD, Morello A (2014) Robust two-qubit gates for donors in silicon controlled by hyperfine interactions. Phys Rev X 4:021044
Kamtekar KT, Monkman AP, Bryce MR (2010) Recent advances in white organic light-emitting materials and devices (WOLEDs). Adv Mater 22:572
Kane BE (1998) A silicon-based nuclear spin quantum computer. Nature 393:133
Kato Y, Yu CI, Goto T (1970) The effect of exchange interaction on the exciton bands in CuCl-CuBr solid solutions. J Phys Soc Jpn 28:104
Kazimierczuk T, Fröhlich D, Scheel S, Stolz H, Bayer M (2014) Giant Rydberg excitons in the copper oxide Cu_2O. Nature 514:343
Kittel C (1963) Quantum theory of solids. Wiley, New York, p 131
Kittel C (1966) Introduction to solid state physics. Wiley, New York
Knill E, Laflamme R, Milburn GJ (2001) A scheme for efficient quantum computation with linear optics. Nature 409:46
Knox RS (1984) Introduction to exciton physics. In: DiBartolo B, Danko J (eds) Collective excitations in solids. Plenum Press, New York, p 183
Köhler A, Bässler H (2015) Electronic processes in organic semiconductors. Wiley-VCH, Weinheim
Koteles ES, Jagannath C, Lee J, Vassell MO (1987) Uniaxial stress as a probe of valence subband mixing in semiconductor quantum wells. Proc SPIE 0792:168
Kudlek G, Presser N, Pohl UW, Gutowski J, Lilja J, Kuusisto E, Imai K, Pessa M, Hingerl K, Sitter A (1992) Exciton complexes in ZnSe layers: a tool for probing the strain distribution. J Cryst Growth 117:309

Kulakovskii VD, Lysenko VG, Timofeev VB (1985) Excitonic molecules in semiconductors. Sov Phys Usp 28:735
Kurizki G, Bertet P, Kubo Y, Mølmer K, Petrosyan D, Rabl P, Schmiedmayer J (2015) Quantum technologies with hybrid systems. Proc Natl Acad Sci 112:3866
Kurtsiefer C, Mayer S, Zarda P, Weinfurter H (2000) Stable solid-state source of single photons. Phys Rev Lett 85:290
Lambrecht WRL, Rodina AV, Limpijumnong S, Segall B, Meyer BK (2002) Valence-band ordering and magneto-optic exciton fine structure in ZnO. Phys Rev B 65:075207
Lampert MA (1958) Mobile and immobile effective-mass-particle complexes in nonmetallic solids. Phys Rev Lett 1:450
Landau LD (1933) Electron motion in crystal lattices. Phys Z Sowjetunion 3:664
Li S, Luo J, Liu J, Tang J (2019) Self-trapped excitons in all-inorganic halide perovskites: fundamentals, status, and potential applications. J Phys Chem Lett 10:1999
Liu J, Su R, Wei Y, Yao-Silva B, Yu Y, Iles-Smith J, Srinivasan K, Rastelli A, Li J, Wang X (2019) A solid-state source of strongly entangled photon pairs with high brightness and indistinguishability. Nat Nanotechnol 14:586
Lo HK, Curty M, Tamaki K (2014) Secure quantum key distribution. Nat Photon 8:595
Loudon R (1963) Theory of first-order Raman effect in crystals. Proc R Soc Lond A275:218
Lounis B, Orrit M (2005) Single-photon sources. Rep Prog Phys 68:1129
MacFarlane GG, McLean TP, Quarrington JE, Roberts V (1957) Fine structure in the absorption-edge spectrum of Ge. Phys Rev 108:1377
Mahan GD, Hopfield JJ (1964) Piezoelectric polaron effects in CdS. Phys Rev Lett 12:241
McCallum JC, Johnson BC, Botzem T (2021) Donor-based qubits for quantum computing in silicon. Appl Phys Rev 8:031314
McLean TP (1963) Excitons in germanium. In: Kuper CG, Whitfield GD (eds) Polarons and excitons. Oliver and Boyd, London, p 367
Mikhnenko OV, Blom PWM, Nguyen T-Q (2015) Exciton diffusion in organic semiconductors. Energy Environ Sci 8:1867
Miller RC, Kleinman DA (1985) Excitons in GaAs quantum wells. J Lumin 30:520
Miller DAB, Chemla DS, Damen TC, Gossard AC, Wiegmann W, Wood TH, Burrus CA (1985) Electric field dependence of optical absorption near the band gap of quantum-well structures. Phys Rev B 32:1043
Moskalenko SA (1958) The theory of Mott exciton in alkali-halide crystals. Zh Opt Spektrosk (USSR) 5:147
Mott NF (1938) Conduction in polar crystals. II. The conduction band and ultra-violet absorption of alkali-halide crystals. Trans Faraday Soc 34:500
Muhonen JT, Dehollain JP, Laucht A, Hudson FE, Sekiguchi T, Itoh KM, Jamieson DN, McCallum JC, Dzurak AS, Morello A (2014) Storing quantum information for 30 s in a nanoelectronic device. Nat Nanotechnol 9:986
Muljarov EA, Zhukov EA, Dneprovskii VS, Masumoto Y (2000) Dielectrically enhanced excitons in semiconductor-insulator quantum wires: theory and experiment. Phys Rev B 62:7420
Müller AS, Avlasevich YS, Müllen K, Bardeen CJ (2006) Evidence for exciton fission and fusion in a covalently linked tetracene dimer. Chem Phys Lett 421:518
Müller M, Bounouar S, Jöns KD, Glässl M, Michler P (2014) On-demand generation of indistinguishable polarization-entangled photon pairs. Nat Photon 8:224
Najafov H, Lee B, Zhou Q, Feldman LC, Podzorov V (2010) Observation of long-range exciton diffusion in highly ordered organic semiconductors. Nat Mater 9:938
Nielsen MA (2004) Optical quantum computation using cluster states. Phys Rev Lett 93:040503
Ogawa T, Takagahara T (1991) Optical absorption and Sommerfeld factors of one-dimensional semiconductors: an exact treatment of excitonic effects. Phys Rev B 44:8138
Ollivier H, Maillette de Buy Wenniger I, Thomas S, Wein SC, Harouri A, Coppola G, Hilaire P, Millet C, Lemaître A, Sagnes I, Krebs O, Lanco L, Loredo JC, Antón C, Somaschi N, Senellart P (2020) ACS Photonics 7:1050

Orieux A, Versteegh MAM, Jöns KD, Ducci S (2017) Semiconductor devices for entangled photon pair generation: a review. Rep Prog Phys 80:076001
Paik H, Schuster DI, Bishop LS, Kirchmair G, Catelani G, Sears AP, Johnson BR, Reagor MJ, Frunzio L, Glazman LI, Girvin SM, Devoret MH, Schoelkopf RJ (2011) Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. Phys Rev Lett 107:240501
Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love PJ, Aspuru-Guzik A, O'Brien JL (2014) A variational eigenvalue solver on a photonic quantum processor. Nat Commun 5:4213
Phillips RT, Lovering DJ, Denton GJ, Smith GW (1992) Biexciton creation and recombination in a GaAs quantum well. Phys Rev B 45:4308
Ploog K, Döhler GH (1983) Compositional and doping superlattices in III-V semiconductors. Adv Phys 32:285
Poem E, Kodriano Y, Tradonsky C, Lindner NH, Gerardot BD, Petroff PM, Gershoni D (2010) Accessing the dark exciton with light. Nature Physics 6:993
Pohl UW (2008) InAs/GaAs quantum dots with multimodal size distribution. In: Wang ZM (ed) Self-assembled quantum dots. Springer, New York, p 43
Pohl UW (2020) Epitaxy of Semiconductors – Physics and fabrication of heterostructures, 2nd edn. Springer, Cham
Pohl UW, Strittmatter A, Schliwa A, Lehmann M. Niermann T, Heindel T, Reitzenstein S, Kantner M, Bandelow U, Koprucki T, Wünsche H-J (2020) Stressor-induced site control of quantum dots for single-photon sources. In Kneissl M, Knorr A, Reitzenstein S, Hoffmann A (eds) Semiconductor nanophotonics. Springer Nature, Cham, pp 52–90
Pope M, Swenberg CE (1982) Electronic processes in organic crystals. Oxford University Press, Oxford, UK
Raussendorf R, Briegel HJ (2001) A one-way quantum computer. Phys Rev Lett 86:5188
Reynolds DC, Collins TC (1981) Excitons: their properties and uses. Academic Press, New York
Robert A, Barkoutsos PK, Woerner S, Tavernelli I (2021) Resource-efficient quantum algorithm for protein folding. npj Quantum Inf 7:38
Rodina AV, Dietrich M, Göldner A, Eckey L, Hoffmann A, Efros AL, Rosen M, Meyer BK (2001) Free excitons in wurtzite GaN. Phys Rev B 64:115204
Rodt S, Schliwa A, Pötschke K, Guffarth F, Bimberg D (2005) Correlation of structural and few-particle properties of self-organized InAs/GaAs quantum dots. Phys Rev B 71:155325
Rodt S, Reitzenstein S, Heindel T (2020) Deterministically fabricated solid-state quantum-light sources. J Phys Condens Matter 32:153003
Rössler U (1979) Fine structure, lineshape, and dispersion of Wannier excitons. In: Treusch J (ed) Festkörperprobleme. Advances in solid state physics, vol 19. Vieweg, Braunschweig, p 77
Rudin S, Reinecke TL, Segall B (1990) Temperature-dependent exciton linewidths in semiconductors. Phys Rev B 42:11218
Sasaki M, Fujiwara M, Ishizuka H, Klaus W, Wakui K, Takeoka M, Miki S, Yamashita T, Wang Z, Tanaka A, Yoshino K, Nambu Y, Takahashi S, Tajima A, Tomita A, Domeki T, Hasegawa T, Sakai Y, Kobayashi H, Asai T, Shimizu K, Tokura T, Tsurumaru T, Matsui M, Honjo T, Tamaki K, Takesue H, Tokura Y, Dynes JF, Dixon AR, Sharpe AW, Yuan ZL, Shields AJ, Uchikoga S, Legré M, Robyr S, Trinkler P, Monat L, Page JB, Ribordy G, Poppe A, Allacher A, Maurhart O, Länger T, Peev M, Zeilinger A (2011) Field test of quantum key distribution in the Tokyo QKD Network. Opt Express 19:10387
Schliwa A, Winkelnkemper M, Bimberg D (2007) Impact of size, shape, and composition on piezoelectric effects and electronic properties of In(Ga)As/GaAs quantum dots. Phys Rev B 76:205324
Schwartz I, Schmidgall ER, Gantz L, Cogan D, Bordo E, Don Y, Zieliński M, Gershoni D (2015) Deterministic writing and control of the dark exciton spin using single short optical pulses. Phys Rev X 5:011009

Schwartz I, Cogan D, Schmidgall ER, Gantz L, Don Y, Zieliński M, Gershoni D (2015b) Deterministic coherent writing of a long-lived semiconductor spin qubit using one ultrafast optical pulse. Phys Rev B 92:201201
Seguin R, Schliwa A, Rodt S, Pötschke K, Pohl UW, Bimberg D (2005) Size-dependent exciton fine-structure splitting in self-organized InAs/GaAs quantum dots. Phys Rev Lett 95:257402
Shields AJ, Pepper M, Ritchie DA, Simmons MY (1995a) Influence of excess electrons and magnetic fields on Mott-Wannier excitons in GaAs quantum wells. Adv Phys 44:47
Shields AJ, Osborne JL, Simmons MY, Pepper M, Ritchie DA (1995b) Magneto-optical spectroscopy of positively charged excitons in GaAs quantum wells. Phys Rev B 52:R5523
Shinada M, Sugano S (1966) Interband optical transitions in extremely anisotropic semiconductors. I. Bound and unbound exciton absorption. J Phys Soc Jpn 21:1936
Shinar J (ed) (2004) Organic light-emitting devices: a survey. Springer, New York
Singh J (1984) The dynamics of excitons. In: Ehrenreich H, Turnbull D (eds) Solid state physics, vol 38. Academic Press, Orlando/New York, p 295
Singh R, Bester G (2010) Lower bound for the excitonic fine structure splitting in self-assembled quantum dots. Phys Rev Lett 104:196803
Singh J, Birkedal D, Lyssenko VG, Hvam JM (1996) Binding energy of two-dimensional biexcitons. Phys Rev B 53:15909
Slussarenko S, Pryde GJ (2019) Photonic quantum information processing: A concise review. Appl Phys Rev 6:041303
Solovyev VV, Kukushkin IV (2009) Measurement of binding energy of negatively charged excitons in GaAs/$Al_{0.3}Ga_{0.7}As$ quantum wells. Phys Rev B 79:233306
Someya T, Akiyama H, Sakaki H (1996) Enhanced binding energy of one-dimensional excitons in quantum wires. Phys Rev Lett 76:2965
Song KS, Williams RT (1993) Self-trapped excitons. Springer series in solid-state sciences, vol 105. Springer, Berlin
Stébé B, Ainane A (1989) Ground state energy and optical absorption of excitonic trions in two dimensional semiconductors. Superlattice Microstruct 5:545
Stievater TH, Li X, Steel DG, Gammon D, Katzer DS, Park D, Piermarocchi C, Sham LJ (2001) Rabi oscillations of excitons in single quantum dots. Phys Rev Lett 87:133603
Tamor MA, Wolfe JP (1980) Drift and diffusion of free excitons in Si. Phys Rev Lett 44:1703
Thewalt MLW, Rostworowski JA (1978) Biexcitons in Si. Solid State Commun 25:991
Thoma A, Schnauber P, Böhm J, Gschrey M, Schulze JH, Strittmatter A, Rodt S, Heindel T, Reitzenstein S (2017) Two-photon interference from remote deterministic quantum dot microlenses. Appl Phys Lett 110: 011104
Thomas GA, Rice TM (1977) Trions, molecules and excitons above the Mott density in Ge. Solid State Commun 23:359
Thomas GA, Timofeev VB (1980) A review of $N = 1$ to ∞ particle complexes in semiconductors. In: Moss TS, Balkanski M (eds) Handbook on semiconductors, vol 2. North Holland Publishing, Amsterdam, p 45
Tomiki T (1969) Optical constants and exciton states in KCl single crystals III. The spectra of conductivity and of energy loss. J Phys Soc Jpn 26:738
Toyozawa Y (1980) Electrons, holes, and excitons in deformable lattice. In: Kubo R, Hanamura I (eds) Excitons. Springer, Berlin
Türck V, Rodt S, Stier O, Heitz R, Engelhardt R, Pohl UW, Bimberg D, Steingrüber R (2000) Effect of random field fluctuations on excitonic transitions of individual CdSe quantum dots. Phys Rev B 61:9944
Ueta M, Nishina Y (eds) (1976) Physics of highly excited states in solids. Lecture notes in physics, vol 57. Springer, New York
Uihlein C, Fröhlich D, Kenklies R (1981) Investigation of exciton fine structure in Cu_2O. Phys Rev B 23:2731
Unrau W, Quandt D, Schulze J-H, Heindel T, Germann TD, Hitzemann O, Strittmatter A, Reitzenstein S, Pohl UW, Bimberg D (2012) Electrically driven single photon source based

on a site-controlled quantum dot with self-aligned current injection. Appl Phys Lett 101:211119
Vogl P (1976) Microscopic theory of electron-phonon interaction in insulators or semiconductors. Phys Rev B 13:694
Vouilloz F, Oberli DY, Dupertuis M-A, Gustafsson A, Reinhardt F, Kapon E (1997) Polarization anisotropy and valence band mixing in semiconductor quantum wires. Phys Rev Lett 78:1580
Vouilloz F, Oberli DY, Dupertuis M-A, Gustafsson A, Reinhardt F, Kapon E (1998) Effect of lateral confinement on valence-band mixing and polarization anisotropy in quantum wires. Phys Rev B 57:12378
Wagner MR, Callsen G, Reparaz JS, Kirste R, Hoffmann A, Rodina AV, Schleife A, Bechstedt F, Phillips MR (2013) Effects of strain on the valence band structure and exciton-polariton energies in ZnO. Phys Rev B 88:235210
Wang X-L, Voliotis V (2006) Epitaxial growth and optical properties of semiconductor quantum wires. J Appl Phys 99:121301
Wang Y, Dolde F, Biamonte J, Babbush R, Bergholm V, Yang S, Jakobi I, Neumann P, Aspuru-Guzik A, Whitfield JD, Wrachtrup J (2015) Quantum simulation of helium hydride cation in a solid-state spin register. ACS Nano 9:7769
Wang X-L, Chen L-K, Li W, Huang H-L, Liu C, Chen C, Luo Y-H, Su Z-E, Wu D, Li Z-D, Lu H, Hu Y, Jiang X, Peng C-Z, Li L, Liu N-L, Chen Y-A, Lu C-Y, Pan J-W (2016) Experimental ten-photon entanglement. Phys Rev Lett 117:210502
Wannier GH (1937) The structure of electronic excitation levels in insulating crystals. Phys Rev 52:191
Washington MA, Genack AZ, Cummins HZ, Bruce RH, Compaan A, Forman RA (1977) Spectroscopy of excited yellow exciton states in Cu_2O by forbidden resonant Raman scattering. Phys Rev B 15:2145
Weisbuch C, Benisty H, Houdré R (2000) Overview of fundamentals and applications of electrons, excitons and photons in confined structures. J Lumin 85:271
Wendin G (2017) Quantum information processing with superconducting circuits: a review. Rep Prog Phys 80:106001
Wicksted J, Matsushita M, Cummins HZ, Shigenari T, Lu XZ (1984) Resonant Brillouin scattering in CdS. I. Experiment. Phys Rev B 29:3350
Wootters WK, Zurek WH (1982) A single quantum cannot be cloned. Nature 299:802
Yamada Y, Sakashita T, Watanabe H, Kugimiya H, Nakamura S, Taguchi T (2000) Optical properties of biexcitons in ZnS. Phys Rev B 61:8363
Yu PY (1979) Study of excitons and exciton-phonon interactions by resonant Raman and Brillouin spectroscopies. In: Cho K (ed) Excitons. Springer, Berlin, p 211
Yu PW, Reynolds DC, Bajaj KK, Litton CW, Klem J, Huang D, Morkoc H (1987) Observation of monolayer fluctuations in the excited states of GaAs-$Al_xGa_{1-x}As$ multiple-quantum-well. Solid State Commun 62:41
Zhang J, Wildmann JS, Ding F, Trotta R, Huo Y, Zallo E, Huber D, Rastelli A, Schmidt OG (2015) High yield and ultrafast sources of electrically triggered entangled-photon pairs based on strain-tunable quantum dots. Nature Commun 6:10067
Zieliński M (2021) Dark-bright excitons mixing in alloyed InGaAs self-assembled quantum dots. Phys Rev B 103:155418
Zieliński M, Don Y, Gershoni D (2015) Atomistic theory of dark excitons in self-assembled quantum dots of reduced symmetry. Phys Rev B 91:085403
Zieliński M, Gołasa K, Molas MR, Goryca M, Kazimierczuk T, Smoleński T, Golnik A, Kossacki P, Nicolet AAL, Potemski M, Wasilewski ZR, Babiński A (2015) Excitonic complexes in natural InAs/GaAs quantum dots. Phys Rev B 91:085303

Part V

Defects

Crystal Defects

15

Karl W. Böer and Udo W. Pohl

Contents

1 Classification of Defects 596
1.1 Defect Types 596
1.2 Defect Notation and Charged Point Defects 598
2 Point Defects 602
2.1 Density of Intrinsic Point Defects 602
2.2 Frozen-in Intrinsic Defect Density 607
2.3 Defect-Formation Energy 609
2.4 Defect Chemistry 611
2.5 Changing of Stoichiometry and Compensation 615
2.6 Brouwer Approximation 617
3 Diffusion of Lattice Defects 620
3.1 Fick's Laws of Diffusion 620
3.2 Types of Diffusion 624
4 Line Defects 627
4.1 Edge and Screw Dislocations, Burgers Vector 628
4.2 Dislocations in Compound Semiconductors 632
4.3 Motion and Creation of Dislocations 635
5 Planar Defects 639
5.1 Stacking Faults and Antiphase Domains 639
5.2 Grain Boundaries 641
6 Summary 643
References 644

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_15

Abstract

Semiconducting properties of most interest are predominantly caused by crystal defects. They are classified into point, line, and planar defects. Some defects are beneficial, such as donors, acceptors, or luminescence centers. These defects determine the desired electronic and optical properties of the semiconductor. Other defects promote nonradiative carrier recombination, carrier trapping, or excessive carrier scattering and are detrimental to device performance.

Native *(intrinsic) point defects* and associates of these defects are formed at elevated temperature and may be frozen-in with decreasing temperature. Their creation is interrelated – among each other and also to the presence of *extrinsic* (impurity) defects – and governed by the conservation of particles and quasi-neutrality. The mobility of defects is provided by various diffusion mechanisms and affected by their charge. *Line defects* involve rows of atoms. Most important are edge and screw dislocations, which affect crystal growth and accommodate strain in semiconductors. Dislocations are characterized by their Burgers vector and its angle to the dislocation line, and their mobility is provided by glide and climb processes. *Planar defects* comprise stacking faults, grain and twin boundaries, inversion-domain boundaries, and interfaces between different semiconductors or between a semiconductor and a metal.

Keywords

Acceptor · Antisite defect · Antiphase domain · Brouwer approximation · Burgers vector · Crystal defects · Compensation · Defect-formation energy · Diffusion mechanisms · Donor · Doping · Edge dislocations · Fick diffusion · Frenkel defects · Frozen-in defects · Grain boundary · Intrinsic defect · Interstitial · Inversion-domain boundary · Jog · Kink · Line defects · Native defect · Partial dislocations · Point defects · Twin boundary · Screw dislocation · Stacking fault · Polytype · Quasi-neutrality · Schottky disorder · Vacancy

1 Classification of Defects

A crystal defect is any region where the microscopic arrangement of atoms differs from that of a perfect crystal. A large variety of crystal defects determine the electronic and optical properties of semiconductors. Some of them provide donors or acceptors. Others are responsible for carrier scattering or recombination; still others trap carriers and influence the space charge that determines the carrier transport.

1.1 Defect Types

Defects are classified into zero-, one-, or two-dimensional types, depending on whether the imperfect region is bounded in three, two, or one dimension, respectively. We distinguish:

- Point defects, such as single impurity atoms or vacancies in an otherwise ideal lattice
- Line defects (dislocations) involving rows of atoms
- Planar defects relating to surfaces or internal crystal interfaces

In addition, volume defects, relating to usually small, three-dimensional inclusions (precipitates) or defect associates, may be distinguished. Some types of defects are illustrated in Fig. 1.

Point Defects Point defects, considered in Sect. 2, are the main class of defects that act as donors or acceptors or, when their energy levels are farther separated from the band edges, as traps or recombination centers. They are also important as scattering centers, especially when charged with respect to atoms of the host lattice.

We distinguish intrinsic and extrinsic defects. *Intrinsic* or *native* defects are lattice defects, which do not involve foreign atoms. These are *vacancies*, i.e., missing atoms on a lattice site, and *interstitials*, i.e., additional host atoms within the lattice (Fig. 1b, upper part). In compound semiconductors, also *antisite defects* occur, e.g., in an *AB* compound where an *A* atom occupies a *B* site. *Extrinsic* defects are impurities located at a lattice site (substitutional impurity; Fig. 1b, lower part) or at an interstitial site.

Small defect associates are also counted as point defects such as defect pairs[1] and higher defect associates. A large variety of defects can be distinguished according to their geometrical arrangement with respect to each other, forming an anisotropic center or forming distant pairs, distinguished by their distance within the lattice while still interacting. Finally, all defects containing an impurity are distinguished by their chemical identity, which relates to the defect center and to the host lattice.

Line Defects Line defects, discussed in Sect. 4, extend through the crystal along a straight or bended line. Most important are *dislocations*, such as the edge

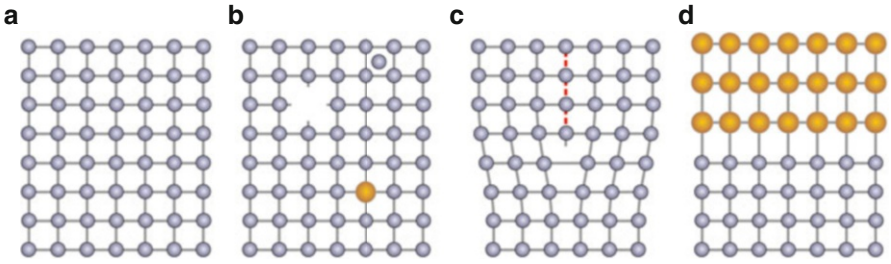


Fig. 1 (**a**) Two-dimensional representation of an ideal cubic lattice. (**b**) Point defects: a vacancy, an interstitial, and an impurity atom. (**c**) Line defect: cross section perpendicular to a dislocation line located at the end of the inserted lattice plane marked in *red*. (**d**) Planar defect: interface at the heterojunction of two dissimilar solids

[1]These comprise divacancies, an impurity associated with an intrinsic defect, and two impurities associated with each other.

dislocation depicted in the cross section in Fig. 1c. The extra lattice plane inserted into the lattice creates at its edge a row of atoms with a dangling bond, yielding pronounced electronic properties. Line defects are inevitably introduced in strained heteroepitaxial structures with a layer thickness exceeding a critical value; for more information, see Sect. 1.2 of ▶ Chap. 16, "Crystal Interfaces". Since line defects are detrimental to electronic devices, their density should generally be low (with a mutual spacing below the diffusion length of carriers).

Planar Defects Planar defects comprise intrinsic faults treated in Sect. 5 and interfaces discussed in ▶ Chap. 16, "Crystal Interfaces". Intrinsic planar faults are *stacking faults*, low- and high-angle *grain boundaries*, *twin boundaries*, and crystal *surfaces*. Interfaces are extrinsic planar faults and comprise heterojunctions to a dissimilar solid (Fig. 1d) or to a metal.

Planar defects have a significant influence on the electronic properties of semiconductors. Most of this influence is caused by defect levels at the boundaries; they are usually charged and attract compensating charges in the adjacent crystal volume (*space charges*). The control of space-charge regions is a major topic of electronic devices.

This list of defects given above is by no means complete but gives an impression of the great abundance of defects present in real crystals. Each of these defects contributes to the wealth of electronic eigenstates, most of them as levels in the bandgap. Therefore, it is not surprising that many of these defects are not unambiguously identified. For a survey on point defects, see Lannoo and Bourgoin (1981); for point defects in silicon see Pichler (2004). Experimental methods of defect identification are described in Bourgoin and Lannoo (1983). Line defects, particularly dislocations, are of basic importance for epitaxy; for more information, see Ayers (2007) and Pohl (2020); interfaces are discussed in more detail in ▶ Chap. 16, "Crystal Interfaces".

1.2 Defect Notation and Charged Point Defects

Both, intrinsic (native) and extrinsic (impurity) defects may change their charge state. This feature is of fundamental importance for the electronic properties of semiconductors. We consider some point defects to illustrate the basic mechanism, which is discussed in more detail later in this chapter.

Extrinsic Donors and Acceptors Foreign atoms which substitute for an intrinsic lattice atom are most easily described when they are from an adjacent column of the periodic system of elements with respect to the replaced atom, e.g., a phosphorus or boron atom replacing a silicon atom in a silicon crystal. The P atom is pentavalent and thus has one electron more in its outer shell than the host lattice atoms (Fig. 2a). When the P atom replaces one of the atoms of a Si lattice, this extra electron is only loosely bound and can easily be separated from the P ion core. Therefore, the P acts as a *donor* in the Si host crystal. In a similar fashion, the incorporation of a trivalent B atom on a Si site causes the deficiency of an electron (Fig. 2b). This missing electron can be regarded as a hole bound to the B atom. The hole can be replaced by an adjacent valence electron and thereby move through the crystal; the B atom acts as an *acceptor*.

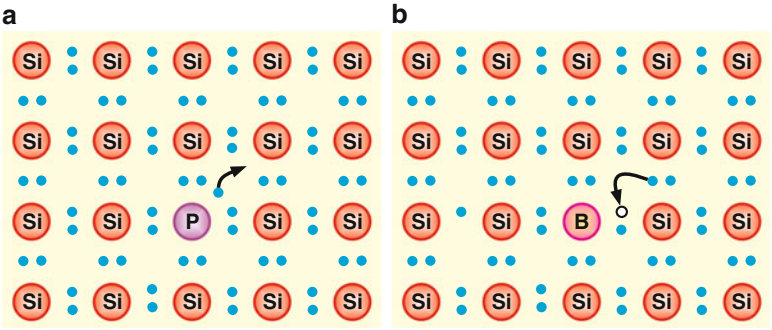


Fig. 2 Two-dimensional representation of a covalently bound silicon crystal with ion cores (*red*) and valence electrons (*blue*). (**a**) Substitutional phosphorus P as donor with extra electron (*blue dot with arrow*). (**b**) Substitutional boron B as acceptor with missing electron (*circle*) to complete bonding at the B atom; the *arrow* indicates the ionization

The extra electron of the P donor is much less tightly bound than a valence electron. Consequently much less energy than the bandgap energy E_g is required to make it a quasi-free electron: it is easily emitted by thermal ionization into the conduction band. Therefore, the donor energy-level (of this P electron) lies close below the edge of the conduction band as shown in Fig. 3a. Similarly, in the hole picture, the B atom becomes ionized (Fig. 3b); it produces a free hole and acts as an *acceptor* (of a valence-band electron), with a level close to the valence band. The energies E_d and E_a are assigned to the unionized donor and acceptor. The energy of defect states is discussed in more detail in Sect. 1 of ▶ Chap. 18, "Shallow-Level Centers".

It is often appropriate to distinguish isovalent impurities (from the same column of the periodic system – also called isoelectronic), isocoric impurities[2] (from the same row of the periodic table, i.e., having the same core), substitutionals with $\Delta z = \pm 1$ (z is the chemical valency) such as P and B in the example considered above, substitutionals with $\Delta z = \pm 2$ or more, amphoteric defects,[3] in which the sign of Δz can change, and transition metals; the latter are identified as a separate group because of their outer shell screening, giving more individuality to the defect and mostly yielding levels deep in the bandgap (see Sect. 3 of ▶ Chap. 19, "Deep-Level Centers").

Defect Notation within the Host Lattice When the donor is ionized, it becomes positively charged. The change in charge character plays an important role in later discussions on carrier capture and scattering. Such charge relations can be followed

[2]An isocoric P in a Si lattice can be thought of as "created" by adding to a lattice atom a proton, i.e., a point charge, and an extra electron (the donor electron), thereby creating the most ideal hydrogen-like defect. Any other hydrogen-like donor, e.g., As or Sb in Si, is of different size, causing more lattice deformation and a substantially different core potential (see Sect. 1 of ▶ Chap. 18, "Shallow-Level Centers").

[3]That is, a defect that can act as a donor or acceptor depending on the chemical potential of the lattice (influenced, e.g., by optical excitation or other doping).

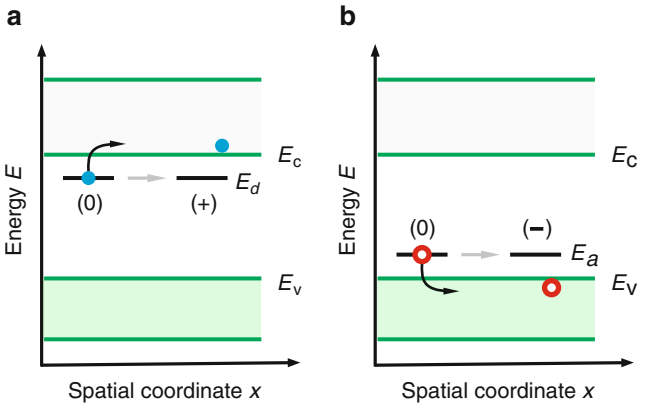


Fig. 3 Band model of a typical semiconductor with (**a**) a donor and (**b**) an acceptor, indicating the respective energy level and the relative charge character with and without charge carrier

easily in a chemical representation: the "reaction" between a donor (D) and an electron (e) can be written as

$$D^0 \longleftrightarrow D^+ + e^-, \tag{1}$$

where 0, +, and − represent neutral, positive, and negative charge characters with respect to the lattice. In literature, also the *Kröger-Vink notation*[4] is found: × for the neutral, • for the positive, and ′ for the negative net charge with respect to the perfect lattice (see Hayes and Stoneham 1985); it is less used in more recent literature. If we want to be more descriptive with respect to the chemical nature of the defect, we write

$$P_{Si}^0 \longleftrightarrow P_{Si}^+ + e^-, \tag{2}$$

which indicates the atomic defect (P) with its position within the lattice (i.e., substituting a Si atom) as the subscript. In general, point defects are identified by the symbol of the defect, with the lattice site on which the defect is located as a subscript. For example, a chlorine vacancy (V) in a NaCl lattice is identified as V_{Cl}, a potassium ion replacing a sodium ion in the same lattice as K_{Na}, and a copper interstitial as Cu_i.

In a similar fashion, the recharging of an acceptor can be described by

$$A^0 \longleftrightarrow A^- + h^+, \tag{3}$$

or

[4]The notation of charges with respect to the neutral lattice was introduced by Kröger, Vink, and Schottky (see Schottky and Stockmann 1954). This notation allows to distinguish from the charge identification used in an ionic lattice, e.g., Na^+Cl^-. Inclusion of a Cd^{++} instead of a Na^+ ion makes the cadmium ion singly positively charged with respect to the neutral lattice.

$$B_{Si}^{0} \longleftrightarrow B_{Si}^{-} + h^{+}, \tag{4}$$

where h^{+} represents a hole in the valence band.

Substitutionals in *AB* Compounds Substitutionals in an *AB*-host lattice act similarly to substitutionals in element semiconductors, except one distinguishes whether an anion or a cation is being replaced. For instance, replacement of a divalent Cd ion in a CdS crystal by a trivalent In ion, denoted by In_{Cd}, yields

$$In_{Cd}^{0} \longleftrightarrow In_{Cd}^{+} + e^{-}. \tag{5}$$

This results in a shallow donor, as does the replacement of a sulfur ion from group VI with a halogen ion from group VII. On the other hand, the replacement of Cd in CdS with an alkali metal ion like Li, or of S with a group V element like P or As, produces an acceptor. However, the incorporation of one type of defect (in CdS, the incorporation of donors) is often easier than for the oppositely charged defect (acceptors in CdS), rendering the material preferably *n*- or *p*-type (CdS is *n*-type). Intrinsic compensation (see Sect. 2.5) is one reason for this preference. An interesting case occurs for some trivalent compounds. For example, in GaAs, the replacement of Ga with the tetravalent Si results in

$$Si_{Ga}^{0} \longleftrightarrow Si_{Ga}^{+} + e^{-} \tag{6}$$

acting as a shallow donor, while the same Si, replacing a pentavalent As ion,

$$Si_{As}^{0} \longleftrightarrow Si_{As}^{-} + h^{+}, \tag{7}$$

results in a shallow acceptor. Depending on the growth condition, one or the other is preferred, and the material becomes either *n*- or *p*-type.

Vacancies and Interstitials Vacancies and interstitials are denoted in a similar fashion. Metal-ion interstitials usually act as donors. For example, in CdS,

$$Cd_{i}^{0} \longleftrightarrow Cd_{i}^{+} + e^{-}. \tag{8}$$

Metal-ion vacancies act as acceptors:

$$V_{Cd}^{0} \longleftrightarrow V_{Cd}^{-} + h^{+}. \tag{9}$$

Nonmetal-ion vacancies usually act as donors:

$$V_{S}^{0} \longleftrightarrow V_{S}^{+} + e^{-}. \tag{10}$$

A first and rather simplified judgment as to whether the defect more readily becomes positively or negatively charged is easily rendered in ionic compounds. Semiconductors with mixed bonding character can also be judged by following the 8-*N* rule (see Sect. 3.1 of ▶ Chap. 3, "The Structure of Semiconductors"). The rule maintains that elements tend to complete their outer shell by sharing electrons with neighbor atoms; surplus electrons are donated to the lattice, and missing electrons are attracted from elsewhere in the lattice. Interstitial metal ions tend to donate their valence electron(s), consequently reducing their radius, and cause less lattice deformation. For more basic information, see Seeger (1997) or Li (2007). Other, more complex aspects of point defects will be discussed in the following.

2 Point Defects

Point defects, such as vacancies and interstitials, are inevitably incorporated into a given crystal lattice by thermodynamic creation, while dopants are usually intentionally incorporated by diffusion from a source.[5] Both types of incorporation will be discussed in this section.

2.1 Density of Intrinsic Point Defects

With increased temperature, lattice vibrations (discussed in ▶ Chap. 4, "Elasticity and Phonons") become more vigorous, which makes the creation of *intrinsic lattice defects*, i.e., vacancies and interstitials, more probable. In addition, because of the anharmonicity of lattice vibrations, the lattice expands and thereby facilitates such a defect creation.[6] We briefly consider two such types of intrinsic disorder.

One type of thermodynamic defect is created by an atom moving from a lattice site into an interstitial position and leaving behind a vacancy. In thermodynamic equilibrium, the density of these defects is determined by the equality of thermal generation of these defect pairs and recombination, i.e., an interstitial finds a vacancy and "recombines". From size consideration, cation interstitials and vacancies are more probable. These are referred to as *Frenkel pairs* and the corresponding lattice disorder as *Frenkel disorder* (Frenkel 1926) (see Fig. 4a).

[5]In contrast, *purification* can be accomplished by diffusion of impurities into a sink. With the solubility of impurities being a function of the temperature, a temperature gradient can be used as a driving force for purification. A more effective means is the use of the boundary between the liquid and solid phase, using the fact that the solubilities in these two phases are substantially different (a measure of which is the *segregation coefficient*). Zone refining is a well-established technique to achieve such a purification (see de Kock 1980).

[6]Besides thermal creation, intrinsic point defects may result from recoils due to particle irradiation (e.g., in detectors or from space environment), leading particularly to vacancy-interstitial Frenkel pairs. Respective threshold displacement energies for silicon were reported by Holmström et al. (2008), for GaAs by Chen et al. (2017).

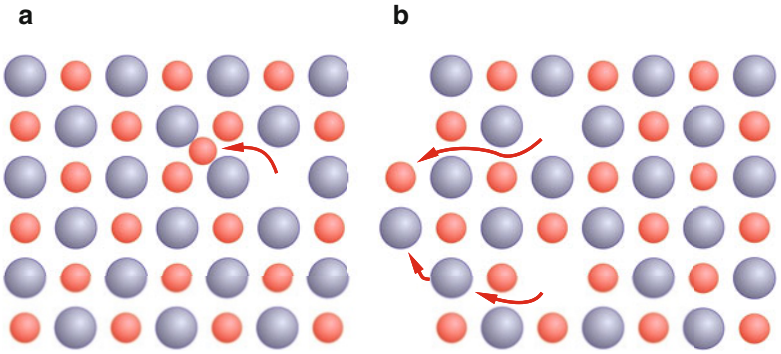


Fig. 4 Intrinsic types of disorder in an ionic compound: (**a**) Frenkel disorder comprising cation vacancies and an equal amount of cation interstitials. (**b**) Schottky disorder with cation and anion vacancies in equal densities distributed throughout the crystal bulk

Another type of disorder is generated at the surface of a crystal, where an atom from the bulk moves to the surface, creating a vacancy which in turn diffuses deeper into the crystal bulk. In an ionic AB compound, both types of (A and B) vacancies must be created in equal amounts to avoid preferential charging of the surface or the bulk after vacancy diffusion. In fact, such a charging will occur initially with every vacancy formed but will make it more difficult to form another vacancy of the same type adjacent to it. It will also make it easier to place a vacancy of the opposite type adjacent to it, which in essence balances their density. The resulting disorder, i.e., an equal amount of anion and cation vacancies distributed throughout a crystal, is referred to as *Schottky disorder* (Schottky 1935) (see Fig. 4b).

The density of Frenkel- or Schottky-type *intrinsic defects* can be obtained from thermodynamic considerations: the *Helmholtz free energy*,

$$F = U - TS, \tag{11}$$

due to these defects must be minimized. As a simple example, we will discuss Schottky defects in an elemental crystal, i.e., single vacancies. Assuming that the concentration of these vacancies is low enough so that they are created independently of each other, the energy of n of these vacancies is given by

$$U = nE_{\text{Schottky}}, \tag{12}$$

where E_{Schottky}, the *Schottky energy*, is the energy required to take a single atom from the crystal volume and put it at the crystal surface (for more details, see Sect. 2.3). The configurational entropy[7]

[7]For simplicity, we have neglected here the vibrational part of the entropy ($S = S_{\text{config}} + S_{\text{vib}}$). These contributions are considered later in this section. The vibrational part results in an increase of the intrinsic defect density.

$$S_{\text{config}} = k \ln W \tag{13}$$

is described by the total number of possibilities of selecting n indistinguishable atoms and moving them to the surface from a crystal containing N atoms:

$$W = \frac{N(N-1)\ldots(N-n+1)}{n!} = \frac{N!}{(N-1)!\ n!}. \tag{14}$$

Using the *Stirling approximation*, valid for large n,

$$\ln n! \cong n\ \ln n - n, \tag{15}$$

we obtain

$$S = k\ln W = k[N\ln N - N - \{(N-n)\ln(N-n) - (N-n)\} - (n\ln n - n)]. \tag{16}$$

When minimizing F at a given temperature, we obtain from Eqs. 11 to 16

$$\left(\frac{\partial F}{\partial n}\right)_T = E_{\text{Schottky}} - kT\ln\frac{N-n}{n} = 0, \tag{17}$$

and for $n \ll N$

$$n = N\exp\left(-\frac{E_{\text{Schottky}}}{kT}\right). \tag{18}$$

Since, for a crystal of unit volume, n is equal to the *density* of vacancies N_{V} and N is the *density* of lattice atoms N_{lattice}, we have for the density of Schottky defects in a monatomic crystal in thermal equilibrium

$$N_{\text{V}} = N_{\text{lattice}}\exp\left(-\frac{E_{\text{Schottky}}}{kT}\right). \tag{19}$$

In a *diatomic* lattice, for neutrality reasons, an equal amount of anion and cation vacancies must be formed.[8] Since the probability of forming ion pairs W_{pair} is the square of the probability of forming single vacancies ($W_{\text{pair}} = W^2$), it follows that

[8]This simple model of a pair-wise defect formation maintaining stoichiometry will be modified later (Sect. 2.5) to permit slight changes in the stoichiometry, thereby making the crystal *n*- or *p*-type. Inversely, the creation of such intrinsic defects can be enhanced or suppressed depending on the position of the Fermi level, i.e., depending on doping.

$$\left(\frac{\partial F}{\partial n}\right)_T = E_{\text{Schottky}} - kT\ln\left(\frac{N-n}{n}\right)^2 = 0, \tag{20}$$

Hence, for $n \ll N$ and for $n = N_{\text{Schottky}}$, the density of Schottky defects in a *diatomic* lattice is given by

$$N_{\text{Schottky}} = N_{\text{lattice}}\exp\left(-\frac{E_{\text{Schottky}}}{2kT}\right). \tag{21}$$

In a similar fashion, the density of *Frenkel pairs* is determined, resulting in

$$N_{\text{Frenkel}} = \sqrt{N_{\text{lattice}}N_{\text{interstitial}}}\exp\left(-\frac{E_{\text{Frenkel}}}{2kT}\right), \tag{22}$$

where $N_{\text{interstitial}}$ is the density of interstitial sites and E_{Frenkel} is the Frenkel energy, i.e., the energy to take one cation and put it at a (distant) interstitial site.

A more sophisticated approach involving the grand canonical ensemble that permits an analysis of more complicated cases can be found in Landsberg and Canagaratna (1984). We will complement the simple model discussed above by lattice relaxation and entropy contributions occurring in semiconductors.

Lattice Relaxation In actual semiconductors, the simplified picture given above needs to be modified to take into consideration:

- The different actual positions within the lattice which can be occupied by an interstitial
- The relaxation of the lattice surrounding a vacancy or an interstitial

For example, within the Si lattice, there are three *different sites* (H, T, and B) to place a self-interstitial, shown for a $(0\bar{1}1)$ plane in Fig. 5a. They have different formation enthalpies, which also depend on the occupancy, i.e., the charge character of the center (Car et al. 1984, Van de Walle and Neugebauer 2004).

The *relaxation* of the surrounding lattice may be minor or can be very substantial. The lattice-relaxation energy is the energy difference between the unrelaxed defect (illustrated for a vacancy in Fig. 5b), where all atoms near the defect remain in their ideal lattice sites, and the relaxed defect (Fig. 5c); in the latter, bonds rearrange to minimize the number of dangling bonds and to maximize bonding energy.

Entropy Contribution There are two contributions for an increased entropy of realistic intrinsic defects:

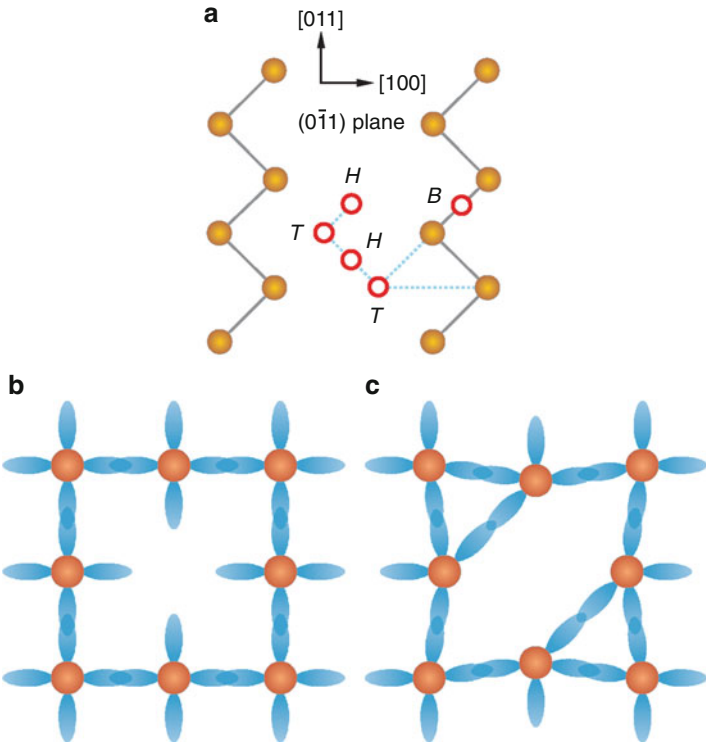


Fig. 5 (**a**) Different sites of self-interstitials in Si shown in a $(0\bar{1}1)$ plane. Open *circles* indicate interstitials with tetragonal (T) or hexagonal (H) symmetry; B denotes a bond-centered site. (**b**, **c**) Vacancy (b) in an unrelaxed state with dangling bonds and (c) in a relaxed state without dangling bonds

- The extended relaxation of the lattice surrounding such a defect
- The change in oscillatory frequencies of the surrounding lattice

The *extended relaxation* permits numerous similar configurations with nearly identical energy. This causes a substantial increase in the configurational entropy and thereby yields a much larger pre-exponential factor determining the density of such defects in equilibrium (see Lannoo and Bourgoin 1981).

The *vibrational part* of the entropy is given in its most simple form as

$$\Delta S_{\text{vib}} = k \ln \left(\sum_i \frac{\nu_{i0}}{\nu_i} \right), \tag{23}$$

where ν_{i0} and ν_i are the original and changed frequencies of the undisturbed lattice and the lattice surrounding the defect. There are two major modes considered near the defect: the *breathing mode*, with all surrounding atoms moving in and out in phase (Fig. 2 of ▶ Chap. 17, "Optical Properties of Defects"), and a *vector-like motion*, including the defect (interstitial). When considering a somewhat reduced binding of

the lattice surrounding the defect, we expect $\nu_{i0} > \nu_i$ and therefore a positive entropy contribution. This contribution also increases the pre-exponential factor and therefore causes a somewhat larger density of intrinsic defects (Talwar et al. 1980). For a review of ab initio calculation of native defects in zincblende III–V and II–VI semiconductors, see Jansen and Sankey (1989), Freysoldt et al. (2014), Lyons et al. (2021); for defects in III-nitrides see Van de Walle and Neugebauer (2004), for lead halide perovskites see Kang et al. (2021).

2.2 Frozen-in Intrinsic Defect Density

The analysis presented in Sect. 1.1 assumes thermodynamic equilibrium but does not ask how long it takes to achieve equilibrium. At sufficiently high temperatures, equilibrium is usually reached faster than the temperature changes upon cooling. While cooling a crystal, the density of defects[9] will start to lag behind the value corresponding to the temperature. Finally, at low enough temperatures, it will not decrease appreciably within the time of the experiment. This "frozen-in" defect density $N_{\text{freezing-in}}$ identifies – via Eqs. 19, 21, or 22 – a *freezing-in temperature* $T_{\text{freezing-in}}$. For example, for Frenkel defects

$$N_{\text{freezing-in}} = \sqrt{N_{\text{lattice}} N_{\text{interstitial}}} \exp\left(-\frac{E_{\text{Frenkel}}}{2kT_{\text{freezing-in}}}\right). \tag{24}$$

Freezing-in occurs because of the reduced mobility of lattice defects with decreasing temperature, which finally makes it impossible for them in the given time frame to find a defect partner for recombination, e.g., for an interstitial to find a vacancy.

For example, the interatomic potential distribution of a Frenkel defect in a simple cubic monatomic lattice is shown in Fig. 6. In the figure, $x = x_{\text{V}}$ is the position of the vacancy; $x_{\text{i1}}, x_{\text{i2}}, x_{\text{i3}}, \ldots$ are the positions of interstitial sites, and ΔE_{i} is the activation energy for interstitial motion. The time required for an atom to move from one to the adjacent interstitial site can be estimated from

$$\tau_{\text{i}} = \frac{2\pi}{\omega_{\text{i}}} \exp\left(\frac{\Delta E_{\text{i}}}{kT}\right), \tag{25}$$

where ω_{i} is an effective oscillation frequency of the interstitial atom. In a crystal with a density of N_{Frenkel} Frenkel defects, the average time τ_{r} for a recombination event to occur is given by

[9] In materials in which two types of defects need to be considered, two different freezing-in temperatures appear, and, because of conservation and neutrality considerations, a more complex behavior is expected (Hagemark 1976).

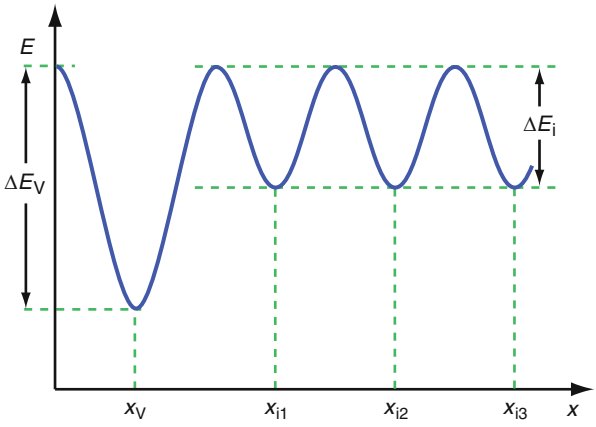


Fig. 6 Atomic potential for an atom on a vacancy site x_V or on an interstitial position x_i

$$\tau_r = \frac{1}{4\pi\ D_i r_i N_{Frenkel}}, \tag{26}$$

where r_i is the reaction distance, which here is equal to the distance between adjacent interstitial sites a_i. When using Eq. 80 for the diffusion constant D_i with $\alpha = 1/6$, we obtain

$$\tau_r = \frac{3}{a_i^3\,\omega_i\sqrt{N_{interstitial}N_{lattice}}}\exp\left(\frac{\Delta E_i + E_{Frenkel}/2}{kT}\right). \tag{27}$$

With ω_i on the order of 10^{13} s^{-1}, $\Delta E_i = 1$ eV, and $E_{Frenkel} = 2$ eV, we obtain a recombination time on the order of 1 year at 500 K. After increasing the temperature to 700 K, the time constant is reduced to about 10 s, and the defect density comfortably follows a sufficiently slowly decreasing temperature.

The freezing-in temperature is a function of the rate of cooling. Freezing-in is reached when this cooling rate equals the rapidly decreasing rate of recombination defined by $dT/d\tau_r$:

$$\frac{dT}{dt} = \frac{dT}{d\tau_r} = -\frac{kT}{\Delta E_i + E_{Frenkel}/2}\,\frac{T}{\tau_r}. \tag{28}$$

From Eqs. 27 and 28, we obtain a well-defined freezing-in temperature contained in τ_r, which increases with increasing cooling rate:

$$T_{freezing\text{-}in} = -\frac{\Delta E_i + E_{Frenkel}/2}{k} \times \frac{1}{\ln\left(\dfrac{kT_{freezing\text{-}in}}{\Delta E_i + E_{Frenkel}/2}\,\dfrac{a_i^3\,\omega_i\sqrt{N_{interstitial}N_{lattice}}\,T_{freezing\text{-}in}}{3\left(-\frac{dT}{dt}\right)}\right)}. \tag{29}$$

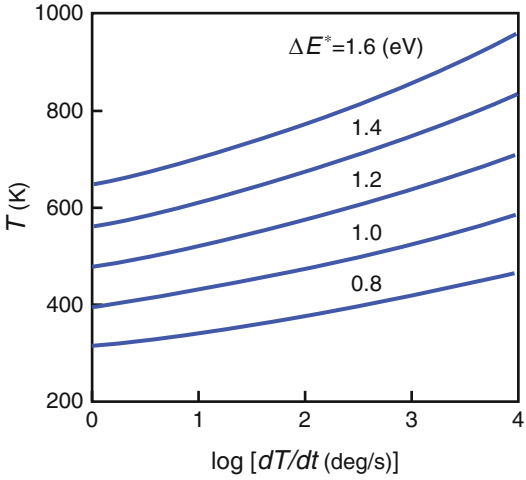


Fig. 7 Freezing-in temperature as a function of the cooling rate with ΔE^* as the family parameter

With the values of the above-given example and a cooling rate of 1 deg/s, the value of the logarithmic term is ~25. Thus, we obtain as a rough estimate, with $\Delta E^* = \Delta E_i + E_{Frenkel}/2$:

$$T_{\text{freezing-in}} \cong 400 \;\; \Delta E^*(\text{eV}) \cdot \left(1 + \frac{1}{30} \ln \left|\frac{dT}{dt}\right|\right)(\text{K}). \tag{30}$$

Since $T_{\text{freezing-in}}$ is logarithmically dependent on the cooling rate, the freezing-in temperature changes little with a normal variation of cooling rates except for rapid quenching (see Fig. 7).

2.3 Defect-Formation Energy

Formation energies of defects control equilibrium concentrations of intrinsic and extrinsic point defects or defect complexes. The calculation of these energies, most commonly using density functional theory (DFT), is today a standard tool to address defect and impurity structures, their concentrations, and their effect on doping (Drabold and Estreicher 2007, Freysoldt et al. 2014); different formation energies calculated by various DFT approaches are addressed by Freysoldt et al. (2016). We briefly point out some general principles and a few results for prominent defects in silicon.

Vacancies The formation energy[10] of vacancies E_{Schottky} is obtained at sufficiently high temperature from the exponential change of the electrical conductivity with temperature when these vacancies are electrically active as the dominant acceptors (see Sect. 2.5 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"). The formation energies are typically in the 2 ... 6 eV range (Berding et al. 1990), but depend significantly on the charge state and the Fermi level.

A rough estimate about the formation energy E_{Schottky} of vacancies can be made from the microscopic equivalent of a macroscopic cavity:

$$E_{\text{Schottky}} = A \; H_{\text{Schottky}}, \tag{31}$$

where A is the surface area of the cavity of the vacancy volume and H_{Schottky} is the specific macroscopic surface energy (Brooks 1963; Friedel 1966). This simple model provides fair agreement with the experiment and can be improved by considering the anisotropy of the cavity, which is not spherical, and of the surface energy (van Vechten 1980). The same cavity model gives some indication of a slight disparity in the density of V_A and V_B in an AB compound, since the formation energy increases with increasing cavity radius. As a net result, we observe a strong preference of n- or p-type semiconductivity when $r_A > r_B$ or $r_A < r_B$, respectively, in lattices with a substantial ionic bonding component; here, A is the cation, and B is the anion.

In Table 1, additional information is given for some of the II–VI compounds. These show substantial preference of nonstoichiometry because of the asymmetry of the formation energy of cation and anion vacancies.

For a more thorough evaluation of the formation energy, we must consider:

- Dangling bonds of the surrounding lattice atoms
- Reconstruction of the surrounding lattice
- Lattice relaxation beyond nearest neighbors

These effects are lattice specific and require an individual analysis for each type of vacancy (see literature cited in Pantelides 1987). For example, the Si vacancy is

Table 1 Melting point T_{m} (K), minimum vapor pressure P_{min} (atm), equilibrium constant log K_p, and surplus element of some II–VI compounds (After Lorenz 1967)

II–VI Compound	T_{m}	$P_{\text{min}}(T_{\text{m}})$	log K_p	Surplus
ZnS	2,103	3.7	0.85	Zn
ZnSe	1,703	0.53	−1.65	Zn
ZnTe	1,568	0.64	−1.4	Te
CdS	1,748	3.8	0.9	Cd
CdSe	1,512	0.41	−2.0	Cd
CdTe	1,365	0.23	−2.75	Te

[10]Often, the enthalpy H_{Schottky} is cited rather than the energy; however, with $\Delta H_{\text{Schottky}} = \Delta E_{\text{Schottky}} + P\Delta V$ and negligible volume changes in the solid, both are almost identical at room temperature. Near the melting point, a $\Delta V/V \cong (1/3)(\Delta l/l) \cong 2\%$ change in volume (Eq. 17 of ▶ Chap. 5, "Phonon-Induced Thermal Properties") may be considered.

Table 2 Activation energies for vacancies and interstitials of Si, P, and Al or combined with vacancies in intrinsic, *n*- and *p*-type silicon in eV (After Car et al. 1985)

Species	E_{int}	Species	E_{int}	E_n	Species	E_{int}	E_p
V_{Si}^{2+}	5.1	$(P_{Si}V_{Si})^{+}$	3.2	3.8	$(Al_{Si}V_{Si})^{+}$	3.9	3.2
V_{Si}^{+}	4.7	$(P_{Si}V_{Si})^{0}$	3.0.	3.0	$(Al_{Si}V_{Si})^{0}$	3.5	3.5
V_{Si}^{0}	4.2	$(P_{Si}V_{Si})^{-}$	3.2	2.6	$(Al_{Si}V_{Si})^{-}$	3.5	4.1
V_{Si}^{-}	4.5	$(P_{Si}Si_i)^{0}$	5.7	5.7	$(Al_{Si}Si_i)^{0}$	4.0	4.0
Si_i^{2+}	5.6	P_i^{0}	5.1	5.1	Al_i^{2+}	4.3	3.0
Si_i^{+}	5.5	P_i^{-}	5.2	4.6	Al_i^{+}	3.9	3.1
Si_i^{0}	5.7	P_i^{2-}	5.6	4.5	Al_i^{0}	3.9	3.9

discussed by Watkins (1986). In contrast to vacancies in other materials, it cannot be frozen-in by rapid quenching at measurable densities because of its high mobility. It is efficiently produced by radiative electron damage at low temperatures. In order to avoid rapid recombination with the coproduced and highly mobile interstitial, trapping of the interstitial at group III atoms in *p*-type Si can be employed (Watkins 1974). For the electron structure of the Si vacancy, see Sect. 2.2.1 of ▶ Chap. 19, "Deep-Level Centers". For an estimation of the energy of vacancy formation, see Car et al. (1985) and Table 2.

Divacancies Vacancies in a covalent crystal are attracted to each other. This can be interpreted as being caused by the reduction of the surface area, which thereby reduces their surface energy. This process is similar to that which causes air bubbles in water to coalesce. Estimated accordingly, the binding energy of divacancies in Si is 1.03 eV, compared to a measured value of 1.2 eV (Ammerlaan and Watkins 1972). In a microscopic model, one realizes that in a divacancy only six bonds are broken, rather than eight in two separate vacancies.

Antisite Defects Antisite defects are intrinsic lattice atoms in a compound placed on a wrong lattice site, e.g., a *B* atom on an *A* site B_A in an *AB* lattice. Such antisite defects may be regarded as substitutional impurities except that they are supplied by a large reservoir of sublattice *B* atoms. Their energy of formation has two contributions: one from the reduction of the bandgap due to disorder and one from an electronic contribution causing a shift in the Fermi energy – the latter depending on doping.

2.4 Defect Chemistry

The creation of different types of defects is often interrelated (see Sect. 2.1). For instance, for Schottky disorder of an *AB* compound, vacancies of both *A and B* ions are produced in equal densities. The quasi-neutrality relation prevents the creation of one charged defect in substantial excess over the others. The creation of lattice defects can be described similarly to the creation of a chemical compound in the

presence of other reaction partners. The governing relation is the *mass-action law*, which, for a reaction

$$A + B \overleftrightarrow{\longrightarrow} AB, \tag{32}$$

can be written as

$$[A] + [B] = K_{AB} \cdot [AB], \tag{33}$$

where the square brackets indicate concentrations and K_{AB} is the mass-action law constant:

$$K_{AB} = K_{AB,0} \ \exp\left(-\frac{\Delta G}{kT}\right), \tag{34}$$

with ΔG as the change in Gibbs free energy[11]:

$$\Delta G = \Delta U + P\Delta V - T\Delta S = \Delta H - T\Delta S, \tag{35}$$

where ΔU is the change in *internal energy* due to changes in the interaction potential of the vacancy with the surrounding lattice and ΔV is the often negligible change in volume. Both are usually combined as the change in *enthalpy* $\Delta H = \Delta U + P\Delta V$. ΔS is the change in *entropy*:

$$\Delta S = \Delta S_{\text{config, basic}} + \Delta S_{\text{config, extend}} + \Delta S_{\text{vib}}. \tag{36}$$

The basic configurational entropy part and the parts dealing with extended lattice relaxation and with lattice vibration are described in Sect. 2.1.

In solids, defect-chemistry description is advantageous because it illustrates the interconnection of the different defects. We will now take into consideration that these defects "react" with each other and thereby change their charge and position in the lattice, i.e., their "defect chemical" composition.

For example, for the Schottky disorder in the Ge crystal, the defect-chemistry notation can be written as

$$\text{Ge}_{\text{Ge}} \overleftrightarrow{\longrightarrow} \text{V}_{\text{Ge}} + \text{Ge}_{\text{surface}}. \tag{37}$$

Neglecting the difference between Ge_{Ge} and $\text{Ge}_{\text{surface}}$, they cancel in Eq. 37, yielding

[11] In Sect. 2.1, the Helmholtz free energy F was used, which is related to the Gibbs free energy G by $G = F+PV$. Since some of the reactions involve an interaction with a gas atmosphere, the more general notation is used here.

$$0 \overleftrightarrow{\longrightarrow} \mathrm{V_{Ge}}. \tag{38}$$

The corresponding mass-action law reads

$$N_{\mathrm{Schottky}} = [\mathrm{V_{Ge}}] = K_{\mathrm{V}} = K_{\mathrm{V},0} \exp\left(-\frac{\Delta G}{kT}\right). \tag{39}$$

Disregarding a change in volume and with $\Delta H = E_{\mathrm{Schottky}}$ as the Schottky energy, Eq. 39 can be rewritten as

$$N_{\mathrm{Schottky}} = K_0 K_{\mathrm{V}} \exp\left(-\frac{\Delta H}{kT}\right). \tag{40}$$

This is equivalent to Eq. 19, derived earlier, with $K_V = N_{\mathrm{lattice}}$ and $K_0 = \exp(\Delta \mathrm{S_{vibr}}/k)$, a factor accounting for the vibrational entropy part in ΔG.

Defect Chemistry Involving Neutral and Ionized Defects Solid-state reactions take place between crystal defects, carriers, and external partners, such as a gas atmosphere. Types of reactions to be considered in solid-state defect chemistry include:

- Intrinsic defect formation
- Changing of stoichiometry by interaction with a gas atmosphere of one of the components
- Doping with foreign defects
- Defect associate formation
- Formation of ionized defects
- Creation of free carriers

A few examples of some typical reactions are given below, along with an analysis that shows the interdependence of the different reactions. For more details, see Kröger (1964).

For example, a Frenkel disorder generates metal-ion (M) interstitials which act as donors and metal-ion vacancies (V) which act as acceptors. The reaction equation for the Frenkel disorder is

$$M_M^0 + \mathrm{V_i^0} \overleftrightarrow{\longrightarrow} M_{\mathrm{i}}^0 + \mathrm{V}_M^0; \quad (E_{\mathrm{Frenkel}}). \tag{41}$$

A "0" is added to the chemical symbol of each defect, representing its neutral state; the energy necessary to achieve this transition is appended in parentheses. These defects can be ionized according to

$$M_{\mathrm{i}}^{0} \overset{\longleftarrow}{\longrightarrow} M_{\mathrm{i}}^{+} + e^{-}; \quad \left(\tilde{E}_{\mathrm{d}}\right) \tag{42}$$

and

$$\mathrm{V}_{M}^{0} \overset{\longleftarrow}{\longrightarrow} \mathrm{V}_{M}^{-} + h^{+}; \quad \left(\tilde{E}_{\mathrm{a}}\right) \tag{43}$$

with $\tilde{E}_{\mathrm{d}} = E_{\mathrm{c}} - E_{\mathrm{d}}$ and $\tilde{E}_{\mathrm{a}} = E_{\mathrm{a}} - E_{\mathrm{v}}$. In addition, we create intrinsic carriers (see Sect. 1.3 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"):

$$0 \overset{\longleftarrow}{\longrightarrow} e^{-} + h^{+}; \quad \left(E_{\mathrm{g}}\right). \tag{44}$$

These reactions can be described by the following set of mass-action law equations:

$$\frac{\left[M_{\mathrm{i}}^{0}\right]\left[\mathrm{V}_{M}^{0}\right]}{\left[M_{M}^{0}\right]\left[\mathrm{V}_{\mathrm{i}}^{0}\right]} = K_{M,\mathrm{i}} \propto \exp\left(-\frac{E_{\mathrm{Frenkel}}}{kT}\right), \tag{45}$$

$$\frac{\left[M_{\mathrm{i}}^{+}\right] n}{\left[M_{\mathrm{i}}^{0}\right]} = K_{\mathrm{D}} \propto \exp\left(-\frac{\tilde{E}_{\mathrm{d}}}{kT}\right), \tag{46}$$

$$\frac{\left[\mathrm{V}_{M}^{-}\right] p}{\left[\mathrm{V}_{M}^{0}\right]} = K_{\mathrm{A}} \propto \exp\left(-\frac{\tilde{E}_{\mathrm{a}}}{kT}\right), \tag{47}$$

$$\frac{np}{N_{\mathrm{c}} N_{\mathrm{v}}} = K_{\mathrm{i}} \propto \exp\left(-\frac{E_{\mathrm{g}}}{kT}\right), \tag{48}$$

with the connecting quasi-neutrality condition:

$$n + \left[\mathrm{V}_{M}^{-}\right] = p + \left[M_{\mathrm{i}}^{+}\right]. \tag{49}$$

The value of the defect chemical approach becomes apparent when we recognize that a crystal defect can participate in various reactions – e.g., V_M in Eqs. 41 and 43 – and can form different species, V_M^0 and V_M^-. Without ionization, the density of all vacancies is given by Eqs. 22 or 45; with ionization, only the density of neutral vacancies is given by Eq. 45; ionized vacancies drop out of Eq. 45 and are governed by Eq. 47. Therefore, the total amount of vacancies increases.

In general, several of these reactions influence each other and cause a shift in the density of the reaction partners. In order to obtain the densities of all partners, the system (45–49) must be solved simultaneously. In Sect. 2.6, we will introduce an instructive approximation for accomplishing this.

2.5 Changing of Stoichiometry and Compensation

In this example, we will show how the heat treatment of an AB compound in an atmosphere of one of the components can change the stoichiometry of this compound.

For example, with Schottky defects we have[12]

$$0 \overset{\longleftarrow}{\longrightarrow} \mathrm{V}_A^0 + \mathrm{V}_B^0; \quad \left(E_{\mathrm{Schottky}}\right). \tag{50}$$

The treatment of the AB compound in a gas of A makes AB grow and consequently causes an increase in the relative density of B vacancies:

$$A_{\mathrm{gas}} \overset{\longleftarrow}{\longrightarrow} A_A + \mathrm{V}_B^0; \quad \left(E_{A,\,\mathrm{gas}}\right), \tag{51}$$

which yields for equilibrium

$$\left[\mathrm{V}_A^0\right]\left[\mathrm{V}_B^0\right] = K_{\mathrm{Schottky}} \propto \exp\left(-\frac{E_{\mathrm{Schottky}}}{kT}\right) \tag{52}$$

and

$$\frac{\left[A_A^0\right]\left[\mathrm{V}_B^0\right]}{P_A} = K_{A,\,\mathrm{gas}} \propto \exp\left(-\frac{E_{A,\,\mathrm{gas}}}{kT}\right). \tag{53}$$

Therefore, we obtain for the change in stoichiometry δ, with $AB \rightarrow A_{1+\delta}B$, the expression

$$\delta = \left[A_A^0\right] - \left[B_B^0\right] = \left[\mathrm{V}_B^0\right] - \left[\mathrm{V}_A^0\right], \tag{54}$$

which can be evaluated using Eqs. 52 and 53.

Another possibility is the reduction of A vacancies according to $A_A^0 \overset{\longleftarrow}{\longrightarrow} A_{\mathrm{gas}} + \left[\mathrm{V}_A^0\right]$ and consequently an increase in A interstitials or of antisites A_B. In actual compound crystals, the densities of both vacancies are not the same, since their formation energy differs. This difference must be taken into account for any real crystal (see Sect. 2.3).

Doping in Equilibrium with a Foreign Gas In a similar fashion, the treatment of a semiconductor in a vapor of a foreign atom C with a partial pressure P_C causes a change of the intrinsic defect density. The interaction takes place via the

[12] Or, for an A_nB_m compound, we have $0 \leftrightarrows n\mathrm{V}_A^0 + m\mathrm{V}_B^0$ with $\left[\mathrm{V}_A^0\right]^n\left[\mathrm{V}_B^0\right]^m = K_{\mathrm{Schottky}}$.

neutrality condition. For instance, if the foreign atom is a metal atom replacing a lattice atom

$$C_{\text{gas}} \overset{\longleftarrow}{\longrightarrow} C_A^0 \quad \text{with} \quad \frac{[C_A^0]}{P_A} = K_{C_{\text{gas}}} \tag{55}$$

and acting as a donor

$$C_A^0 \overset{\longleftarrow}{\longrightarrow} C_A^+ + e^- \quad \text{with} \quad \frac{n[C_A^+]}{[C_A^0]} = K_D, \tag{56}$$

it will tend to reduce the density of intrinsic defects with the same charge and to increase the density of intrinsic defects with the opposite charge, since they are interrelated by

$$n + [\mathrm{V}_B^0] = p + [\mathrm{V}_A^+] + [C_A^+]. \tag{57}$$

In more general terms, a foreign atom incorporated as a donor tends to reduce the intrinsic donor densities and to increase the density of intrinsic acceptors. Vice versa, intrinsic and extrinsic acceptors cause an increase in the equilibrium density, i.e., the solubility of the foreign donors.

This important interrelationship is called *compensation*. It is used to increase doping densities otherwise limited by low solubility; this is technically important, e.g., for tunnel diodes. In devices fabricated under significant depart from thermodynamic equilibrium using epitaxy (Sect. 3.3 of ▶ Chap. 1, "Properties and Growth of Semiconductors"), doping limits given by Eq. 57 may be exceeded to some extent by kinetic barriers.

Formation of Defect Associates When crystal defects are able to form defect associates, these associates constitute a new species and influence the balance of the unassociated defects. For instance, in an *AB* compound with Schottky disorder, the metal vacancies may form *dimers*, i.e., two vacancies as nearest neighbors:

$$2\mathrm{V}_A^0 \overset{\longleftarrow}{\longrightarrow} \mathrm{V}_{A_2}^0; \quad (E_2), \tag{58}$$

with E_2 as the binding energy of these dimers.[13] With the formation of $\mathrm{V}_{A_2}^0$, a new species is created involving metal vacancies. The total amount of A vacancies is thus increased. Therefore, the balance between V_A and V_B in the Schottky disorder is shifted.

[13]With proper charging, these dimers could be regarded as equivalent to a nonmetal molecule B_2 sitting on a lattice site of a cluster of four vacancies. Such molecules are often covalently bonded and therefore have a substantial binding energy; hence, they have a high probability of occurring.

Compensation with Fixed Density of Donors In the following example, an impurity is present at a *fixed* density within a monatomic semiconductor together with Schottky defects acting as acceptors. The extrinsic substitutional dopant acts as a donor with a density $[F_M]_{\text{total}}$. The following reactions take place, following the corresponding mass-action laws:

$$[F_M]_{\text{total}} = \text{const} \tag{59}$$

$$0 \overset{\longleftarrow}{\longrightarrow} \mathrm{V}_M^0; \quad (E_{\text{Schottky}}) \quad \left[\mathrm{V}_M^0\right] = K_0 \exp\left(-\frac{E_{\text{Schottky}}}{kT}\right) \tag{60}$$

$$\mathrm{V}_M^0 \overset{\longleftarrow}{\longrightarrow} \mathrm{V}_M^- + h^+; \quad (\tilde{E}_a) \quad p\left[\mathrm{V}_M^-\right] = N_{\mathrm{v}}\left[\mathrm{V}_M^0\right]\exp\left(-\frac{\tilde{E}_a}{kT}\right) \tag{61}$$

$$F_M^0 \overset{\longleftarrow}{\longrightarrow} F_M^- + e^-; \quad (\tilde{E}_d) \quad n\left[F_M^0\right] = N_{\mathrm{c}}\left[F_M^0\right]\exp\left(-\frac{\tilde{E}_d}{kT}\right) \tag{62}$$

$$0 \overset{\longleftarrow}{\longrightarrow} np; \quad (E_{\mathrm{g}}) \quad np = N_{\mathrm{c}}N_{\mathrm{v}}\exp\left(-\frac{E_{\mathrm{g}}}{kT}\right) \tag{63}$$

Two conditional equations, the *conservation of mass equation*:

$$[F_M]_{\text{total}} = \left[F_M^0\right] + \left[F_M^+\right] \tag{64}$$

and the *quasi-neutrality equation*:

$$n + \left[\mathrm{V}_M^-\right] = p + \left[F_M^+\right], \tag{65}$$

force an interaction between the different reaction partners, i.e., force compensation. We will analyze this interaction in the following section after introducing the useful Brouwer approximation.

2.6 Brouwer Approximation

Before discussing the solution to the problem given above, we will indicate here how to analyze the typical system of equations encountered in defect chemistry. The governing equations for defect chemical reactions are coupled with conservation of mass or charge equations. They contain the sum of densities, e.g., the neutrality equation Eq. 49:

$$n + \left[\mathrm{V}_M^-\right] = p + \left[M_{\mathrm{i}}^+\right]. \tag{66}$$

Since each of these densities varies exponentially with temperature, mostly with different slopes, they are usually of substantially different magnitudes. Therefore, we can neglect one of the terms on each side and thereby can distinguish four cases:

$$\left[V_M^-\right] = \left[M_i^+\right] \tag{67}$$

$$n = \left[M_i^+\right] \tag{68}$$

$$\left[V_M^-\right] = p \tag{69}$$

$$n = p \tag{70}$$

Depending upon the relative magnitude of the different activation energies, we have different temperature ranges in which the validity of these neutrality approximations changes over from one to the other case. For example, at low temperatures, Eq. 68 holds; with increasing temperature, Eq. 70 becomes valid. In each of these temperature ranges, the governing equation can be given explicitly. For instance, for $n = \left[M_i^+\right]$, we obtain from Eq. 46[14]

$$n = \left[M_i^+\right] = \sqrt{\left[M_i^+\right] K_D} = \sqrt{\left[M_i^+\right]} \exp\left(-\frac{\tilde{E}_d}{2kT}\right), \tag{71}$$

and have $p = \left[V_M^-\right]$ with Eq. 47, if $E_a < E_d$. At higher temperature, we obtain from $n = p$ with Eq. 48:

$$n = p = \sqrt{N_c N_v K_i} = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2kT}\right). \tag{72}$$

In each of these temperature ranges, the densities can be represented as straight lines in an Arrhenius plot (ln[ ] vs. $1/T$), which shows the dominant interaction. In the intermediate temperature range, a smooth transition occurs. Such a transition is neglected and replaced by sharp breaks in the Brouwer approximation (Brouwer 1954). An illustration for such a relation is shown in Fig. 8 for the example given above.

In the Brouwer diagram, majority (n) and minority (p) carrier densities show a split with the same angle (α) in the semilogarithmic plot (Fig. 8). The obtained results are similar to those obtained in Sect. 2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"; however, the carrier depletion range is neglected here.

Brouwer Diagram of Compensation with Fixed Donor Density When carriers *and* atomic reaction partners are involved, the resulting Brouwer diagram becomes more instructive. For instance, in a monatomic semiconductor with Schottky defects and a fixed density of donors given by Eq. 64, several of these defects interact. We see from Eqs. 59, 60, 61, 62, 63, 64, 65 that five out of eight possible temperature ranges occur:

[14] In order to avoid exact compensation, we must also consider some extrinsic donors to make the donors predominant (see Sect. 2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers").

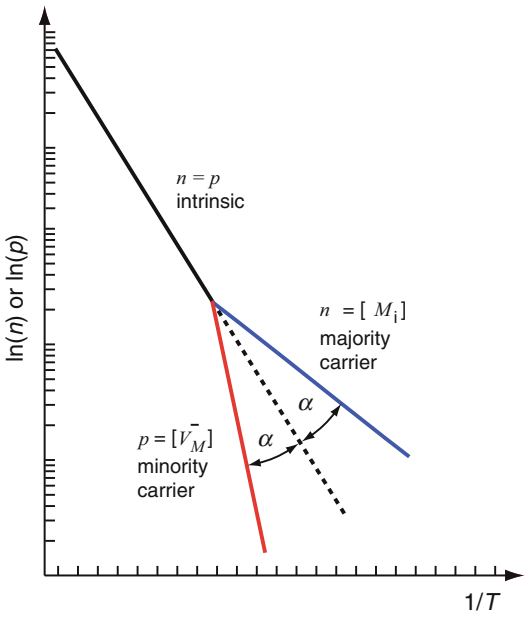


Fig. 8 Brouwer diagram for a simple example of a semiconductor with one dominant intrinsic donor and free carriers

$$
\begin{array}{lrl}
\text{I} & [F_M]_{\text{total}} = [F_M^+]; & \left[\text{V}_M^-\right] = [F_M^+] \\
\text{II} & = \left[F_M^0\right]; & \left[\text{V}_M^-\right] = [F_M^+] \\
\text{III} & = \left[F_M^0\right]; & n = [F_M^+] \\
\text{IV} & = [F_M^+]; & n = [F_M^+] \\
\text{V} & = [F_M^+]; & n = p
\end{array}
\tag{73}
$$

The corresponding Brouwer diagram is given in Fig. 9. It shows that the density of the neutral intrinsic vacancy $[\text{V}_M^0]$ is not influenced by any of the other reactions. It extends as a straight line through all temperature ranges. All other reactions show interdependencies.

In the temperature ranges I, IV, and V, essentially all foreign atoms are ionized ($[F_M]_{\text{total}} = [F_M^+]$). At low temperatures (range I), $[F_M^+]$ is equal to the density of ionized acceptors $\left[\text{V}_M^-\right]$. At higher temperatures (range III), $[F_M^+]$ is equal to the free electron density; this is the depletion range. In range V, we observe $n > [F_M^+]$; this is the intrinsic range with $n = p$. There are two intermediate ranges where the density of ionized donors is depressed (II and III); the density of charged acceptors starts to decline (II) because of competition with electrons in the quasi-neutrality condition.

The density of charged intrinsic acceptors dramatically increases at low temperatures (I) because of the incorporation of foreign donors, permitting neutrality via $[F_M^+] = \left[\text{V}_M^-\right]$. When the electron density increases (II and III), the density of ionized acceptors decreases until n takes over (IV) in the neutrality condition. With n larger than the donor density (i.e., when $n = p$), $\left[\text{V}_M^-\right]$ approaches the

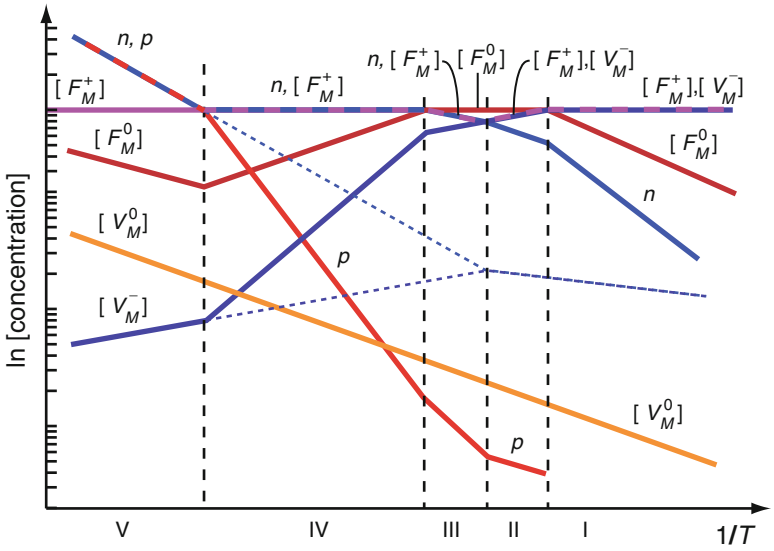


Fig. 9 Brouwer diagram for a monatomic semiconductor with fixed density of foreign donors $[F_M]_{\text{total}} = \text{const}$ and intrinsic vacancies acting as acceptors (see text). *Dotted* curve occurs if no foreign doping is present $\left([F_M]_{\text{total}} = 0\right)$ (After Kröger 1964)

density it would have in the undoped case. The recombination of free holes also causes a reduction in $\left[F_M^-\right]$. Therefore, the neutral donor density increases again.

This example shows the rather complex interdependency of the different densities and presents a good illustration of an instructive analysis. The examples presented in these sections, which dealt with defect chemistry, have been simplified somewhat in order to indicate the principles involved, rather than to illustrate actual material behavior. For further reading, see original literature in Kröger (1964).

3 Diffusion of Lattice Defects

The basic concepts of the diffusion of lattice defects are summarized here briefly since they are prerequisites for actual doping – where foreign atoms are supplied by diffusion from the surface or when intrinsic Schottky defects are created by outdiffusion of lattice atoms to the surface.

3.1 Fick's Laws of Diffusion

Diffusion is determined by *Fick's first law*, which relates the diffusion current j_i to the diffusion tensor D_{ik} and the density gradient:

$$j_i = -\sum_{k=1}^{3} D_{ik} \frac{\partial N}{\partial x_k}, \tag{74}$$

where N is the density of the diffusing lattice defect. Diffusion must also follow the continuity equation (*Fick's second law*):

$$\frac{\partial N}{\partial t} = -\sum_{i} \frac{\partial}{\partial x_i} j_i = \sum_{ik} \frac{\partial}{\partial x_i} D_{ik} \frac{\partial N}{\partial x_k}. \tag{75}$$

For small defect densities, we can disregard defect interaction, i.e., D_{ik} is independent of N, and we have

$$\frac{\partial N(\mathbf{x}, t)}{\partial t} = \sum_{ik} D_{ik} \frac{\partial^2 N(\mathbf{x}, t)}{\partial x_i \partial x_k}, \tag{76}$$

with D as a symmetrical second-rank tensor following the point-group symmetry of the crystal. For an isotropic (cubic) crystal, the diffusivity D (also referred to as diffusion coefficient) is a scalar, and we have

$$\frac{\partial N(\mathbf{x}, t)}{\partial t} = D \frac{\partial^2 N(\mathbf{x}, t)}{\partial x^2}. \tag{77}$$

The diffusion equation can be integrated most easily by employing a Laplace transformation. For a simple case of initial conditions, with the dopant deposited at the surface ($x = 0$) and no depletion of dopants for a one-dimensional *semi-infinite sample*, we have

$$\begin{array}{l} N = N_0 \ \text{ at } \ x = 0 \ \text{ for } \ t \geq 0 \\ N = 0 \ \text{ at } \ x > 0 \ \text{ for } \ t = 0 \end{array}, \tag{78}$$

yielding as the solution of Eq. 77

$$N(x, t) = N_0 \left[1 - \text{erf}\left(\frac{x}{\sqrt{4Dt}} \right) \right], \tag{79}$$

the well-known error-function distribution of dopants as the diffusion experiment proceeds (see Fig. 10).

In a very simple example of foreign atoms diffusing on interstitial sites, we relate the diffusivity to the spacing between interstitial sites a and the jump frequency ν_j by

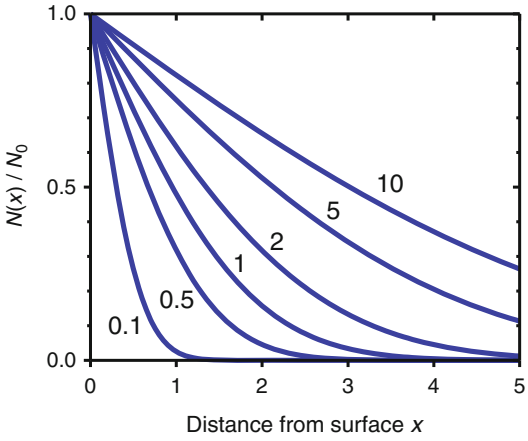


Fig. 10 Density distribution of dopants $N(x, t)$ supplied from a sample surface without depletion and shown for varied values of the diffusion time Dt

$$D = \alpha a^2 \nu_j, \tag{80}$$

where α is a geometry-related factor. For a cubic lattice and interstitial diffusion, one has $\alpha = 1/6$. The jump frequency depends, via a Boltzmann factor, on the temperature:

$$\nu_j = \nu_0 \exp\left(-\frac{\Delta E_i}{kT}\right), \tag{81}$$

where ΔE_i is the height of the saddle point between adjacent lattice atoms over which the diffusion proceeds (Fig. 6). We express the diffusivity as

$$D = D_0 \exp\left(-\frac{\Delta E_i}{kT}\right), \quad \text{where} \quad D_0 = \alpha a^2 \nu_0 \tag{82}$$

and ν_0 as an effective jump frequency.

The path of each of the diffusing atoms is that of a random walk, ignoring correlation effects, with $\overline{R^2}$ the mean square total displacement given by

$$\overline{R^2} = \hat{n} a^2 = \alpha D t, \tag{83}$$

where $\hat{n}$ is the number of interstitial jumps. The average distance traveled by the diffusing atom is simply $\sqrt{\overline{R^2}}$; the factor α is often included in published values of D, yielding for the average distance $\sqrt{Dt}$.

The diffusion coefficients for intrinsic and a number of extrinsic defects are given in Table 3.

Table 3 Self- and impurity diffusion coefficients for some semiconductors in terms of diffusion constant D_0 and energy barrier ΔE_i

	GaAs		InAs		Si		Ge	
Impurity	D_0 (cm^2/s)	ΔE_i (eV)	D_0 (cm^2/s)	ΔE_i (eV)	D_0 (cm^2/s)	ΔE_i (eV)	D_0 (cm^2/s)	ΔE_i (eV)
As	0.7	3.2	$3 \cdot 10^7$	4.45	60	4.20	10.3	2.51
Ga	$1 \cdot 10^7$	5.6			3.6	3.51	40	3.15
In	$5 \cdot 10^{-11}$	at 1,000 °C	$6 \cdot 10^5$	4.0	16.5	3.91	33	3.03
Sb					12.9	3.98	3.2	2.42
Ag	25	2.27	$7.3 \cdot 10^{-4}$	0.26				
Au	29	2.64	$5.8 \cdot 10^{-4}$	0.65				
Cu	0.03	0.53	$3.6 \cdot 10^{-3}$	0.52				
Mg	$2.6 \cdot 10^{-2}$	2.7	$2 \cdot 10^{-6}$	1.17				
O	$2 \cdot 10^{-3}$	1.1						
S	$1.8 \cdot 10^{-2}$	2.6	6.78	2.20				
Se	$8 \cdot 10^{-3}$	4.16	12.6	2.20				
Sn	$3.8 \cdot 10^{-2}$	2.7	$1.49 \cdot 10^{-8}$	1.17				

3.2 Types of Diffusion

In the previous section, a simple diffusion of a foreign or lattice atom from interstitial to adjacent interstitial position was assumed, following a potential as shown in Fig. 6. In actual crystals, such a path is only one of the several paths which contribute to the diffusivity (see Flynn 1972). Additional possibilities exist, and one distinguishes the following types of lattice diffusion:

(a) A simple exchange mechanism
(b) A ring-type exchange mechanism
(c) A vacancy-induced diffusion (vacancy mechanism)
(d) A lattice-relaxation mechanism
(e) A simple interstitial diffusion (as previously described)
(f) An interstitial diffusion with collinear displacement
(g) An interstitial diffusion with noncollinear displacement (interstitialcy mechanism)
(h) A dumbbell interstitial mechanism
(i) A crowdion mechanism
(j) Motion with alternating recharging (athermal motion)

Nine simple examples of these mechanisms are illustrated schematically for a two-dimensional square lattice in Fig. 11; for reviews, see Gösele 1986, Tan 2017. For a comprehensive review of some earlier concepts, see Flynn 1972.

In actual semiconductors, the diffusion processes are more involved since in all of them lattice relaxation, and in most of them various charge states, vacancies, and interstitials are involved, making diffusion possible at much lower temperatures. In addition, there are usually several interstitial positions possible, offering different paths for interstitial diffusion (see Fig. 5).

Diffusion in Covalent Semiconductors
In silicon, diffusion is studied since long due to its technological importance. In the self-diffusion of Si and in the diffusion of dopants and other substitutional impurities, both vacancies (V) and interstitials (I) participate. V and I contribute in neutral as well as charged states depending on the Fermi level, and their concentration is altered by chemical reactions at the Si surface.

Silicon Self-Diffusion The diffusion of Si self-atoms in thermal equilibrium is described by the diffusivity

$$D_{\mathrm{Si,self}} = D_{\mathrm{I}}\tilde{N}_{\mathrm{I,eq}} + D_{\mathrm{V}}\tilde{N}_{\mathrm{V,eq}}, \tag{84}$$

with the diffusivities D and equilibrium concentrations $\tilde{N}$ of interstitials and vacancies (normalized into atomic fractions). Since both I and V may exist in several charge states, each summand of Eq. 84 is a sum on its part, comprising the different diffusivities and concentrations of I and V in their contributing charge

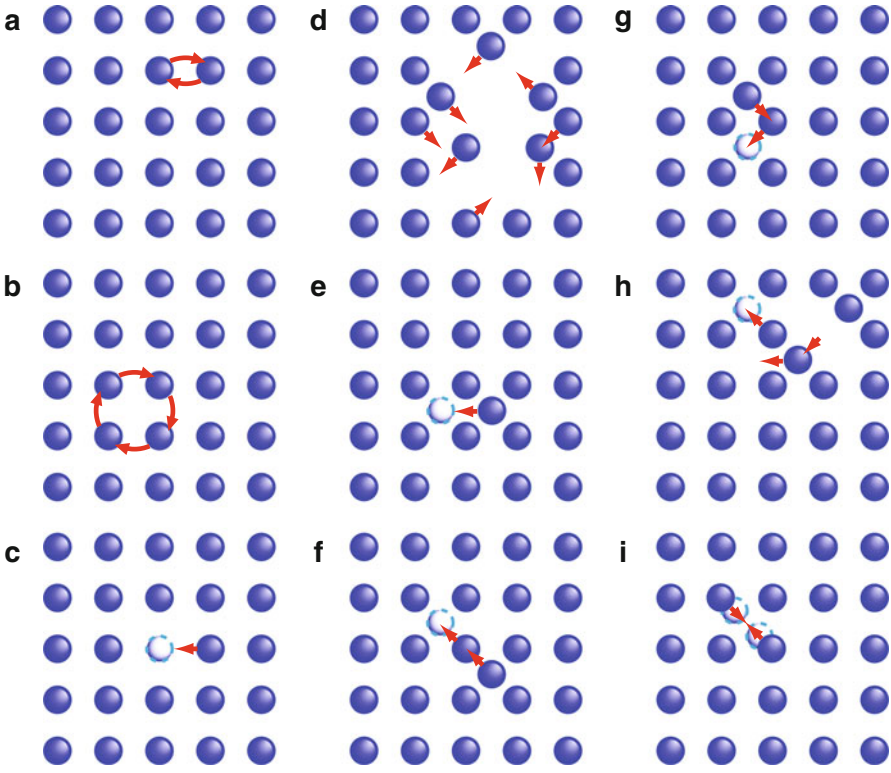


Fig. 11 Different types of lattice diffusion for (**a–i**) given in the text

states. D_I and D_V can hence both be described accurately in terms of Eq. 82 only over a limited temperature range, but a fairly good description of $D_{\mathrm{Si,\ self}}$ is obtained over a large range using $D_\mathrm{I}\tilde{N}_{\mathrm{I,eq}} = 2175.4\ \mathrm{cm}^2/\mathrm{s} \times \exp(-4.95\mathrm{eV}/(kT))$ and $D_\mathrm{V}\tilde{N}_{\mathrm{V,eq}} = 0.0011\ \mathrm{cm}^2/\mathrm{s} \times \exp(-3.52\mathrm{eV}/(kT))$ (Kube et al. 2013; Tan 2017); thereby the Si self-diffusion shows a slightly concave shape in an Arrhenius plot expressing the dominance of I at higher and of V at lower temperature.

Dopant Diffusion in Silicon All substitutional impurities diffuse faster than Si self-diffusion and with a smaller activation enthalpy. The reason is a pair-complex formation of charged I or V point defects with the dopant. Due to the dependence on the charge state, the diffusivities D of dopants in Si depend on the Fermi level. Usually, the charge states $-2 \ldots +1$ are contributing, where neutral point defects are dominant and – depending on the specific dopant – some terms may be negligibly small. For *intrinsic* Si, the resulting total diffusivities show an Arrhenius dependence (see Fig. 12). Under extrinsic conditions, the Fermi-level effect and additional effects at high dopant concentration require a detailed discussion (Tan 2017).

An analysis of Nichols et al. (1989), using interstitial injection from the surface of Si by oxidation (Tan et al. 1983), indicates that the diffusion of B, P, and As in Si

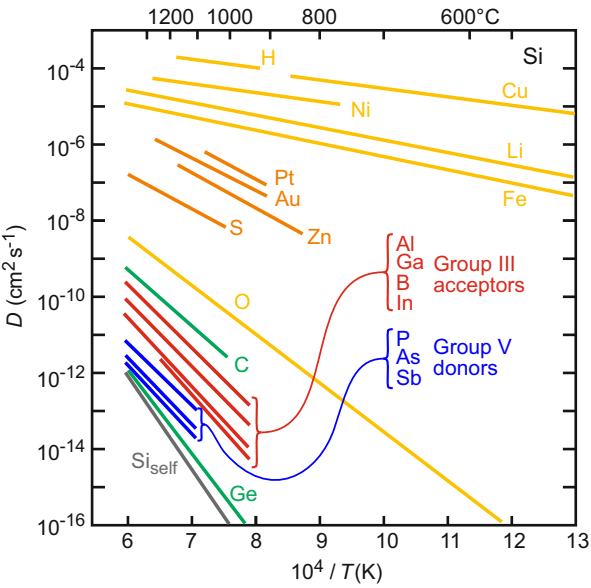


Fig. 12 Diffusivities of dopants in intrinsic silicon. The *gray line* represents Si self-diffusion (Data from Gösele 1988 and Bracht 2000)

is interstitial mediated, while the diffusion of Sb seems to be vacancy mediated. All of these have an activation energy of approximately 2.5 eV (Nichols et al. 1989).

Deep centers, especially some of the transition metal ions, are known to be fast diffusers (Weber 1983). They easily form associates with intrinsic defects mediating their diffusions, e.g., gold associates with interstitials, vacancies, or other dopants (see Stolwijk et al. 1983 and Lang et al. 1980).

Hydrogen as H^+ is a fast diffuser in *p*-type Si. It preferentially moves on an interstitial *B*-*B* path (see Fig. 5), meandering between the relaxing (by 0.4 Å) Si atoms. In *n*-type Si, hydrogen in the H^- charge state may be more stable with diffusion along a *T*-*H* path (see Van de Walle et al. 1989).

Diffusion in Germanium Ge has recently gained interest due to its electron and hole mobilities exceeding those of Si; together with new dielectric high *k* insulators, a Ge FET might outperform a short-channel Si MOSFET. Diffusion in Ge can be described by mechanisms involving only the dopant and vacancies. For *intrinsic* conditions, the diffusivities of dopants are given by Arrhenius dependences similar to those of Si (Bracht and Brotzmann 2006). Under *extrinsic* conditions, Ge self-diffusion is enhanced for *n*-type doping and decreases for *p*-type doping. Enhanced diffusivity in *n*-type Ge is also found for As doping. The observed quadratic dependence of self-diffusion and dopant diffusion on the electron concentration *n* can be explained by the dominance of a doubly negatively charged intrinsic point defect, which could be assigned to the acceptor-type vacancy V^{2-} (Tan 2017).

Diffusion in Gallium Arsenide

Self-diffusion and diffusion of dopants in compound semiconductors is governed by native point defects. There are more species participating as in elemental

semiconductors, and the Fermi-level effect is usually much more pronounced. There are many anomalies observed in diffusion experiments which cannot be described by Fick's law (Eq. 75), and the diffusion mechanisms are generally poorly understood.

GaAs shows a dominant Fermi-level effect, and also a substantial dependence on the As vapor pressure. Within the thermodynamically allowed equilibrium composition range around the stoichiometric $Ga_{0.5}As_{0.5}$ composition, there exist V_{Ga}, V_{As} vacancy, I_{Ga}, I_{As} self-interstitial, and Ga_{As}, As_{Ga} antisite species in thermal equilibrium.

All substitutional species in GaAs, including self-atoms and dopants, diffuse via charged point defects involving various species. For Ga self-diffusion and the Al-Ga interdiffusion in AlAs/GaAs superlattices, the dependence on a cubic electron concentration indicates the participation of a triply negatively charged point defect in the studied temperature range. From the dependence of the diffusivity on the As_4 vapor-phase pressure, the point defect was assigned to the vacancy V_{Ga}^{3-}, where at high temperature V_{Ga}^{2-} species may contribute (Tan 2017). This (these) species also govern(s) the diffusion of Si, the main *n*-type dopant of GaAs. In contrast, the main *p*-type dopants Zn and Be diffuse via a pairing mechanism involving an interstitial, a process which is also observed in other III-V compounds. Such correlated V–dopant and I–dopant mechanisms were directly imaged in bulk AlN (Ishikawa et al. 2014).

Diffusion mechanisms in GaAs and other compound semiconductors are complex and still far from being understood. A detailed discussion is given in Tan (2017).

Athermal Diffusion

When intrinsic defects in Si are produced by electron radiation at low temperatures (~20 K), they are observed to migrate with very low activation energies of ~0.2 eV (Watkins 1986). This is in contrast to high-temperature (thermal) diffusion data, which indicates generation and motion of intrinsic defects in Si with an activation energy on the order of 3 . . . 5 eV (Frank 1981). Such vacancies or interstitials are known to act as deep-level defects. With sufficient free electrons and holes available from the preceding electron irradiation, they can be alternately recharged as they move to different sites, in turn stimulating the next jump – *Bourgoin and Corbett mechanism* (Bourgoin and Corbett 1972).

Other alternate diffusion processes which proceed at very low temperatures may also involve dopants, e.g., aluminum (Troxell et al. 1979) or boron (Troxell and Watkins 1980), Zn-O pairs in GaP, or Fe-B pairs in Si (see Pantelides (1987) for more information).

4 Line Defects

Line defects are one-dimensional imperfections of the crystalline order, which often extend throughout the entire lattice. Along the line characterizing the defect, the symmetry of the crystal may be distorted either with respect to translation or to

rotation. Line defects referring to *translation* are called *dislocations*. Dislocations have a major influence on semiconductor interfaces, on certain types of crystal growth, and on electron transport and are therefore discussed here (see Friedel (1964), Hull (1975), and You and Johnson (2009)). Other important influences include mechanical deformation (yield strength) and the formation of grain boundaries (see Sect. 5.2).

Distortions of the *rotational* symmetry are called *disclinations*. Like dislocations they are sources of internal strain. Disclinations can be used to describe frustrated systems resulting in curved lattice regions which are common in amorphous or organic semiconductors. They are of limited interest in single crystals, where they can move only by diffusion climb, but they appear in polycrystals at the edges of grain boundaries (Hirth et al. 2019). For more information, see the review of Kleman and Friedel (2008).

There are two main types of dislocations:

- Edge dislocations
- Screw dislocations

Other ways to distinguish dislocations will be discussed below (see Read (1953), Weertman and Weertman (1960), Cottrell (1964), Pirouz (1989), or Hao et al. (1998)); for theoretical analyses, see Nabarro (1967) and Hirth and Lothe (1982).

4.1 Edge and Screw Dislocations, Burgers Vector

4.1.1 Edge and Screw Dislocations

Edge dislocations (Taylor 1934; Orowan 1934) are formed by an extra lattice plane inserted into a part of a crystal as illustrated in Fig. 13a. The *dislocation line* extends along the terminating edge of the inserted plane. The edge dislocation is called *positive*, if the extra half plane is added *above* the dislocation line, and negative for an insertion below. Such dislocations are produced when a crystal is exposed to nonuniform mechanical stress large enough to result in plastic deformation along a glide plane or when two semiconductors with different lattice constants form a common interface. In Fig. 13a, an added {001} plane is indicated. Another preferred glide plane in cubic crystals is the {110} plane.

Dislocations are identified by the dislocation line and their Burgers vector (see below). For an edge dislocation, the Burgers vector **b** is at an angle of 90° to the dislocation line **l** (Fig. 13a).

A *screw dislocation* (Burgers 1939 – Fig. 13b) shows a step at the outer surface, while only a slight lattice deformation exists surrounding the dislocation within the crystal. The screw dislocation is important for crystal growth (Frank 1949a) since adherence of atoms is substantially enhanced at an inside surface edge (see Sect. 3.1.5 and Fig. 8 of ► Chap. 1, "Properties and Growth of Semiconductors"). A screw dislocation continuously maintains such an edge during growth by forming a

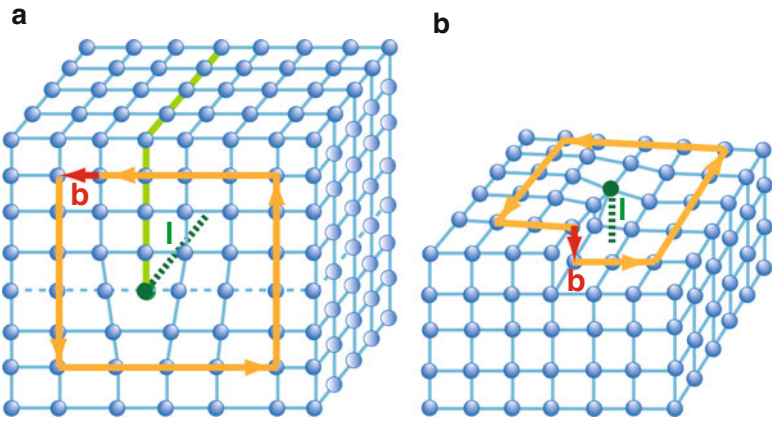


Fig. 13 (**a**) Edge dislocation with an additional plane (marked in *green*) partially inserted into a simple cubic crystal, ending at the dislocation line **l**. The glide plane is indicated by a *dashed line*. (**b**) Screw dislocation in a simple cubic crystal producing a step at the crystal surfaces. The Burgers vector **b** is identified by a *red arrow*. *Orange arrows* mark a Burgers circuit about the dislocation line

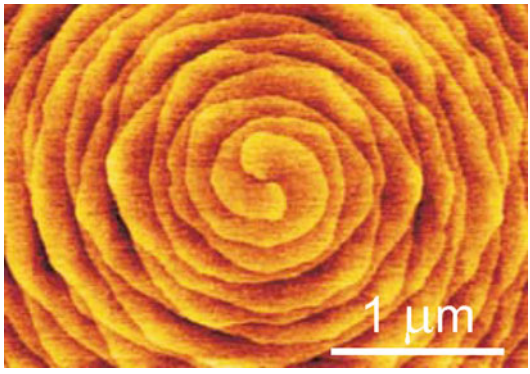


Fig. 14 Spiral growth shown on the (0001) surface of a GaN crystal with the core of a screw dislocation in the center of the picture (After Akasaka and Yamamoto 2014)

growth spiral (Fig. 14). For a screw dislocation, the Burgers vector **b** is parallel to the dislocation line **l** (Fig. 13b).

A *mixed dislocation* is formed when the Burgers vector is at an angle $0 < \theta < 90^\circ$ to the dislocation line. Such a dislocation has a step and a screw component; an example is shown in Fig. 15. Mixed dislocations are far more common than pure edge and screw dislocations; they are often termed after the angle between Burgers vector **b** and dislocation line **l**.

4.1.2 The Burgers Vector

The *Burgers vector* and its angle to the dislocation line identify the dislocation, and its square yields the strain energy of the dislocation per unit length. In order to

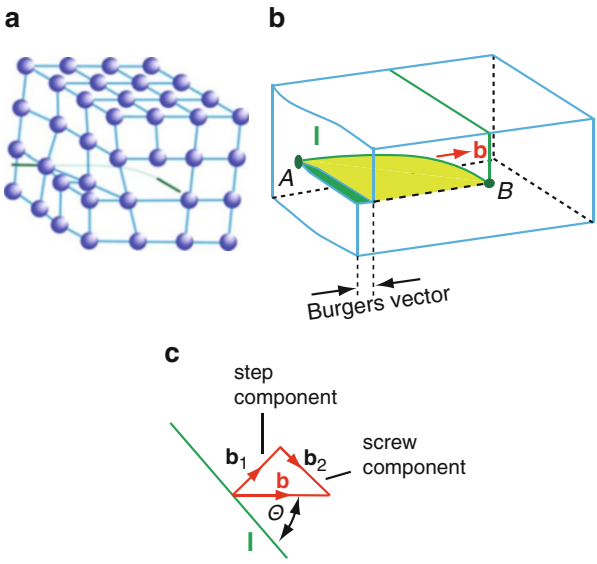


Fig. 15 Mixed dislocation, with bend dislocation line **l**. (**a**) Lattice model, (**b**) varying angel between **b** and **l**, yielding pure screw at *A* and pure edge at *B*. (**c**) The Burgers vector **b** involves an edge ($\mathbf{b}_1$) and a screw ($\mathbf{b}_2$) component

define the Burgers vector, we construct a polygon with an equal number of lattice steps on each side of the polygon.[15] The path is closed in a perfect crystal. However, when this path surrounds a dislocation, it is no longer closed. The lattice vector needed for completion of the polygon is the Burgers vector **b** (see Fig. 13).

For a pure edge dislocation, the Burgers vector is orthogonal to the dislocation line, and for a pure screw dislocation, it is parallel. If the dislocation line is a straight line, the Burgers vector is constant, and the type of dislocation does not change. When the dislocation line bends, the step and screw components of such a mixed dislocation continuously change (see Fig. 15).

A dislocation line can neither begin nor end within a crystal (Hirth and Lothe 1982). Consequently, the dislocation line forms a closed loop within the crystal, or it begins and ends at an interface (surface) or grain boundary of the crystal.

The displacement of atoms near the dislocation line is connected to a strain energy per unit length. The strain energy increases quadratically with strain in the linear stress-strain relation of Hooke's law, $E_{\text{dislocation}}/\text{length} \propto Gb^2$, where G is the shear modulus; the proportionality constant varies only little for different types of dislocations (less than a factor of two). To minimize strain in the crystal, a dislocation line can move as outlined in Sect. 4.3, or it can split: the strain energy

[15]When a mirror-symmetry plane exists normal to the dislocation line, an arbitrariness in the sense of this line cannot be avoided. There are hence different conventions to define the sign of the dislocation line and the Burgers circuit (clockwise or reverse), yielding different signs for the Burgers vector (see Hirth and Lothe 1982). For a finish-start/left-hand (FS/LH) convention with a counterclockwise circuit, **l** pointing from surface to bulk and **b** drawn from the finish to the start point to close the circuit, the screw dislocation in Fig. 13b is defined left-handed.

of a dislocation with Burgers vector **b** is lowered, if the dislocation is divided into two (or more) partial dislocations with smaller Burgers vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ fulfilling $\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2$ and *Frank's rule*:

$$|\mathbf{b}_1|^2 + |\mathbf{b}_2|^2 < |\mathbf{b}|^2. \tag{85}$$

According to Frank's rule, dislocations with shortest Burgers vector are stable. If the Burgers vector is a translation vector of the crystal, the corresponding dislocation is termed *perfect dislocation*. In an fcc structure and the related zincblende structure, primitive translations represent vectors of the type $\frac{a_0}{2}\langle 110\rangle$. Burgers vectors which do *not* correspond to a primitive translation refer to *partial dislocations*. In the fcc structure, a partial dislocation leads to a change of the *ABC* stacking order along the {111} direction, resulting in a planar stacking fault (see Sect. 4.2). An example for a dissociation of a perfect dislocation into two more favorable partial dislocations according to Frank's rule is the reaction

$$\frac{a_0}{2}\left[\bar{1}01\right] \rightarrow \frac{a_0}{6}\left[\bar{2}11\right] + \frac{a_0}{6}\left[\bar{1}\,\bar{1}2\right]. \tag{86}$$

The two created partials are Shockley dislocations with Burgers vectors of length $a_0/\sqrt{6}$, while the perfect dislocation has $a_0/\sqrt{2}$ in length, thus obeying Eq. 85.

4.1.3 Dislocation Imaging and Counting

Dislocations can be made visible by several methods (Hull 1975). One of the easiest is via surface etching, i.e., removing parts of the lattice surrounding the dislocation, which are under stress and therefore dissolve more easily. The surface etching method is shown schematically in Fig. 16. Since most of the dislocations extend throughout the entire crystal, we obtain their density by counting the *etch pits* at the surface. The best Si crystals contain less than 1 etch pit/cm^2. Typical commercial semiconductors have a dislocation density of 10^2–10^4 etch pits/cm^2. Direct imaging of dislocations and the related strain distribution is achieved by diffraction in scanning-electron microscopy: back-scatter electron diffraction (BSED) and electron channelling contrast imaging (ECCI), see Wilkinson (2000), Schulze (2018). Other methods for direct observation of dislocations include transmission electron

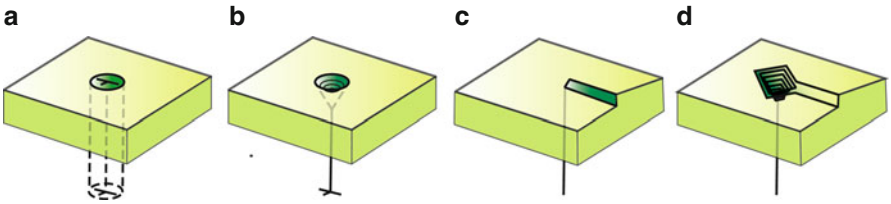


Fig. 16 Formation of etch pits at the surface surrounding an edge (**a**, **b**) and a screw dislocation (**c**, **d**)

microscopy (TEM) of thin layers; EBIC,[16] relating to their electronic properties; reflection or transmission imaging, relating to mechanical stress of the surrounding. Three-dimensional imaging is achieved by TEM-based dislocation tomography (Feng et al. 2020). In addition, X-ray diffraction and X-ray tomography are deployed. Indirect observation is possible by using decoration, which uses precipitates along dislocation lines to increase contrast and also to permit light microscopic observation. For recombination-generation behavior of decorated defects in Si, see Berg et al. (1992).

4.2 Dislocations in Compound Semiconductors

In a binary (*AB*) compound with *zincblende structure*, edge dislocations in the [110] direction are energetically favored but require *two* extra (110) planes: an *a* plane and a *b* plane as shown in Fig. 17a. Again, the Burgers vector is a lattice vector. There is the possibility of splitting these adjacent planes into two separate partial dislocations, as shown in Fig. 17b. The separation of planes maintains the correct *AB* sequence above and below the dislocation line, except for a jump of composition between both partial dislocations. Such partial dislocations have a lower energy than the unit edge dislocation and therefore are favored. In more complex crystals, more than two extra planes are required to restore periodic order. Therefore, the larger the Burgers vector becomes, the more complex the crystal structure is.

In *wurtzite AB* compounds, common Burgers vectors are of the type $\frac{a}{3}\langle 11\bar{2}0\rangle$ with slip (see Sect. 4.3) in the basal plane (0001), but slip on $\{1\bar{1}00\}$ and $\{11\bar{2}0\}$ also occurs. For perfect dislocations in wurtzite lattices, see Osipiyan and Smirnova (1968).

The preferred slip plane in tetragonally bonded semiconductors is the {111} plane which could lie either between the closely spaced planes, called *glide set*, or between the wider spaced planes, called *shuffle set* (*Ba* or *bB* in Fig. 18). These are 60° dislocations; they are prominent in tetrahedrally bonded semiconductors and can best be visualized by cutting out a lattice slab and rejoining the displaced atoms along the dashed lines 1-5-6-4 for the glide set or along 1-2-3-4 for the shuffle set. With such an operation, an extra lattice plane is inserted below 5-6. For more information, see Bauer et al. (1993), Branchu et al. (1999), Inoue et al. (1998), and Brochard et al. (1998).

Along such dislocations, a row of dangling bonds would appear, as shown in a perspective view in Fig. 19a.[17] Dangling bonds are expected to effectively trap electrons with major influence on electrical properties (Labusch and Schröter

[16]Electron beam-induced conductivity is used in a scanning electron microscope (Heydenreich et al. 1981).

[17]The 60° dislocation in Fig. 19a is drawn for zincblende structure; to obtain this defect for the diamond structure of silicon, consider *all* atoms to be identical.

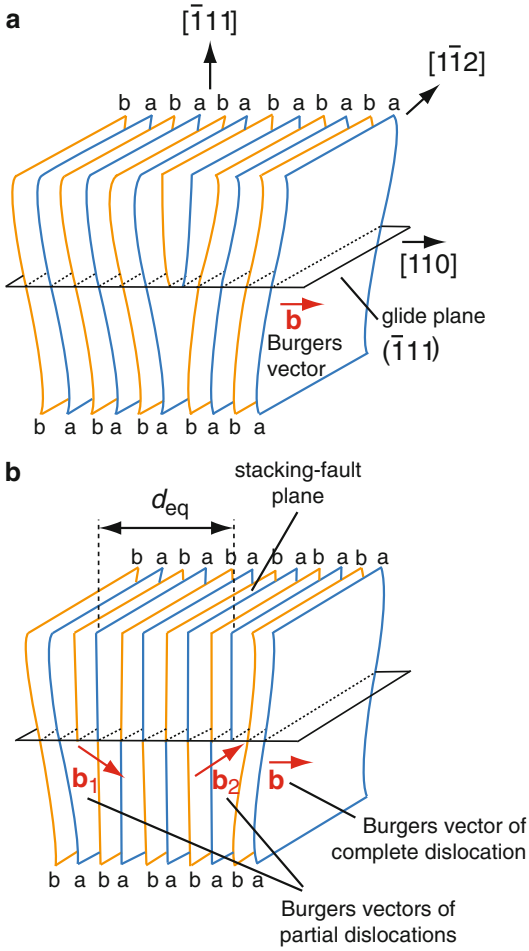


Fig. 17 (**a**) Unit edge dislocation in an *AB* compound with Burgers vector along the [110] direction of a face-centered cubic crystal. (**b**) Extended dislocation of the same compound of two Shockley partial dislocations separated by a stacking fault (plane *b* continues as plane *a* and plane *a* as plane *b* above the partial dislocation plane)

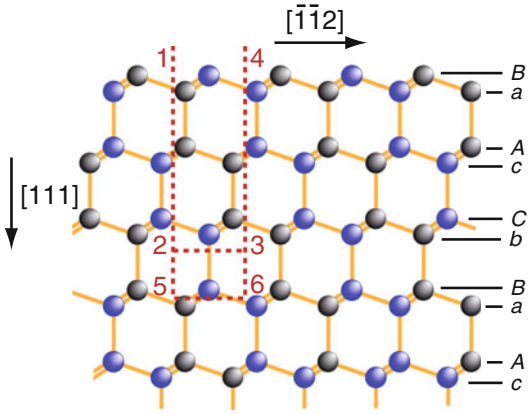


Fig. 18 Projection of a Si lattice in the $(1\bar{1}0)$ plane. *Blue circles* represent atoms in the paper plane; *dark circles* lie in the next plane below or above, as indicated also by two bonds. The (111) plane is normal to the paper (also normal to the direction [111]) and would appear as a horizontal trace. The *dashed red lines* indicate a cutout of a lattice slab for creating a 60° dislocation just below it

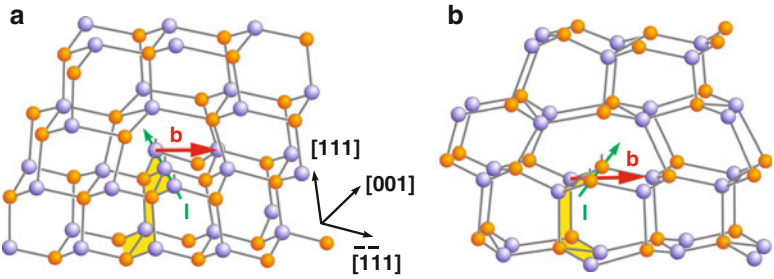


Fig. 19 60° dislocation in a (**a**) zincblende and a (**b**) wurtzite crystal with a row of dangling bonds at the dislocation line **l** (*green*). The Burgers vector is drawn in *red*, and the inserted extra plane is marked in *yellow*

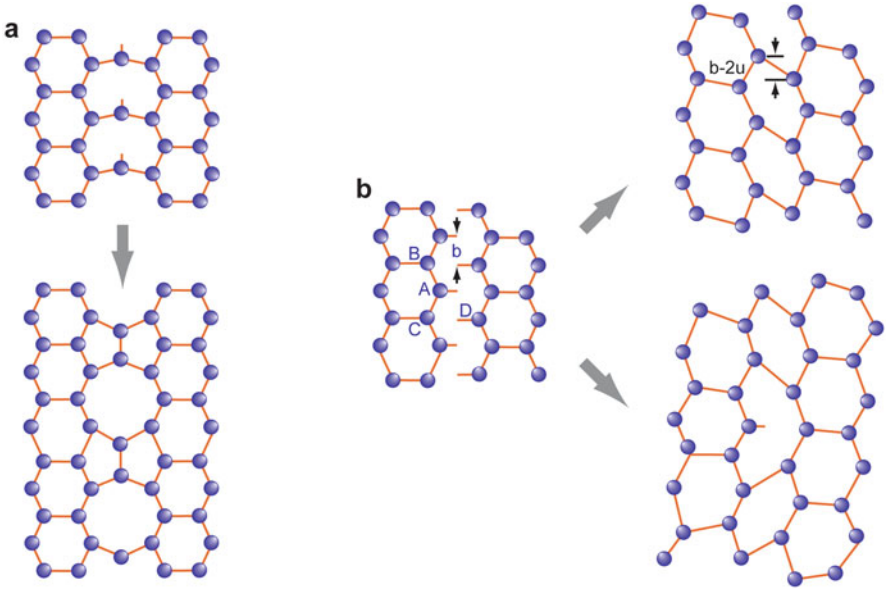


Fig. 20 (**a**) Top, 30° glide partial dislocation in Si shown in a (111) plane; *bottom*, reconstructed. (**b**) 90° glide partial dislocation in Si in a (111) plane; *top right*, reconstructed; *bottom right*, antiphase defect (After Jones 1981)

1980). However, reconstruction of the dislocation core eliminates most of these dangling bonds with substantial reduction of the electrical influence of such dislocations (Hirsch 1985). For imaging of such dislocations, see Ning and Huvey (1996) and Gutakovskii et al. (1995).

The 60° dislocation consists of two partials (a 30° and a 90° partial), each of which is capable of reconstruction. Figure 20a shows the reconstruction of the 30° glide partial; whereas Fig. 20b shows dangling bonds along a 90° glide partial and

its reconstructed core. The 90° glide partial shows an interesting alternative of bonding (upper and lower middle part of the structure shown at the bottom right in Fig. 20b) which has the same energy. In the transition region, a dangling bond is created. Such a defect is also referred to as an *antiphase defect* and can move as a soliton along the reconstructed glide partial (Heggie and Jones 1983; for more details, see Hirsch 1985).

4.3 Motion and Creation of Dislocations

All dislocations can move conservatively if the motion is parallel to the Burgers vector or, more generally, when the dislocation moves within its slip plane. Such a motion is called *glide*,[18] shown in a two-dimensional representation in Fig. 21, and is the essential part in a plastic deformation of a crystal.

The motion of an edge dislocation normal to both its Burgers vector and its dislocation line is called *climb*. It is nonconservative; meaning, it causes a change of the height of the inserted plane, i.e., atoms must move to or from the dislocation core (dislocation line) (see Fig. 22a). A gliding dislocation that encounters an obstacle – such as another crossing dislocation, a vacancy, a foreign atom, or other crystal defects – can become pinned at such an obstacle, i.e., the obstacle hinders dislocation movement. For crossing dislocations, see Justo et al. (1997). For the example given in Fig. 22b, the dislocation must "climb over" the obstacle in order to continue moving. Therefore, glide and climb are important in plastic deformation of crystals (see Cottrell 1958 or Hirth and Lothe 1982).

4.3.1 Dislocation Kinks and Jogs

Dislocation lines in real crystals are not straight lines; they contain short segments, where the dislocation line is displaced on an atomic scale. A *kink* is formed when part of the dislocation is shifted within the same glide plane by one lattice plane from its original position (Fig. 23a, b). At these kinks, dangling bonds exist where reconstruction cannot be completed (Hirsch 1985). A kink does not impede glide of the dislocation.

If the dislocation line moves from one atomic slip plane to another, a *jog* is formed (Fig. 23c, d). Jogs are segments of the dislocation line that have a component normal to the glide plane. If this segment extends over more than one interplanar spacing, it is termed *superjog*. The jog on a screw dislocation has an edge component; it can only glide in the plane containing the dislocation line before and after the jog (paper plane in Fig. 23c).

Both kinks and jogs can nucleate in pairs on an initially straight dislocation line (Fig. 24). Such pairs (of opposite sign) are formed by thermal fluctuations in the crystal and can initiate movement of the dislocation line.

[18]The terms *glide* and *slip* are generally used to describe, respectively, the motion of a single dislocation and many dislocations.

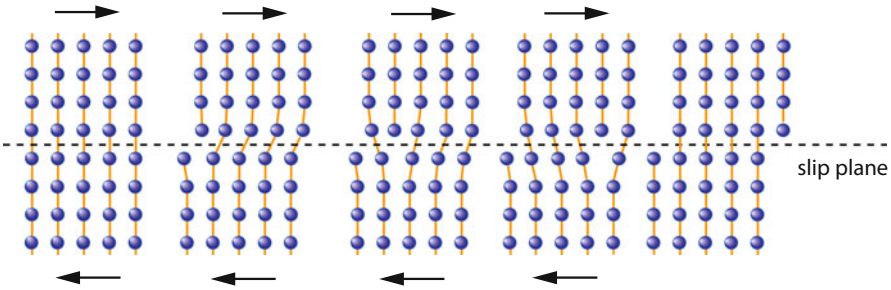


Fig. 21 Glide motion of an edge dislocation along a slip plane. *Arrows* indicate stress exerted on the crystal

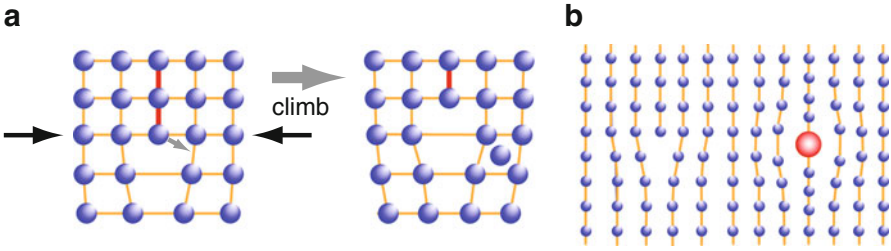


Fig. 22 (**a**) Edge dislocation performing a climb process and releasing atoms thereby. (**b**) Strain field surrounding a foreign atoms hinders motion of a dislocation along a glide plane

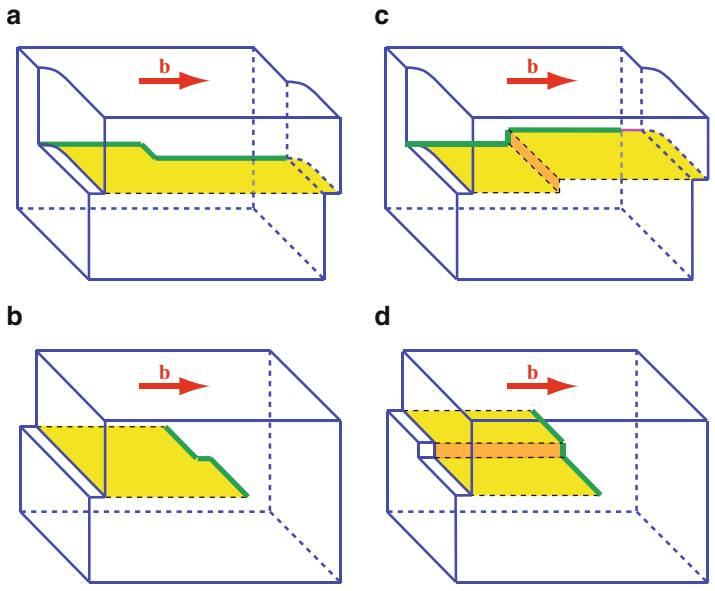


Fig. 23 Dislocation line with a kink (**a**, **b**) and a jog segment (**c**, **d**) on a screw (**a**, **c**) and an edge dislocation (**b**, **d**). The glide plane is marked in *yellow*

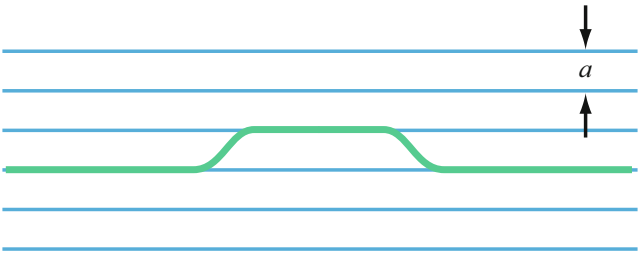


Fig. 24 Top view on a glide plane: movement of a dislocation line **l** in the glide plane by nucleation of a pair of kinks

Dislocation Climb If the last line of atoms along an edge dislocation is removed, e.g., by diffusion to the crystal surface, then the dislocation has climbed by one atomic spacing. A *climb* is a motion out of the glide plane. However, the diffusion of only a few atoms from the edge is more likely. This results in a climb of only a fraction of the dislocation, with a jog to the undisturbed part. Climb by nucleation of jog pairs is equivalent to dislocation motion in perpendicular direction by nucleation of kink pairs. Widening of the distance between jog pairs usually requires diffusion of the interspacing line of atoms to the surface.

4.3.2 Dislocation Velocity

When a shear stress is exerted parallel to the Burgers vector of an edge dislocation, it is unlikely that the entire dislocation moves step by step as shown in Fig. 21. When drawing the edge dislocation in a plane normal to that shown in Fig. 21, we can picture a more probable sequence of events, in which part of the dislocation moves one lattice spacing and connects to the remaining part with kinks (Fig. 24).

The dislocation velocity is then determined by the nucleation rate and motion of double kinks, as well as by the interaction of such kinks with localized lattice defects. Such an interaction is known to depend on the electrical charge of these defects and is influenced by the position of the Fermi (or quasi-Fermi) level (Patel and Chaudhuri 1966). For instance, the velocity of dislocations under otherwise identical conditions is ~50 times larger in *n*-type than in *p*-type Ge. Other deep-level centers, e.g., oxygen, cause pinning of the dislocation. This dislocation motion is more complicated when partial dislocations need to be considered. Generation of double kinks in the two partials may be correlated (Wessel and Alexander 1977). The velocity v of dislocations can generally be described by the empirical relation

$$v = v_0(\sigma/\sigma_0)^m \exp(\Delta E/(kT)), \tag{87}$$

where σ is the applied stress, T is the temperature, and the activation energy ΔE and characteristic parameters v_0 and σ_0 are experimentally determined. Some partials (e.g., with a Burgers vector inclined to a stacking fault) cannot glide; these are called *sessile* (Frank 1949b). Other dislocations which can move easily are called *glissile*.

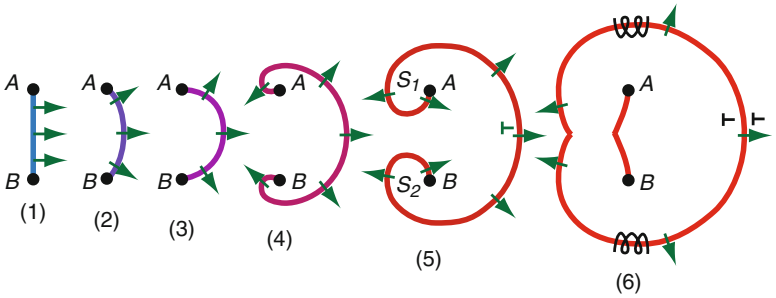


Fig. 25 (1) Frank-Read source with progressively growing dislocation loop (2–5) and separation (6) to repeat this process (1–6)

Edge dislocations in semiconductors are generally at least partially charged, since in addition to some remaining dangling bonds, the stress field surrounding the dislocation creates traps for carriers, and the termination of an extra layer in an ionic crystal carries an inherent charge (Petrenko and Whitworth 1980). The movement of such a dislocation is therefore influenced by other charges.

The Frank-Read Source The application of sufficient shear stress produces additional dislocations from a dislocation pinned at the two ends (*Frank-Read source*, Frank and Read 1950) and forced to bow by mechanical deformation perpendicular to the direction of the dislocation (Fig. 25). In this way, *dislocation loops* can be formed sequentially, and the process repeats itself periodically as long as the stress persists (see Read 1953). For observation of dislocation loops in Si-doped GaAs, see Chen et al. (1992).

4.3.3 Dislocations and Electronic Defect Levels

The dangling-bond states introduced by dislocations are electronic defect states within the bandgap. Nonreconstructed dislocations result in a half-filled defect band in the bandgap. Their presence affects the carrier density; for instance, plastic deformation on *n*-type Ge can render the material *p*-type (Labusch and Schröter 1980).

Core reconstruction reduces the density of dangling bonds dramatically so that the remaining ones act as isolated deep-level point defects. In addition, the lattice deformation near the reconstructed core causes defect levels in the bandgap.

Electron paramagnetic resonance and its angular dependence yield information about the density and orientation of defects with unpaired spins, i.e., dangling bonds. Experiments with plastically deformed Si at 650 °C indicate that only 0.2 ... 2 % of the available sites are not reconstructed and have unpaired spins at dangling bond centers (Weber and Alexander 1983). Kinks, curved dislocations, or other special features, such as jogs or nodes, are the suspected sites of the remaining dangling bonds (Osip'yan 1983).

Deep-level transient spectroscopy (DLTS) indicates that filling of deep levels at dislocations, which lie in closer proximity to each other, produces a Coulomb barrier. This impedes consequent trapping of adjacent centers and can be recognized in the DLTS signal (Kveder et al. 1982). Acceptor levels at $E_c - E_a \cong 0.35$ and 0.54 eV, and a donor level at $E_d - E_v \cong 0.4$ eV, have been detected in plastically deformed Si (Kimerling and Patel 1979; Weber and Alexander 1983).

Luminescence and photoconductivity measurements yield additional information about deep levels associated with (mostly reconstructed) dislocations, kinks, jogs, etc. For reviews on Si defects, see Weber and Alexander (1983) and Mergel and Labusch (1982), for SiC see Galeckas et al. (2006), Kawahara et al. (2014), for GaN see Yao et al. (2017).

The unique identification of specific dislocation-related defects is difficult: with deformation at elevated temperatures, other defects and defect associates are formed, the signatures of which cannot easily be separated.

5 Planar Defects

Planar faults are *two-dimensional* defects. There are several types of such faults:

- Stacking faults
- Low-angle grain boundaries, twin boundaries, and crystallite boundaries
- Inversion-domain boundaries
- Heterointerfaces and metal/semiconductor interfaces
- Surfaces (ideal and with absorption or chemisorption layers)

These defects have a significant influence on the electrical properties of real semiconductors. Most of this influence is caused by defect levels at the boundaries (interfaces, surfaces), which are usually charged and attract compensating charges in the adjacent crystal volume (*space charges*). Heterointerfaces between two semiconductors and metal/semiconductor interfaces are treated in ▶ Chap. 16, "Crystal Interfaces". The other planar defects will only be mentioned briefly here since they deal with profound inhomogeneities which are not the topic of this book.

5.1 Stacking Faults and Antiphase Domains

Stacking Faults Stacking faults are the least disordered types of interface defects. They occur by alternating between wurtzite and zincblende structures: to form the lattice from atomic layers stacked on top of each other in the **c** direction,[19] we start from a hexagonal ("close-packed") layer (*A*) and deposit the next identical layer (*B*)

[19] Along [0001] in the hexagonal wurtzite lattice, corresponding to the [111] direction of the cubic zincblende lattice

offset and turned by 60° to touch each second interspace at $\frac{1}{4}\frac{1}{4}\frac{1}{4}$. For the deposition of a similarly offset third layer, we have the choice of turning it back by 60° to the original position (*A*), resulting in a wurtzite lattice, or turning it forward by another 60° to offset it from the bottom layer by 120° (*C*), resulting in a zincblende lattice (see Figs. 9 and 11 of ▶ Chap. 3, "The Structure of Semiconductors"). The layer sequence *ABABAB* therefore characterizes wurtzite; the layer sequence *ABCABC* characterizes zincblende. The sequences *ABABABCABAB* and *ABCABCBCABC* consequently identify *stacking faults* in the wurtzite and the zincblende structure, respectively. A thin slice of a stacking fault can be seen as being imbedded between two partial dislocations, i.e., the boundary of a stacking fault is a partial dislocation (see Fig. 26; for more details, see Nabarro 1967 and Hirth and Lothe 1982).
ZnS and SiC are examples for materials that show such stacking faults. The reason is a small difference of lattice energy between wurtzite and zincblende structure in these semiconductors (Yeh et al. 1992). There can be a long-range periodicity in these stacking faults, up to a few hundred Å in SiC, which produces so-called *polytypes*.

Antiphase Domains An antiphase domain is a region in a polar crystal where the atoms are located on regular lattice sites but in the opposite order with respect to the undisturbed crystal. In binary crystals with zincblende or wurtzite lattices, hence anions occupy cation sites and vice versa within the domains (see Fig. 27). Such a disorder nucleates particularly at the heterointerface of a polar semiconductor grown on a nonpolar semiconductor.[20] Here, antiphase domains are created during growth at monoatomic steps on the nonpolar semiconductor; if growth starts with an anion layer, growth on a one monolayer higher region yields an inequivalent crystal orientation. Such a mismatch can be avoided by double steps on the nonpolar semiconductor.[21] Boundaries between domains of reversed orientation correspond to stacking faults and are also termed *inversion-domain*

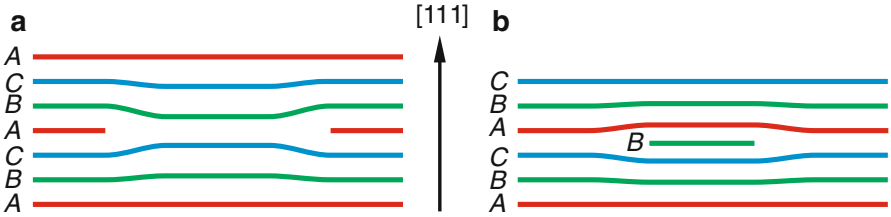


Fig. 26 (**a**) Intrinsic and (**b**) extrinsic stacking fault in a zincblende lattice with a removal or an insertion of a layer between two dislocations, respectively. *A*, *B*, and *C* denote hexagonal close-packed layers stacked along the [111] direction

[20]An example is the well-lattice-matched InP/Si(001), a prominent material for integrating optoelectronics of InP-based devices with the established Si electronics.

[21]On Si(001) substrates, such double steps can be achieved applying ∼6° offcut orientation to create single-step terraces and a thermal treatment for double-step formation.

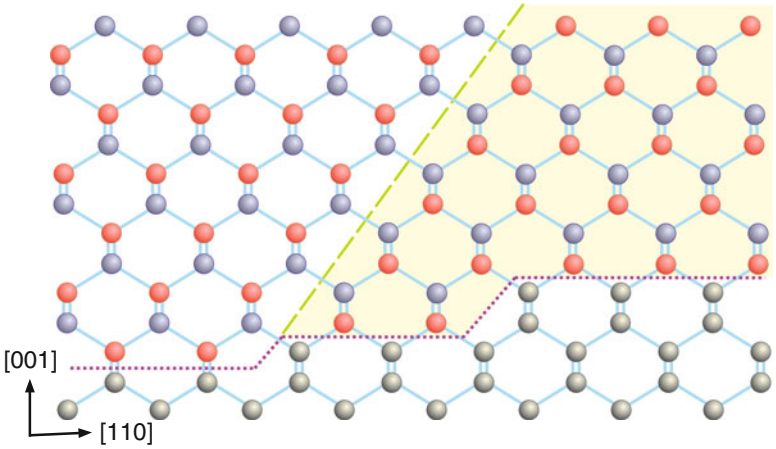


Fig. 27 Creation of an antiphase domain boundary at a monoatomic step at the interface of a nonpolar to a polar semiconductor. No such boundary forms at a double step. The crystal orientations in the non-shaded and shaded regions are inverse. The *green dashed line* marks an inversion-domain boundary

boundaries. At these boundaries, bonds among equally charged ions occur (see Fig. 27). Heterointerfaces are discussed in more detail in ▶ Chap. 16, "Crystal Interfaces".

Inversion-domain boundaries (IDBs) generally create states in the bandgap. They introduce nonradiative carrier recombination and degrade the performance of electronic and optoelectronic devices. IDBs which are inclined to the interface may lead to a restoring of the undisturbed order: a further monoatomic step in the right part of Fig. 27 induces a second IDB which can be inclined to the left; at the line of intersection of the two IDB planes, the disorder annihilates, yielding an antiphase *domain* bounded by these planes.

5.2 Grain Boundaries

One distinguishes several types of grain boundaries, the more important ones will be discussed in this section.

Low-angle Grain Boundaries Low-angle grain boundaries show a very small tilt between two regular crystallites and can be regarded as an array of dislocations (Bragg and Burgers 1940), as shown in Fig. 28. The angle between the two grains (at most a few degrees) is given by

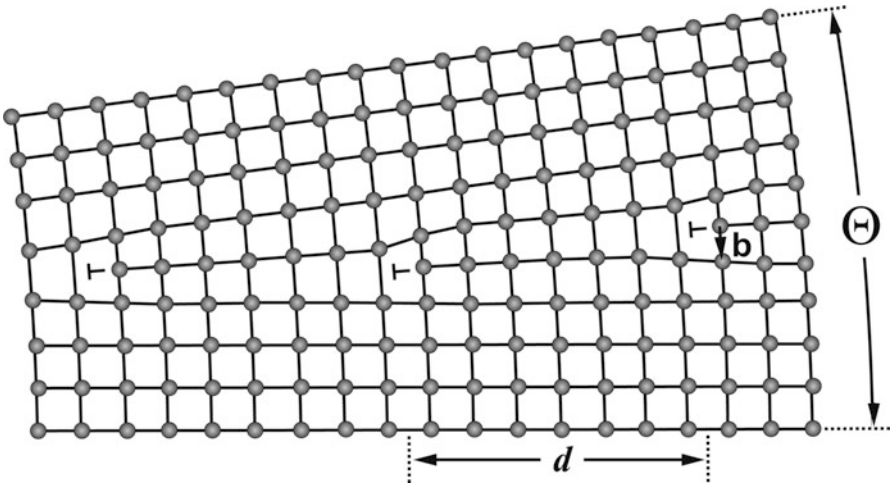


Fig. 28 Low-angle grain boundary with an array of edge dislocations separated by seven lattice planes. The symbol ⊥ indicates the insertion of an extra lattice plane

$$\theta \cong b/d, \tag{88}$$

where b is the length of the Burgers vector and d is the dislocation spacing.

Large-angle Grain Boundaries *Twin boundaries* are special angle boundaries between two identical crystallites. The least disturbed twin boundary is that of a stacking fault, as discussed in Sect. 5.1: in a *zincblende or diamond* lattice, the stacking order may reverse at a single fault yielding *ABCABCBACBA*; the sequence in a *wurtzite* lattice may proceed from an *AB* sequence to a *BC* sequence yielding *ABABCBCBC*. In either case, there is a single plane of atoms – the twinning plane – shared by two undisturbed crystals with mutual reflection symmetry.

Other twin boundaries are under an angle at which solely each second (or third) atom falls onto a lattice site as shown in Fig. 29. It can therefore be regarded as an array of vacancy lines similar to an array of edge dislocations in low-angle grain boundaries.

Other crystallite boundaries may occur under a wide variety of angles and may incorporate at the interface a variety of disorders, including vacancies, dislocations, and liquid-like structures with a high degree of local stress (Fig. 30).

Grain boundaries can have a major influence on semiconducting properties by trapping carriers and creating compensating space-charge layers. Some of the effects related to carrier mobility will be discussed in Sect. 4.3 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields".

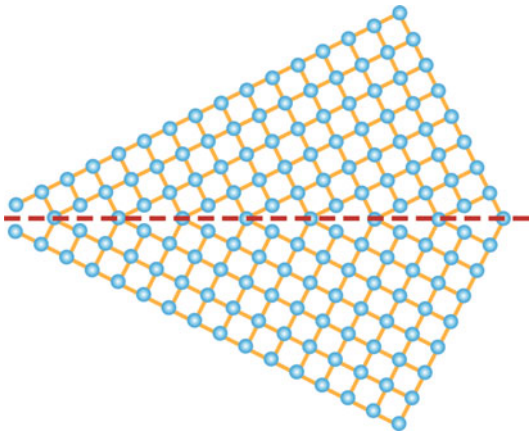


Fig. 29 Twin-grain boundary (*dashed line*) in the (112) surface of a cubic crystal

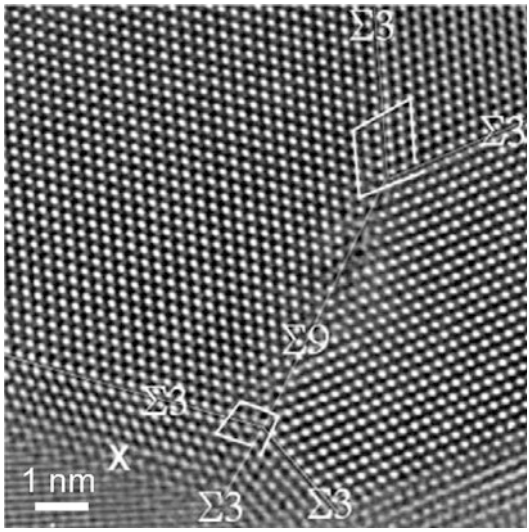


Fig. 30 Grain boundaries of nanocrystalline palladium (fcc structure) taken with high-resolution transmission electron microscopy along [110], making the atomic array visible. Burgers circuits at the triple and quadruple points show a projected translation of $\frac{a_0}{4}\langle 112\rangle$, representing the edge part of 60° mixed dislocations with $\frac{a_0}{2}\langle 110\rangle$ Burgers vector (Rösner et al. 2011)

6 Summary

Crystal defects are classified into point defects, line defects, and planar defects. Intrinsic *point defects*, and associates of these defects, are formed at elevated temperatures in thermodynamic equilibrium. The density of vacancies and interstitials increases following the Arrhenius law with an activation energy for typical semiconductors in the 2 . . . 5 eV range and with a very large pre-exponential factor indicating substantial entropy contribution. With decreasing temperature, one reaches a freezing-in temperature below which the annealing of these defects can no longer follow any reasonable cooling rate, leaving a residual density of these

intrinsic defects which are frozen-in; almost all electronic devices contain inhomogeneous doping distributions which must remain frozen-in during the life of the device. The thermodynamic approach of estimating the density of point defects can be extended to a defect-chemistry approach in which various interactions can be taken into account. Such interactions include reactions between extrinsic and intrinsic defects, associate formation, ionization of these defects, and interrelation with free carriers. All of these force an interrelation through conservation of particles and quasi-neutrality.

Defect incorporation often involves diffusion from or to outer surfaces or between different defects. Such a diffusion is described by a random walk of atoms between neighboring sites; it may also involve more complicated site exchanges of atoms during each step, including recharging of defects during alternating steps, and associate formation. The diffusion is measured by a diffusion constant that has matrix form in anisotropic crystals and can have vastly different magnitude dependent on activation energies, temperature, and crystal structure.

Line defects are created during crystal growth (epitaxy) or under external stress. Most important are dislocations, which are characterized by their Burgers vector and its angle to the dislocation line. Mixed dislocations comprise edge and screw components, and any bending of the dislocation line changes their fraction. Perfect dislocations may dissociate into partial dislocations with smaller Burgers vectors. Dislocations move conservatively on glide planes and nonconservatively by climb processes. The motion may commence by thermally activated pairs of kinks and jogs.

Planar defects are two-dimensional faults representing extended inhomogeneities in the crystal. A disturbed order of close-packed atom layers in tetrahedrally bonded semiconductors leads to stacking faults or, by local variations of the stacking order in compound semiconductors, to antiphase boundaries. A single stacking-fault plane may constitute a large-angle twin boundary separating perfect crystals of different orientations. A low-angle grain boundary is formed between two regular crystallites by a series of dislocations.

Defects are present in very small concentrations in any semiconductor device, where they determine the electronic and optical properties. Materials, described by chemists as very pure, often contain too many impurities for semiconductor devices to work properly. Purification with respect to certain impurities to a concentration of less than one atom in 10^8 is sometimes required, corresponding to 99.999999 % purity (8-niner material), while 6-niner semiconducting grade is common. The control altogether of point, line, and planar defects, the identification of desirable defects which can be introduced in appropriate concentrations, and the recognition and avoidance (or electrical neutralization) of undesirable defects are essential.

References

Akasaka T, Yamamoto H (2014) Nucleus and spiral growth mechanisms of nitride semiconductors in metalorganic vapor phase epitaxy. Jpn J Appl Phys 53:100201

Ammerlaan CAJ, Watkins GD (1972) Electron-paramagnetic-resonance detection of optically induced divacancy alignment in silicon. Phys Rev B5:3988
Ayers JE (2007) Heteroepitaxy of semiconductors. CRC Press Taylor & Francis, Boca Raton
Bauer S, Rosenauer A, Link P, Kuhn W, Zweck J, Gebhardt W (1993) Misfit dislocations in epitaxial ZnTe/GaAs (001) studied by HRTEM. Ultramicroscopy 51:221
Berding MA, Sher A, Chen A-B (1990) Vacancy formation and extraction energies in semiconductor compounds and alloys. J Appl Phys 68:5064
Berg A, Brough I, Evans JH, Lorimer G, Peaker AR (1992) Recombination-generation behaviour of decorated defects in silicon. Semicond Sci Technol 7:A263
Bourgoin J, Corbett JW (1972) A new mechanism for interstitial migration. Phys Lett A 38:135
Bourgoin J, Lannoo M (1983) Point defects in semiconductors II: experimental aspects. Springer, Berlin
Bracht H (2000) Diffusion mechanisms and intrinsic point-defect properties in silicon. Mater Res Soc Bull 25:22
Bracht H, Brotzmann S (2006) Atomic transport in germanium and the mechanism of arsenic diffusion. Mater Sci Semicon Process 9:471
Bragg WL, Burgers WG (1940) Slip in single crystals: discussion. Proc Phys Soc London 52:54
Branchu S, Pailloux F, Garem H, Rabier J, Demenet JL (1999) Partial dislocation source in InSb: a new mechanism. Phys Stat Sol A 171:59
Brochard S, Rabier J, Grilhé J (1998) Nucleation of partial dislocations from a surface-step in semiconductors: a first approach of the mobility effect. Eur Phys J Appl Phys 2:99
Brooks H (1963) Binding in metals. Trans Metall Soc AIME 227:546
Brouwer G (1954) A general asymmetric solution of reaction equations common in solid state chemistry. Philips Res Rep 9:366
Burgers JM (1939) Betrachtungen über die auf Grund von Verlagerungen im regulären Krystallgitter auftretenden Spannungsfelder. I. Untersuchung der geometrischen Beziehungen bei den Verschiebungen in einfachen Krystallen unter dem Einfluss von Spannungen. Proc Kon Ned Acad Wetenschap. 42:293; II. Lösungen der Elastizitätsgleichungen für anisotrope Substanzen mit regulärer Symmetrie. ibid 42:378 (Consideration of stress fields due to shifts in a regular crystal lattice; I Investigation on the geometric relation of displacements in simple crystals under the influence of stress; II Solutions of elasticity equations for anisotropic matter with regular symmetry; in German)
Car R, Kelly PJ, Oshiyama A, Pantelides ST (1984) Microscopic theory of atomic diffusion mechanisms in silicon. Phys Rev Lett 52:1814
Car R, Kelly PJ, Oshiyama A, Pantelides ST (1985) Microscopic theory of impurity-defect reactions and impurity diffusion in silicon. Phys Rev Lett 54:360
Chen TP, Chen LJ, Huang TS, Guo YD (1992) Transmission electron microscope investigation of dislocation loops in Si-doped GaAs crystals. Semicond Sci Technol 7:A300
Chen N, Gray S, Hernandez-Rivera E, Huang D, LeVan PD, Gao F (2017) Computational simulation of threshold displacement energies of GaAs. J Mater Res 32:1555
Cottrell AH (1958) Dislocations and plastic flow in crystals. Oxford University Press, London
Cottrell AH (1964) Theory of crystal dislocations. Gordon & Breach, New York
de Kock AJR (1980) Crystal growth of bulk crystals: purification, doping and defects. In: Moss TS, Keller SP (eds) Handbook of semiconductors, vol 3. North-Holland, Amsterdam, pp 247–333
Drabold DA, Estreicher SK (2007) Theory of defects in semiconductors. Springer, Berlin/Heidelberg
Feng Z, Fu R, Lin C, Wu G, Huang T, Zhang L, Huang X (2020) TEM-based dislocation tomography: challenges and opportunities. Curr Opin Solid State Mater Sci 24:100833
Flynn CP (1972) Point defects and diffusion. Claredon Press, Oxford
Frank FC (1949a) The influence of dislocations on crystal growth. Discuss Faraday Soc 5:48
Frank FC (1949b) Sessile dislocations. Proc Phys Soc A 62:202

Frank FC, Read WT Jr (1950) Multiplication processes for slow moving dislocations. Phys Rev 79:722
Frank W (1981) Self-interstitials and vacancies in elemental semiconductors between absolute zero and the temperature of melting. In Treusch J (ed), Festkörperprobleme/Advances in Solid State Physics, vol 26. Vieweg, Braunschweig pp 221–242
Frenkel JI (1926) Über die Wärmebewegung in festen und flüssigen Körpern. Z Phys 35:652 (On the thermal motion in solids and liquids, in German)
Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, Janotti A, Van de Walle CG (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86:253
Freysoldt C, Lange B, Neugebauer J, Yan Q, Lyons JL, Janotti A, Van de Walle CG (2016) Electron and chemical reservoir corrections for point-defect formation energies. Phys Rev B 93:165206
Friedel J (1964) Dislocations. Addison-Wesley, Reading
Friedel J (1966) Theory of crystal defects. In: Gruber B (ed) Proc summer school Hrazany, Czechoslovakia. Academic Press, New York, p 415
Galeckas A, Hallén A, Majdi S, Linnros J, Pirouz P (2006) Combined photoluminescence-imaging and deep-level transient spectroscopy of recombination processes at stacking faults in 4H−SiC. Phys Rev B 74:233203
Gösele UM (1986) Point defects and diffusion mechanisms in crystalline semiconductors, vol 26, Festkörperprobleme/Advances in Solid State Physics. Vieweg, Braunschweig, p 89
Gösele UM (1988) Fast diffusion in semiconductors. Annu Rev Mater Sci 18:257
Gutakovskii AK, Fedina LI, Aseev AL (1995) High resolution electron microscopy of semiconductor interfaces. Phys Stat Sol A 150:127
Hagemark KI (1976) Frozen-in native defects in semiconductor compounds. J Chem Phys Sol 37:461
Hao M, Sugahara T, Sato H, Morishima Y, Naoi Y, Romano LT, Sakai S (1998) Study of threading dislocations in wurtzite GaN films grown on sapphire by metalorganic chemical vapor deposition. Jpn J Appl Phys 37:L291
Hayes W, Stoneham AM (1985) Defects and defect processes in nonmetallic solids. Wiley, New York
Heggie M, Jones R (1983) Microscopy of semiconducting materials. Inst Phys Conf Ser 67:45
Heydenreich J, Blumtritt H, Gleichmann R, Johansen H (1981) Combined application of SEM(EBIC) and TEM for the investigation of the electrical activity of crystal defects in silicon. In: Becker P, Johari O (eds) Scanning electron microscopy I. SEM Inc AMF O'Hare, Chicago, p 351
Hirsch PB (1985) Dislocations in semiconductors. In: Loretto MH (ed) Dislocations and properties of real materials. Institute of Metals, London, p 333
Hirth JP, Lothe J (1982) Theory of dislocations, 2nd edn. Wiley, New York
Hirth JP, Hirth G, Wang J (2019) Disclinations and disconnections in minerals and metals. PNAS 117:196
Holmström E, Kuronen A, Nordlund K (2008) Threshold defect production in silicon determined by density functional theory molecular dynamics simulations. Phys Rev B 78:045202
Hull D (1975) Introduction to dislocations. Pergamon Press, Oxford
Inoue M, Suzuki K, Amasuga H, Nakamura M, Mera Y, Takeuchi S, Maeda K (1998) Reliable image processing that can extract an atomically-resolved line shape of partial dislocations in semiconductors from plan-view high-resolution electron microscopic images. Ultramicroscopy 75:5
Ishikawa R, Mishra R, Lupini AR, Findlay SD, Taniguchi T, Pantelides ST, Pennycook SJ (2014) Direct observation of dopant atom diffusion in a bulk semiconductor crystal enhanced by a large size mismatch. Phys Rev Lett 113:155501
Jansen RW, Sankey OF (1989) Theory of relative native- and impurity-defect abundances in compound semiconductors and the factors that influence them. Phys Rev B 39:3192
Jones R (1981) Reconstructed dislocations in covalently bonded semiconductors. In: Cullis AG, Joy DC (eds) Microscopy of semiconducting materials, vol 60:45, Inst Phys Conf Ser. IOP Publishing, Bristol
Justo JF, Bulatov VV, Yip S (1997) Core effects in dislocation intersection. Scr Mater 36:707

Kang J, Li J, Wie S-H (2021) Atomic-scale understanding on the physics and control of intrinsic point defects in lead halide perovskites. Appl Phys Reviews 8:031302
Kawahara C, Suda J, Kimoto T (2014) Identification of dislocations in 4H-SiC epitaxial layers and substrates using photoluminescence imaging. Jpn J Appl Phys 53:020304
Kimerling LC, Patel JR (1979) Defect states associated with dislocations in silicon. Appl Phys Lett 34:73
Kleman M, Friedel J (2008) Disclinations, dislocations, and continuous defects: A reappraisal. Rev Mod Phys 80:61
Kröger FA (1964) The chemistry of imperfect crystals. North Holland Publ, Amsterdam
Kube R, Bracht H, Hüger E, Schmidt H, Hansen JL, Larsen AN, Ager III JW, Haller EE, Geue T, Stahn J (2013) Contributions of vacancies and self-interstitials to self-diffusion in silicon under thermal equilibrium and nonequilibrium conditions. Phys Rev B 88:085206
Kveder VV, Osipyan YA, Schröter W, Zoth G (1982) On the energy spectrum of dislocations in silicon. Phys Stat Sol A 72:701
Labusch R, Schröter W (1980) Electrical properties of dislocations in semiconductors. In: Nabarro FRN (ed) Dislocations in solids, vol 5. North Holland Publ, Amsterdam, pp 127–191
Landsberg PT, Canagaratna SG (1984) The grand partition function in defect statistics. Phys Stat Sol B 126:141
Lang DV, Grimmeiss HG, Meijer E, Jaros M (1980) Complex nature of gold-related deep levels in silicon. Phys Rev B 22:3917
Lannoo M, Bourgoin J (1981) Point defects in semiconductors. Springer, Berlin
Li SS (2007) Semiconductor physical electronics, 2nd edn. Plenum Press, New York
Lorenz MR (1967) Thermodynamics, materials preparation and crystal growth. In: Aven M, Prener JS (eds) Physics and chemistry of II–VI compounds. North Holland Publishing, Amsterdam, p 75
Lyons JL, Wickramaratne D, Van de Walle CG (2021) A first-principles understanding of point defects and impurities in GaN. J Appl Phys 129:111101
Mergel D, Labusch R (1982) Optical excitations of dislocation states in silicon. Phys Stat Sol A 69:151
Nabarro FRN (1967) Theory of crystal dislocations. Claredon Press, Oxford
Nichols CS, Van de Walle CG, Pantelides ST (1989) Mechanisms of equilibrium and nonequilibrium diffusion of dopants in silicon. Phys Rev Lett 62:1049
Ning XJ, Huvey N (1996) Observation of twins formed by gliding of successive surface-nucleated partial dislocations in silicon. Philos Mag Lett 74:241
Orowan E (1934) Zur Kristallplastizität III, Über den Mechanismus des Gleitvorganges. Z Phys 89:634 (On the plasticity of crystals III, On the mechanism of gliding, in German)
Osip'yan YA (1983) Electrical properties of dislocations in plastically deformed float zone silicon. J Phys Colloq (Orsay Fr) 44(C4, Suppl 9):103
Osipiyan YA, Smirnova IS (1968) Perfect dislocations in the wurtzite lattice. Phys Stat Sol 30:19
Pantelides ST (1987) The effect of hydrogen on shallow dopants in crystalline silicon. In: Engström O (ed) Proceedings 18th international conference on physics of semiconductors. World Scientific Publishing, Singapore, p 987
Patel JR, Chaudhuri AR (1966) Charged impurity effects on the deformation of dislocation-free germanium. Phys Rev 143:601
Petrenko VF, Whitworth RW (1980) Charged dislocations and the plastic deformation of II–VI compounds. Philos Mag A 41:681
Pichler P (2004) Intrinsic point defects, impurities, and their diffusion in silicon. Springer, Wien/New York
Pirouz P (1989) On twinning and polymorphic transformations in compound semiconductors. Scr Metall 23:401
Pohl UW (2020) Epitaxy of Semiconductors – Physics and fabrication of heterostructures, 2nd edn. Springer, Cham
Read WT Jr (1953) Dislocations in crystals. McGraw-Hill, New York

Rösner H, Kübel C, Ivanisenko Y, Kurmanaeva L, Divinski SV, Peterlechner M, Wilde G (2011) Strain mapping of a triple junction in nanocrystalline Pd. Acta Mater 59:7380
Schottky W (1935) Über den Mechanismus der Ionenbewegung in festen Elektrolyten. Z Phys Chem B 29:335 (On the mechanism of ion motion in solid electrolytes, in German)
Schottky W, Stöckmann F (1954) Vergleichende Betrachtungen über die Natur der Störstellen in Halbleitern und Phosphoren. Halbleiterprobleme 1:80 (Comparative considerations on the nature of impurities in semiconductors and phosphors, in German)
Schulze A, Strakos L, Vystavel T, Loo R, Pacco A, Collaert N, Vandervorsta W, Caymax M (2018) Non-destructive characterization of extended crystalline defects in confined semiconductor device structures. Nanoscale 10:7058
Seeger K (1997) Semiconductor physics: an introduction, 6th edn. Springer, New York
Stolwijk NA, Schuster B, Hölzl J, Mehrer H, Frank W (1983) Diffusion and solubility of gold in silicon. Physica B+C 116:335
Talwar DN, Vandevyver M, Zigone M (1980) Impurity induced Raman scattering spectra in zincblende-type crystals: application to mixed indium pnictides. J Phys C 13:3775
Tan TY, Gösele U, Morehead FF (1983) On the nature of point defects and the effect of oxidation on substitutional dopant diffusion in silicon. Appl Phys A 31:97
Tan TY (2017) Point defects and diffusion in semiconductors. In: Paul A, Divinski S (eds) Handbook of solid state diffusion, vol 1. Elsevier, Amsterdam
Taylor GI (1934) The mechanism of plastic deformation of crystals, part I, theoretical. Proc R Soc London A 145:362
Troxell JR, Watkins GD (1980) Interstitial boron in silicon: a negative-*U* system. Phys Rev B 22:921
Troxell JR, Chatterjee AP, Watkins GD, Kimerling LC (1979) Recombination-enhanced migration of interstitial aluminum in silicon. Phys Rev B 19:5336
Van de Walle CG, Denteneer PJH, Bar-Yam Y, Pantelides ST (1989) Theory of hydrogen diffusion and reactions in crystalline silicon. Phys Rev B 39:10791
Van de Walle CG, Neugebauer J (2004) First-principles calculations for defects and impurities: applications to III-nitrides. J Appl Phys 95:3851
van Vechten JA (1980) A simple man's view of the thermochemistry of semiconductors. In: Moss TS, Keller SP (eds) Handbook of semiconductors, vol 3. North Holland Publishing, Amsterdam, pp 1–111
Watkins GD (1974) Lattice defects in semiconductors. Inst Phys Conf Ser 23:1
Watkins GD (1986) The lattice vacancy in silicon. In: Pantelides ST (ed) Deep centers in semiconductors. Gordon and Breach, New York, p 147
Weber ER (1983) Transition metals in silicon. Appl Phys A 30:1
Weber ER, Alexander H (1983) Deep level defects in plastically deformed silicon. J Phys Colloq (Orsay Fr) 44(C4):C4-319–C4-328
Weertman J, Weertman JR (1960) Elementary dislocation theory. Macmillan, New York
Wessel K, Alexander H (1977) On the mobility of partial dislocations in silicon. Philos Mag 35:1523
Wilkinson AJ (2000) Advances in SEM-based diffraction studies of defects and strains in semiconductors. J Electron Microsc 49:299
Yao Y, Ishikawa Y, Sudo M, Sugawara Y, Yokoe D (2017) Characterization of threading dislocations in GaN (0001) substrates by photoluminescence imaging, cathodoluminescence mapping and etch pits. J Cryst Growth 468:484
Yeh C-Y, Lu ZW, Zunger A (1992) Zinc-blende – wurtzite polytypism in semiconductors. Phys Rev B 46:10086
You JH, Johnson HT (2009) Effect of dislocations on the electrical and optical properties of GaAs and GaN. In: Ehrenreich H, Spaepen F (eds) Solid state physics, vol 61. Academic Press, New York, pp 143–261

Crystal Interfaces

16

Karl W. Böer and Udo W. Pohl

Contents

1 Structure of Heterointerfaces 650
1.1 Pseudomorphic Layers 651
1.2 Strain Relaxation 652
2 Electronic Properties of Heterointerfaces 656
2.1 Issues for Band Alignment 656
2.2 Band-Alignment Models 662
2.3 Interface Dipole 669
2.4 Ab Initio Approaches 674
3 Metal-Semiconductor Interfaces 680
3.1 Classification of Semiconductor-Metal Interfaces 680
3.2 The Classical Schottky-Mott Model 682
3.3 Experimental Results on Schottky Barriers 688
3.4 Interface-Dipol Models 694
4 Summary 698
References 698

Abstract

Interfaces to other semiconductors, producing a heterojunction, or to conductors, acting as contacts or barriers, are important parts of almost every semiconductor device. Their basic interface properties are a decisive element of the device operation and its performance. Layers of different semiconductors with a common interface may be coherently strained up to a to a critical layer thickness, which is roughly inverse to the mismatch of their in-plane lattice parameters. In thicker layers strain is at least partially relaxed by misfit dislocations.

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_16

The electronic properties of semiconductor heterojunctions and metal-semiconductor contacts are governed by the alignment of their electronic bands. Early models describe band offsets and barrier heights as the difference of two bulk properties. While related chemical trends are found for certain conditions, the band lineup cannot be predicted with sufficient accuracy by a single universal model. Interdiffusion on an atomic scale, defects located at the interface, and leaking out of eigenfunctions from one into the other material create interface dipoles, modify alignments guessed from simple properties of the bulk materials. Additional shifts originate from strain. Various models exist, each describing certain groups of materials forming interfaces. *Linear models* define reference levels *within* each material such as charge neutrality levels or branch-point energies within the bandgap, or average interstitial potentials, or they use localized states of impurities lying deep in the bandgap. *Nonlinear models* account for the formation of interface dipoles; such a charge accumulation can be induced by band states near the interface in one semiconductor lying in the bandgap of the other or by disorder at the interface inducing gap states. More recent first-principle approaches model hetero-interfaces explicitly or align bands with respect to the vacuum level by including surfaces.

Keywords

Anderson model · Band alignment · Band bending · Band lineup · Bardeen model · Blocking contact · Debye length · Diffusion potential · Disorder-induced gap states · Electron affinity · Heterointerfaces · Interface-dipol theory · Interfaces · Metal-induced gap states · Metal-semiconductor interfaces · Model-solid theory · Ohmic contact · Pseudomorhic layer · Schottky barrier · Schottky-Mott model · Space-charge region · Strain relaxation · Transition-metal reference level · Valence-band offset

1 Structure of Heterointerfaces

The heteroboundary between two crystals of different composition is of great technical importance for heterojunction devices. It is characterized by a mismatch in several parameters, most importantly the lattice constant, the *electron affinity*[1] (see Bauer et al. 1983), and the bandgap energy. The mismatch of expansion coefficients is of interest when a device is produced at temperatures substantially different from

[1]The electron affinity of a semiconductor is defined as the energy difference from the lower edge of the conduction band to the vacuum level, i.e., the energy gained when an electron is brought from infinity into the bulk of a crystal, resting at E_c. It should be distinguished from the electron affinity of an atom, which is equal to the energy gained when an electron is brought from infinity to attach to an atom and forms an anion.

room temperature or is exposed to temperature cycling during operation. It may generate major interface defects and cause curling of thin layers or, in extreme cases, result in delamination. This chapter first outlines the structural properties of heterointerfaces and then presents the electronic properties; eventually the properties of the interface of a semiconductor to a metal are pointed out.

1.1 Pseudomorphic Layers

Thin semiconductor layers may be grown by epitaxy (Sect. 3.3 of ▶ Chap. 1, "Properties and Growth of Semiconductors") on a semiconductor of a different composition without defects at the common interface. Figure 1 illustrates a layer with the same (cubic) crystal structure as the substrate beneath but – in absence of the common interface – with a different *unstrained* lattice constant a_L. If the difference in lattice constants is not too large (below ~1%), the layer adopts an in-plane lattice constant $a_{L\|}$ identical[2] to that of the substrate a_S. If $a_L < a_S$ the layer is tensely strained in lateral direction and consequently experiences a compressive deformation in the vertical direction to approximately maintain its bulk density; vice versa, $a_L > a_S$ leads to $a_{L\|} < a_L$ and $a_{L\perp} > a_L$. Such a heterostructure is called *pseudomorphic*, and the layer is designated *coherently* strained.

The lateral strain (in-plane strain) in a pseudomorphic layer is given by the *lattice mismatch*[3] (or *misfit*) f:

$$\varepsilon_{\|} = f \equiv \frac{a_S - a_L}{a_L}. \tag{1}$$

Often alloys, e.g., $In_xGa_{1-x}P/GaAs$, are used as layer material with a composition parameter chosen to match the lattice constant of the substrate

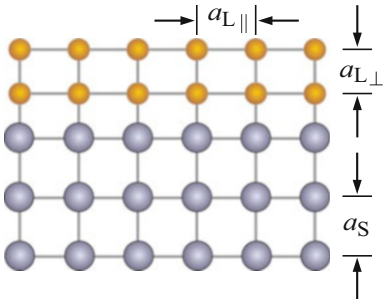


Fig. 1 Coherently strained layer (*yellow atoms*) with an unstrained lattice constant a_L on a substrate (*blue atoms*) with a different lattice constant a_S

[2]The substrate remains almost unstrained due to its large thickness.

[3]It must be noted that also other definitions for f are used in literature, particularly $f = (a_L–a_S)/a_S$ or $f = (a_L–a_S)/a_L$.

(here $x = 0.49$ for the GaAs substrate); still such layers are not perfectly matched due to differences in thermal expansion and the high growth temperatures. The vertical strain $\varepsilon_\perp$ is given by

$$\varepsilon_\perp = -D \ \varepsilon_{||}, \tag{2}$$

where the distortion factor D depends on the crystal structure and the orientation of the interface with respect to the crystallographic axes of the layer. The assumption of a *constant* ratio $D = -\varepsilon_\perp/\varepsilon_{||}$ in strained layers is based on the assumed validity of Hooke's law (Eq. 10 of ▶ Chap. 4, "Elasticity and Phonons"), i.e., a linear stress-strain relation. For cubic crystals and biaxial strain (i.e., $\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{||}$), the distortion factors for the main orientations are

$$\begin{aligned} D_{001} &= \frac{2C_{12}}{C_{11}}, \\ D_{110} &= \frac{C_{11} + 3C_{12} - 2C_{44}}{C_{11} + C_{12} + 2C_{44}}, \\ D_{111} &= \frac{2C_{11} + 4C_{12} - 4C_{44}}{C_{11} + 2C_{12} + 4C_{44}}. \end{aligned} \tag{3}$$

For biaxially strained hexagonal crystals and a basal-plane interface the distortion factor reads

$$D_{0001} = \frac{2C_{13}}{C_{33}}. \tag{4}$$

Since the stiffness constants C_{ij} are positive quantities, vertical and lateral strains have always opposite sign; a tensile lateral strain leads to a vertical compression and vice versa.

The bulk density of the strained layer is not perfectly preserved despite the counteracting vertical strain. The relative volume change is given by the trace of the strain tensor,

$$\Delta V/V = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}. \tag{5}$$

1.2 Strain Relaxation

A coherently strained layer contains an elastic strain energy. This homogeneous strain energy increases in the harmonic approximation of Hooke's law quadratically with the strain $\varepsilon_{||}$, i.e., with the misfit f for a pseudomorphic layer. Furthermore, it increases linearly with the layer thickness t_L. At some *critical layer thickness* t_c, this strain energy gets larger than the energy required to form structural defects, such as misfit dislocations, which *plastically* relax part of the strain (see Fig. 2). The plastic relaxation reduces the overall strain; at the same time the dislocation energy

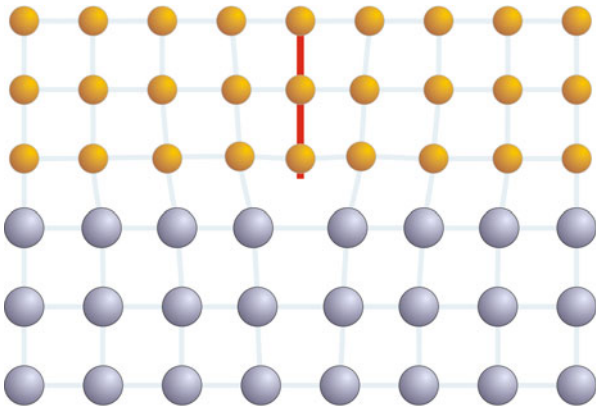


Fig. 2 Cross-section scheme of an edge dislocation (*red line*) introduced into an initially coherently strained layer (*yellow atoms*) for plastic strain relaxation

increases from zero to a value determined by the particular kind of the introduced dislocation.

Dislocations (Sect. 4 of ▶ Chap. 15, "Crystal Defects") in the active part of a semiconductor device introduce dangling bonds and inhomogeneities; the resulting enhanced nonradiative relaxation and scattering of carriers seriously degrade the device performance. The thickness of all layers in a device are therefore kept below the critical value t_c for each individual layer.[4]

The dependence of the critical thickness t_c of a pseudomorphic layer on the misfit f (Eq. 1) was calculated assuming a balance of forces (Matthews and Blakeslee 1974) or areal strain energies (Frank and Van der Merve 1949; Van der Merve 1962; Matthews 1975) in thermodynamic equilibrium. The formation of a dislocation requires a formation energy E_D associated with the locally highly strained region at the dislocation line. In the energy balance model, this energy cost is balanced by the elastic release of homogeneous strain energy E_H in the layer outside the core region of the dislocation. The elastic energy E_I at the interface is given by the sum of $E_H + E_D$. The remaining mismatch f_r of the partially relaxed layer refers to an *average* lattice constant of the layer. In the presence of dislocations, f_r is less than the natural misfit f of Eq. 1, which is defined by the *unstrained* lattice constants. To a good approximation, f_r is given by the sum of f and the residual strain

$$\varepsilon = \frac{a_{L,\,r} - a_L}{a_L}$$

which may have opposite sign; a_L and $a_{L,r}$ denote the unstrained and partially relaxed lateral lattice parameters of the layer. The dependence of the elastic energy at the interface from ε is then given by

[4]In a stack of layers, strain compensation may be applied by using different layer materials with counteracting positive and negative strain to keep the accumulated overall strain low.

$$\begin{aligned} E_{\mathrm{I}} &= E_{\mathrm{H}} + E_{\mathrm{D}}, \\ E_{\mathrm{H}} &\propto t_{\mathrm{L}}\ \varepsilon^2, \\ E_{\mathrm{D}} &\propto \varepsilon + f. \end{aligned} \tag{6}$$

The relation for the interface energy given by Eq. 6 is illustrated in Fig. 3 for a $GaAs_{0.9}P_{0.1}$ layer on GaAs substrate with a misfit $f = -0.36\%$ and a thickness below, just at, and above the critical value for plastic relaxation (Ball and Van der Merve 1983). The homogeneous strain energy-density disappears at zero strain and increases quadratically with ε (blue curves), while the dislocation energy-density gets zero at $f_{\mathrm{r}} = \varepsilon + f = 0$ (red curves). In any case the energy at the interface E_{I} (green curves) tends to attain a minimum. The criterion for the critical thickness t_{c} is

$$\partial(E_{\mathrm{I}})/\partial(|\varepsilon|) = 0 \tag{7}$$

evaluated at $|\varepsilon| = |f|$. For $t_{\mathrm{L}} > t_{\mathrm{c}}$ the homogeneous strain of the layer gets larger than the misfit $|f|$, and dislocations introduce a strain of opposite sign, thereby reducing $|\varepsilon|$ and hence E_{I}.

For the evaluation of the critical thickness in a given heterostructure, the geometry of the strain-relaxing dislocations must be specified. Applying an explicit expression for E_{D} in Eqs. 6 and 7, we obtain a transcendent equation (Ball and Van der Merve 1983), which can be calculated numerically:

$$t_{\mathrm{c}}(|f|) \;=\; b\frac{(1 - v\cos^2\alpha)}{8\pi|f|(1+v)\sin\alpha\cos\beta}\ln\left(\frac{\rho t_{\mathrm{c}}}{b}\right). \tag{8}$$

Here, b is the absolute value of the Burgers vector, $v = C_{12}/(C_{11}+C_{12})$ is Poisson's ratio, α is the angle between the Burgers vector and the dislocation line, β is

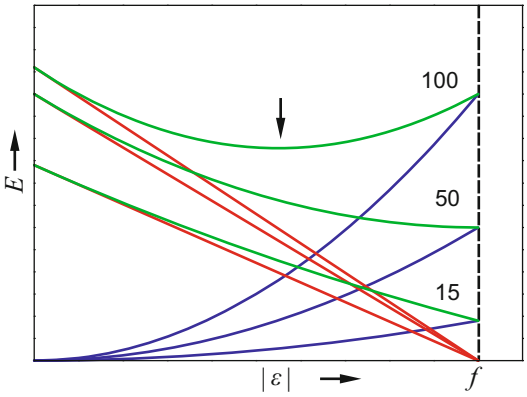


Fig. 3 Energy densities occurring at a biaxially strained layer with a thickness of 15, 50, and 100 times the substrate lattice constant a_{S}. ε denotes lateral strain, and $f = -0.36\%$ is the assumed misfit. The *blue*, *green*, and *red curves* are homogeneous strain, strain at the interface, and dislocation energy, respectively. The arrow denotes an energy minimum of the thick layer attained by plastic strain relaxation

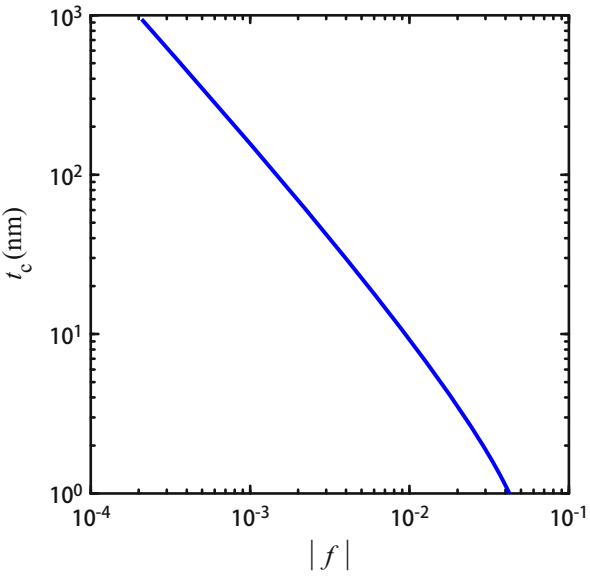


Fig. 4 Critical thickness t_c of a pseudomorphic layer with zincblende structure for accommodating 60° dislocations. An unstrained lattice constant $a_L = 5.657$ Å is assumed

the angle between the glide plane of the dislocation and the interface, and the factor $\rho \cong 2 \ldots 4$ accounts for the strain energy of the dislocation core.[5] In the solution shown in Fig. 4, accommodating 60° dislocations in a biaxially strained layer with zincblende (or diamond) structure and (001) interface plane, $a_L = 5.657$ Å, $\nu = 0.33$, and $\rho = 2.72$ are assumed.

The critical layer thickness is approximately inverse to the misfit f. For crystalline metal layers, the experimentally determined values agree reasonably well with the predictions by equilibrium theory. For semiconductor layers, however, also significantly larger values were observed. Due to kinetic barriers for the generation and movements of misfit dislocations, the critical thickness calculated from equilibrium models provides a lower limit.

Above the critical layer thickness, the lattice mismatch is partially relieved by a *network of mismatch dislocations*, which form a more or less regular grid at the interface. The average spacing is inverse to the remaining lateral misfit f_r and may be anisotropic as shown in Fig. 5 due to the effect of different kinds of dislocations. The dislocation density decreases linearly down to a minimum value with the distance from the heterojunction (Fig. 6), resulting from an annihilation of misfit dislocations. The density of threading dislocations at the surface was found to be approximately inverse to the layer thickness (Sheldon et al. 1988) and depends on the heterogeometry. Usually, it results in an opposite strain at the outer surfaces.

[5]Linear elasticity theory is not applicable at the highly strained dislocation core. Only the elastic strain energy outside a somewhat arbitrary cutoff radius r_{cutoff} about the dislocation line is hence calculated and a term comprising r_{cutoff} is added to account for nonlinear elastic energy and dangling bonds in the core. Usually $r_{cutoff} = b/\rho$ is chosen yielding in Eq. 8 the term $\ln(t_c/r_{cutoff}) = \ln(\rho\, t_c/b)$, also written $\ln(t_c/b)+const$ with $const = \ln(\rho)$.

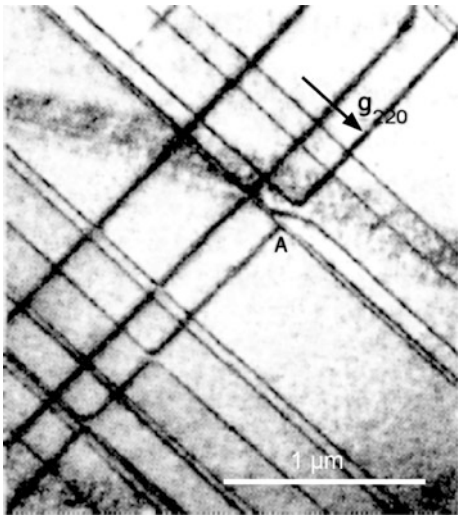


Fig. 5 Plan view transmission-electron micrograph of a 20 nm thick partially relaxed $In_{0.2}Ga_{0.8}As$/GaAs(001) layer with $f = -1.4\%$. Average spacings in orthogonal ⟨110⟩ directions are 200 and 400 nm (After Dixon and Goodhew 1990)

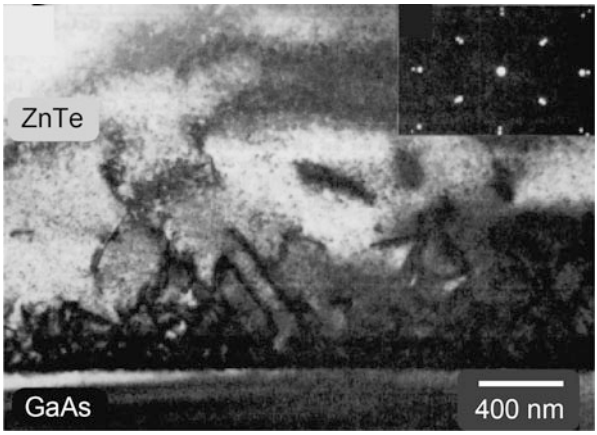


Fig. 6 Cross-section transmission-electron micrograph of a 1.8 μm thick ZnTe/GaAs(001) layer with $f = -7{,}5\%$, imaged along [110]. Double reflections in the diffraction pattern from the interface region (*inset*) refer to GaAs (*outer spots*) and ZnTe (*inner spots*) (After Bauer et al. 1993)

This can best be seen for a lattice mismatch in one direction within the interface plane. It results in a bending of the crystal, with a neutral plane near its center. It has compression and expansion at the heterointerface and the surface, respectively.

2 Electronic Properties of Heterointerfaces

2.1 Issues for Band Alignment

At the interface from one semiconductor to another or to a metal the connection of the bands, the so-called band alignment (also band lineup, band offset, or band discontinuity) is the most interesting property of its electronic structure. It has become clear that the general behavior which determines the band lineup is rather

similar between semiconductor-to-semiconductor, semiconductor-to-metal, and metal-to-metal interfaces. Therefore, aspects which are important for heterojunctions or for contacts are introduced in a parallel description.

2.1.1 Experimental Results for Valence-Band Offsets

The band alignment is caused by a number of factors, not all of which are well controlled or even understood to date. Some of the contributing factors are the difference between the two electron affinities (or workfunctions) of the involved semiconductors (or metals), the interface defect levels and an induced interface dipole, and the lattice misfit f at the interface. The experimental results that are obtained for abrupt heterojunctions can be summarized as follows (Flores and Tejedor 1987):

1. There is *commutativity for the valence-band offset* for a wide variety of well-deposited semiconductors (Margaritondo 1983). For instance,

$$\Delta E_{\mathrm{v}}[\mathrm{AlAs/GaAs(110)}] = -\Delta E_{\mathrm{v}}[\mathrm{GaAs/AlAs(110)}] \pm 0.05 \ \mathrm{eV} \tag{9}$$

   applies (in this example the first semiconductor is deposited on top of the second); see also Katnani (1987).
2. There is *transitivity of the valence-band offset* for a wide variety of semiconductors (Katnani and Bauer 1986; Katnani 1987):

$$\Delta E_{\mathrm{v}}(A \rightarrow B) + \Delta E_{\mathrm{v}}(B \rightarrow C) = -\Delta E_{\mathrm{v}}(C \rightarrow A). \tag{10}$$

3. In many heterojunctions there is only a minor dependency of the valence-band offset on the *crystal-face orientation* of any of the semiconductors with good interface match (Grant et al. 1987; Chiaradia et al. 1984).
4. The *common-anion rule* indicates that the valence-band offset is very small for heterojunction couples of compounds having the same anion. This rule is best fulfilled for compounds with a large fraction of ionic bonding (Frensley and Kroemer 1977; Kroemer 1984) but not well fulfilled for III–V compounds, e.g., AlAs/GaAs, or even some II–VI compounds, such as HgTe/CdTe.

These experimental results can be used as a guideline within a certain band of accuracy which is typically on the order of ± 0.1 eV and with caution for possible exceptions. Items (1–3) are consistent with linear theories. The summary of the experimental results given above indicates the typical observation of an *insensitivity* of the band offsets to a large variety of interface manipulation. From this experimental basis of rather weak general trends, it is difficult to develop a common ground for testing theoretical models. Therefore, it is not surprising that several of widely different models often yield agreement within the band of experimental uncertainties, and it is difficult to weed out insufficient models. Only recently first-principles calculations have become feasible to account for the formation of chemical bonds, atom relaxations, and charge redistribution at the interface

Table 1 Experimental values for valence and conduction-band offset. $\Delta E_v > 0$ for *A-B* heterojunctions means $E_v(B) > E_v(A)$; *A/B* means *A* deposited on *B*. sg and av denote staggered gap and average value; data in parenthesis are estimated from $\Delta E_c + \Delta E_v = \Delta E_g$. Experimental methods: photoemission (PH), *I–V* characteristics (IV), other methods (OT) (Data of crystalline materials only; for individual references see original, after Margaritondo and Perfetti 1987)

Heterojunction	Layer/substrate	ΔE_v	ΔE_c	Exp. method
Si-Ge	Ge/Si(111)	0.17	(0.27)	PH
	Si/Ge(111)	0.4	(0.04)	PH
AlAs-Ge	Ge/AlAs	0.9	(0.63)	PH
	AlAs/Ge(100)	0.78	(0.75)	PH
AlAs-GaAs	GaAs/AlAs(110)	0.4	(0.45)	PH
	AlAs/GaAs(110)	0.15	(0.7)	PH
	Either on (100)	0.38	(0.47)	PH
AlSb-GaSb	Superlattice	0.4	(0.5)	OT
GaAs-Ge	Ge/GaAs(110)	av 0.49	av (0.18)	PH
	GaAs/Ge(110)	0.34	0.39	OT
	Ge/GaAs(111)Ga	0.48	(0.20)	PH
	Ge/GaAs($\bar{1}\,\bar{1}\,\bar{1}$)As	0.60	(0.08)	PH
	GaAs/Ge(110)	0.24	(0.34)	PH
	GaAs/Ge(100)	0.44	(0.24)	PH
	GaAs/Ge(100)	0.68	0.05	IV
GaAs-InAs	GaAs/InAs(100) sg	0.17	(−0.09)	PH
GaP-Si	Si/GaP(110)	0.80	(0.33)	PH
ZnSe-Ge	Ge/ZnSe(110)	av 1.46	av (0.45)	PH
	ZnSe/Ge(110)	1.29	(0.62)	PH
ZnSe-GaAs	ZnSe/GaAs(110)	0.96	(0.27)	PH

(Weston et al. 2018, Tung and Kronik 2016). Such self-consistent calculations are demanding but provide realistic values within the above-mentioned accuracy.

Table 1 lists the experimental values for the band-edge offset of a number of heterojunctions compiled by Margaritondo and Perfetti (1987).

2.1.2 Space-Charge Regions

When describing the contact between two materials in this section, we will use a scale that extends typically several hundred nm from the interface. The interface presents boundary conditions, e.g., in its charge character that needs compensating charges further inside the material. Depending on the conductivity and other material parameters, such compensating charges are distributed (i.e., they are *space charges*) over a region of several Debye lengths[6] which in semiconductors are typically on the order of 100 nm. Connected with such charges is a characteristic

[6]The Debye length L_D is a function of the carrier density. It is typically on the order of $30\sqrt{10^{16}\text{cm}^3/n}$ (nm) with n (cm^{-3}) as the density of free carriers.

bending of bands. It is this aspect which permits some helpful classification, particularly for interfaces to metals (Sect. 3.1).

In order to identify interactions that produce the interconnection of the bands at the interface and the band bending in the space-charge region, we consider two materials that are separated by a wide enough distance in vacuum, as shown in Fig. 7 for a semiconductor-semiconductor pair (*a*) and for a metal-semiconductor pair (*d*). Each pair is aligned through the common vacuum level E_0. Within each material there are a number of relevant energies[7] which will play an important role:

- The band edges E_c and E_v
- The Fermi level E_F
- The workfunction $\phi = E_0 - E_F$ counted from the Fermi level
- The electron affinity $\chi = E_0 - E_c$ counted from the conduction-band edge
- The ionization energy $I = E_0 - E_v$ counted from the valence-band edge
- The bandgap energy $E_g = E_c - E_v = I - \chi$.

Let us assume first that such a pair is suspended in a vacuum vessel by an insulating filament and kept at a high enough temperature so that some electrons can exit through the surface of the semiconductor and the metal. These electrons will flow between both materials, charging one with respect to the other until equilibrium is reached and the current becomes the same in one direction as in the opposite direction. Now, both Fermi levels have become aligned; the two materials are in thermodynamic equilibrium. The amount of the resulting potential shift between the two materials is referred to as the Volta potential or contact potential[8] $V_c = \Delta E_0/e$, with

$$\Delta E_0 = \phi_2 - \phi_1. \tag{11}$$

Such an alignment is shown in the second row Fig. 7b, e.

When narrowing the gap between the two materials, their relative charge will be attracted to each other, and as free carriers move toward the interface, they create a

[7]The workfunction ϕ and the electron affinity χ are in literature also defined in terms of potentials, yielding an additional factor e to obtain an energy.

[8]The contact potential can be measured by the Kelvin method, i.e., by shaping the two materials into the plates of a capacitor and vibrating these plates against each other. With an induced areal charge

$$Q_A = \varepsilon_0 V_c/d$$

and d oscillating in time, the charge must also oscillate since V_c does not depend on the distance d between both plates; hence an *ac* current will flow between the plates when externally connected with a wire. The current vanishes when a counterpotential is applied which is equal and opposite to the contact potential. This null method is a convenient one to directly determine the contact potential.

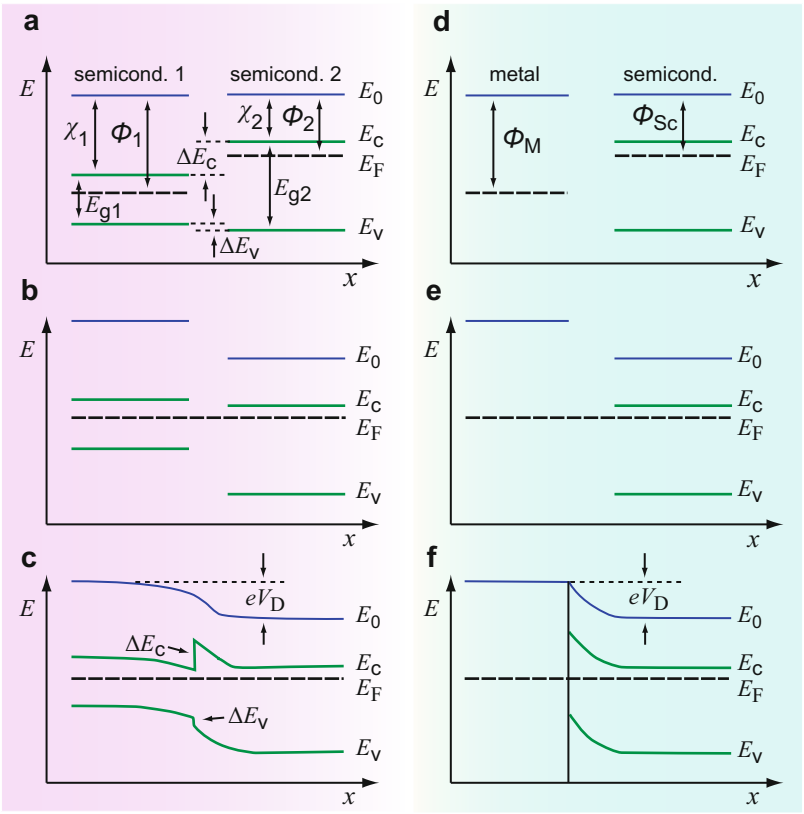


Fig. 7 Schematics of band diagrams for heterojunctions (*left*, **a**, **b**, **c**) and metal-semiconductor contacts (*right*, **d**, **e**, **f**) with aligned vacuum level E_0 (**a**, **d**), brought into thermodynamic equilibrium resulting in an aligned Fermi level E_F (**b**, **e**) and brought into contact with compensating space-charge regions (**c**, **f**)

more localized dipole layer across a thin gap that is left open. These carriers leave behind a partially depleted region that now has the opposite charge (e.g., caused by depleted donors or acceptors) with respect to the carriers which drifted toward the interfaces. This partially depleted region comprises the space-charge region; its exact width and charge distribution is not important for this discussion and will be analyzed in Sect. 3.2.

When closing the gap, the double layer and the space-charge region will readjust to maintain equilibrium, i.e., without external forces applied; the current across the interface will become zero after adjustment, and within each space-charge region, the gradient carriers will be stabilized by a built-in field (corresponding to the slope of the bands) that produces a drift current equal and opposite in sign to the diffusion caused by the gradient.

Such *band bending* is shown in the last row of panels (*c* and *f*) of Fig. 7. It extends into both sides of the heterojunction for a few Debye lengths (subfigure *c*)

but is invisible on the metal side of the metal-semiconductor interface in the scale drawn because the Thomas-Fermi length[9] is much shorter. The total band bending is given by the diffusion potential V_D, also referred to as built-in potential.

We will now focus on the different relative energy offsets at this interface that permit a classification of the various semiconductor heterojunctions; metal-to-semiconductor interfaces are considered in Sect. 3.

2.1.3 Classification of Interfaces

A classification of semiconductor heterojunctions is conventionally made with respect to interfaces, disregarding space-charge regions, and with respect to their electrical performance by including space-charge effects.

Classification Disregarding Space Charges Disregarding the band deformation due to space-charge regions, one can distinguish several types of semiconductor heterojunctions or metal-contact interfaces dependent on the relative values of electron affinities or workfunctions. For metal-semiconductor pairs, see Sect. 3. For semiconductor heterojunctions, the main types of interfaces are shown in Fig. 8. The type designation is used to distinguish low-dimensional structures like quantum wells; see ▶ Chap. 8, "Bands and Bandgaps in Solids" (Sect. 3 and Fig. 30).[10] The band aligning sketched in Fig. 8 corresponds to the conventional classifications type I–III.

Classification Including Space-Charge Regions The classification between different types of heterojunctions becomes a bit more involved when space-charge regions are included. There are several possibilities for arranging heterojunctions depending on the relative size and sign of ΔE_c and ΔE_v and the doping of both semiconductors. For an overview, see Milnes and Feucht (1972). One has to distinguish between interfaces that have simple steps, as shown in Fig. 7c, or spikes, either in the conduction band or in the valence band or in both.

An example for a *p-n* heterojunction is shown for the AlSb/GaSb heterojunction in Fig. 9.[11] This figure also indicates how a heterojunction can change from one behavior to the other with spikes appearing in different bands by simply changing the sequence of doping (*n* to *p* in subfigure (*a*) or *p* to *n* in subfigure (*b*)).

For semiconductor couples that contain a spike in the energy diagram at the interface, an impediment in the carrier transport results and carrier trapping can occur. Such a trapping can become part of a charge redistribution at the interface which modifies the simple Anderson model (see next section).

[9]The Thomas-Fermi length in a metal, in which the carriers are constrained by the Fermi-Dirac distribution, is the equivalent of the Debye length in a semiconductor.

[10]$E(z)$ energies within a low-dimensional nanostructure are usually drawn without slope or bending in absence of electric fields, because the nanostructure dimensions are commonly much smaller than Debye lengths.

[11]The valence-band lineup for InSb/GaSb/AlSb/InAs is 0, −0.45, −0.15 eV (Milnes, 1991, private communication).

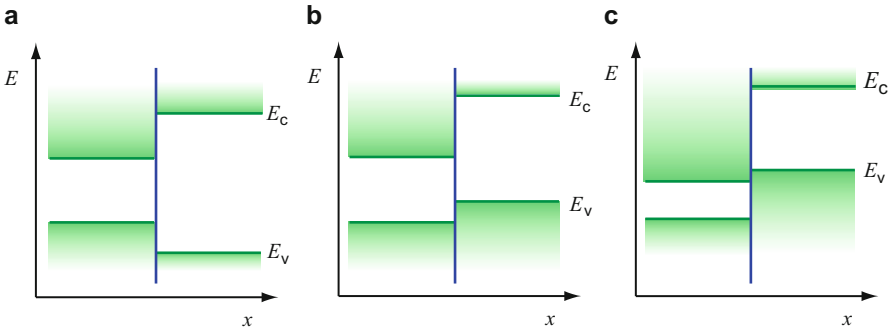


Fig. 8 Schematics of the different types of possible band lineups of semiconductor heterojunctions, neglecting band bending due to space charges near the interface. (**a**) Type I alignment (straddled configuration), (**b**) type II (staggered), (**c**) type III (misaligned, broken gap)

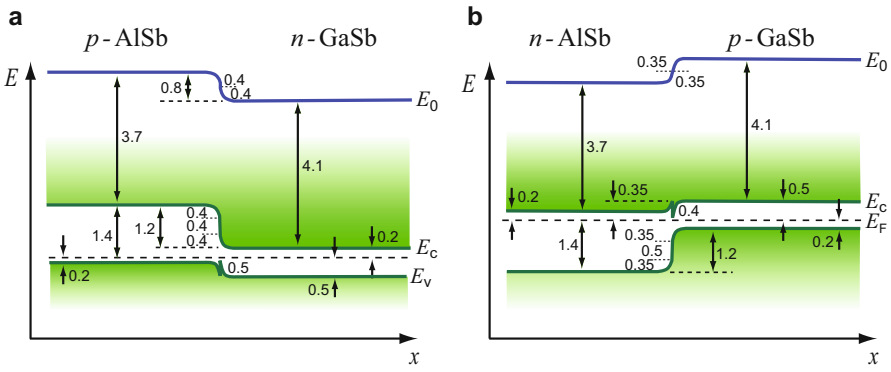


Fig. 9 AlSb/GaSb heterojunction with energy values given in eV. When AlSb is *p*-type (**a**) the spike occurs in the valence band; when AlSb is *n*-type (**b**) the spike occurs in the conduction band (After Milnes and Feucht 1972)

2.2 Band-Alignment Models

2.2.1 The Anderson Model for Heterojunctions

A simple electron-affinity rule for the band alignment of a heterojunction proposed by Anderson (1962) can be used for some orientation when the following conditions are fulfilled:

1. The interface represents an abrupt transition from one to the other semiconductor with no interdiffusion.
2. There is a negligible density of interface states; dipole effects at the interface can be neglected.

3. There is no interface strain due to lattice mismatch.
4. Band offsets are considered independent on interface orientation, interface bonding, arrangements of atoms, and doping of both semiconductors.

For such heterojunctions, the Anderson model is valid, and the band offsets are a property of the bulk material; the conduction-band and valence-band offsets are then given by the difference of the electron affinities and ionization energies, respectively:

$$\Delta E_c = \chi_{\text{semicond2}} - \chi_{\text{semicond1}}, \tag{12}$$

$$\Delta E_v = I_{\text{semicond2}} - I_{\text{semicond1}}, \tag{13}$$

where $I = E_0 - E_v$ is the ionization energy and E_0 is the vacuum level. The sum of both band offsets is equal to the difference in bandgap energy of the two semiconductors:

$$\Delta E_g = \Delta E_v + \Delta E_c. \tag{14}$$

This extremely simple rule has enjoyed great popularity and is often used for estimates of new heterojunctions (see, e.g., Lyu et al. 2020). However, mostly one or the other of conditions (1–4) are not fulfilled – particularly no. 2 and 4, and the validity of the electron affinity rule is far from being universal. Electron affinities for some semiconductors are given in Table 2. Values for the band-edge offset calculated using various approaches for a number of heterojunctions as compiled by Bechstedt and Enderlein (1988) are listed in Table 3.

Theoretical reasons for a need to modify the Anderson model were already pointed out by Kroemer (1975). Experimental discrepancies became obvious for large ΔE_v couples, e.g., for Ge/ZnSe which individually show a difference in ionization energies of $\Delta(E_0 - E_v) = 2.21$ eV, while the interface valence-band

Table 2 Electron affinity χ for some semiconductors, modifications indicated in brackets are 3C, zincblende; 2H, wurtzite; 6H, hexagonal polytype (After Adachi 2005)

Semiconductor	χ (eV)	Semiconductor	χ (eV)
Ge	4.14	InP	4.50
Si	4.05	InAs	5.06
SiC (3C)	3.83	InSb	4.72
SiC (6H)	3.34	ZnS (3C)	3.9
AlAs	3.5	ZnSe	4.06
AlSb	3.65	ZnTe	3.68
GaN (2H)	3.3	CdS (2H)	4.5
GaP	3.75	CdSe (2H)	4.95
GaAs	4.15	CdTe	4.28
GaSb	4.21		

Table 3 Valence-band offsets ΔE_v calculated using various approaches (After Bechstedt and Enderlein 1988)

Heterojunction	Andersen model	Harrison model	Frensley-Kroemer model	Tersoff interface-dipole model	Hasegawa-Ohno model
Ge-Si	0.30	0.38		0.18	0.29
Ge-GaAs	0.67	0.41	0.71	0.32	0.66
Ge-ZnSe	2.02	1.46	1.82	1.52	2.01
Si-GaP	0.73	0.50	0.96	0.45	0.69
GaAs-AlAs	0.18	0.06	0.00	0.55	0.12
GaSb-AlSb	0.51	−0.02		0.38	0.09
GaAs-ZnSe	1.35	1.05	1.11	1.20	1.35
InAs-GaSb	−0.60	−0.53	−0.49	−0.43	−0.33
InAs-GaAs	0.16	0.32		0.00	−0.13
$Ga_{0.48}In_{0.52}As$/InP	0.30	0.28	0.40	0.26	0.36
GaAs/$Ga_{0.52}In_{0.48}P$	0.29	0.30	0.38	0.29	0.37

offset was measured as $\Delta E_v = 1.44$ eV. Such a discrepancy is indeed too large to be accounted for by experimental errors and clearly gives an example for the breakdown of the Anderson model (Niles and Margaritondo 1986).

2.2.2 Linear Models for Band-Edge Offsets

Many of the early attempts to model the band offset at the interface were historically influenced by the belief that electron affinities and workfunctions that determine the exit of electrons from semiconductors or metals can be described as intrinsic bulk properties. Therefore, it was suggestive to search for means to describe the band interconnection between two solids by the difference of two numbers, each one related to the bulk of one solid:

$$\Delta E_v = E_{ref,A} - E_{ref,B}, \tag{15}$$

with $E_{ref,A}$ and $E_{ref,B}$ as the appropriate reference levels of semiconductors A and B. The ensuing theories are the *linear theories* and are supported by the experimental items (1–3) in Sect. 2.1.1.

All of the linear theories neglect any nonlinear influence of the interface, such as defect-induced modifications. These are introduced either by an additive term to Eq. 15

$$\Delta E_v = E_{ref,A} - E_{ref,B} + eV_{dipole} \tag{16}$$

with eV_{dipole} the correction due to an interface dipole, or by a compression factor S in

$$\Delta E_{\mathrm{v}} = S \times \left(E_{\mathrm{ref},A} - E_{\mathrm{ref},B}\right) \tag{17}$$

that reduces the bulk-related band offset by an interface adjustment (see Sect. 3.3).

The first type of the linear theories comprises the Anderson and Schottky-Mott theories, which use the differences between workfunction and electron affinities. The limited success of these theories sparked the search for improvements that will be discussed in the following sections.

Several attempts were made to describe the band-edge offsets by a linear combination of bulk properties of both semiconductors: All of them depend on finding an appropriate intrinsic energy reference-level E_{ref} in the bulk of each semiconductor, relative to the vacuum level from which the valence-band offset can be determined (Eq. 15). A wide variety of suggestions has been made to determine such a level that is characteristic for the bulk. The choice of Anderson, using the electron affinity χ with $E_{\mathrm{ref}} = \chi + E_0$, i.e., including the electron transfer through the surface to vacuum for such a definition, is not a good one, as was experimentally shown, e.g., by Niles and Margaritondo (1986).

Other suggestions for a bulk-related reference level were:

- The intrinsic Fermi level (Adams and Nussbaum 1979), suffering from substantial shortcomings (Lee 1985)
- The mean energy of atomic p orbitals of cations and anions from LCAO calculations (Harrison 1977)
- The average interstitial potential computed from pseudopotentials (Frensley and Kroemer 1977) or from first-principles DFT calculations (Weston et al. 2018, Choe et al. 2018, 2021, Di Liberto and Pacchioni 2021)
- The branch-point energy or the charge neutrality-level considered by Tersoff (1984a, b, 1985) Flores and Tejedor (1979), and Robertson (2013)
- Localized states of transition-metal impurities (Zunger 1986; Langer and Heinrich 1985a) or interstitial hydrogen (Van de Walle and Neugebauer 2003)
- A local reference in a model solid composed of neutral atoms (Van de Walle 1989)

The linear models must be expanded to contain specific microscopic elements of the interface which include the creation of bonds and dipoles, the micro-diffusion of ions to adjust such interfaces (see Sect. 2.3.1), and the interface strain. These nonlinear elements of the theory can be appended by:

- Simply adding a dipole part (Eq. 16), as suggested by Frensley and Kroemer (1977) and by Mailhiot and Duke (1986)
- Modifying the bulk-related band offset by a pinning strength parameter, similar as it is done successfully for the Schottky barriers (Tersoff 1985)
- Introducing a strain-induced deformation of the band edge at the interface (Van de Walle 1989)

In Sect. 2.3, we will delineate in more detail some of these nonlinear models. In the following subsections, we will first describe the linear models.

2.2.3 The Harrison Valence-Band Offsets

The most direct way to obtain the band-edge offsets was proposed by Harrison (1977). The method resorts to the top edge of the valence band in the bulk which is less than the conduction-band edge influenced by perturbations and can be estimated from tabulated values of atomic p-state orbitals.

Harrison uses an expression suggested by Chadi and Cohen (1975) that computes the top of the valence band (Γ_8^v) of a diatomic compound semiconductor from

$$E_v = \frac{E_{p,\,\text{cation}} + E_{p,\,\text{anion}}}{2} - \sqrt{V_{\text{covalent}}^2 + V_{\text{polar}}^2} \tag{18}$$

with $E_{p,\text{cation}}$ and $E_{p,\text{anion}}$ as the atomic term values of the p orbitals of cations and anions, respectively, and V_{covalent} as the covalent energy

$$V_{\text{covalent}} = -1.28\,\hbar^2/\left(m_0 d^2\right) \tag{19}$$

with d as the nearest anion-cation distance. V_{polar} is the polar energy

$$V_{\text{polar}} = \left(E_{p,\,\text{cation}} - E_{p,\,\text{anion}}\right)/2. \tag{20}$$

The valence-band offset at the heterointerface is then obtained from the difference of the values of E_v (Eq. 18) for the two materials. For a listing of differences in the atomic term values of the p orbitals, see Table 3 and Harrison (1980). Figure 10 shows a comparison between experimental values and the prediction of the Harrison values. The difference between electron affinities can be obtained experimentally from photoemission measurements, where the threshold energy directly yields the difference from the top of the valence band to the vacuum level. There is reasonable agreement (between ± 0.1 eV) for couples with low lattice mismatch, but only fair agreement is observed and a deviation of up to 1 eV for semiconductor couples with larger lattice mismatches. For a review, see Katnani and Margaritondo (1983).

2.2.4 Transition-Metal or Hydrogen Levels as Reference

Transition Metals as Reference Another reference energy is that of deep defect levels of transition-metal impurities. Because of the screening from outer shell electrons and the d-like localized character of their wavefunction, it was suggested that their defect levels are more related to the vacuum level (like free atoms) than are adjusted relative to the band edges of the host semiconductor (see Ledebo and Ridley 1982; Vogl and Baranowski 1985; Caldas et al. 1984).

When used as a reference level, one adjusts the valence (and conduction) bands so that the deep levels of transition metals are matched at the same energy, independent of the host. In Fig. 11a the so-adjusted band offsets are held constant and the defect-level energy of six transition-metal dopants is shown within three

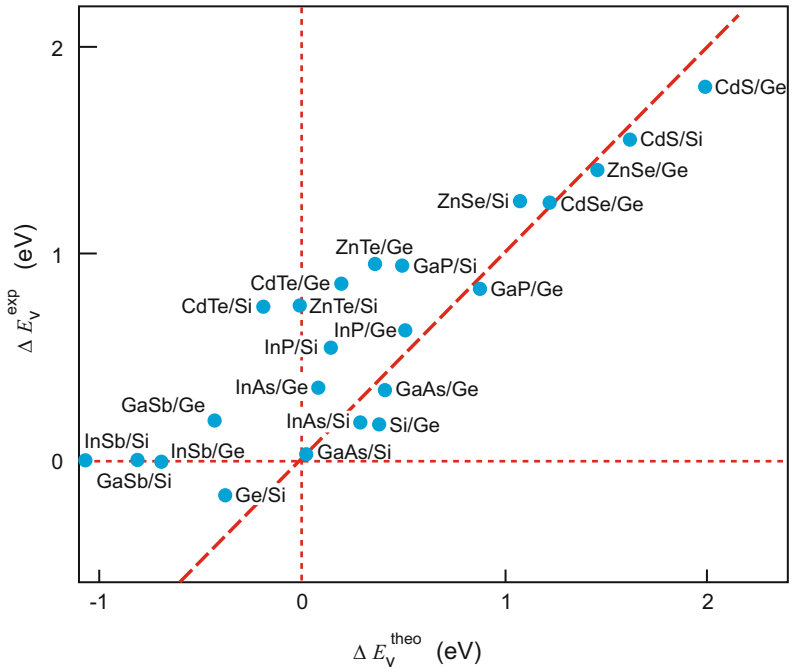


Fig. 10 Experimental values for the jump of the valence band of various heterojunctions obtained from photoemission (ΔE_v^{exp}), compared with the calculation from the Harrison model (ΔE_v^{theo}). Best agreement is obtained for junctions with high degree of lattice match (After Bauer et al. 1983)

semiconductors: GaAs, InP, and GaP. The figure confirms excellent matching of the defect-level energy. This procedure, proposed by Langer and Heinrich (1985a) and Zunger (1986), can be used to deduce the valence-band offset at actual (ideal) interfaces.[12] An indication in favor of this scheme can also be obtained by plotting the valence- and conduction-band edges (solid lines) of $Ga_{1-x}Al_xAs$ of various compositions in relation to the assumed constant level of an Fe acceptor (Fig. 11b). The band edges shift with composition x in good agreement with the observed shifts of the valence- or conduction-band offset, indicated by symbols in Fig. 11b.

Hydrogen as Reference A comparable universal alignment of deep impurity levels as described above was also reported for hydrogen (Van de Walle and Neugebauer 2003). By computing the position of the Fermi energy where the stable charge state of interstitial hydrogen changes from the H^+ donor state to the H^-

[12]That is, such interfaces in which no other effects, e.g., strain, atomic defects, or dislocations, modify the bands or an interface dipole exists (see also Heinrich and Langer 1986).

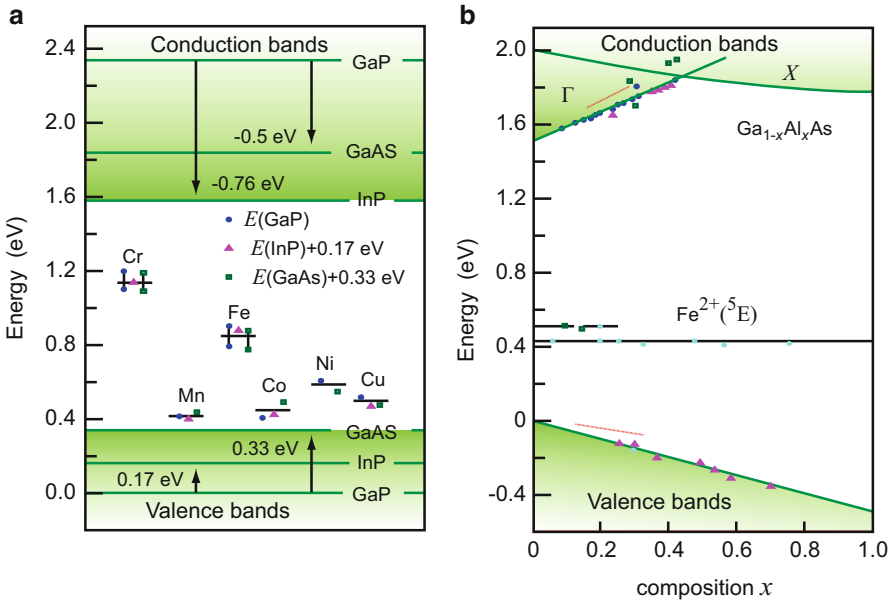


Fig. 11 (**a**) Energy levels of transition-metal dopants in three III–V semiconductors obtained by a vertical shift of each band diagram so that the deviation of the average experimental value of each acceptor level is minimized. (**b**) Band edges of various $Ga_{1-x}Al_xAs$ samples plotted relative to the energy of the Fe acceptor, assumed to remain at the same level (symbols at Fe^{2+} levels). Plotted also are the observed $Ga_{1-x}Al_xAs$/GaAs band-edge offsets measured for interfaces between couples of different compositions x (symbols at valence and conduction bands) (After Langer and Heinrich 1985a, b)

acceptor state, predictions of band alignments for a wide range of host zincblende and wurtzite compound semiconductors were given.

The assignments were based on the finding of a rather constant +/- transition level of H at 4.5 eV below the vacuum level, hence representing a common reference level for tetrahedrally bonded semiconductors. The reason for this behavior is referred to dangling bonds created by hydrogen. In the positive charge state, H^+ bonds to the anion (e.g., As, Se) by breaking a cation–anion bond of the host; this creates a dangling bond on the cation. The dangling bond has an energy near the conduction-band minimum and is hence favorably unoccupied and positively charged. Conversely, an H^- bond to the cation (Ga, Zn) by breaking the cation–anion bond leaves an anion dangling bond with an energy level near the valence-band maximum; this level is preferentially occupied and negatively charged. If the Fermi level is located midway between the two dangling-bond energies, H^+ and H^- have equal formation energy. This location of E_F corresponds to the charge-neutrality level discussed in Sect. 2.3.2, where it is employed to line up semiconductor band structures.

2.2.5 The Frensley-Kroemer Model

Frensley and Kroemer (1977) suggested to take the average interstitial potential $V_{\text{interstitial}}$ as a reference potential, given by the average electrostatic potential at the midpoint between adjacent atoms. Relative to this interstitial potential, they computed the position of the valence-band edge from pseudopotential bandstructure calculations, including the exchange interaction using a Slater approximation. The agreement of this simple model with the experiment is surprisingly good (Tables 1 and 3). However, Frensley and Kroemer then added a term due to an interface dipole $\Delta E_{\text{v}} = E_{\text{v}}(2) - E_{\text{v}}(1) + V_{\text{dipole}}$ and assumed a V_{dipole} given by the average Philips electronegativities of all component elements. Unfortunately, the inclusion of this dipole term usually worsens the agreement with experimental results. One therefore needs to look for better ways to account for the nonlinearities induced by the interface. This will be discussed in the next section.

2.3 Interface Dipole

2.3.1 Heterovalent Interfaces

The interface between two semiconductors with different ionicity of the bonding, e.g., ZnSe/GaAs(001) or Ge/GaAs(001), creates polar charges. In order to understand the general principles involved, we simplify the interface by assuming ideal tetrahedral bonding and negligible bond-polarization effects.

The Ge/GaAs(001) interface can be arranged in several different ways, with significant consequences for the interface potential. This results in a net charging of the Ga and As atom of $+e$ and $-e$ with respect to the Ge atom. When proceeding with a probe in [110] direction from the Ge through the Ge/GaAs(110) interface, the average potential (as calculated by integrating the Poisson equation) remains the same as in Ge (here assumed to be zero): the atomic planes in GaAs parallel to the interface contain an equal amount of cations and anions and are consequently not charged on average; see Fig. 12a.

When now turning the crystal and looking at the Ge/GaAs(00l) interface, we recognize that each atom plane parallel to the interface is charged (Fig. 12b). Here a potential gradient occurs when proceeding from Ge and crossing the first charged Ga layer; the gradient becomes zero (the potential remains constant) after crossing the oppositely charged As plane and reappears at the following Ga plane. Consequently, one obtains a staircase-like increase of the potential starting from the interface. Obviously, this is impossible to maintain for any larger sequence of steps since the built-in field would create a dielectric breakdown (Grant et al. 1987).

This problem can be easily eliminated by an atomic rearrangement shown in Fig. 13a with an alternating protrusion of Ge and Ga atoms across the interface. This causes the charge at this interface to be one half of the flat interface with half the potential step height as a result. The next (full) As layer then overcompensates the first charge and causes a potential step-down. The following alternating full layers result in an oscillation of the potential with an average potential step at the interface of δ, representing a net dipole of about 0.1 eV there.

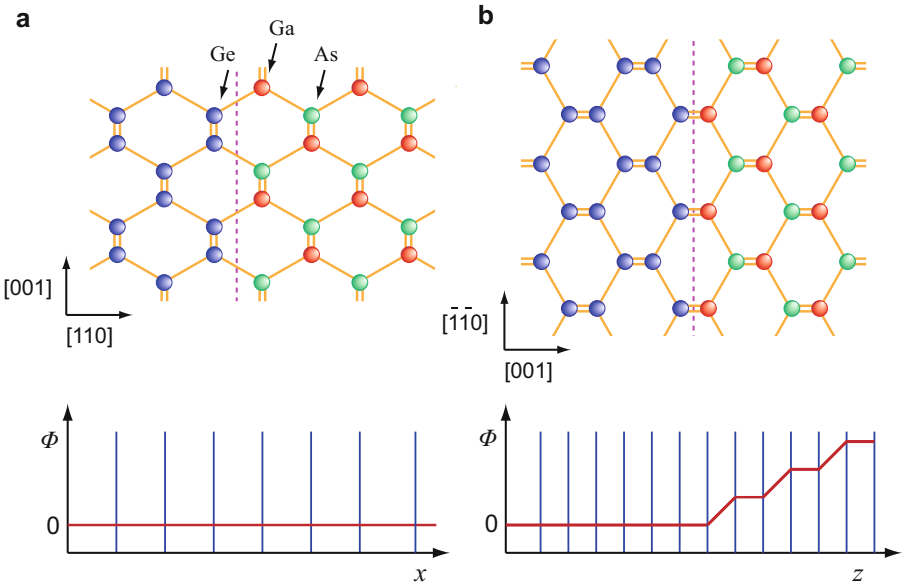


Fig. 12 Ge/GaAs heterojunction. The *double lines* shown are two tetrahedral bonds projected onto the plane of the figure. The *lower part* of the figure gives the potential integrated by probing normal to the interface (from *left* to *right*) with the average potential in Ge assumed to be zero. (**a**) The Ge/GaAs(110) heterojunction with a perfectly flat interface; (**b**) the Ge/GaAs(001) heterojunction with a perfectly flat interface starting with a Ga layer. The potential averaged over planes parallel to the interface shows a staircase-type increase when proceeding into the GaAs lattice (After Harrison et al. 1978)

A third alternative is shown in Fig. 13b for an interface which is corrugated by three layers, providing the possibility to imbed 1/16 of a full layer of As, followed by 1/2 of a layer of Ga, and then followed by a layer that contains 1/16 of a full layer of Ge, replacing As. The result of this corrugation is the elimination of the interface dipole, as shown at the bottom of Fig. 13b. It is suggestive that even during the most careful sequential MBE deposition, the self-regulating forces to minimize the interface energy (i.e., here to minimize the interface dipole) are not operating to full perfection. The probability of obtaining an atomic interface misalignment is indeed larger than that for generating intrinsic point defects which are basically controlled by nearest-neighbor interaction. The controlling forces at a polar interface extend over larger lattice distances and probably can be expected to cause on an average the dipole to vanish, however, with large local fluctuations (Kroemer 1983). This may explain the substantial fluctuation of ± 0.1 eV of the band offsets that are observed even under the most careful deposition of heterojunctions.

2.3.2 Interface-Dipole Theory

The dominant role of charge accumulation at the interface rather than the effect of bulk properties was applied by Tersoff for calculating heterojunction-band offsets

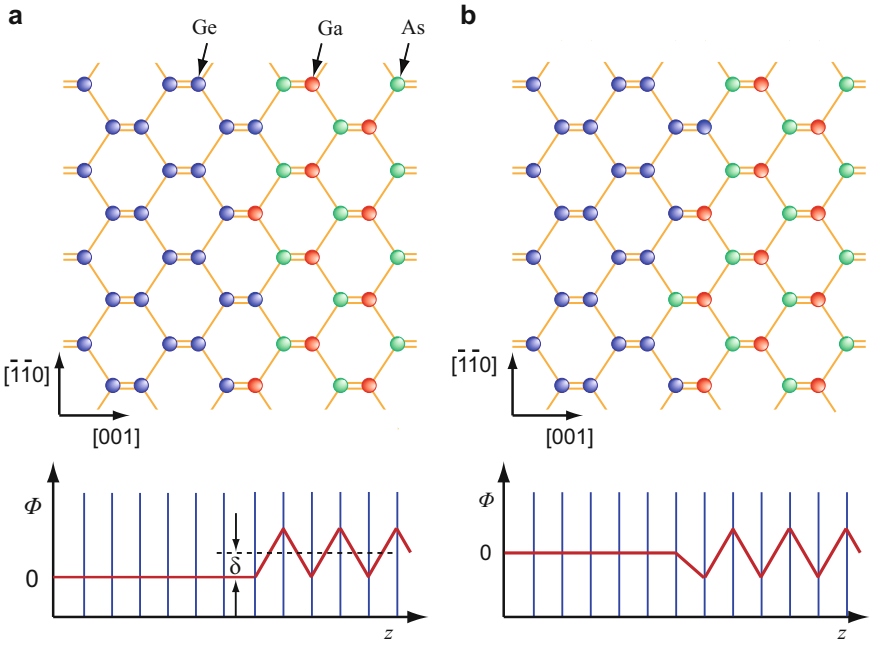


Fig. 13 Ge/GaAs(001) heterojunction as in Fig. 12. (**a**) Corrugated interface with Ge and Ga atoms alternatingly exchanged, resulting in an interface dipole of δ; (**b**) corrugated interface involving three layers with As and Ge penetrating one layer each beyond the average position of the interface. The resulting average potential shows a vanishing dipole (After Harrison et al. 1978)

(Tersoff 1984a, 1986) and barrier heights at a metal-semiconductor interface (Tersoff 1984b). According to this model, there is generally a dipole at a heterointerface associated with electronic states in the bandgap, which are induced by the band discontinuity; these states induced in one of the semiconductors are analogous to the metal-induced gap states discussed in Sect. 3.3.3. States lying near the conduction-band edge at side *A* of the interface have exponentially decaying tails into side *B*; at side *B* they lie in the gap of the semiconductor (Fig. 14). Any state in the gap has a mixture of valence- and conduction-band character. Occupying such a state leads locally to an excess charge, according to its degree of conduction character. Filling a state which lies near the top of the gap gives a large excess charge of almost one electron due to a large conduction character. Leaving that state empty gives an only slight charge deficit. Conversely filling a state near the bottom of the gap at side *A* results in a slight excess charge in proportion to its little conduction character, while leaving it empty leads to a charge deficit of almost one electron; states at the bottom of the bandgap are occupied, and those at the top are unoccupied. The band discontinuity at the interface hence induces a net dipole. The resulting dipole always acts to oppose the leaking out

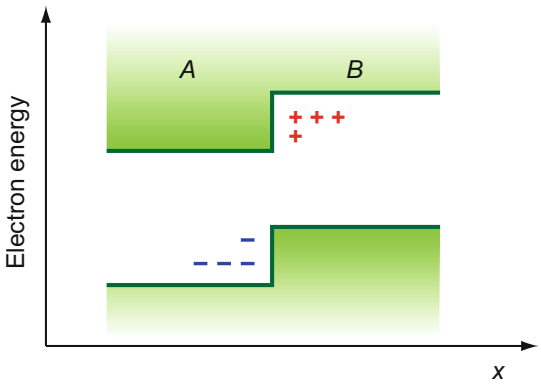


Fig. 14 Scheme of band alignment at a semiconductor heterojunction. (+) and (–) represent net charges of unoccupied states with electron deficit and occupied states with electron excess, respectively

of carriers; it consequently reduces the band offset, i.e., the potential step is screened by a factor of ε, the bulk dielectric constant.

The dipole drives the lineup toward a minimum dipole moment. A condition for zero-dipole lineup is the alignment of the characteristic energies (*effective midgap energies* or *branch-point energies*) separating the bonding and antibonding character of the states in the two semiconductors. Such reference levels are the effective midgap energies E_B for each semiconductor, which are calculated from the band structure. The band alignment between two semiconductors A and B is then given by

$$\Delta E_v = E_B(B) - E_B(A). \tag{21}$$

For a heterojunction E_B plays a role in analogy to the Fermi energy in metals: E_B is aligned for the respective semiconductors. Results of the interface-dipole theory are given in Table 3 for some heterojunctions.

2.3.3 Disorder-Induced Gap-State Model

Another means to align the valence bands is by using an average hybrid-orbital energy E_{HO} which for numerous semiconductor-insulator and semiconductor-metal interfaces is observed to lie at

$$E_{HO} - E_0 \cong 5 \pm 0.1 \text{ eV} \tag{22}$$

in reference to the vacuum level E_0 (Hasegawa and Ohno 1986; Hasegawa et al. 1986). Analogous to Harrison's model (Sect. 2.2.3), the reference energy E_{HO} is derived from atomic s and p states,

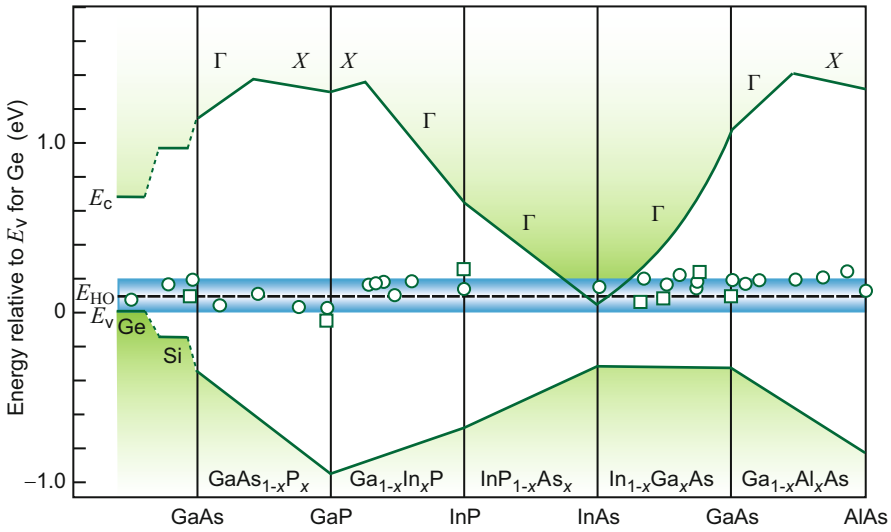


Fig. 15 Lineup of experimental valence and conduction band offsets in respect to a reference level E_{HO} that is 0.1 eV above E_v(Ge) and 5.0 eV below the vacuum level E_0. This level is identified as *horizontal dashed line*. Experimental values of the Fermi-level pinning observed at heterojunction interfaces (*squares*) and metal-semiconductor Schottky barriers (*circles*) show a fluctuation of less than ±0.1 eV around E_{HO} (After Hasegawa and Ohno 1986)

$$E_{HO} = -\left(\frac{E_s + 3E_p}{4}\right)_{average} + \Delta U, \tag{23}$$

where ΔU is a correction accounting for electron-electron interaction (Hasegawa and Ohno 1986; see also Harrison and Tersoff 1986). Figure 15 gives an illustration of the observed lineup of valence-band offsets after calibration to the observed pinning of the Fermi level in *p*-type Ge, taken as $E = 0$. The observed values for Fermi-level pinning for other semiconductor-insulator or Au-semiconductor interfaces lie then within a band of ±0.1 eV around the $E = 0$ level.

Hasegawa and Ohno (1986) assume that a thin disordered layer at the interface always causes disorder-induced gap states (DIGS) that pin the position of the Fermi level within the gap. This thin layer is characterized by fluctuation of bond lengths and angles and possibly other defects, including interface stress, and may cause a sufficient level density for Anderson localization (Sects. 1 and 2.2 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors") yielding a DIGS continuum, consisting of bonding and antibonding states. The Fermi level is determined by charge neutrality, i.e., at the neutrality level which is identified here as E_{HO}. Such DIGS were more recently assigned to the origin of instabilities in GaN-based MOS transistors (Matys et al. 2017).

2.4 Ab Initio Approaches

Ab initio calculations provide another access to the valence-band offset. It can be shown within these calculations that a dependence of ΔE_v on interfacial strain and, to a lesser extent, on crystal orientation exists, as demonstrated for pseudomorphic interfaces by Van de Walle and Martin (1985, 1986), and Tung and Kronik (2016, 2019), using a self-consistent local density-functional theory in conjunction with nonlocal or hybride functionals. Strain seems to have generally a strong effect.

With an improved theoretical basis, supercell calculations are performed: dipoles at the interface can be extracted from valence-electron distribution, which are found to be only three to four lattice planes apart from the interface different from the corresponding bulk values. Therefore, a supercell of more than five monolayers, calculating the electron structure of such a superlattice, is often sufficient to obtain the band offsets.

There are largely five first-principle approaches for the calculation of band alignment, where more recent calculations combine density-functional theory and many-body GW calculations; for reviews see Robertson (2013), Tung and Kronik (2016).

1. Explicit modeling of heterointerfaces (a) using potential alignment from superlattice calculations combined with bulk band-structure calculations, and including atomic relaxations (Weston et al. 2018). Former modeling (b) also used natural valence-band offsets with a band alignment assuming transitivity by averaging offset values for various atomic configurations at the interface (Van de Walle and Martin 1987; Li et al. 2009).
2. Explicit heterointerface modeling by minimizing the charge density at the interface. By separating the electrostatic effect into neutral polyhedra about every anion and cation, local electron densities are generated. The valence-band offset then follows from the difference of the two bulk terms plus an interface dipol representing the difference between the plane-averaged electrostatic potential energies across the interface (Tung and Kronik 2016).
3. Calculation of charge neutrality levels (branch-point energies) from bulk band-structures to define an internal reference energy; at such a level the character of electronic states changes from valence-band-like to conduction-band-like character (Tersoff 1984a; Mönch (1996), Schleife et al. (2009).
4. Calculation of localized levels of impurities located deep in the bandgap, such as levels of transition metals (Ledebo and Ridley 1982; Caldas et al. 1984) or hydrogen (Van de Walle and Neugebauer 2003).
5. Alignment of valence and conduction bands with respect to the vacuum level implying the ionization potential and electron affinity from surface calculations (Höffling et al. 2012; Hinuma et al. 2014; Stevanovic et al. 2014).

In the explicit interface modeling of layer A upon substrate B (1a) by Weston et al. (2018) atomic slabs of sufficient thickness with alternating A and B crystals are fixed to the bulk in-plane lattice constant of B, and the out-of-plane lattice parameter of A is relaxed. The valence-band maxima are obtained from the electronic band structure calculated for this superlattice, which allows for aligning the electrostatic potentials in the bulk-like regions of A and B. The *model solid* (1b) constructed by Van de Walle and Martin (1987) for lattice-matched systems and mismatched systems (Van de Walle 1989) is an approach different from (2) and (3) using an internal reference energy; the model is able to reproduce closely the results of their supercell calculations in a simplified manner.

The neutral-polyhedra theory (2) introduced by Tung and Kronik (2016), Tung (2014) constructs a model solid distinct from (1b) by filling the entire crystal volume with anion-centered or cation-centered Wigner-Seitz cells, each individually neutral. A model solids composed of such building blocks has well-defined band-edge positions. The overlapping volume of two neighboring anion- and cation-centered cells contains a single sp^3 bond; a sheet of such volumes has no net charge, but a net dipole moment that leads to a difference in potential energy and is applied for band alignment at an interface.

The approaches (1) and (2) are discussed in more detail below. The idea of branch-point energies (3) was pointed out in Sect. 2.3.2, and the model of using localized deep-level impurities (4) was already introduced in Sect. 2.2.3. For a review on different theoretical approaches, see Franciosi and Van de Walle (1996), Hinuma et al. (2014).

2.4.1 The Model-Solid Approach

The model-solid approach suggested by Van de Walle and Martin (1987), Van de Walle (1989) emulates the classical electron affinity rule Eqs. 12 and 13 by constructing a *local* reference level and avoiding dipoles. Within this approach the charge density in a semiconductor is composed by a superposition of *neutral atoms*. The potential outside each such sphere goes exponentially to zero. This zero is taken as reference level. The construction leads to a well-defined electrostatic potential with respect to the vacuum level *in each atom*. By superposition the average electrostatic potential in a model solid composed of such atoms is hence specified on an absolute energy scale. The electron configuration of an atom in the solid is determined from a tight-binding calculation. This leads for, e.g., one Si atom in a silicon bulk crystal to 1.46 s and 2.54 p electrons, meaning that a part of the two s electrons of a Si atom are excited into the p band.

The result of the calculation is the position of the valence band on some absolute energy scale, allowing to relate it to the respective value of another semiconductor. For semiconductors with zincblende or diamond structure, a simplified one-band approach is used by averaging out valence-band splitting, yielding $E_{\mathrm{v,av}}$ with the uppermost valence band at

$$E_{\mathrm{v}} = E_{\mathrm{v,av}} + \frac{\Delta_0}{3}, \tag{24}$$

Δ_0 being the spin-orbit parameter; spin-orbit effects are added a posteriori. The change of the valence-band edge at the interface due to the hydrostatic part of the strain can be estimated from

$$\Delta E_{\mathrm{v,av}}^{\mathrm{strain}} = a_{\mathrm{v}} \frac{\Delta V}{V} = a_{\mathrm{v}} \left(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}\right), \tag{25}$$

with a_{v} as the hydrostatic deformation potential for the valence band and ε_{ii} as the diagonal elements of the strain tensor (Sect. 1.1 of ▶ Chap. 4, "Elasticity and Phonons"). In non-isotropic semiconductors, the shear strain causes additional band-edge deformation and splitting that need to be considered (Van de Walle 1989). The total valence-band offset is then given by

$$\Delta E_{\mathrm{v}}^{\mathrm{tot}} = \left(E_{\mathrm{v}}^{(A)} - E_{\mathrm{v}}^{(B)}\right) + \left(\Delta E_{\mathrm{v,\,av}}^{\mathrm{strain}(A)} - \Delta E_{\mathrm{v,\,av}}^{\mathrm{strain}(B)}\right), \tag{26}$$

with the second bracket caused by lattice strain. The offset of conduction bands cannot be calculated by a mere shift of each valence-band edge by the bandgap $E_{\mathrm{g}}^{(i)}$ but must also consider the strain, requiring an estimate corresponding to Eq. 25:

$$\Delta E_{\mathrm{c}}^{\mathrm{strain}} = a_{\mathrm{c}} \left(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}\right) \tag{27}$$

with a_{c} as the conduction-band deformation potential. Table 4 lists the values for the relevant parameters of various semiconductors. The band-edge offset therefore depends to a large extent on the substrate or any strain-accommodating interlayer onto which a strained layer is deposited.

2.4.2 Density-Functional Theory Combined With Many-Body Perturbation

Combining density-functional theory (DFT) and the many-body perturbation theory has significantly improved the precision of ab initio models (Aulbur et al. 2000). The perturbing interaction of an electron with its environment is included by applying a quasiparticle concept (Hedin 1965, Van de Walle 2019): quasiparticles composed of electrons and their respective positive polarization clouds interact via a screened Coulomb interaction W; see Sect. 2.2.3 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals".

The energy difference between the bare particle states and those of the quasiparticle are described by the self-energy. Various approaches were developed to account for this nonlocal energy, which comprises all exchange and correlation effects. The widely applied *GW* approximation (*GWA*) is, e.g., complemented by

Table 4 Average valence-band energies $E_{v,av}$ and deformation potentials of the valence band a_v and the conduction band a_c, calculated from the model-solid theory. Δ_0 and E_g denote measured spin-orbit splitting and energy gap at 0 K, respectively. All values are given in eV (After Van de Walle 1989)

Solid	$E_{v,av}$	a_v	a_c^{dir}	E_g^{dir}	E_c^{dir}	a_c^{indir}	E_g^{indir}	E_c^{indir}	Δ_0
Si	−7.03	2.46	1.98	3.37	−3.65	4.18	1.17	−5.85	0.04
Ge	−6.35	1.24	−8.24	0.89	−5.36	−1.54	0.74	−5.51	0.30
AlP	−8.09	3.15	−5.54	3.63	−4.46	5.12	2.51	−5.58	0.07
GaP	−7.40	1.70	−7.14	2.90	−4.47	3.36	2.35	−5.02	0.08
InP	−7.04	1.27	−5.04	1.42	−5.58				0.11
AlAs	−7.49	2.47	−5.64	3.13	−4.27	4.09	2.23	−5.17	0.28
GaAs	−6.92	1.16	−7.17	1.52	−5.29				0.34
InAs	−6.67	1.00	−5.08	0.41	−6.13				0.38
AlSb	−6.66	1.38	−6.97	2.32	−4.12	3.05	1.70	−4.74	0.65
GaSb	−6.25	0.79	−6.85	0.75	−5.32				0.82
InSb	−6.09	0.36	−6.17	0.24	−5.58				0.81
ZnS	−9.15	2.31	−4.09	3.84	−5.29				0.07
ZnSe	−8.37	1.65	−4.17	2.83	−5.40				0.43
ZnTe	−7.17	0.79	−5.83	2.39	−4.48				0.91

first-order vertex corrections in the screened Coulomb interaction ($GW\Gamma^1$) (Grüneis et al. 2014). Further improvements comprise the Perdew-Burke-Ernzerhof (*PBE*) semilocal functional (Perdew et al. 1996) and the Heyd-Scuseria-Ernzerhof (*HSE*) hybrid functional (Heyd et al. 2003, 2006). Hybrid functionals mix a portion of the nonlocal exact Hartree-Fock exchange with local-density approximation (LDA) or GGA exchange in the short-range exchange potential, applying typically a mixing parameter α of 0.25 for the amount of Hartree-Fock exchange. The optimum mixing parameter may be determined by fitting to the experimental bandgap energy (Alkauskas et al. 2008). Results of calculations comparing the PBE and HSE models are reported by Hinuma et al. (2014) and Weston et al. (2018).

The more precise hybrid-functional (HSE) calculations are computationally very demanding for large-scale systems such as interfaces. Since the average electrostatic potential of a solid is not well defined on an *absolute* energy scale, band-structure calculations for individual bulk crystals must be complemented by explicit interface calculations for aligning the average electrostatic potential at the contact of the two solids. This can be achieved by explicitly calculating a superlattice structure of sufficient thickness[13] to contain bulk-like regions of each layer, and examining the average of the electrostatic potential V determined in these bulk-like regions (Van de Walle and Martin 1987).

[13]Typically, the electrostatic potential converges fast (a few lattice planes) as one moves away from the interface.

An approach of reasonable precision (energy within 50 meV) at acceptable computational cost is achieved by combining GGA and HSE calculations (Weston et al. 2018). Valence- and conduction-band edges of the infinite bulk of the two individual solids (on a relative scale) were obtained from HSE calculations, applying mixing parameters α determined for each solid by fitting to the experimental bandgap energy. The valence-band offset ΔE_v at the contact for A/B (i.e., A on B) is given by

$$\Delta E_\mathrm{v} = E_{\mathrm{v},B} - E_{\mathrm{v},A} + e\Delta V, \tag{28}$$

where ΔV is the potential alignment; a positive value of ΔE_v indicates that the valence-band maximum is higher in material B. This quantity is obtained from an explicit calculation of a fully relaxed superlattice. It turned out that this step can be performed by the much less demanding GGA calculations, which agree within 50 meV with alignments obtained from HSE calculations (Weston et al. 2018). ΔV is a macroscopic average computed from the varying electrostatic potential $V(\mathbf{r})$ in two steps; first a planar average is formed by

$$V(z) = (1/A)\int_A V(\mathbf{r})\, dx\, dy, \tag{29}$$

with the area A of a unit cell in the interface plane. $V(z)$ exhibits periodic oscillations along z due to the ionic cores as shown in Fig. 16. In a second step, the quantity V (constant in the bulk) is obtained from

$$V = (1/L)\int_{-L/2}^{L/2} V(z)\, dz; \tag{30}$$

here, L is the spatial period of one oscillation along z. The potential alignment ΔV is given by the difference $V_B - V_A$ of the two solids B and A; the electrostatic potential

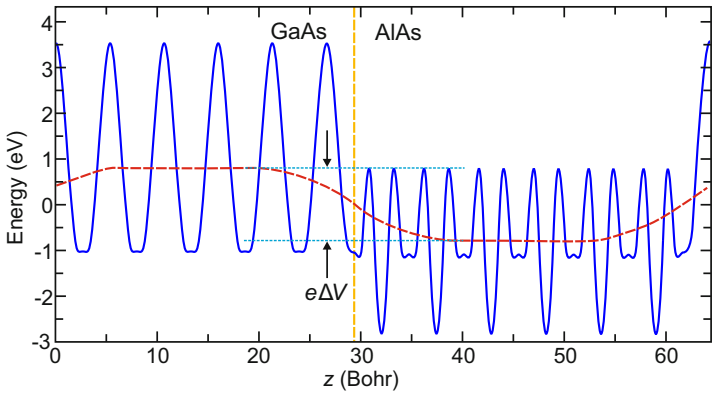


Fig. 16 Computed planar averaged (*blue curves*) and macroscopically averaged (*red curve*) electrostatic energy of a (001)-oriented GaAs/AlAs superlattice. The yellow line indicates the interface plane containing As atoms. (After Weston et al. 2018)

Table 5 Valence-band offsets in eV for various *A*/*B* heterojunctions (i.e., *A* on *B*); positive values refer to $E_v(A) > E_v(B)$. *f* is the lattice mismatch defined by $(a_B\text{-}a_A)/a_A$, first and second values of hexagonal structures refer to *a* and *c* axes. *Heterojunction* indicates the interface plane being polar (001) or nonpolar (110) for cubic or *m*-plane $(10\bar{1}0)$ for hexagonal structures; GGA (of PBE), HSE, and GWΓ^1 denote models indicated in the text. (After Hinuma et al. 2014 (a) and Weston et al. 2018 (b))

Heterojunction *A*/*B*	*f* (%)	Theory $\Delta E_{v,calc}$ GGA	HSE	(a) HSE + GWΓ^1 (b) HSE + GGA	Experiment $\Delta E_{v,exp}$
Ge/Si(001)[b]	−4.01	−0.74	−0.86	−0.82 [b]	−0.7...−0.8
Si/GaP(110)[a]	0.37	0.32	0.40	0.51 [a]	0.80
AlP/GaP(001)[b]	−0.24	0.39	0.60	0.63 [b]	0.57
Ge/AlAs(110)[a]	0.05	0.95	1.10	1.19 [a]	0.95 ± 0.2
Ge/ZnSe(110)[a]	0.19	0.96	1.52	1.85 [a]	1.52
GaAs/AlAs(110)[a]	0.12	0.45	0.47	0.55 [a]	0.40
AlAs/GaAs(001)[b]	−0.16	0.39	0.51	0.50 [b]	0.55
GaSb/AlSb(001)[a]	0.66	0.42	0.45	0.48 [a]	0.40 ± 0.15
AlN/GaN$(10\bar{1}0)$[b]	2.51/4.15	0.32	0.62	0.62 [b]	0.70
ZnO/GaN$(10\bar{1}0)$[b]	−1.85/−0.29	0.27	1.28	1.25 [b]	0.8...1.0

of the GaAs/AlAs structure in Fig. 16 is calculated applying HSE functionals. Band offsets for cubic and hexagonal semiconductors obtained from this approach are given in Table 5; band alignments of Zn-IV-nitrides are reported in Adamski et al. (2020).

2.4.3 The Neutral-Polyhedra Theory

A model solid distinct from the approach above was constructed for tetrahedrally bonded zincblende semiconductors by Tung and Kronik (2016, 2014). In this approach, the entire crystal volume is filled with Wigner-Seitz cells centered, e.g., about anions. An anion-centered cell has ¼ negative cation at four of the eight corners. Such a cell has vanishing net charge, dipole moment, and quadrupole moment; it may hence constitute a building block for a model solid with well-defined average internal potential and band-edge energies (Tung and Kronik 2016). The same applies analogous for cation-centered cells. Two neighboring anion- and cation-centered cells have an overlapping volume, which contains ¼ anion, ¼ cation, and two valence electrons. The main character of this volume ("*quadrant*") is assumed to reflect that of the single comprised sp^3 bond between the neighboring anion and cation. A sheet of such quadrants has no net charge, but a net dipole moment that leads to a difference in potential energy across the sheet.

The dipole term of a quadrant sheet at an interface between two dissimilar crystals corresponds to the potential alignment term in Eq. 28. In an explicit modeling of a hetero-interface, the valence-band offset follows from the difference

of the two bulk terms plus this plane-averaged interface-dipol term. The calculations were performed within DFT, using the PBE exchange-correlation functionals for unrelaxed and relaxed heterojunctions with different orientations. Both the relaxation and the type of interface bonds (interface at anion or cation plane) are generally found to play a substantial role; a difference of ~0.2–0.8 eV is found in the valence-band offset of either unrelaxed or relaxed heterojunctions. On the other hand, the computed offset is largely insensitive to the interface orientation and interface bond identity (Tung and Kronik 2016). Results for band offsets of some zincblende junctions are given in Table 5; offsets for perovskite oxide heterojunctions are reported in Tung and Kronik (2019).

3 Metal-Semiconductor Interfaces

Metal-semiconductor interfaces are special types of heterojunctions. The electrical properties of all *metal contacts* are determined by this interface.[14] The most important parameter of a metal/semiconductor interface is its barrier height ϕ_B or the related contact potential (Sect. 2.1.2), which determines the electron density n at this interface. Like the difference in electron affinities at a heterojunction, there are many factors which determine the barrier height, such as the chemical nature of metal and semiconductor and the actual structure of the interface.

3.1 Classification of Semiconductor-Metal Interfaces

For metal-semiconductor boundaries one distinguishes so-called *blocking* (or *rectifying*) *contacts*, which show a barrier layer at the interface as given in Fig. 17a, d and *injecting* (or *ohmic*) *contacts*, shown in subfigures b, c. They are determined in the simple Schottky model by the relative difference of workfunctions: for a connection to *n-type* semiconductors, the ratio of the electron density at the metal-semiconductor interface is given by

$$n_c = N_c \exp\left(\frac{\phi_M - \chi_{Sc}}{kT}\right), \tag{31}$$

where N_c is the effective level density at the conduction-band edge of the semiconductor (Sect. 1.3 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"); the density of electrons in the bulk of an *n*-type semiconductor is

[14]This applies not for *point contacts*; here, high-field tunneling effects (see ▶ Chap. 29, "Carrier Generation") and phonon coupling determine the electrical behavior.

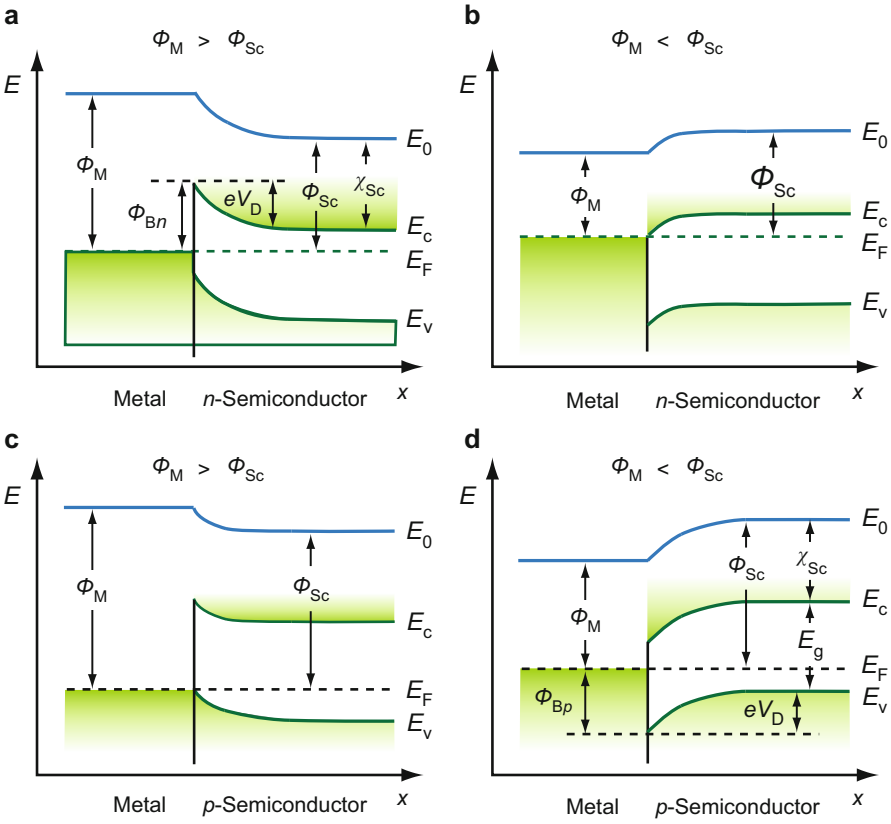


Fig. 17 Energy-band diagrams of ideal junctions between metal and an either *n*-type semiconductor (**a**, **b**) or a *p*-type semiconductor (**c**, **d**). Metal workfunctions ϕ_M being larger (**a**, **c**) or smaller (**b**, **d**) than the workfunctions of the semiconductors ϕ_{Sc} are assumed

$$n_0 = N_c \exp\left(\frac{E_c - E_F}{kT}\right). \tag{32}$$

For a *blocking contact*, this ratio is $n_c/n_0 \ll 1$, and for an *injecting contact*, it is $n_c/n_0 \gg 1$. These are shown in Fig. 17 as panels *a* and *b*, respectively.

In a similar fashion one has blocking or injecting contacts for *p-type* semiconductors depending on whether the ratio of hole densities $p_c/p_0 \ll 1$ or $p_c/p_0 \gg 1$, respectively, with p_c and p_0 defined analogous to Eqs. 31 and 32. In *p*-type semiconductors, *holes* are the majority carriers and the blocking barrier occurs in the valence band; see Fig. 17d.

3.2 The Classical Schottky-Mott Model

The Schottky-Mott model of the metal-semiconductor contact goes back to Schottky et al. (1931). It was further developed by Schottky (1939), Mott (1938, 1939), and Davydov (1938) with most of the attention paid to the carrier transport through the space-charge region and control by the barrier at the interface.

We assume the conditions presumed for the Anderson model (Sect. 2.2.1) and consider an *n*-type semiconductor connected to a metal of a sufficiently high workfunction as illustrated in Fig. 17a. When forming the contact, electrons from the semiconductor leak out into the adjacent metal.[15] The electron density at the interface between the metal electrode and the semiconductor is reduced below its equilibrium bulk value n_0, and thereby a positive space-charge region is created within the semiconductor near the metal contact. The corresponding negative charge to render the total device neutral is located at the metal contact-surface. The space-charge layer in the semiconductor results in a field ramp and a potential step (Fig. 17a), referred to as the *Schottky barrier*.[16]

At the metal-semiconductor interface, the band bending extends to a value eV_D, which is the same as the difference of the Fermi levels before equilibrium or the difference of the workfunctions $\phi_M - \phi_{Sc}$. The barrier height counted from the Fermi level is given for an *n*-type semiconductor by

$$\phi_{Bn} = \phi_M - \phi_{Sc} + (E_c - E_F)_{bulk} = \phi_M - \chi_{Sc}, \tag{33}$$

as shown in Fig. 17a or for a *p*-type semiconductor by

$$\phi_{Bp} = E_g + \chi_{Sc} - \phi_M = I_{Sc} - \phi_M \tag{34}$$

shown in Fig. 17d. ϕ_{Bn} or ϕ_{Bp} are identified as the Schottky barrier height.

Workfunctions ϕ_M for a number of metals as contacts to some typical semiconductors are given in Table 6. Since the values sensitively depend on cleanliness and microscopic structure of the surface, reported data show a considerable scatter; listed values may be regarded as typical for nominally clean surfaces. For electron affinities ϕ_{Sc} of the semiconductor, see Table 2.

The electron density at the metal-semiconductor interface is given by

[15]Even though the electron density inside a metal is much higher than in the semiconductor, at its boundary to the semiconductor, this density is substantially reduced according to its effective workfunction. It is this electron density which causes a reduction of *n* in the semiconductor at the interface.

[16]A similar Schottky barrier appears in *p*-type semiconductors near a metal electrode with low workfunction (Fig. 17d), again when the hole density near the electrode is much smaller than in the bulk. Here the space-charge region is negatively charged and the resulting field is positive.

Table 6 Workfunction ϕ_M of metals determined photoelectrically except for: *TE* thermionic emission, *FE* field emission (After Hayes et al. 2013)

Metal	Face	ϕ_M (eV)	Metal	Face	ϕ_M (eV)
Ag	(100)	4.64	Nb (TE)	(100)	4.02
	(110)	4.52		(110)	4.87
	(111)	4.74		(111)	4.36
Al	(100)	4.20	Ni	(100)	5.22
	(110)	4.06		(110)	5.04
	(111)	4.26		(111)	5.35
Au	(100)	5.47	Pd	(111)	5.6
	(110)	5.37		Polycr.	5.22
	(111)	5.31	Pt	(110)	5.84
Cu (FE)	(100)	5.10		(111)	5.93
	(110)	4.48		Polycr.	5.64
	(111)	4.94	Ta (TE)	(100)	4.15
Fe	(100)	4.67		(110)	4.80
	(111)	4.81		(111)	4.00
Hg	Liquid	4.475	Ti	Polycr.	4.33
Ir	(100)	5.57	U	(100)	3.73
	(110)	5.42		(110)	3.90
	(111)	5.76		(111)	3.67
Mo	(100)	4.53	W (FE)	(100)	4.63
	(110)	4.95		(110)	5.22
	(111)	4.55		(111)	4.45
			Zn	Polycr.	3.63

$$n_c = n(x = 0) = N_c \exp\left(-\frac{\phi_{MS}}{kT}\right), \tag{35}$$

where ϕ_{MS} ($= \phi_M - \chi_{SC}$ in the linear model) is the *metal-semiconductor workfunction* (see Eq. 31) and N_c is the effective level density[17] at the metal-semiconductor interface. This density n_c is assumed here to be independent of current and applied voltage. The electron density in the bulk is given by the density of the shallow, uncompensated donors

$$n_0 \cong N_D. \tag{36}$$

When n_c is *lower* than the electron density in the bulk, a depletion region results near the contact to the metal producing a *blocking* contact. We discuss an example with a donor concentration $N_D = 10^{17}$ cm^{-3} and a lower electron density at the boundary of $n_c = 10^{10}$ cm^{-3}. In Fig. 18a we show the electron distribution computed from Eqs. 38–40 for $T = 300$ K, $\varepsilon = 10$, and a mobility $\mu = 100$ cm^2/(Vs). Because

[17]This is slightly different from N_c within the semiconductor bulk because of a different effective mass at the interface.

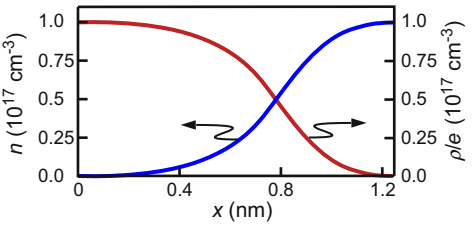


Fig. 18 Typical electron density (*blue curve*) and space-charge distribution (*red*) in a Schottky barrier computed from Eqs. 38, 39, and 40

of the large ratio of the bulk-to-surface carrier densities n_0 and n_c, the electron density in the space-charge region rapidly decreases to values very small compared to the donor density N_D, thus rendering the space charge

$$\rho(x) = e[p_D - n(x)] \cong e[N_D - n(x)] \cong eN_D \quad \text{for } 0 \le x < x_D, \tag{37}$$

independent of n in a substantial fraction of this junction region; p_D is the density of positively charged, ionized donors.

In using this *constant* space charge[18] within the entire width of the Schottky barrier (i.e., $0 \le x < x_D \cong 80$ nm) in this example, the resulting approximation – referred to as the *Schottky approximation* – permits a major simplification of the governing set of equations:

$$\frac{dn}{dx} = \frac{j_n - e\mu_n nF}{\mu_n kT} \tag{38}$$

$$\frac{dF}{dx} = \frac{eN_D}{\varepsilon_{stat}\varepsilon_0} \tag{39}$$

$$\frac{d\psi_n}{dx} = F \tag{40}$$

where ψ_n is the electrostatic potential and F is the electric field. This allows a decoupling of the Poisson equation from the transport equation. Integration of Eq. 39 yields

$$F(x) = F_c + \frac{eN_D}{\varepsilon_{stat}\varepsilon_0}x; \tag{41}$$

the field decreases in the Schottky approximation linearly with increasing distance from the metal-semiconductor interface (see dashed line in Fig. 19b), with F_c, the

[18]The error encountered at the boundary of this range (here 80 nm) seems to be rather large (factor 2) when judging from the linear plot of Fig. 18. The accumulative error, when integrating from the metal/semiconductor interface, however, is tolerable, as shown in Fig. 19. The substantial simplification in the mathematical analysis justifies this seemingly crude approach.

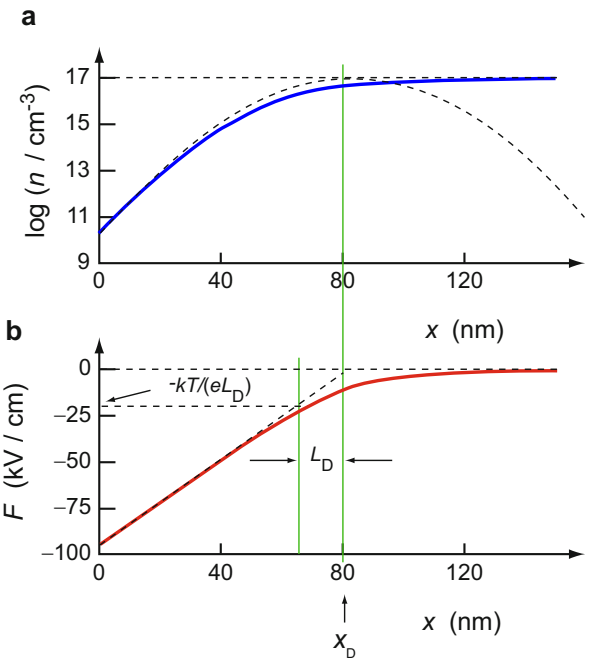


Fig. 19 (**a**) Electron density and (**b**) electric field of a Schottky barrier for parameters given in the text. *Solid curves*, exact solutions of Eqs. 38–40; *dashed curves*, Schottky approximation computed from Eqs. 41 and 47

maximum value of the field at $x = 0$, used here as the integration constant. From the integration of Eq. 40 after insertion of Eq. 41, one obtains the electrostatic electron potential

$$\psi_n(x) = \psi_{n,\mathrm{D}} + F_\mathrm{c}\, x + \frac{eN_\mathrm{D}}{2\varepsilon_\mathrm{stat}\varepsilon_0} x^2, \tag{42}$$

which decreases parabolically with increasing x. As integration constant we have used the *electron diffusion-potential* $\psi_{n,\mathrm{D}}$ ($= V_\mathrm{D}$ in Fig. 17) which is appropriate for zero current. For a finite current, the solutions $F(x)$ and $\psi_n(x)$ have exactly the same form (Eqs. 39 and 40 do not depend on j_n), however, with integration constants that are current dependent. This results in an essentially parallel shift of $F(x)$ and $\psi_n(x)$ in x with changing j_n.

For a positive space charge ($+eN_\mathrm{D}$), i.e., for an n-type semiconductor, F_c is negative and $\psi_{n,\mathrm{D}}$ is positive; their values are calculated below. When inserting F_c from Eq. 51 (see below), the potential distribution can also be written as

$$\psi_n(x) = \frac{1}{\sqrt{2}} \frac{kT}{e} \left[\sqrt{\frac{2e\psi_{n,\mathrm{D}}}{kT}} - \frac{x}{L_\mathrm{D}} \right]^2, \tag{43}$$

an expression that is sometimes helpful. L_D is the *Debye length* (Eq. 52), which is a characteristic length for the variation of $\psi_n(x)$ and $F(x)$.

Figure 19 shows a comparison between the approximate (dashed) and the exact solutions (solid curves) obtained by numerical integration of Eqs. 38–40 with $\rho(x) = e[N_D - n(x)]$ in the Poisson equation. Near the electrode, this approximation is quite satisfactory and consequently is mostly used.

The Schottky approximation permits the definition of a barrier layer-thickness x_D from the linear extrapolation of $F(x)$ with $F(x_D) = 0$, as indicated in Fig. 19b. For a computation of x_D, see Eq. 53.

Zero-Current Solution of the Electron Distribution The electron density-distribution can easily be obtained for $j_n = 0$ from the transport equation (Eq. 38)

$$\frac{dn}{dx} = -\frac{e\,n\,F}{kT}. \tag{44}$$

After replacing $F(x)$ by $d\psi_n/dx$ one obtains by integration the Boltzmann distribution with an exponential in the electron potential $\psi_n(x)$ (see Eq. 43):

$$n(x) = n_0 \exp\left(-\frac{e\psi_n(x)}{kT}\right). \tag{45}$$

When inserting $F(x)$ from Eq. 42 and using as a convenient parameter

$$\frac{1}{L_D^2} = \frac{e}{kT}\,\frac{eN_D}{\varepsilon_{stat}\varepsilon_0}, \tag{46}$$

where L_D is the *Debye length* (see below), one obtains

$$n(x) = n_0 \exp\left(-\frac{e\psi_{n,D}}{kT} - \frac{eF_c\ x}{kT} - \frac{x^2}{2L_D^2}\right) \tag{47}$$

which is shown as dashed curve in Fig. 19a.

This holds for zero currents or, as a good approximation, as long as the net current is small compared to both drift and diffusion currents: $j_n \ll (j_{n,drift}, j_{n,diffusion})$. This range is referred to as the *Boltzmann range*.

Diffusion Potential and Junction Field The solutions (Eqs. 41 and 42) contain two integration constants, the electron potential and the electric field at the metal-semiconductor interface. These can easily be evaluated for zero current. The electron-potential step between bulk and metal-semiconductor interface is obtained from Eq. 45 by setting $x = 0$, yielding with $n(x = 0) = n_c$:

$$\psi_{n,\mathrm{D}} = \frac{kT}{e} \ln\left(\frac{n_0}{n_\mathrm{c}}\right). \tag{48}$$

The electron diffusion-potential depends only on the ratio of the bulk and interface densities of carriers. The diffusion potential is also given by the product of maximum barrier field and barrier width:

$$\psi_{n,\mathrm{D}} = -\frac{F_\mathrm{c}\ x_\mathrm{D}}{2}. \tag{49}$$

For the barrier field[19] at $x = x_\mathrm{D}$, one obtains from Eq. 41:

$$F_\mathrm{c} = -\frac{eN_\mathrm{D}\, x_\mathrm{D}}{\varepsilon_\mathrm{stat}\varepsilon_0}; \tag{50}$$

after combining Eqs. 49 and 50 and eliminating x_D, one can express the *barrier field at zero current* as a function of $\psi_{n,\mathrm{D}}$:

$$F_\mathrm{c} = -\sqrt{\frac{2eN_\mathrm{D}\psi_{n,\mathrm{D}}}{\varepsilon_\mathrm{stat}\varepsilon_0}}. \tag{51}$$

For reasonable values of doping and of the electron potential, F_c is on the order of 40 kV/cm.

Debye Length and Barrier Width The Debye length is introduced from Eq. 46 as the distance from x_D in which the electron potential has increased by a factor $kT/(2e)$:

$$L_\mathrm{D} = \sqrt{\frac{\varepsilon_\mathrm{stat}\varepsilon_0\ kT}{e^2 N_\mathrm{D}}} \tag{52}$$

and typically is on the order of a few ten or hundred nm (Fig. 20a).

The barrier layer thickness (Fig. 19) can be expressed in terms of L_D by combining Eqs. 49 and 50, while eliminating F_c:

[19] That is, the maximum field which lies in this approximation at the metal-semiconductor boundary (neglecting image forces)

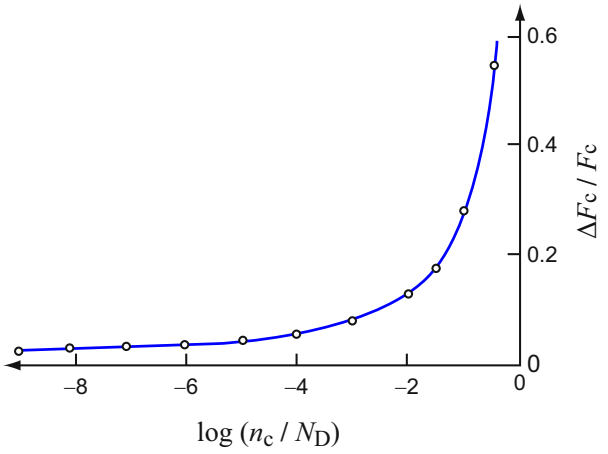


Fig. 20 Computed relative error of $F_c = F(z{=}0,\ j = 0)$ between the exact solution and the Schottky approximation as function of n_c/N_D

$$x_D = L_D\sqrt{\frac{2e\psi_{n,D}}{kT}}, \tag{53}$$

which means that x_D is usually a few (typically 3–6) Debye lengths thick, since $\psi_{n,D}$ is typically on the order of 10 kT/e.

Accuracy of the Schottky Approximation In the part of the junction near $x = x_D$, the Schottky approximation is not satisfactory, since ρ has not yet reached its constant value eN_D (see Figs. 18 and 19). The error made by computing the maximum field F_c or the barrier width x_D, using the expressions Eq. 51 or 53, respectively, can be substantial when n_c is not (at least) several orders of magnitude smaller than N_D. The error for the barrier field F in the Schottky approximation is plotted in Fig. 20 as function of n_c/N_D; it is less than 5% for $n_c/N_D < 10^{-5}$ but increases rapidly above 10% when n_c/N_D increases above 10^{-2}.

3.3 Experimental Results on Schottky Barriers

3.3.1 The Bardeen Model

The Schottky model assumes – similar to the Anderson model (Sect. 2.2.1) – an alignment of the bulk Fermi levels. Bardeen (1947) was the first who pointed out that major discrepancies of experimental results on metal-semiconductor contacts could not be explained with a simple shift of the Fermi levels, as determined by

bulk properties. These observations centered around the fact that for several semiconductors the barrier height was essentially *independent* of the chemical identity (i.e., the workfunction) of the contacting metal. Bardeen explained this by assuming that the Fermi level at the semiconductor surface is *pinned* by surface defects and no longer free to adjust to the contacting metal. The barrier height is then given by

$$\phi_{Bn} = E_g - E_s \tag{54}$$

or

$$\phi_{Bp} = E_s, \tag{55}$$

where E_s is the energy of the surface state responsible for the Fermi-level pinning and is conventionally counted from the top of the valence band. A thin insulating film then needs to be assumed between semiconductor and metal in Fig. 21 to accommodate the additional potential drop. This film must be thin enough to allow perfect tunneling for currents with nonvanishing bias. The film contains the additional dipole layer that is required to equalize the Fermi level across the entire system in equilibrium.

The effect of pinning can be understood qualitatively in terms of localized interface states with an energy E_B called charge-neutrality level. States below E_B are assumed neutral if they are filled with electrons, and states lying above are assumed neutral if they are empty. If the density of such interface states is large, then adding electrons to the semiconductor or extracting them from the semiconductor does not alter the position of the Fermi energy. The Fermi level is *pinned*. When the contact between the metal and the semiconductor is made, both addition

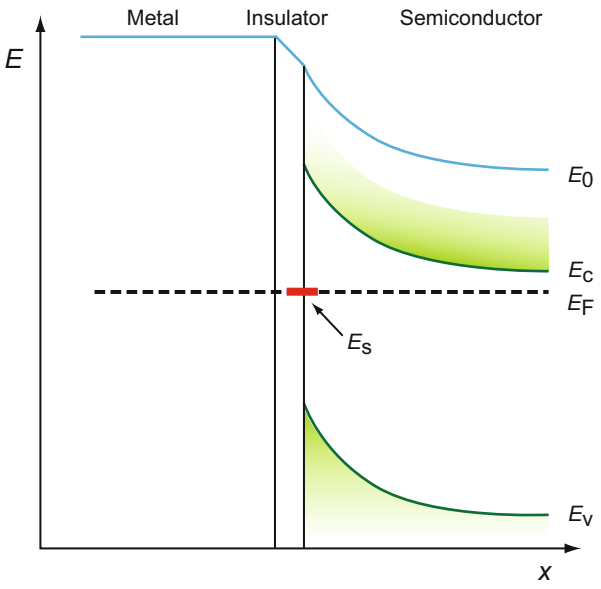


Fig. 21 Schematics of a Schottky barrier with an insulating interlayer that accommodates the necessary voltage drop to provide the lining up of different metals to a Fermi level pinned to a surface state of energy E_s

of electrons and extraction of electrons are accommodated by the interface states, leaving the Fermi level virtually unchanged.

With improved experimental accuracy and an increased database, it was recognized that neither the Schottky-Mott nor the Bardeen model describes the experimental results of metal-semiconductor contacts sufficiently. In actuality, there is an interface-dipole layer, but it is modified to account for an *intermediate* behavior. An illustration of the observed behavior is given in Fig. 22 which presents as an example the dependence of the barrier height at a Si(111)2 × 1 interface to various metals as a function of the workfunction of these metals. It shows a "compressed" linear relationship (compare with Eq. 33):

$$\phi_{Bn} = S\,(\phi_M - \chi_{Sc}) + C, \tag{56}$$

with $0 < S < 1$. The Schottky-Mott model corresponds to $S = 1$, and the Bardeen model requires $S = 0$; (in the example of Fig. 22, $S \cong 0.2$). The compression factor S is also referred to as the *pinning strength parameter* (for $S > 0$ the pinning is not complete), *index of interface behavior*, or the *slope parameter*. For many semiconductors and insulators, the chemical trend of S is described by the simple empirical equation (Mönch 1996)

$$S = 1/\left(1 + 0.1(\varepsilon_\infty - 1)^2\right); \tag{57}$$

ε_∞ is the macroscopic optical dielectric constant.

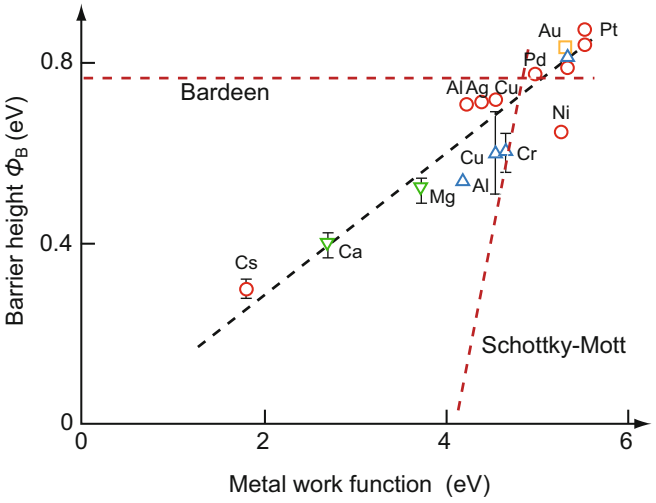


Fig. 22 Measured barrier height of Schottky contacts, applied in situ on Si(111)2 × 1 cleaved surfaces and plotted against the metal workfunction (After Mönch 1970)

3.3.2 Chemical Trends for Schottky Barriers

There is a large body of publications dealing with experimental observations on metal-semiconductor interfaces, most of them relating to Schottky barrier (rectifying) contacts (Fig. 17a). General reviews are given by Rhoderick (1978), Henisch (1984), and Sharma (1984). Numerous review papers deal with experimental methods, e.g., Brillson (1982), Calandra et al. (1985), Flores and Tejedor (1987), LeLay et al. (1987), and Capasso and Margaritondo (1987). Many of these observations are handicapped by an inadequate description of the actual interface, i.e., when electrodes are applied under insufficiently controlled ambient conditions onto semiconductor surfaces that lack proper identification in terms of their atomic structure and adsorbate composition. Furthermore, several experimental methods to determine the barrier height are limited in accuracy and produce variances which interfere with a trend analysis.

The basic experimental results that must be explained by a valid Schottky barrier model can be summarized as follows (Flores and Tejedor 1987):

1. The position of the Fermi level at the metal-semiconductor interface depends very little on the metal, except for very few cases; but even for these, the changes are usually limited to ± 0.2 eV for elemental and III–V semiconductors; the Schottky-Mott relation (Eq. 33) is severely compressed (Eq. 56), with S typically in the 0.1–0.2 range.
2. The sensitivity of the barrier height to the metal workfunction increases sharply (i.e., S increases) with increasing ionicity or heat of formation.
3. Schottky barriers are fully established at less than monolayer coverage with a barrier height that does not change with increased metal layer-thickness.
4. A wide variety of chemical reactions involving the semiconductor and metal layer near their interface, but leaving the cover layer (formerly the metal) highly conductive, do not change the barrier height significantly.

In Sect. 3.3.1, we have described the severe compression of the Schottky-Mott relation (Eq. 33) that can be described with Eq. 56 by a compression factor $S < 1$. The magnitude of this compression shows a substantial trend toward higher values with increased heat of formation ΔH_{F} of the AB compound, shown in Fig. 23 (Kurtin et al. 1969). An identical dependence of S is observed on the difference of electronegativity $\Delta\chi$ of the ions forming an AB compound semiconductor.

There are several other trends which are barely visible in covalent semiconductors because of the high compression but become clearly discernible in more ionic compounds. An example of such a trend is the dependence of the barrier height on the interface heat of reaction, shown for a number of n-type semiconductor-metal barriers in Fig. 24.

A reaction between the electrode metal M and the anion B of the AB compound semiconductor is assumed according to

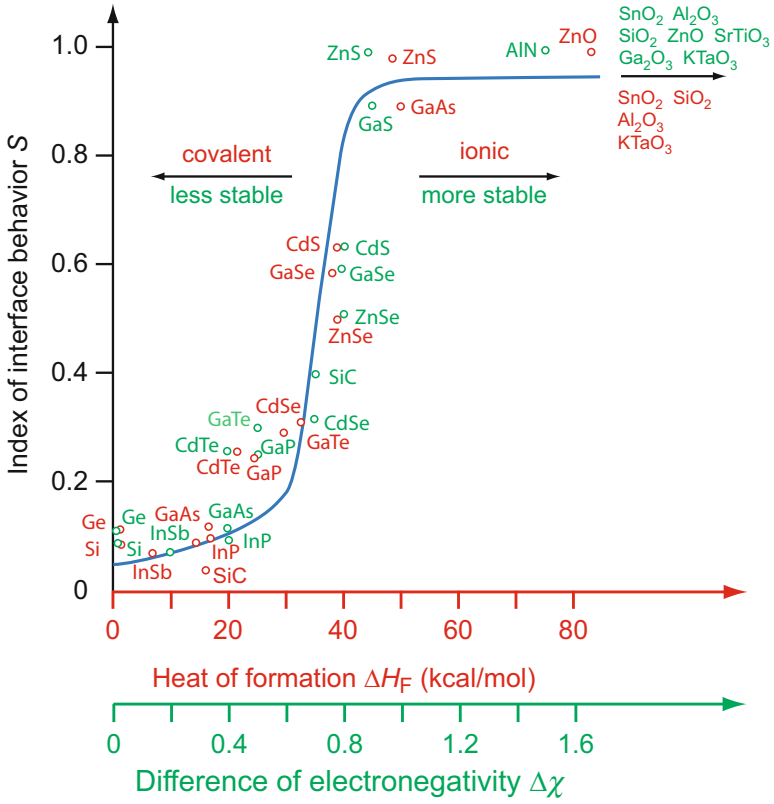


Fig. 23 The index of interface behavior S (Eq. 53) as a function of the heat of formation of the semiconductor (After Brillson 1978); an identical dependence is found on the difference in ion electronegativity of AB semiconductors, i.e., the ionicity of the bonding

$$M + \frac{1}{x}AB \rightarrow \frac{1}{x}[M_xB + A] \tag{58}$$

with

$$\Delta H_R = \frac{1}{x}[H_F(AB) - H_F(M_xB)]. \tag{59}$$

H_F is the heat of formation of the AB compound and that of the most stable metal-anion compound M_xB, normalized per metal atom M. For a wide range of semiconductors, this trend has been confirmed (Brillson 1982). It indicates substantially higher barriers in nonreactive (noble) metals than in highly reactive ones with a steep change in barrier height near $\Delta H_R \cong 0$ (Fig. 24).

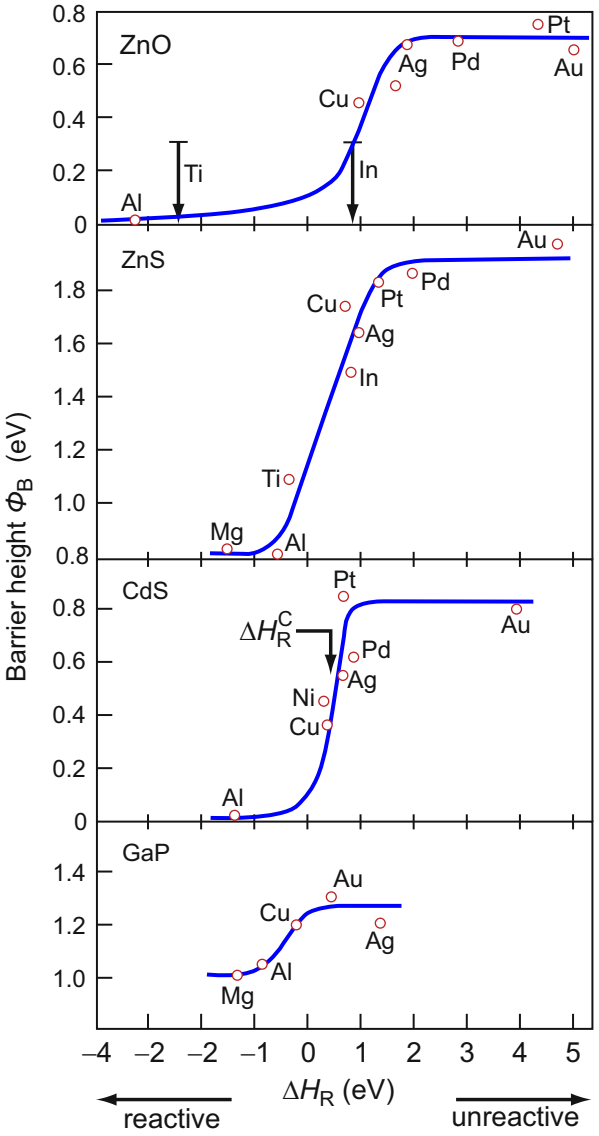


Fig. 24 Barrier height as a function of the heat of interface chemical reaction ΔH_R for various metals on four *n*-type semiconductors as indicated in the individual panels. Compound formation is assumed between metal and anion (After Brillson 1978)

Similar trends are observed for barriers in *p*-type semiconductors (Fig. 25). The barrier height is here measured from the top of the valence band at the interface to the Fermi level. The barrier height of various *p*-type III–V and II–VI compounds shown in Fig. 25 relates to the heat of reaction of their anion partner with the contact metals Cu, Ag, and Au. The trend is obvious; however, care should be exercised in the interpretation of the data since various other parameters also change

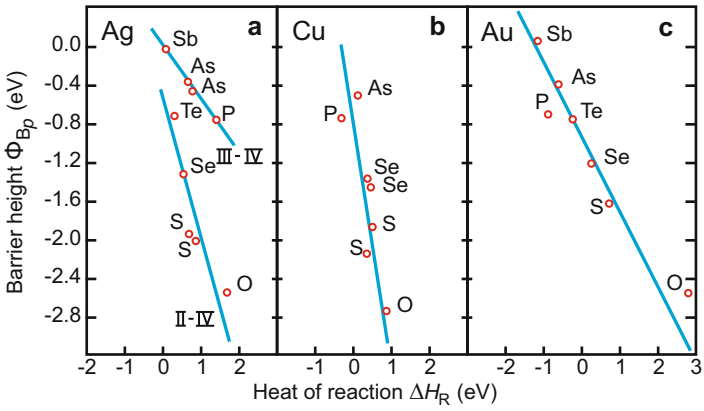


Fig. 25 Barrier height for *p*-type III–V and II–VI compound semiconductors with anions identified, as a function of the heat of reaction ΔH_R with Cu, Ag, and Au (After Brillson 1982)

monotonically with the heat of reaction such as, e.g., the electronegativity (Fig. 23), and for some models could be used more appropriately.

3.4 Interface-Dipol Models

3.4.1 Model of Interface-Induced States

Two basically different early models on the physical nature of the interface states introduced in the Bardeen model (Sect. 3.3.1) have been proposed: the *model of metal-induced gap states* (MIGS, also referred to as ViGS, *virtual gap states*) and the *defect model*. They both are based on the same mechanism for Fermi-level pinning. Energy levels in the gap of the semiconductor near the interface accommodate carriers flowing across the interface for equilibration of the electronegativity when the contact is made. The charge transferred between the metal and the semiconductor creates an interface dipole, which pins the Fermi level and hence controls the Schottky barrier height.

The MIGS model assumes that the wavefunctions of the metal electrons have exponentially decaying tails into the semiconductor (Heine 1965; Tersoff 1984b). The same model was applied in the interface-dipole theory (Sect 2.3.2) to describe the alignment of a heterojunction. The MIG states are located in the bandgap and decay on an atomic scale with a charge decay-length of some Å, making the first few layers of the semiconductor locally metallic: The local density of states in the semiconductor bandgap is filled with a smooth density of gap states. The gap states are related to those bands of the semiconductor that are nearest in energy. MIGS which are related to the valence band are then occupied, and those with conduction-band character are empty. At an effective midgap point E_B gap states change from primarily valence character to conduction character; see Sect 2.3.2.

The Fermi level is pinned at or near this energy E_B, yielding local charge neutrality. A relatively low number of MIGS (about one per 100 atoms at the interface) is required to produce the pinning effect. Without adjustable parameters, the MIGS model could reasonably predict experimentally observed pinned Schottky barrier heights for a number of metal-semiconductor combinations and explain why more ionic semiconductors do not show a universal barrier height.

The defect model assumes that the Fermi-level pinning originates from localized electronic states originating from defects near the interface (Spicer et al. 1979). Such defects are, e.g., vacancies in the semiconductor or the disorder-induced defects considered in the model of Hasegawa and Ohno (1986); see Sect. 2.3.3. The energy for the formation of the defect can be created by the heat of condensation of surface adatoms or from the heat of formation of compounds made from metal and semiconductor atoms forming at the interface. A low number of defects (order of 1 per 100 interface atoms) are required to pin the Fermi energy, analogous to the ViGS model. There exist numerous experimental and theoretical studies on the microscopic nature of such defects and quite a number of related detailed models. Many chemical trends could be explained for specific junctions of semiconductors to metals. Sometimes both ViGS model *and* defect model are needed to explain the data. The density of interface states is largely governed by experimental imperfections. Recent developments in theoretical approaches account at least for much of the rich variety of phenomena.

3.4.2 An Initio Approaches on Schottky-Barrier Modeling

We already pointed out that similar physical principles affect the band alignment at semiconductor heterointerfaces and metal/semiconductor contacts. This is also reflected in more recent ab initio calculations, which model metal/semiconductor junctions on atomic level.

Modeling of MIGS

In a GGA-based first-principles calculation of a contact between a metal (Al, Au) and an elemental semiconductor (Ge, Si, Sn), various defects (e.g., V_{Ge}, Ge_i, Al_i, Al_{Ge}) were placed at various positions near the interface; the study implied (001) and (111) interface orientations (Sasaki and Nakayama 2016). There was generally a substantial decrease of the defect formation-energy found for defect locations approaching the interface within a range of typically 10 Å, independent of interface orientation and type of metal, semiconductor, or defect. The decrease of formation energy is related to an exponential increase of the defect density near the interface (cf. Sect. 2 of ▶ Chap. 15, "Crystal Defects"). The formation energy of defects was found to decrease further by hybridization with metal-induced gap states; MIGS hence promote the formation of disorder-induced gap states.

A strong increase of gap states by defects was also reported for semiconductor interfaces to compound metals (Robertson et al. 2020). The studied metal silicides, rare-earth arsenides, and Sb interfaces have an underlying covalent lattice, which does not perfectly match that of covalent semiconductors; this creates a periodic

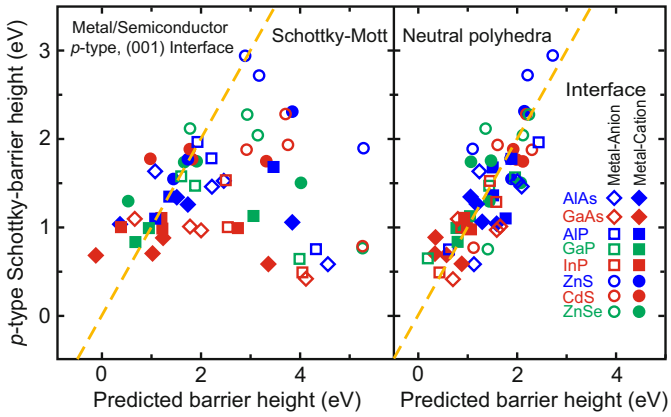


Fig. 26 Calculated Schottky-barrier heights of metal/*p*-type semiconductor junctions as a function of predictions of the Schottky-Mott model (*left*) and the neutral-polyhedra theory (*right*). *Open and filled symbols* denote anion or cation plane of the semiconductor at the interface, respectively, *yellow lines* indicate perfect agreement (After Tung and Kronik 2021b)

array of coordination defects. The GGA-based calculations found bonding configurations like dangling bonds, lateral bonds, or miscoordinated sites which create defect gap states localized at the interface. These states add to the MIGS density of states with an even larger DOS and cause an apparent weakening of Fermi-level pinning (i.e., a larger *S* parameter). The gap states depend on the orientation and termination of the interface, leading to a significant difference of the defect states and hence in Schottky-barrier heights.

A dependence of the Schottky-barrier height on interface termination was previously also reported for junctions of Al on (001)-oriented Ge, GaAs, AlAs, and ZnSe (Berthod et al. 2003). In ab initio pseudopotential calculations, larger barrier values were generally found for anion-terminated surfaces compared to cation termination. The difference could be explained by a microscopic dipole created by the screened charge of the polar (001) semiconductor surface and its image charge at the metal surface. The screening effect was recently assigned directly to the metallic nature of interface bonds at metal-cation interfaces (Tung and Kronik 2021a). The study based on DFT supercell calculations points out that the dependence of the barrier height on the metal is weakened by a screening effect of the metallic bond, and the Fermi level is apparently pinned. The calculations could in fact well reproduce the experimentally measured average *p*-type Schottky-barrier height for metal-cation (001) interfaces of the zincblende semiconductors studied also in Fig. 26, and for Ge and Si.

Modeling by Neutral Polyhedra

The model solid constructed by neutral polyhedra introduced in Sect. 2.4.3 was recently applied to provide quantitative values of Schottky-barrier heights for a large number of epitaxial interfaces between metals and *p*-type zincblende

semiconductors; see Tung and Kronik (2021b) and the review on models by Tung (2014).

The DFT-GGA calculations based on PBE and HSE exchange-correlation functionals modeled relaxed epitaxial (110) interfaces, and relaxed (001) interfaces with either metal-anion or metal-cation bonds; mostly metals with cubic structure and with lattice mismatches below 2% were considered. Starting with the Schottky-Mott model Eq. 34, the study considers the charge redistribution at the interface by adding a term Δ_{IID}, yielding the barrier height

$$\phi_{\mathrm{B}_p} = I_{\mathrm{Sc,NPT}} - \phi_{\mathrm{M,NPT}} + \Delta_{\mathrm{IID}}. \tag{60}$$

Here, $I_{\mathrm{Sc,NPT}}$ is the ionization potential of the semiconductor (index *Sc*) calculated using the neutral-polyhedra theory (*NPT*), $\phi_{\mathrm{M,NPT}}$ is the accordingly calculated work function of the metal (*M*), and Δ_{IID} is the correction due to charge rearrangement at the interface termed irreducible interface dipole (*IID*). The formation of the Schottky-barrier height is assumed to be controlled by the bond formation at the interface over a single interplanar distance; in the interface-dipol term Δ_{IID}, two contributions are considered: (1) The dipole arising from bond polarization; this part is assumed proportional to the product of the average electronegativity difference across the interface, the interface distance, and the areal density of bonds. A common single proportionality constant is used for all studied interfaces. (2) The dipole due to charge-transfer, created by differences of the charge-density $\rho_{\mathrm{M}} - \rho_{\mathrm{Sc}}$ across the interface; this difference is scaled by a factor accounting for the different shapes and sizes of the neutral polyhedra for metals and semiconductors (Tung and Kronik 2021b).

Results of the NPT calculations of (001) interfaces are compared to predictions of the Schottky-Mott model in Fig. 26. Both results are plotted versus Schottky-barrier heights extracted from standard DFT supercell calculations. In this procedure, the barriers were taken from the difference between the Fermi level and the position of the bulk valence-band maximum, mapped onto the average electrostatic potential in the semiconductor portion of the supercell; the semiconductor ionization potential and the metal work function for the corresponding free surfaces were independently obtained from free-slab calculations. Fig. 26 shows a clear trend of the barrier height in the NPT calculations for both metal-anion and metal-cation interface bonds; the latter required a shift of the interface plane towards the next anion plane due to charge transfer on a longer scale (the cations are metallic themselves; the "effective metal" hence ends only with the first layer of cations included). Such a clear trend is also found in respective results on nonpolar (110) interfaces. The calculations indicate that the charge transfer at the interface can be assumed extremely short ranged.

4 Summary

The interface between two semiconductors or a semiconductor and a metal are of essential importance to most semiconductor devices. In semiconductor heterojunctions, the band offset governs electrical and electro-optical effects in electronic and optoelectronic devices. The metal-semiconductor interface determines the barrier height and thereby the electrical properties of any device contact.

Early models describe band offsets and barrier heights as the difference of two bulk properties such as the Fermi-level or band positions. These *linear models* show only limited agreement with the experiment. Often, one observes a compression of scale when assuming a band-edge offset or barrier height given from a linear model. This is an indication that the specific interface plays an important role causing various interface defects and modifying interface dipoles. *Nonlinear models* account for the formation of such interface dipoles. Disorder at the interface or defects induced by the close proximity of materials with different band energies cause leaking out of eigenfunctions from one into the other material. Charge-neutrality levels or branch-point energies within the bandgap derived from such states yield reasonable descriptions for many experimental observations on both types of interfaces for small interface strain. When a larger strain can coherently be maintained, shifts described by deformation potentials become important. Recent ab initio calculations allow for explicitly modeling entire interface structures, confirming the strong effect of relaxation and yielding reasonable precision for many interfaces.

The interface between semiconductors or between metal and semiconductor provides a host of challenging problems, in spite of almost a century of research that has steadily accelerated its pace and of the introduction of rather sophisticated experimental methods and theoretical analyses. We still encounter major problems in predicting the interface band-lineups observed in experiments with an accuracy in excess of ± 0.1 eV. Designing optimized devices in respect to these interfaces will thus generally include empirical studies.

References

Adamski NL, Wickramaratne D, Van de Walle CG (2020) Band alignments and polarization properties of the Zn-IV-nitrides. J Mater Chem C8:7890

Adachi S (2005) Properties of Group IV, III–V and II–VI semiconductors. Wiley, Chichester

Adams MJ, Nussbaum A (1979) A proposal for a new approach to heterojunction theory. Solid State Electron 22:783

Alkauskas A, Broqvist P, Devynck F, Pasquarello A (2008) Band offsets at semiconductor-oxide interfaces from hybrid density-functional calculations. Phys Rev Lett 101:106802

Anderson RL (1962) Experiments on Ge-GaAs heterojunctions. Solid State Electron 5:341

Aulbur WG, Jönsson L, Wilkins JW (2000) Quasiparticle calculations in solids. Solid State Phys 54:1

Ball CAB, Van der Merve JH (1983) The growth of dislocation-free layers. In: Nabarro FRN (ed) Dislocations in solids, vol 6. North Holland Publishing, Amsterdam, pp 121–141

Bardeen J (1947) Surface states and rectification at a metal semi-conductor contact. Phys Rev 71:717
Bauer RS, Zurcher P, Sang HW Jr (1983) Inequality of semiconductor heterojunction conduction-band-edge discontinuity and electron affinity difference. Appl Phys Lett 43:663
Bauer S, Rosenauer A, Link P, Kuhn W, Zweck J, Gebhardt W (1993) Misfit dislocations in epitaxial ZnTe/GaAs (001) studied by HRTEM. Ultramicroscopy 51:221
Bechstedt F, Enderlein R (1988) Semiconductor surfaces and interfaces. Akademie-Verlag, Berlin
Berthod C, Binggeli N, Baldereschi A (2003) Schottky barrier heights at polar metal/semiconductor interfaces. Phys Rev B 68:085323.
Brillson LJ (1978) Transition in Schottky barrier formation with chemical reactivity. Phys Rev Lett 40:260
Brillson LJ (1982) The structure and properties of metal-semiconductor interfaces. Surf Sci Rep 2:123
Calandra C, Bisi O, Ottavian G (1985) Electronic properties on silicon-transition metal interface compounds. Surf Sci Rep 4:271
Caldas MJ, Fazzio A, Zunger A (1984) A universal trend in the binding energies of deep impurities in semiconductors. Appl Phys Lett 45:671
Capasso F, Margaritondo G (eds) (1987) Heterojunction band discontinuities. North Holland, Amsterdam
Chadi DJ, Cohen ML (1975) Tight-binding calculations of the valence bands of diamond and zincblende crystals. Phys Status Solidi B 68:405
Choe D-H, West D, Zhang S (2018) Band alignment and the built-in potential of solids. Phys Rev Lett 121:196802
Choe D-H, West D, Zhang S (2021) Revealing the vacuum level in an infinite solid by real-space potential unfolding. Phys Rev B 103:235202
Chiaradia P, Katnani AD, Sang HW Jr, Bauer RS (1984) Independence of Fermi-level position and valence-band edge discontinuity at GaAs-Ge(100) interfaces. Phys Rev Lett 52:1246
Davydov B (1938) The rectifying action of semiconductors. J Tech Phys USSR 5:87
Di Liberto G, Pacchioni G (2021) Band offset in semiconductor heterojunctions. J Phys C Cond Matter 33:415002
Dixon RH, Goodhew PJ (1990) On the origin of misfit dislocations in InGaAs/GaAs strained layers. J Appl Phys 68:3163
Flores F, Tejedor C (1979) Energy barriers and interface states at heterojunctions. J Phys C Solid State Phys 12:731
Flores F, Tejedor C (1987) On the formation of semiconductor interfaces. J Phys C Solid State Phys 20:145
Franciosi A, Van de Walle CG (1996) Heterojunction band offset engineering. Surf Sci Rep 25:1
Frank FC, Van der Merve JH (1949) One-dimensional dislocations. I. Static theory. Proc Roy Soc Lond A 198:205, One-dimensional dislocations. II. Misfitting monolayers and oriented overgrowth. Proc Roy Soc Lond A 198:216; One-dimensional dislocations. III. Influence of the second harmonic term in the potential representation, on the properties of the model. Proc Roy Soc Lond A 200:125
Frensley WR, Kroemer H (1977) On the formation of semiconductor interfaces. Phys Rev B 16:2642
Grant RW, Kraut EA, Waldrop JR, Kowalczyk SP (1987) Interface contributions to heterojunction band discontinuities: x-ray photoemission spectroscopy investigations. In: Capasso F, Margaritondo G (eds) Heterojunction band discontinuities. North Holland, Amsterdam, pp 167–206
Grüneis A, Kresse G, Hinuma Y, Oba F (2014) Ionization potentials of solids: the importance of vertex corrections. Phys Rev Lett 112:096401
Harrison WA (1977) Elementary theory of heterojunctions. J Vac Sci Technol 14:1016
Harrison WA (1980) Electronic structure and the properties of solids. WH Freeman, San Francisco
Harrison WA, Tersoff J (1986) Tight-binding theory of heterojunction band lineups and interface dipoles. J Vac Sci Technol B 4:1068

Harrison WA, Kraut EA, Waldrop JR, Grant RW (1978) Polar heterojunction interfaces. Phys Rev B 18:4402
Hasegawa H, Ohno H (1986) Unified disorder induced gap state model for insulator–semiconductor and metal–semiconductor interfaces. J Vac Sci Technol B 4:1130
Hasegawa H, Ohno H, Sawada T (1986) Orbital energy for heterojunction band lineup. Jpn J Appl Phys 25:L265
Hayes WM, Lide DR, Bruno TJ (eds) (2013) CRC handbook of chemistry and physics, 94th edn. CRC Press, Boca Raton
Hedin L (1965) New method for calculating the one-particle Green's function with application to the electron-gas problem. Phys Rev 139:A796
Heine V (1965) Theory of surface states. Phys Rev 138:A1689
Heinrich H, Langer JM (1986) Band offsets in heterostructures. In: Grosse P (ed) Festkörperprobleme/advances in solid state physics, vol 16. Vieweg, Braunschweig, pp 251–275
Henisch HK (1984) Semiconductor contacts. Claredon Press, Oxford
Heyd J, Scuseria GE, Ernzerhof M (2003) Hybrid functionals based on a screened Coulomb potential. J Chem Phys 118:8207
Heyd J, Scuseria GE, Ernzerhof M (2006) Erratum: „Screened hybrid density functionals applied to solids" J Chem Phys 118:8207. J Chem Phys 124:219906
Hinuma Y, Grüneis A, Kresse G, Oba F (2014) Band alignment of semiconductors from density-functional theory and many-body perturbation theory. Phys Rev B 90:155405
Höffling B, Schleife A, Rödl C, Bechstedt F (2012) Band discontinuities at Si-TCO interfaces from quasiparticle calculations: comparison of two alignment approaches. Phys Rev B 85:035305
Katnani AD (1987) Trends in semiconductor heterojunctions. In: Capasso F, Margaritondo G (eds) Heterojunction band discontinuities. North Holland, Amsterdam, pp 115–166
Katnani AD, Bauer RS (1986) Commutativity and transitivity of GaAs-AlAs-Ge(100) band offsets. Phys Rev B 33:1103
Katnani AD, Margaritondo G (1983) Microscopic study of semiconductor heterojunctions: photoemission measurement of the valance-band discontinuity and of the potential barriers. Phys Rev B 28:1944
Kroemer H (1975) Problems in the theory of heterojunction discontinuities. Crit Rev Solid State Sci 5:555
Kroemer H (1983) Heterostructure devices: a device physicist looks at interfaces. Surf Sci 132:543
Kroemer H (1984) Barrier control and measurements: abrupt semiconductor heterojunctions. J Vac Sci Technol B 2:433
Kurtin S, McGill TC, Mead CA (1969) Fundamental transition in the electronic nature of solids. Phys Rev Lett 22:1433
Langer JM, Heinrich H (1985a) Deep-level impurities: a possible guide to prediction of band-edge discontinuities in semiconductor heterojunctions. Phys Rev Lett 55:1414
Langer JM, Heinrich H (1985b) On a direct connection of the transition metal impurity levels to the band edge discontinuities in semiconductor heterojunctions. Physica B 134:444
Ledebo L-Å, Ridley BK (1982) On the position of energy levels related to transition-metal impurities in III-V semiconductors. J Phys C Solid State Phys 15:L961
Lee RJ (1985) Some comments on Nussbaum's heterojunction lineup theory. IEEE Electron Device Lett 6:130
LeLay G (1983) Electronic and atomic structure of Ag–Si(111) and Ag–Ge(111). J Vac Sci Technol B 1:354
LeLay G, Derien J, Boccara N (eds) (1987) Semiconductor interfaces: formation and properties. Springer, Berlin
Li Y-H, Walsh A, Chen S, Yin W-J, Yang J-H, Li J, Da Silva JLF, Gong XG, Wei S-H (2009) Revised ab initio natural band offsets of all group IV, II-VI, and III-V semiconductors. Appl Phys Lett 94:212109
Lyu S, Lambrecht WRL (2020) Band alignment of III-N, ZnO and II–IV-N_2 semiconductors from the electron affinity rule. J Phys D Appl Phys 53:015111

Mailhiot C, Duke CB (1986) Many-electron model of equilibrium metal-semiconductor contacts and semiconductor heterojunctions. Phys Rev B 33:1118
Margaritondo G (1983) Microscopic investigations of semiconductor interfaces. Solid State Electron 26:499
Margaritondo G, Perfetti P (1987) The problem of heterojunction band discontinuities. In: Capasso F, Margaritondo G (eds) Heterojunction band discontinuities. North Holland, Amsterdam, pp 59–114
Matthews JW (1975) Coherent interfaces and misfit dislocations. In: Matthews JW (ed) Epitaxial growth, Part B. Academic Press, New York, pp 559–609
Matthews JW, Blakeslee AE (1974) Defects in epitaxial multilayers I. Misfit dislocations. J Cryst Growth 27:118
Matys M, Kaneki S, Nishiguchi K, Adamowicz B, Hashizume T (2017) Disorder induced gap states as a cause of threshold voltage instabilities in Al_2O_3/AlGaN/GaN metal-oxide-semiconductor high-electron-mobility transistors. J Appl Phys 122:224504
Milnes AG, Feucht DL (1972) Heterojunctions and metal-semiconductor junctions. Academic Press, New York
Mönch W (1970) On metal-semiconductor surface barriers. Surf Sci 21:443
Mönch W (1996) Empirical tight-binding calculation of the branch-point energy of the continuum of interface-induced gap states. J Appl Phys 80:5076
Mönch W (1996) Chemical trends of barrier heights in metal-semiconductor contacts: on the theory of the slope parameter. Appl Surf Sci 92:367
Mott NF (1938) Note on the contact between a metal and an insulator or semi-conductor. Proc Cambridge Phil Soc 34:568
Mott NF (1939) The theory of crystal rectifiers. Proc R Soc (London) A 171:27
Niles DW, Margaritondo G (1986) Heterojunctions: definite breakdown of the electron affinity rule. Phys Rev B 34:2923
Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865
Rhoderick EH (1978) Metal-semiconductor contacts. Claredon Press, Oxford
Robertson J (2013) Band offsets, Schottky barrier heights, and their effects on electronic devices. J Vac Sci Technol A 31:050821
Robertson J, Guo Y, Zhang Z, Li H (2020) Extending the metal-induced gap state model of Schottky barriers. J Vacuum Science & Technology B 38:042208.
Sasaki S, Nakayama T (2016) Defect distribution and Schottky barrier at metal/Ge interfaces: Role of metal-induced gap states. Jpn J Appl Phys 55:111302.
Schleife A, Fuchs F, Rödl C, Furthmüller J, Bechstedt F (2009) Branch-point energies and band discontinuities of III-nitrides and III-/II-oxides from quasiparticle band-structure calculations. Appl Phys Lett 94:012104
Schottky W (1939) Zur Halbleitertheorie der Sperrschicht- und Spitzengleichrichter. Z Phys 113:367 (On the semiconductor theory of junction and point rectifiers, in German)
Schottky W, Störmer R, Waibel F (1931) Über die Gleichrichterwirkungen an der Grenze von Kupferoxydul gegen aufgebrachte Metallelektroden. Z Hochfrequenztech 37: 162 (On the rectifying action of cuprous oxide at the junction to metal electrodes, in German)
Sharma BL (ed) (1984) Metal-semiconductor schottky barrier junctions and their applications. Plenum Press, New York, ch 3
Sheldon P, Jones KM, Al-Jassim MM, Yacobi BG (1988) Dislocation density reduction through annihilation in lattice-mismatched semiconductors grown by molecular-beam epitaxy. J Appl Phys 63:5609
Spicer WE, Chye PW, Skeath PR, Su CY, Lindau I (1979) New and unified model for Schottky barrier and III-V insulator interface states formation. J Vac Sci Technol 16:1422

Stevanovic V, Lany S, Ginley DS, Tumas W, Zunger A (2014) Assessing capability of semiconductors to split water using ionization potentials and electron affinities only. Phys Chem Chem Phys 16:3706
Tersoff J (1984a) Theory of semiconductor heterojunctions: the role of quantum dipoles. Phys Rev B 30:4874
Tersoff J (1984b) Schottky barrier heights and the continuum of gap states. Phys Rev Lett 52:465
Tersoff J (1985) Schottky barriers and semiconductor band structures. Phys Rev B 32:6968
Tersoff J (1986) Band lineups at II-VI heterojunctions: failure of the common-anion rule. Phys Rev Lett 56:2755
Tung RT (2014) The physics and chemistry of the Schottky barrier height. Appl Phys Rev 1:011304
Tung RT, Kronik L (2016) Band offset formation at semiconductor heterojunctions through density-based minimization of interface energy. Phys Rev B 94:075310
Tung RT, Kronik L (2019) Bulklike band-offset mystery solved through energy minimization: Lessons from perovskite oxide heterojunctions. Phys Rev B 99:115302
Tung RT, Kronik L (2021a) Fermi level pinning for zinc-blende semiconductors explained with interface bonds. Phys Rev B 103:085301
Tung RT, Kronik L (2021b) Quantitative explanation of the Schottky barrier height. Phys Rev B 103:035304.
Van de Walle CG (1989) Band lineups and deformation potentials in the model-solid theory. Phys Rev B 39:1871
Van de Walle CG (2019) Polarons get the full treatment. Physics 12:68
Van de Walle CG, Martin RM (1985) Theoretical study of Si/Ge interfaces. J Vac Sci Technol B 3:1256
Van de Walle CG, Martin RM (1986) Theoretical calculations of semiconductor heterojunction discontinuities. J Vac Sci Technol B 4:1055
Van de Walle CG, Martin RM (1987) Theoretical study of band offsets at semiconductor interfaces. Phys Rev B 35:8154
Van de Walle CG, Neugebauer J (2003) Universal alignment of hydrogen levels in semiconductors, insulators and solutions. Nature 423:626
Van der Merve JH (1962) Crystal interfaces II. Finite overgrowth. J Appl Phys 34:123
Vogl P, Baranowski JM (1985) Electronic structure of 3d-transition metal impurities in semiconductors. Acta Phys Pol A 67:133
Weston L, Tailor H, Krishnaswamy K, Bjaalie L, Van de Walle CG (2018) Accurate and efficient band-offset calculations from density functional theory. Comput Mater Sci 151:174
Wei S-H, Zunger A (1993) Band offsets at the CdS/$CuInSe_2$ heterojunction. Appl Phys Lett 63:2549
Zunger (1986) Electronic structure of 3d transition-atom impurities in semiconductors. Solid State Phys 39:275–464

Optical Properties of Defects

17

Karl W. Böer and Udo W. Pohl

Contents

1 Energy and Strength of Defect Absorption 704
1.1 Energy of Shallow Defects 704
1.2 Deep-Level Defects 706
2 Line Shape of Electronic Defect Transitions 719
2.1 Homogeneous Lines 719
2.2 Inhomogeneous Broadening 722
2.3 Deep-Level Defect Line Broadening 726
2.4 Photoionization Edge Shape 734
3 Optical Absorption of Disordered Crystals 738
3.1 Band Tailing 739
3.2 Heavy Doping and Burstein-Moss Effect 744
4 Summary 748
References 750

Abstract

The optical absorption spectrum of lattice defects provides the most direct information about their electronic properties; general characteristics are obtained from such spectra even without detailed knowledge about the defect structure. Two substantially different types of absorption spectra occur: line spectra near the fundamental band-to-band absorption originating from shallow-level defects and usually broad spectra from deep-level, tight-bonding defects. Shallow defects are described by a hydrogen-like model using a Rydberg energy modified by the effective mass and dielectric constant. Lattice

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_17

coupling and relaxation of these centers is only weak. Deep centers show strong electron-lattice coupling described by the Huang-Rhys factor, which expresses the mean number of emitted phonons during lattice relaxation after electron capture. Electronic and vibronic properties of deep centers are described in a configuration-coordinate diagram, with optical excitation and emission processes represented by vertical transitions according to the Franck-Condon principle. Photoionization refers to an excitation from a defect center into a band and leads to edge-shaped absorption.

At high defect density, the localized wavefunctions overlap and form an impurity band, which at densities exceeding the Anderson-Mott limit merges with band states. such an extended density of states leads to an Urbach tail in the optical absoption near the band edge. In addition, the absorption edge of band-to-band transitions in heavily doped semiconductors is blueshifted due to the Burstein-Moss effect resulting from band filling.

Keywords

Andersen-Mott limit · Band tailing · Band-to-band absorption · Burstein-Moss effect · Configuration-coordinates · Deep-level defects · Defect-center relaxation · Effective-mass defect · Electron-lattice coupling · Franck-Condon principle · Huang-Rhys factor · Hydrogen-like model · Inhomogeneous broadening · Lattice coupling · Line shape · Oscillator strength · Shallow-level defects · Urbach tail

1 Energy and Strength of Defect Absorption

There is a large variety of lattice defects. Two typical classes of prominent point defects are shallow-level and deep-level centers; their electronic properties will be discussed in ▶ Chaps. 18, "Shallow-Level Centers", and ▶ 19, "Deep-Level Centers". This introductory chapter focuses on the optical absorption spectrum which provides direct information on the electronic defect structure. Attention is given to the general optical absorption behavior, rather than to the detail.

Optical transitions at a lattice defect are determined by the ground and excited energy states of the defect center, their oscillator strength, and the influence of the surrounding lattice. This results in an absorption spectrum with lines (bound-to-bound transitions) at a certain energy, with a certain absorption constant (the strength of absorption) and a certain line shape. Each of these will be discussed in the following sections. Eventually, also bound-to-free transitions resulting in an absorption edge will be discussed.

1.1 Energy of Shallow Defects

The most obvious difference between various types of defects is the energy of a specific absorption line. Since an understanding of the optical absorption of lattice

defects requires a knowledge of the basic elements of its electronic structure in relation to the band structure of the surrounding lattice, these elements will be given here in a simplified form for the two major classes of point defects: the shallow-level and deep-level defects.

Point defects (impurities) that result in shallow levels are described by a hydrogen-like model. We focus here on donors; in the description of shallow acceptors also the structure of the valence band must be considered. In the hydrogen-like model of a shallow donor the defect is replaced by a proton – the extra proton of the donor – and one electron, the extra electron that circles the proton at a distance of several lattice constants. Such a quasi-hydrogen donor can easily be discussed in a semiempirical form, analogous to Wannier excitons introduced in Sect. 1 of ▶ Chap. 14, "Excitons". For a more detailed analysis, see Sect. 1 of ▶ Chap. 18, "Shallow-Level Centers".

The eigenstates of the shallow donor can be described similar to those of a hydrogen atom, using a modified field, reduced by the static dielectric constant $\varepsilon_{\text{stat}}$, and with an electron of modified mass m_n instead of m_0. When using the Rydberg energy, which is the ionization energy of the hydrogen atom $R_\infty = E(n=1) - E(n=\infty) = R_\infty\left(\frac{1}{1} - \frac{1}{\infty}\right) = 13.6\,\text{eV}$, the eigenstates of the donor can be expressed by quasi-hydrogen energies

$$E_{\text{qH},n} = -R_\infty \frac{m_n}{m_0} \frac{1}{\varepsilon_{\text{stat}}^2} \frac{1}{n^2}. \tag{1}$$

The corresponding quasi-Bohr radius is

$$a_{\text{qH},n} = a_{\text{H}} \varepsilon_{\text{stat}} \frac{m_n}{m_0} n^2, \tag{2}$$

where n is the principal quantum number and $a_{\text{H}} = 0.529$ Å is the hydrogen Bohr radius.

The use of the quasi-hydrogen model obtained from the effective-mass approximation is justified when the quasi-Bohr radius is large compared to the lattice constant a: the bound electron "circles" the center well within the lattice of the host, and the Coulomb forces at this distance are screened by $\varepsilon_{\text{stat}}$. Here, the extent of the solution in k space is small compared to π/a. Then only k vectors near the band minimum contribute, and the use of a well-defined effective mass is justified.[1]

For typical semiconductors with $m_n \cong 0.1\, m_0$ and $\varepsilon_{\text{stat}} \cong 10$, we obtain a quasi-Bohr radius of $\cong$ 50 Å, which is much larger than a typical lattice constant of 5.5 Å, and a ionization energy of $\cong$ 10 meV, i.e., a very small fraction of the bandgap energy.

[1] The quasi-hydrogen energy can be expressed as a function of the quasi-Bohr radius (for $n = 1$): $E_{\text{qH}} = \hbar^2 k_{\text{r}}^2/(2m_n)$ for $k_{\text{r}} = 1/a_{\text{qH}}$. This is a familiar expression identifying the circling electron in a parabolic and isotropic conduction band.

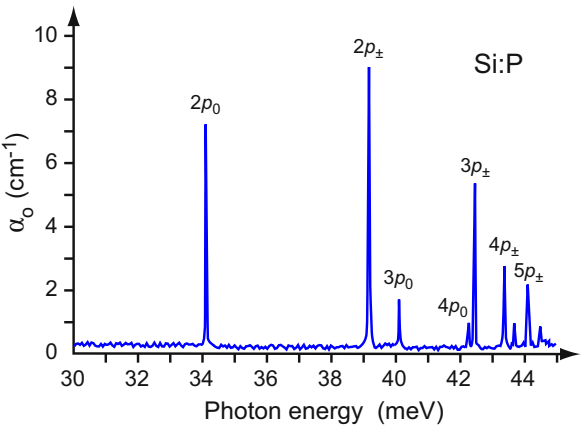


Fig. 1 Hydrogen-like absorption spectrum from the 1 *s* ground state into the excited states of P donors in Si at 4 K (After Jagannath et al. 1981). (See Fig. 7 of ▶ Chap. 18, "Shallow-Level Centers", for the corresponding level spectrum)

Such a center is indeed a *shallow* donor, which is intimately connected to the conduction band.

There are three possibilities for an optical absorption at such a defect:

1. The transition of its electron into an excited state
2. After its electron is removed, the excitation of an electron from the valence band into its ground or excited state
3. Transitions from a localized state of the defect center into the continuum states of the band

The first possibility requires very little energy and lies in the far IR part of the spectrum as shown for a P donor in silicon in Fig. 1. The second possibility lies close to the band edge:

$$h\nu_{\text{low}} = E_{\text{qH}} \quad \text{or} \quad h\nu_{\text{high}} = E_{\text{g}} - E_{\text{qH}}, \tag{3}$$

with the subscripts for low-energy or high-energy transitions to the same center. The transitions of the third kind (see also Sect. 2.4) are edge-shaped and distinctly different from the line-type of absorption listed before.

1.2 Deep-Level Defects

Deep levels (subscript *deep*) are levels which often are closer to the center of the bandgap than to one of the band edges. They are created from defect centers in which their core potential plays a dominant role. This is a tight-binding potential, rather than the far-reaching Coulomb potential. In calculating their eigenstates, both conduction and valence bands have to be considered. Examples of such centers are most of the transition-metal impurities and certain

vacancies and self-interstitials. The chemistry of the center, as well as its lattice surrounding, and its phonons substantially influence its electronic level energy.

The binding energy of deep centers requires a detailed quantum-mechanical analysis, which is discussed in ▶ Chap. 19, "Deep-Level Centers". The results cannot be generalized in a simple form, as possible for shallow centers.

Since the binding energy of deep centers is usually much larger than typical phonon energies, the phonons may play a role in the related transitions. This can occur via two mechanisms:

1. Adding to, or subtracting from, the excitation energy (phonon-assisted transition, phonon replica – see, e.g., Figs. 20b and 23 of ▶ Chap. 18, "Shallow-Level Centers" or Fig. 16 of ▶ Chap. 30, "Carrier Recombination and Noise")

$$h\nu = E_{\text{deep}} \pm m\hbar\omega \tag{4}$$

   for m phonons involved, a process with decreasing probability for each additional phonon emitted (−) or absorbed (+)
2. The center relaxes (by emitting phonons) when an electron is emitted into, or captured from, a band state, thereby changing the charge character of the defect state. Here, the maximum of the emission or absorption line is given by an equation similar to Eq. 4

$$h\nu = E_{\text{deep}} \pm S\hbar\omega_{\text{char}} \tag{5}$$

except that S is the Huang-Rhys factor which gives the number of phonons involved in the center relaxation, and ω_{char} is a characteristic (optical) phonon frequency, typically of a breathing mode – see Q_1 in Fig. 2. For alkali halides, S is very large and on the order of 20.

In contrast to phonon-assisted transitions where phonons need to be supplied (a temperature-dependent process), here phonons are created while the center relaxes. Therefore, the shift given in Eq. 5 is independent of the temperature – see Fig. 11.

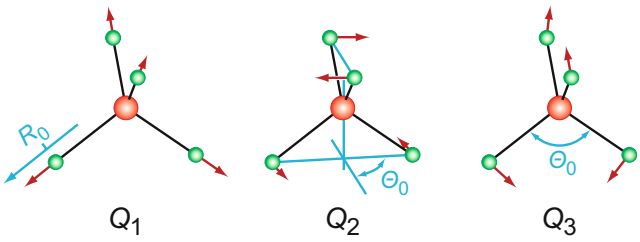


Fig. 2 Normal coordinates Q_1 – Q_3 of a center with tetrahedral symmetry

1.2.1 Defect-Center Relaxation and Configuration Coordinates

The process of defect-center relaxation illustrated in Fig. 5 of ▶ Chap. 15, "Crystal Defects" needs further explanation as it can substantially change the energy of the defect level within the gap. Furthermore, energies between the band edges and the center no longer add up to the bandgap energy, as they do for shallow-level defects indicated in Eq. 3.

These processes can be qualitatively understood by recognizing that a change in the charge of a defect causes a change in the bonding to the neighbor atoms and consequently a change in the relative position of these atoms. This has an influence on the electronic eigenstates of the defect. Such changes take much more time than, e.g., a capture of an electron in the defect. Consequently, the defect *relaxes after* capture and, with the change in atomic configuration, the electronic defect level also changes, as the electronic eigenstate relaxes.

A similar process takes place when an electron within the defect is excited from the ground state into an excited state with a different wavefunction. Consequently, the excited state requires a somewhat larger space, causing a change in bonding to the neighboring atoms, and resulting in a change of the atomic configuration. Another component of this excitation process can be seen as a change from a core configuration toward a more hydrogen-like one. This causes polarization changes, which further result in shifts of the positions of the surrounding atoms. Finally, with defect ionization, the electron now belongs to the conduction band, and the defect level relaxes to the unoccupied state.

While the recombination or optical excitation occurs instantaneously, shifting of the neighboring lattice atoms takes time and is described as a relaxation of the lattice into a new equilibrium position. Since it is the relative atomic distance to the neighbors that is changing, the configuration-coordinate diagram that pictures this relation is most helpful.

Configuration-Coordinates The configuration-coordinate representation is used to depict the coupling between the elastic (vibronic) lattice energy and the electron-lattice interaction energy. The basic elements of the configuration-coordinate diagram are briefly sketched here to prepare an understanding of the influence of lattice relaxation on deep-level defects.

The elastic deformation is represented in terms of normal coordinates Q. The meaning of Q depends on the symmetry of the center and the considered distortion. Figure 2 illustrates three of the nine normal coordinates Q_i of a center with tetrahedral symmetry. They are characterized by the irreducible representation of their symmetry point group, in this case T_d. A prominent distortion is the *breathing mode* Q_1 with A_1 symmetry, which actually keeps the symmetry of the center and just scales the size; Q_1 may hence be described by a simple displacement $u = R - R_0$ of an atom next to the central atom. The modes Q_2 and Q_3 both have E symmetry; Q_2 represents a twist about the vertical axis in opposite directions for the lower and upper two atoms. This distortion may be described by a twisting angle $\theta - \theta_0$. In the distortion Q_3 the bond angle θ_0 is varied and may be described similar to Q_2. The remaining six normal coordinates of the tetrahedron involve more complex atom motions with T symmetry.

In a simple harmonic approximation of the lattice energy, the Hamiltonian is given by

$$H_0 = \frac{1}{2}M\frac{d^2u}{dt^2} + \frac{1}{2}\beta\ u^2 = \frac{1}{2M}\left(P^2 + M^2\omega^2Q^2\right), \tag{6}$$

where u is the lattice distortion and β is the restoring force constant – see ▶ Chap. 4, "Elasticity and Phonons". This is conventionally expressed in normal coordinates P (momentum $P = Mdu/dt$) and Q (distortion Q; e.g., $Q = u = R - R_0$). Assuming a linear perturbation (Huang and Rhys 1950; Lax 1952; Pekar 1953; O'Rourke 1953) of the oscillation mode caused by an electron transition and resulting in a shift of the equilibrium configuration,[2] we have for the total Hamiltonian

$$H = H_0 - VQ, \tag{7}$$

which can be transformed by a simple coordinate shift $\tilde{Q} = Q - V/\beta$ (with V the electronic interaction force) into

$$H = \frac{1}{2}M\dot{\tilde{Q}}^2 + \frac{1}{2}\beta\ \tilde{Q}^2 - \frac{1}{2}\frac{V^2}{\beta}. \tag{8}$$

Consequently, the energy as a function of the distortion coordinate is given by parabolas, the *elastic* energy by the usual parabola with its minimum energy $E = 0$ at $Q = 0$. The *total* energy is given by the displaced $(\tilde{E}, \tilde{Q})$ parabola, with its minimum at $(E - V^2)/2\beta$ at $Q = V/\beta$; this yields

$$E_{\text{relax}} = -\frac{V^2}{2\beta}, \tag{9}$$

the relaxation energy after capture of an electron, as shown in Fig. 3 by the solid parabola.

The electronic interaction force V is proportional to a coupling constant S (the Huang-Rhys factor – see below) and the spring constant β, which is proportional to the square of the breathing mode frequency (close to the LO phonon frequency) ω^2_{char} (see Eq. 32 of ▶ Chap. 4, "Elasticity and Phonons"). Consequently, we obtain

$$Q_{\min} \propto -\frac{S}{\omega^2_{\text{char}}}. \tag{10}$$

[2]The basis for this analysis is the adiabatic approximation describd below. Here, the Hamiltonian is split into a part that deals with electrons and atoms at fixed lattice positions, multiplied by a function of displaced lattice atoms. The Hamiltonian can then be written as a sum of the electronic part and the interacting ionic part. This ionic part results in shifted, harmonic oscillations.

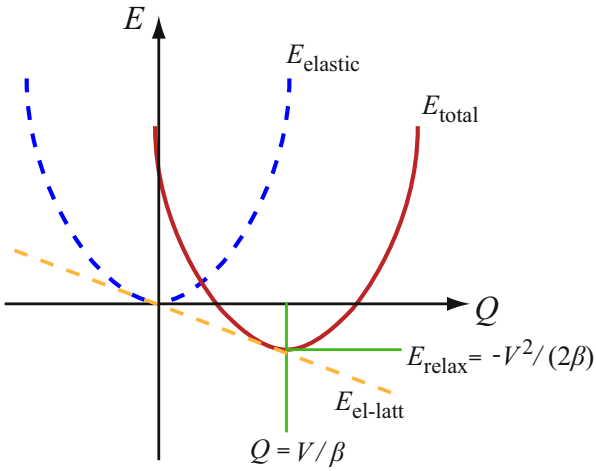


Fig. 3 Configuration-coordinate representation with elastic and electron-lattice interaction energies (*dashed curve*) and resulting total energy (*solid curve*)

In actuality, however, we must take into consideration higher order terms in the lattice distortion and electron-lattice interaction. They cause a deformation of the displaced parabola in addition to its shift and was more recently assigned to the origin of anharmonic carrier capture (Kim et al. 2019).

1.2.2 The Huang-Rhys Factor and Relaxation Process

The electron-lattice coupling in the neighborhood of a tightly bound defect is determined by the interaction between electrons and appropriate phonons and is given by the *Huang-Rhys factor S.* It is related to the coupling constant α_c of a free electron in an ideal lattice – see below. The Huang-Rhys factor can be interpreted as the number of phonons emitted, while a defect center relaxes after it has captured an electron or after it is optically excited and the electron is removed (see Fig. 4; Huang and Rhys 1950 and Henry 1980):

$$S = \frac{E_{\text{relax}} - E_c}{\hbar\omega_{\text{char}}} = \frac{\frac{1}{2}\mu\omega_{\text{char}}^2 (Q_2 - Q_1)^2}{\hbar\omega_{\text{char}}}, \tag{11}$$

where μ is the reduced mass (Eq. 34 of ▶ Chap. 11, "Photon–Phonon Interaction"), E_{relax} is the relaxation energy, $\hbar\omega_{\text{char}}$ is the characteristic energy of the emitted phonons, and Q_2 and Q_1 are the configuration coordinates for the minimum energy in the excited and ground states, respectively – see below.

The Huang-Rhys factor can also be expressed by

$$S = \frac{(Q_2 - Q_1)^2}{2\hbar/(\mu\omega_{\text{char}})}, \tag{12}$$

that is, as the ratio of the square of the total displacement divided by twice the mean square of the amplitude of the zero-point oscillation $\hbar/(\mu\omega_{\text{char}})$.

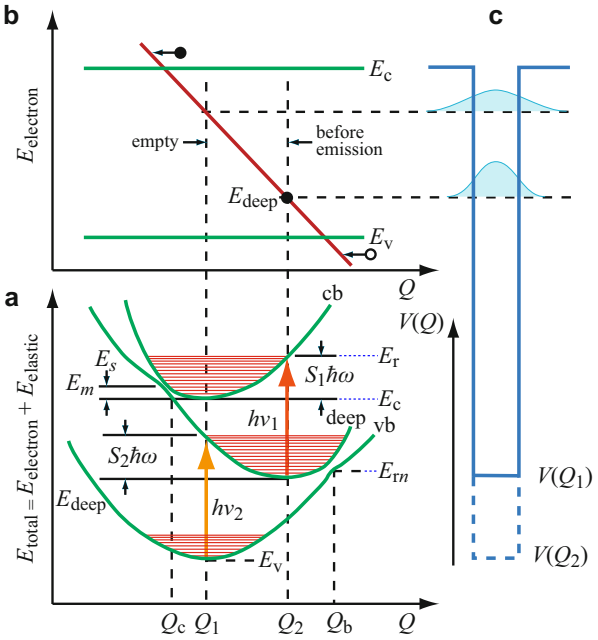


Fig. 4 (**a**) Configuration-coordinate diagram (schematic) of a deep-level center with $E_{deep}(Q)$ minimum at Q_2 between conduction band $E_c(Q)$ and valence band $E_v(Q)$ minima at Q_1 . Optical excitation of the filled deep-level center takes place from its minimum at $E_{deep}(Q_2)$ to the conduction band at $E_c(Q_2)$ with $h\nu_1$, followed by (fast) relaxation with S_1 phonons to $E_c(Q_1)$, the bottom of the conduction band (see Sect. 2.3.1). When the deep center is empty, the excitation occurs from $E_v(Q_1)$ to $E_{deep}(Q_1)$ with $h\nu_2$, and consequent relaxation of S_2 phonons to $E_{deep}(Q_2)$. (**b**) The corresponding band diagram. (**c**) The shape of a simplified core potential with well depths $V(Q_1)$ and $V(Q_2)$ and schematics of the corresponding envelope function (Modified from Henry 1980)

Defect-Level Relaxation In the configuration-coordinate diagram (Fig. 4a), the electronic plus elastic energy of the defect center and its nearest neighbors is plotted as a function of the distance (Q) between the defect and one representative neighbor for a breathing mode. The energy of the defect center, with an electron occupying it, is given by a curve (deep) with minimum at Q_2. When this electron is in the conduction band (cb) and the hole is in the valence band (vb), the minima are at Q_1. Optical (electronic) transitions occur from the relevant minimum without changing Q; consequent relaxation occurs with emission of S_i phonons, thus changing Q until the new minimum is reached; S_i is the corresponding Huang-Rhys factor. A typical example is explained in the caption of Fig. 4. More information on this subject is given in Sect. 2.3.1.

The sum of the optical excitation transitions from the center to the conduction band, and from the valence band to the same center $h\nu_1 + h\nu_2$, can consequently exceed the bandgap energy by a substantial amount:

$$h\nu_1 + h\nu_2 = E_g + S_1\,\hbar\omega_{char} + S_2\,\hbar\omega_{char}. \tag{13}$$

The two Huang-Rhys factors are a measure of the coupling of an electron *trapped* at the center to the lattice. This needs to be distinguished from α_c, which is the coupling factor of a *free electron* to the lattice. Typically, α_c is on the order of 0.1 for predominantly covalent compounds and is much smaller than *S,* which can be on the order of 10 for the same compound, depending on the type of center, i.e., electron localization. The stronger the coupling, the more phonons that can be emitted in the process of relaxation.[3]

1.2.3 First-Principles Calculation of Model Parameters

The configuration-coordinate model with empirical fitting parameters for each center has successfully described the experimentally observed optical bands and nonradiative transitions of many deep-level centers. The parameters of this one-dimensional model are the modal mass M of the effective vibration, the displacement of atoms from the potential energy minima ΔR yielding $\Delta Q = M^{1/2}\Delta R$, and the effective optical phonon frequencies ω_{gnd} and ω_{exc} of the ground and excited states (often simply merged into a single characteristic quantity ω_{char}); in addition, Huang-Rhys factors S are used as given by Eq. 11. It is surprising that such a simple model is applicable although many vibrational modes with very different frequencies participate in the relaxation of quite different centers. Microscopic information about the defect is obscured by individual parameter fitting, though.

The validity of the one-dimensional configuration-coordinate model has more recently been confirmed by first-principles hybrid density-functional theory (Alkauskas et al. 2012). Basically, the multidimensional vibrational problem was mapped onto an effective one-dimensional configuration-coordinate diagram. The calculations focused on the relaxation of atoms close to the defect and on dominant vibrations. Effective frequencies for ground and excited states were calculated from a weighted sum of normal modes, eventually yielding the parameters E_{ZPL}, ΔQ, ω_{gnd} and ω_{exc}; here E_{ZPL} denotes the energy of the zero-phonon line (= the energy difference of vibration ground states of the electronic ground and excited states). Applying a Gaussian smearing with small σ to simulate additional broadening mechanisms, the calculated line shapes agree very well with experimental spectra, as demonstrated, e.g., for cubic diamond and hexagonal 4H SiC (Jin et al. 2021) and crystals listed along with effective parameters for various defect-related luminescence transitions in Table 1.

1.2.4 Recombination-Center Relaxation

The process of capturing a conduction-band electron into such a deep center shall be described here briefly, since it can be understood qualitatively in the configuration-

[3]This coupling is related to the depth of the electron levels. Centers with strong bonding (strong coupling) are more effective in "pushing the surrounding lattice atoms apart" when the electron is excited to a higher energy state. Eigenfunctions of shallow levels have the tendency to "slide over" the surrounding atoms when excited, by permitting the electron to circle within the surrounding lattice, thereby exerting comparatively little force on the surrounding atoms.

Table 1 Calculated (Theo *T*) and experimental (Exp *E*) effective parameters of defect-related transitions in wurtzite semiconductors. E_{ZPL} and FWHM denote the zero-phonon line energy and the full width at half maximum of the phonon-assisted emission band, *M* is the effective mass of the vibrational mode and ΔR is the root mean square displacement of the atoms between two equilibrium configurations. Data for GaN and ZnO are from Alkauskas et al. (2012) and for AlN from Aleksandrov and Zhuravlev (2020)

Solid	Defect transition	Theo/Exp	E_{ZPL} (eV)	FWHM at *T* (eV)	*M* (amu)	ΔR (Å)	$\hbar\omega_{gnd}$ (meV)	$\hbar\omega_{exc}$ (meV)	S_{gnd}	S_{exc}
GaN	$Mg_{Ga}^{0} + e^{-} \rightarrow Mg_{Ga}^{-}$	T	3.24	0.44 at 0 K	45	0.24	47	34	12.2	
		E	3.30	0.36 at 13 K						
	$C_{N}^{0} + e^{-} \rightarrow C_{N}^{-}$	T	2.60	0.35 at 77 K	51	0.22	42	36	11.2	10.3
		E	2.64	0.39 at 77 K			41	40	12.8	13.4
	$V_{N}^{3+} + e^{-} \rightarrow V_{N}^{2+}$	T	3.02	0.33 at 0 K	68	0.45	23	21	36.0	34.8
		E	3.07	0.36 at 5 K						
AlN	$V_{Al}^{2-} + e^{-} \rightarrow V_{Al}^{3-}$	T	2.89	0.43 at 300 K	24.6	0.15	74.1	75.9	5.2	5.3
		E								
	$V_{N}^{2+} + e^{-} \rightarrow V_{N}^{+}$	T	5.03		26.3	0.50	36.1	35.8	28.5	28.2
		E								
	$C_{N}^{0} + e^{-} \rightarrow C_{N}^{-}$	T	4.07	0.47 at 300 K	22.9	0.22	71.1	69.1	9.5	9.2
		E								
ZnO	$N_{O}^{0} + e^{-} \rightarrow N_{O}^{-}$	T	2.20	0.54 at 300 K	48	0.28	40	32	15.3	15.1
		E	2.30	0.55 at 300 K						

coordinate diagram used above to describe the excitation process. This recombination is the inverse to a *thermal excitation,* described first.

When the defect atom is thermally excited to oscillate (curve *deep)* with sufficient amplitude to cross Q_c in Figs. 4a, b, its electronic eigenstate can effectively mix with the excited electron states at E_m within the band. The trapped electron can then pass from the defect state to band states. Inversely, when it is in the conduction band it can pass from band states into the defect state. There, the electron can sequentially emit phonons until a significant fraction of bandgap energy is dissipated, and it has settled into the relaxed state of this defect at Q_2 (Henry 1980). The general models on nonradiative carrier capture are more recently complemented by first principles calculations of specific centers (Alkauskas et al. 2014, 2016).

One of the important *band-to-band recombination* processes takes place via sequential capture of an electron and a hole by the same lattice defect, with relaxation for a substantial fraction of the bandgap energy. The process needs a lattice defect that results in a very deep level with strong lattice coupling. Therefore, such a defect center is also called a *recombination center.*

An electron trapped at the recombination center has the ability to recombine with a hole when thermally excited in the center to E_{rn}, with a transition at Q_b, and consequent relaxation by phonon emission to E_v at Q_1.

This *elastic relaxation* must be distinguished from an *inelastic relaxation* in which the defect center moves into a *metastable* position. Such an inelastic relaxation is related to a photochemical reaction or to DX centers (Sect. 2.7 of ▶ Chap. 19, "Deep-Level Centers"). The inelastic relaxation may also be field enhanced and can be observed by a dielectric loss analysis using *Cole-Cole plot* responses (Hayes and Stoneham 1984).

In the adiabatic limit, the capture cross-section of the recombination center can be evaluated from a semiclassical model (Kayanuma and Fukuchi 1984), as[4]

$$\sigma_r = \sigma_\infty \exp\left(-\frac{E_m}{kT}\right), \quad \text{where} \quad \sigma_\infty = \frac{\omega_{char}}{2\pi}\frac{1}{N_c v_{rms}}. \tag{14}$$

Here σ_∞ is the cross section of the center in a simple gas kinetic approach (see Sect. 2.2 of ▶ Chap. 22, "Carrier-Transport Equations"), E_m is the thermal activation energy, ω_{char} is the lattice-distortion eigenfrequency (in a breathing mode at the recombination center), N_c is the density of states near the band edge, and v_{rms} is the thermal rms velocity of the electrons. This yields (Sumi 1983):

$$\sigma_r(\mathrm{cm}^2) = 6.5 \times 10^{-14}\left(\frac{\hbar\omega_{char}}{40\,\mathrm{meV}}\right)\left(\frac{300\,\mathrm{K}}{T}\right)\left(\frac{m_n}{m_0}\right)\exp\left(-\frac{E_m}{kT}\right). \tag{15}$$

[4]In this chapter cross sections are identified with the commonly used σ. In later chapters treating of transport the symbol s is used instead to distinguish cross sections from conductivity.

When is the Adiabatic Approximation Justified? When an electron is captured from a band state to an excited state of the recombination center, it has a finite probability to be *reemitted* into band states. The net capture probability was evaluated by Henry and Lang (1977) to be 60% of the initial capture transition. In the nonadiabatic limit, reemission of trapped carriers can be neglected (Sumi 1983). The use of the adiabatic approximation, however, is not appropriate, when during the electronic transition within the band some relaxation already has taken place. This lowers the thermal excitation threshold and avoids level crossing.

The decision whether or not to use the adiabatic approximation can be answered by comparing the natural electron lifetime ($\tau_e \cong \hbar/\Delta E_{\mathrm{FWHM}}$, with ΔE_{FWHM} as the full width at half maximum) with the oscillatory time $\tau_{\mathrm{lattice}} \cong 1/\omega_{\mathrm{char}}$, i.e., the time in which the potential of the center changes markedly. With $\tau_e \ll \tau_{\mathrm{lattice}}$, the adiabatic approximation can be used since the electron during its lifetime encounters a certain deformed state of the center rather than a changing one.

The *adiabaticity parameter*

$$\gamma_a = \frac{E_S^2}{\sqrt{S_2 kT\hbar\omega_{\mathrm{char}}}} \tag{16}$$

is a more sophisticated measure of this relation and indicates that at sufficiently high temperatures with $\gamma_a \ll 1$, the adiabatic approximation is acceptable (Sumi 1983).

For the discussion of transitions with major center relaxation, we distinguish three closely related approximations: the *adiabatic approximation;* the *Condon approximation* (Kubo 1952), in which the electron-lattice interaction is linearized; and the *static approach* (Markham 1956), in which the electronic wavefunctions are independent of the defect coordinates, while the lattice wavefunction depends on the electronic state. A listing of earlier literature and a critical discussion of the different approximations is given by Peuker et al. (1982). For further development, see Sumi (1983) and Kayanuma and Fukuchi (1984).

There are other processes which can change the energy of the center, such as Jahn-Teller distortion and the influence of mechanical, electric, and magnetic fields. These effects will be discussed in the appropriate sections of ▶ Chaps. 18, "Shallow-Level Centers", and ▶ 19, "Deep-Level Centers".

Shallow and Deep-Center Spectra The shallow electronic eigenstates are significantly influenced by the lattice environment of the defect. This environment is expressed in terms of the effective mass and dielectric constant. The deep eigenstates are influenced by the chemistry of the defect, the lattice, expressed by its valence and conduction band, and the lattice coupling (i.e., by phonon interaction).

The resulting optical excitation spectrum would be hopelessly complicated were it not for that often – for various reasons – only a few lines appear in the spectrum. Two such reasons are:

1. The lines can have vastly different strengths, causing only the strongest lines to be recognized, while most of the weaker lines disappear in a broad and often unstructured baseline absorption.
2. Many of the lines from excited states of a large variety of defects lie almost on top of each other. These lines are caused by quasi-hydrogen states, which only depend on the host lattice via $\varepsilon_{\text{stat}}$ and m_n and not on the chemistry, i.e., the core potential of the defect (Grimmeiss 1985).

1.2.5 Oscillator Strength of Optical Absorption Lines

A typical absorption spectrum of lattice defects shows a variety of lines with different shapes and amplitudes. The amplitude or *oscillator strength of the absorption* is measured by the absorption constant α_o, which is given by the optical cross section σ_o for the specific transition and the density of the centers:

$$\alpha_o = \sigma_o \; N_d \quad \text{or} \quad \alpha_o = \sigma_o \; J_{fi}. \tag{17}$$

Here N_d is the density of the defect states, and J_{fi} is the joint density of states for an excitation involving bands, e.g., for a transition from a defect level into the band. The cross section for an absorption of a photon by the defect (from initial state (i) to final state (f)) is in turn defined by the ratio of $h\nu$ times the optical transition rate (see Eq. 4 of ▶ Chap. 13, "Band-to-Band Transitions")

$$r_{fi} = \frac{1}{\pi h}\left(\frac{e A_0}{m_0}\right)^2 \left|\mathbf{e}\cdot\mathbf{M}^{(n)}_{fi,qH}\right|^2 \delta(E_f - E_i - h\nu_{fi}) \tag{18}$$

to the energy flux W of the incoming optical radiation; the energy flux is given by

$$W = 2\pi\nu^2 A_0^2 n_r \varepsilon_0 c, \tag{19}$$

which results in an optical cross section of

$$\sigma_o = h\nu \frac{r_{fi}}{W} = \frac{e^2}{2\pi^2 m_0^2 c\, n_r \varepsilon_0 \nu}\left|\mathbf{e}\cdot\mathbf{M}^{(n)}_{fi,qH}\right|^2 \delta(E_f - E_i - h\nu_{fi}). \tag{20}$$

This optical cross section is not very useful for practical evaluation since it assumes an ideal transition represented by a δ-function. Therefore, we need to be more specific as to an actual transition, which involves damping that causes a finite line width and strength; both of them are intimately coupled. We replace the δ function by a more realistic line-shape function, which will be discussed in the next section:

$$g(\nu - \nu_0) = \frac{\gamma h^2}{2\pi}\frac{1}{(h\nu - |E_f - E_i|)^2 + \left(\frac{\gamma h}{2}\right)^2}, \tag{21}$$

where γ is the damping constant ($=1/\tau$ with τ the lifetime in the excited state), yielding for the optical cross section

$$\sigma_{\mathrm{o}} = \frac{e^2 h^2}{4\pi^3 m_0^2 c\, n_{\mathrm{r}} \varepsilon_0 \nu} \left| \mathbf{e} \cdot \mathbf{M}_{\mathrm{fi,qH}}^{(n)} \right|^2 \frac{\gamma}{(h\nu - |E_{\mathrm{f}} - E_{\mathrm{i}}|)^2 + \left(\frac{\gamma h}{2}\right)^2}. \tag{22}$$

Another specific assumption needs to be made to evaluate the momentum matrix element – see Dexter (1958). For instance, for a transition between bound states of a hydrogen-like defect, we have

$$\mathbf{M}_{\mathrm{fi}} = \sum_{\mathbf{k},\mathbf{k}'} \int_V \Psi_m^*(\mathbf{k}') \left(-i\hbar \frac{\partial}{\partial \mathbf{r}}\right) \Psi_n(\mathbf{k}) \; d\mathbf{r}. \tag{23}$$

The impurity function can be expressed as an expansion in Bloch electron eigenfunctions

$$\Psi(\mathbf{k}) = \sum_{\mathbf{k}} c_{\mathrm{c}}(\mathbf{k}) \psi_{\mathrm{c}}(\mathbf{k}, \mathbf{r}) \quad = F(\mathbf{r}) \psi(\mathbf{k}_0, \mathbf{r}), \tag{24}$$

with $F(\mathbf{r})$ as the envelope function. This permits a substantial simplification: Since the Bloch function is always the same, the matrix element changes only with changes in the envelope function. Therefore, we can discuss the strength of the absorption lines from an analysis of the envelope function only and from corresponding selection rules. Equation 23 can then be simplified by separating the matrix element of the unperturbed crystal:

$$\begin{aligned} \mathbf{M}_{\mathrm{fi}} &= \int_V F_{\mathrm{f}}(\mathbf{r}) \psi^*(\mathbf{k}_0, \mathbf{r}) P_A F_{\mathrm{i}}(\mathbf{r}) \psi(\mathbf{k}_0, \mathbf{r}) \; d\mathbf{r} \\ &\cong \int_V |\psi(\mathbf{k}_0, \mathbf{r})|^2 F_{\mathrm{f}}(\mathbf{r}) P_A F_{\mathrm{i}}(\mathbf{r}) \; d\mathbf{r} \end{aligned} \tag{25}$$

with $P_A = \mathbf{A}(-\mathrm{i}\hbar\partial/\partial\mathbf{r})/A_0$. For the envelope function $F(\mathbf{r})$, taken as quasi-hydrogen function

$$F(\mathbf{r}) = \frac{1}{\sqrt{\pi a_{\mathrm{qH}}^3}} \exp\left(\frac{-r}{a_{\mathrm{qH}}}\right) = \sum_k c_{\mathrm{c}}(\mathbf{k}) \exp(\mathrm{i}\mathbf{k} \cdot \mathbf{r}), \tag{26}$$

we obtain coefficients $c_{\mathrm{c}}(\mathbf{k}) = c_{\mathrm{c}}(k)$ by Fourier transform

$$c_{\mathrm{c}}(k) = \frac{8\sqrt{\pi}}{\sqrt{V} a_{\mathrm{qH}}^{5/2} \left[k^2 + \left(\frac{1}{a_{\mathrm{qH}}}\right)^2\right]^2}; \tag{27}$$

here a_{qH} is the quasi-hydrogen radius and V is the normalization volume. We obtain the matrix element of the optical transition at the defect center by multiplying the matrix element for an ideal lattice transition with the coefficient $c_{\mathrm{c}}(k)$ (Eagles 1960; Zeiger 1964):

$$\left|\mathbf{M}^{(n)}_{\mathrm{fi,\,qH}}\right|^2 = |\mathbf{M}_{\mathrm{fi}}|^2\,[c_{\mathrm{c}}(k)]^2. \tag{28}$$

Hence, we observe an enhanced absorption that extends up to k values on the order of $1/a_{\mathrm{qH}}$ (see Eq. 27). The optical absorption caused by such a transition within a hydrogen-like donor has been discussed by Callaway (1963) and Zeiger (1964). The transition is also controlled by selection rules, i.e., by symmetry considerations involving the defect and its surrounding lattice.

The Oscillator Strength Another way of looking at an optical absorption is to start from the free center (an isolated defect atom), which interacts with electromagnetic radiation through its electric ($\mathbf{R} = e\sum_i \mathbf{r}_i$) or magnetic ($\mathbf{J} = \sum_i \mathbf{l}_i + 2\mathbf{s}_i$) dipole moment for i electrons in the center. The matrix elements of these dipole operators relate to the corresponding momentum matrix elements as:

$$|\mathbf{R}_{\mathrm{fi}}| = \int \psi_{\mathrm{f}}\,\mathbf{R}\psi_{\mathrm{i}}dV = 2\pi m\,(\nu_{\mathrm{f}} - \nu_{\mathrm{i}})\int \psi_{\mathrm{f}}\,\mathbf{P}\psi_{\mathrm{i}}dV. \tag{29}$$

For centers with inversion symmetry, the matrix elements are $\neq 0$ only for transitions between states of uneven parity, e.g., from *s* to *p states*. Even-parity transitions are electric dipole-forbidden but magnetic dipole-allowed. These *selection rules* are determined by the symmetry of the center, which can be easily given for an isolated center. The *oscillator strength* of such an isolated center is given by

$$f_{\mathrm{fi}} = \frac{4\pi m\,(\nu_{\mathrm{f}} - \nu_{\mathrm{i}})}{3h}\,|\mathbf{R}_{\mathrm{fi}}|^2 = \frac{1}{3\pi m h\,(\nu_{\mathrm{f}} - \nu_{\mathrm{i}})}\,|\mathbf{P}_{\mathrm{fi}}|^2, \tag{30}$$

and the optical cross section can be expressed as a function of the oscillator strength:

$$\sigma_{\mathrm{o}} = \frac{\pi e^2}{m_0 c} f_{\mathrm{fi}}\,\frac{\gamma}{(h\nu - |E_{\mathrm{f}} - E_{\mathrm{i}}|)^2 + \left(\frac{\gamma h}{2}\right)^2}, \quad \text{where } \gamma = \frac{1}{\tau}. \tag{31}$$

When embedding this isolated center into a lattice, the velocity of light is reduced to c/n_{r} and the effective electric field is increased by the lattice polarization to $F_{\mathrm{eff}} = F_0\left(n_{\mathrm{r}}^2 + 2\right)/3$. When the density N of these centers is small enough to neglect their interaction, the total absorption α_{o} relates to N times f_{fi} (Smakula 1930) as

$$N f_{\mathrm{fi}} = \alpha_{\mathrm{o}}\, n_{\mathrm{r}}\, \gamma \left(\frac{F_0}{F_{\mathrm{eff}}}\right)^2 C^*, \tag{32}$$

where C^* is a constant, and $n_{\mathrm{r}}(F_0/F_{\mathrm{eff}})^2$ is typically on the order of 1. This gives a very simple formula to estimate the density of defect centers from the product of absorption constant $\alpha_{\mathrm{o}}^{\mathrm{max}}$ at the line maximum and line width ΔE, related to γ – see Sect. 2.1:

$$Nf_{\mathrm{fi}} = 1.3 \times 10^{17} \alpha_{\mathrm{o}}^{\max} \left(\mathrm{cm}^{-1}\right) \ \Delta E(\mathrm{eV}). \tag{33}$$

For known defect center densities, the oscillator strength can be determined. For deep centers, f_{fi} is on the order of 1; for F centers in alkali halides $f_{\mathrm{fi}} \cong 0.5$.

2 Line Shape of Electronic Defect Transitions

The optical absorption of defect states by typical resonance absorption has a Lorentzian line shape, with a line width depending on damping. Other composed lattice resonances have Gaussian shape and are discussed in Sect. 2.2. Still other line shapes are related to the specific excitation and deexcitation mechanisms involving phonons and are discussed in Sect. 2.3.

2.1 Homogeneous Lines

All lines created by the superposition of identical contributions from any one of the involved centers are called *homogeneous lines.* Such a superposition is shown in Fig. 5a. Typical examples are listed in this section.

Lifetime Broadening The line shape can be caused by a finite lifetime of the electron in the excited state. It can be described by the damping term in the resonance equation responsible for the absorption, which is of *Lorentzian type* – see Sect. 2.1 and Eq. 46 of ▶ Chap. 11, "Photon–Phonon Interaction". It is given by the line-shape function[5]

$$g(\nu - \nu_0) = \frac{\gamma}{2\pi} \frac{1}{(\nu - \nu_0)^2 + (\gamma/2)^2}. \tag{34}$$

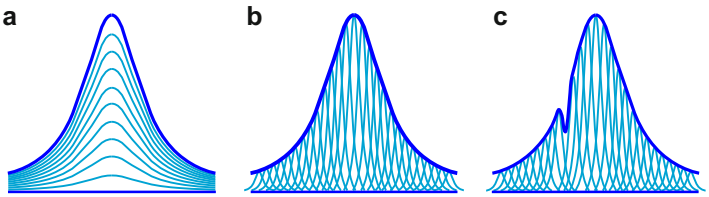


Fig. 5 Schematic representation of line widths: (**a**) Homogeneously broadened line from more and more identical centers. (**b**) Inhomogeneously broadened line from superposition of different homogeneous components. (**c**) Removal of one of the components by photochemical changes (hole burning) or bleaching (After Stoneham 1969)

[5]The line-shape function enters into the absorption cross-section (Eq. 20); that is, $(1/h)g(\nu - \nu_{\mathrm{fi}})$ replaces $\delta(E_{\mathrm{f}} - E_{\mathrm{i}} - h\nu_{\mathrm{fi}})$ after introducing the damping term – see also Eq. 21.

It yields a line width at half of its maximum strength:

$$\mathrm{FWHM} = \gamma \quad \text{(Lorentzian)}. \tag{35}$$

There is another typical line shape, the *Gaussian line shape*, which is observed for deep-center relaxation – see Sect. 2.3. It is often related to inhomogeneous broadening (see Sect. 2.2). The line-shape function for a Gaussian line is given by

$$g(\nu - \nu_0) = \frac{1}{\sqrt{2\pi}\gamma} \exp\left[-\left(\frac{\nu - \nu_0}{2\gamma}\right)^2\right] \tag{36}$$

with a line width at half of its maximum strength

$$\mathrm{FWHM} = 2\sqrt{2\ln 2}\,\gamma = 2.355\,\gamma \quad \text{(Gaussian)}. \tag{37}$$

Both line shapes are distinguished by their tails (the Gaussian line shows a steeper decline) as shown in Fig. 6. The damping factor in Eqs. 32 and 34 can be related to the lifetime of the excited state and in simple cases is given by $\gamma \cong 1/\tau$.

In defect centers, the intrinsic line width is determined by the natural lifetime in the excited state. This lifetime in turn can be determined by

- Recombination
- Nonradiative bound-carrier-phonon interaction

Both contributions depend on the actual type of defect center. For all practical reasons, the first contribution is usually small compared to the phonon interaction.

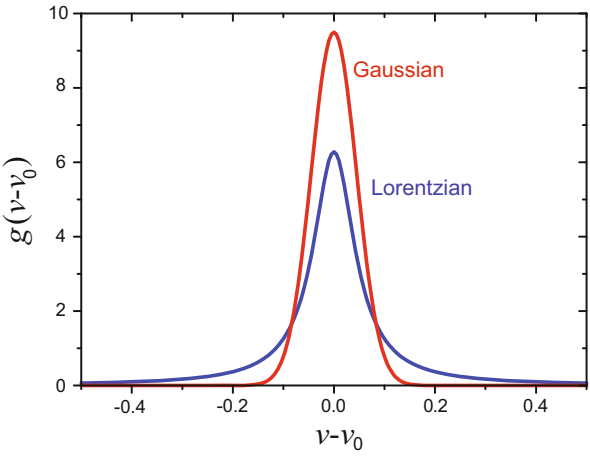


Fig. 6 Line-shape function according to Eqs. 34 and 36 with $\nu_0 = 10^{12}\ \mathrm{s}^{-1}$ and $\gamma = 1\ \mathrm{s}^{-1}$

In *shallow* centers and at low temperatures (<20 K) the emission of acoustic phonons is the limiting process. For instance, a transition from $2p_0$ to 1 *s* of a quasi-hydrogen defect yields for such a process a line width of

$$\Delta\nu \cong 10^{-3} \times \left(\Xi^2/\rho\, a_{\mathrm{qH}}^3\, v_{\mathrm{s}}^3\right) \nu_{\mathrm{fi}}, \tag{38}$$

where ρ is the density of the semiconductor, Ξ is the deformation potential, a_{qH} is the quasi-hydrogen radius, and v_{s} is the sound velocity. For a donor in Ge this line width $h\Delta\nu$ is estimated to be on the order of 10 μeV. In stress-free ultrapure Ge with (H,O) donors, it was observed (Haller et al. 1987) to be approximately 8.5 μeV.

In contrast to the homogeneously broadened line, which is rarely seen,[6] we usually observe inhomogeneously broadened lines when slightly shifted narrow lines are superimposed. This is shown in Fig. 5b and will be discussed in Sect. 2.2.

Influence of a Magnetic Field The optical absorption of defect centers depends on a number of external influences. Most of these cause inhomogeneous broadening and will be discussed later. A magnetic field, however, can have an influence on the homogeneous line width. The high magnetic inductance compresses the defect eigenfunctions (see Sect. 3.4 of ▶ Chap. 18, "Shallow-Level Centers"), causing a reduction in the scattering cross-section and thereby an increase in the lifetime of hydrogen-like defects. This causes a reduction in the homogeneous line width.

Such a line narrowing is shown in Fig. 7 for InSb, which has a low effective mass, resulting in extremely large quasi-hydrogen radii. Consequently, this results in a large magnetic compression at only a moderate magnetic inductance. For more detail, see Stradling (1984).

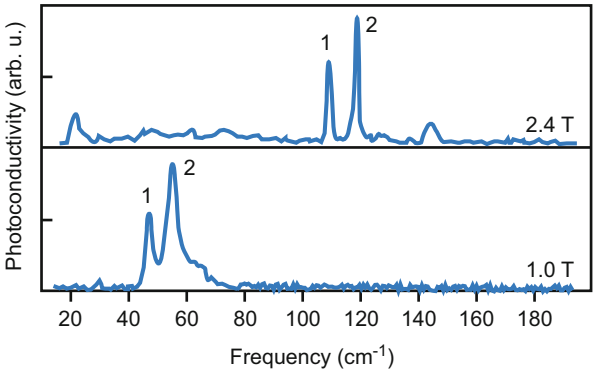


Fig. 7 Cyclotron resonance (1) and 1 $s \rightarrow 2p^+$ transition (2) of a hydrogen-like impurity in InSb at 4.2 K, showing a shift and substantial narrowing of the lines with increased magnetic inductance from 1.0 to 2.4 T (After Stradling 1984)

[6]Only in ultrapure semiconductors with vanishing internal stress and vanishing electric fields, the lines of isolated impurities have their natural width – see Jagannath et al. (1981) and Haller et al. (1987); also the natural isotope mixture contributes substantially to inhomogeneous broadening as demonstrated for isotopically pure ^{28}Si – see Karaiskaj et al. (2003).

2.2 Inhomogeneous Broadening

Most line broadening is due to the superposition of emissions from different centers of the same impurity atom, which differ from each other because of slightly variant lattice environments. Such environments may differ by built-in strain or electric fields, or by proximity to other defects. *Inhomogeneous line broadening* (Stoneham 1969) is illustrated in Fig. 5b, and can be detected by bleaching one subgroup of these centers with a narrow laser line (see Fig. 5c), which can be detected with a consequent probing scan. This bleaching of a narrow line is often referred to as *hole burning* or *spectral hole burning.*

Hole Burning The width of the spectral hole, burned with a narrow line-width laser, depends on the transition probability of the defect center in a statistically arranged surrounding. With a short exposure, followed immediately by probing (echo), only such centers with highest excitation probability are bleached. This results in a narrow line width of the bleached spectral "hole". When given a longer bleaching time, statistical fluctuations permit a wider distribution to be bleached, making the line width of the "hole" wider.

2.2.1 Line Broadening by Electric Fields and Stress

An electric field causes a *Stark effect* shift and splitting of the electronic eigenstates of the defect – see Sect. 3.3 of ▶ Chap. 18, "Shallow-Level Centers". For hydrogen-like defects with large orbitals, this effect is rather large. It can be estimated by including the external field **F** into the Hamiltonian of the effective-mass equation:

$$\left(-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial \mathbf{r}^2}-\frac{e^2}{4\pi\varepsilon_{\text{stat}}\varepsilon_0 r}-V_{\text{ext}}(r)\right)\psi_e \;=\; E\;\psi_e. \tag{39}$$

The eigenvalues for Eq. 39 in a uniform electric field ($V_{\text{ext}} = (\mathbf{F}/e)\cdot\mathbf{r}$) are

$$E_{\text{qH}}(F)=E_{\text{qH}}^{(0)}\left\{-\frac{1}{n^2}+\frac{3}{2}\frac{Fa_{\text{qH}}^{(0)}}{E_{\text{qH}}^{(0)}}n(n_1-n_2)-\left(\frac{Fa_{\text{qH}}^{(0)}}{2E_{\text{qH}}^{(0)}}\right)^2\frac{n^4}{8}\times\left[17n^2-3(n_1-n_2)-9m^2+19\right]+\ldots\right\}, \tag{40}$$

with the first term representing the undisturbed quasi-hydrogen solution and the second and third terms representing the linear and quadratic Stark effects, respectively. The superscript (0) identifies the ground state for quasi-hydrogen solutions, n_1 and n_2 are two integers (≥ 0) fulfilling $n = n_1 + n_2 + |m| + 1$, and m is the orbital quantum number – here the component in **F** direction. Such an external (to the lattice defect) field can also be produced by other charged centers in sufficiently close proximity. Then, $V_{\text{ext}}(r)$ is the potential distribution modified through such charged centers $\left(\sum\{e_i/(4\pi\varepsilon_{\text{stat}}\varepsilon_0|\mathbf{R}_i-\mathbf{r}|)\}\right)$.

The statistical distribution of defects throughout the lattice causes a variation of this effective field from defect to defect. Consequently, it creates an accumulated broadening of the line, instead of a well-defined shift and splitting of the hydrogen-like spectrum.

The probability of finding an electric field F^* is given (Larsen 1976) by

$$P(F^*) = \frac{2}{\pi} F^* \int_0^\infty r \exp\left(\frac{15}{4}(2\pi r)^{3/2}\right) \sin(F^* r)\ dr. \tag{41}$$

Here F^* is expressed by the once-integrated Poisson equation $(\rho \overline{r}/(\varepsilon_{\text{stat}}\varepsilon_0))$. After introducing the average distance between defects $\overline{r} = N_i^{-1/3}$, we obtain

$$F^* = e N_i^{2/3}/(\varepsilon_{\text{stat}}\varepsilon_0), \tag{42}$$

where N_i is the total density of charged centers. The probability function (Eq. 41) is given in Fig. 8a and shows a maximum at $5 e N_i^{2/3}/(\varepsilon_{\text{stat}}\varepsilon_0) \cong 420$ V/cm for $N_i = 10^{13}$ cm^{-3} and $\varepsilon_{\text{stat}} = 10$.

The strong broadening of higher excited states is shown in Fig. 8b for the $1\,s \rightarrow 3p$ transition of a shallow donor, peaking near 42 cm^{-1} (5.2 meV). Their larger eigenfunctions are more prone to overlap into adjacent regions of higher fields from neighboring Coulomb centers when compared to the narrower $1\,s \rightarrow 2p$ transition near 35 cm^{-1}. For further detail, see Larsen (1976).

Spectral Diffusion If statistically random charging of traps (by optical or electrical excitation) occurs in the lattice environment of an emitting center, the *temporally* fluctuating Stark effect leads to energy shifts of the emitter's luminescence. This

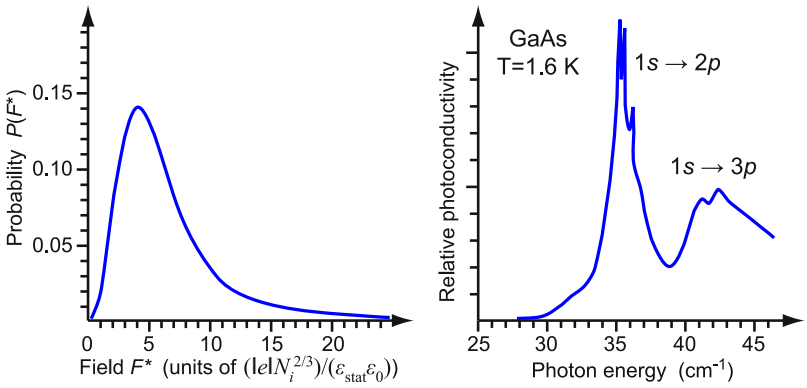


Fig. 8 (**a**) Probability distribution for finding an average field F^* produced by a random distribution of point charges of density N_i at a given lattice point (Holtzmark distribution). (**b**) Photoconductivity distribution of high purity GaAs at 1.6 K (instrument resolution 8.6 μeV = 0.069 cm^{-1}), showing substantial broadening of the 42 cm^{-1} peak of a hydrogenic donor due to electric fields from neighboring Coulomb centers (After Larsen 1976)

phenomenon is referred to as *spectral diffusion* and was first reported for spin resonance experiments (Klauder and Anderson 1962). Such shifts take place on a large variety of energy and time scales, usually well below the time resolution of the recording setup. Consequently the natural line width of a single emitter is inhomogeneously broadened. The effect is observed in many luminescence centers with narrow natural line width, and as well in quantum wells and quantum dots where the interaction is termed *quantum-confined Stark effect*. In quantum dots, spectral diffusion was applied to assign line groups with synchronous jitter to different emissions from a single emitter (Türck et al. 2000); characteristic spectral diffusion times down to the nanosecond range were found (Sallen et al. 2011).

Mechanical Stress Broadening Mechanical stress due to dislocations or other statistically distributed defects surrounding a specific type of lattice defect contributes to the line broadening of this defect in a manner similar to electric field broadening. Such stresses, however, have a stronger influence on the structure of the conduction band. Hence, they will preferably broaden the donor states created by mixing with such band states – see Sect. 3.2 of ▶ Chap. 18, "Shallow-Level Centers". However, other states may be influenced also, as, e.g., nondonor-like states can be perturbed by a large *deformation potential.*

2.2.2 Line Broadening with Compositional Disorder

The line widths of hydrogenic defects (excitons and impurities) are substantially broader in semiconductor alloys than in their unalloyed components. As an example, an exciton line-width is analyzed here. It should be noted that even the natural isotope mixture in pure unalloyed semiconductors leads to a (small) inhomogeneous broadening (Karaiskaj et al. 2003).

The compositional disorder in random alloys causes a local variation of the exciton energy. Depending on the actual composition of the alloy within the volume of the exciton

$$V_{\mathrm{exc}} = \frac{4\pi}{3}\langle r_{\mathrm{exc}}^3 \rangle = 10\pi a_{\mathrm{qH}}^3, \tag{43}$$

one observes a variation of $\varepsilon_{\mathrm{stat}}$ and m_n, which in turn causes a change in a_{qH} and E_{qH} (Goede et al. 1978). The probability of finding j atoms A of the alloy $A_{\bar{x}}B_{1-\bar{x}}C$ in the volume of an exciton is given by

$$P(j, N_{\mathrm{exc}}) = \binom{N_{\mathrm{exc}}}{j} \bar{x}^j (1-\bar{x})^{N_{\mathrm{exc}}-j}, \tag{44}$$

with N_{exc} as the number of cation sites in the exciton volume V_{exc}, and $\bar{x} = j/N_{\mathrm{exc}}$. For sufficiently large N_{exc} one can approximate Eq. 44 with a Gaussian distribution:

$$P(x) = \frac{1}{\sqrt{\pi} N_{\text{exc}} \sigma} \exp\left(-\frac{(x - \overline{x})^2}{\sigma^2} \right), \tag{45}$$

which yields a width at half of its maximum value

$$\text{FWHM} = 2\sqrt{\ln 2}\,\sigma = 2\sqrt{2\ln 2}\sqrt{\frac{\overline{x}(1-\overline{x})}{N_{\text{exc}}}}. \tag{46}$$

This results in a half-width of the exciton line

$$\begin{aligned} \text{FWHM (exciton)} &= 2c\sqrt{\ln 2}\,\sigma \left| \partial E_{\text{g}} / \partial \overline{x} \right| \\ &= 2c\sqrt{2\ln 2}\left\{ \Delta E_{\text{g}} - b(2\overline{x} - 1) \right\} \sqrt{\frac{\overline{x}(1-\overline{x})}{10\pi\sqrt{2}} \left(\frac{a}{a_{\text{qH}}} \right)^3}, \end{aligned} \tag{47}$$

where $\sqrt{2}a^3$ is the volume of the primitive cell, c is a fitting parameter on the order of 1 (here $\cong$ 1.1), and $\Delta E_{\text{g}} = E_{\text{g}}\,(x = 0) - E_{\text{g}}\,(x = 1)$. With the bowing parameter b (here $\cong$ 0.3 eV) the bandgap energy is given by

$$E_{\text{g}}(x) = (1 - x)E_{\text{g}}(0) + xE_{\text{g}}(1) - bx(1 - x). \tag{48}$$

The experimentally observed line broadening of exciton luminescence agrees reasonably well with the given estimates, as shown in Fig. 9 (see also Singh and Bajaj 1986). For line broadening in quantum wells by compositional disorder see Baranovskii et al. (1994).

Line Broadening with Heavy Doping A very similar line broadening due to the statistical fluctuation of interdefect distances is observed with heavy doping. In a simplified approach, the line width can be estimated as

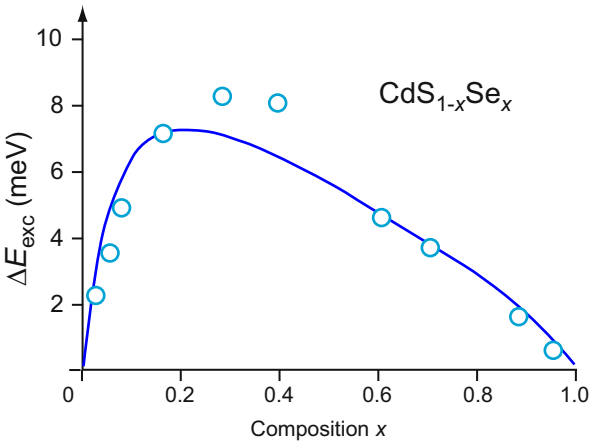


Fig. 9 Line-width broadening of the A exciton in $CdS_{1-x}Se_x$ due to compositional disorder. The curve is calculated from Eq. 47 with $\Delta E_{\text{g}} = 0.728$ eV, $b = 0.31$ eV, $a_{\text{qH}} = 25$ Å, $c = 1.1$ (After Goede et al. 1978)

$$\mathrm{FWHM} = \mathrm{FWHM}\,(0)\ \exp\left(-\frac{\overline{r}}{a_{\mathrm{qH}}}\right), \tag{49}$$

with a_{qH} as the quasi-Bohr radius of the hydrogen-like defect, and $\overline{r} = N_{\mathrm{d}}^{-1/3}$ as the mean distance between the defects. FWHM (0) is the normalized line width (Mott 1974).

2.3 Deep-Level Defect Line Broadening

Deep-level defect centers can show a much larger line broadening when the coupling with lattice phonons is strong. The spectrum of some of these deep-level defects shows, even at low temperatures, a broad Gaussian line shape of typically several $\hbar\omega_{\mathrm{LO}}$ half-widths. This signature is an indication of a strong coupling of the defect center with the surrounding lattice. Strong lattice coupling can yield line widths of several hundred meV. In contrast, narrow line widths of the zero-phonon absorption of $\leq$1 meV, with well-separated phonon replica are observed for other centers with only minor lattice coupling.

The magnitude of the lattice coupling is a function of

1. The host lattice: The coupling is increased with higher ionicity of the lattice binding-forces and lower elastic force constant.
2. The defect center: Point defects show a stronger coupling than defect associates.
3. The electronic eigenstates of the center: Deep defect levels show a stronger coupling than shallow states.

In a simplified picture, we may observe a deep-defect center with strong coupling that tends to push the surrounding lattice atoms apart when an electronic excitation occurs, rather than to extend a hydrogen-like orbit over more of the surrounding lattice. The changes in hydrogen-like orbits have little effect on the position and polarization of the neighboring atoms, while the deformation of the lattice surrounding of an expanding, tightly bound center has a larger effect. In strong interaction with the surrounding lattice, many phonons are created by the relaxation that follows such an electronic excitation and have a major effect on its excitation energy and line width. The basic principles involved are discussed in the following section.

2.3.1 Franck-Condon Principle and Lattice Relaxation

The processes involved during excitation of a defect center with strong lattice coupling are described in a configuration-coordinate diagram as introduced in Sect. 1.2.1. However, the configuration-coordinate diagram is applied here to the ground and excited states of a deep center (Fig. 10). It pictures for a breathing mode the electronic and elastic lattice energies as a function of the relative distance between the defect center and its nearest surrounding lattice atoms. The center's ground state is identified by A_0, within a parabola $E(Q)$ near its minimum. Horizontal

lines indicate specific *phonon-excited* states. With *electronic excitation,* the parabola is displaced. It is shifted in energy by the electronic excitation energy and in interatomic distance to a larger equilibrium distance Q_{B_0}; the excited defect atom has an increased effective radius (Fig. 10b).

In Fig. 10a the vibrational eigenfunction $\phi(Q)$ is indicated from the adiabatic approximation (▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals", Eq. 2). It can be shown that in addition to the normal selection rules for the electronic part, the optical transition probability depends on the vibrational part of the eigenfunctions (see Eq. 58) and is largest when the overlap integral

$$\int \phi_{\mathrm{f}}^{*}(Q)\,\phi_{\mathrm{i}}(Q)\;\,dQ \tag{50}$$

is a maximum. This is the case for an approximately vertical transition (*Franck-Condon principle*; Fig. 10a). For the original discussion of the Frank-Condon principle, see Condon and Morse (1929).

With changes from this optimum transition energy, the transition probability drops off gradually, resulting in a rather broad line width (Fig. 10b).

After optical excitation, the vibrational state is also excited and will relax to its minimum energy (B_0 in Fig. 10b). The difference between the initial excited state and the relaxed state represents the *Franck-Condon shift* (Lax 1952) between optical and thermal excitation. The latter, as an indirect transition, can proceed directly from minimum to minimum and thus needs less energy than the optical excitation. In a

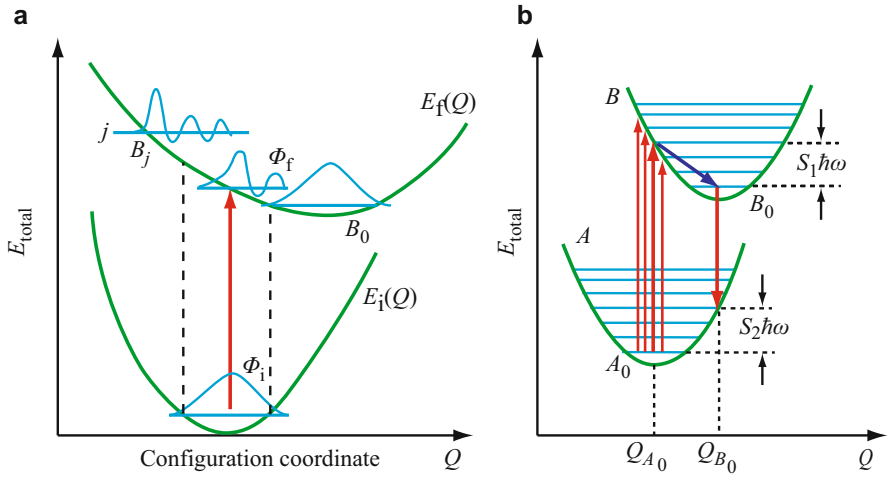


Fig. 10 Configuration-coordinate diagram showing the ground and excited states of a defect with strong lattice coupling and typical excitation and recombination transitions. (**a**) Ground and excited vibrational states indicated. (**b**) Franck-Condon principle showing larger optical energy than needed for thermal excitation from A_0 to B_0 and indicating absorption line shape ($A_0 \rightarrow B$) with multiple transitions possible. *Horizontal lines* indicate vibrational states of the system

similar fashion, light emission (luminescence) proceeds vertically from the relaxed excited state to the ground electronic state (A in Fig. 10b) and then relaxes to its minimum total energy state A_0.

The energy difference between optical excitation ($A_0 \rightarrow B$) and emission ($B_0 \rightarrow A$), *the Franck-Condon transitions,* is called the *Stokes shift.*

2.3.2 Strong, Medium, and Weak Lattice Coupling

The strength of the coupling can be measured by the number of phonons emitted during a relaxation process. This is represented by S_1 after optical excitation, and S_2 after a photon emission (Fig. 10b), with S_i as the Huang-Rhys factor – see below and Sect. 1.2.2.

The total energy of the defect is given by

$$E_e = E_{e_0} + E_Q = E_{e_0} + \tfrac{1}{2}\beta(Q - Q_0)^2, \tag{51}$$

where E_{e0} is the electronic ground-state energy, E_Q is the elastic energy, and β is the force constant. The relaxation energy in state B, following a transition from state A_0, is expressed by (see Fig. 10b)

$$E_Q = \tfrac{1}{2}\beta\left(Q_{A_0} - Q_{B_0}\right)^2 = \tfrac{1}{2}\mu\omega_{\text{char}}^2\left(Q_{A_0} - Q_{B_0}\right)^2, \tag{52}$$

where $\omega_{\text{char}}^2 = \beta/\mu$ and μ is the reduced mass of the center. The Huang-Rhys factor is given by Eq. 11.

When $S > 6$, we speak of strong coupling; when $S \ll 1$, we have weak coupling. The line shape is substantially different in each case.

Strong Coupling For strong coupling, the line-shape function is given by a Poisson distribution of the phonon spikes at low temperatures

$$g(\nu - \nu_0) = \sum_m \left\{\frac{S_1^m}{m!}\exp(-S_1)\right\}\delta\left(\nu - \nu_0 - \frac{\omega_{\text{char}}}{2\pi}m\right), \tag{53}$$

with m as the number of phonons emitted at the specific transition and the delta function representing the zero-phonon line (for $m = 0$) as well as the phonon replica. This results in a spectrum of spikes separated by $\hbar\omega_{\text{char}}$. In actuality, each of these phonon lines is broadened, and all replica melt into one broad feature for high values of the Huang-Rhys factor.

At higher temperatures ($kT > \hbar\omega_{\text{char}}$), the line-shape function converts to a Gaussian distribution

$$g(\nu - \nu_0) = \frac{1}{\sqrt{4\pi S_1 \hbar\omega_{\text{char}} kT}}\exp\left[-\frac{(h\nu - h\nu_0 - S_1\hbar\omega_{\text{char}})^2}{4S_1\hbar\omega_{\text{char}}kT}\right], \tag{54}$$

where $h\nu_0$ is the optical electronic transition ($A_0 \rightarrow B$) in Fig. 10b, and ω_{char} is the frequency of the characteristic phonon mode. Individual phonon lines are not resolved but contribute to a broad line of width

$$\text{FWHM} = \begin{cases} 2\sqrt{2\ln 2}\sqrt{S_1}\hbar\omega_{\text{char}} & \text{for } kT < \hbar\omega_{\text{char}} \\ 2\sqrt{2\ln 2}\sqrt{S_1\hbar\omega_{\text{char}}kT} & \text{for } kT > \hbar\omega_{\text{char}}. \end{cases} \tag{55}$$

It is the strong coupling that permits interaction with a wide phonon spectrum. It broadens each individual line, so that even at very low temperatures only one broad feature appears in absorption and emission. This is shown in absorption and, shifted by approximately twice the relaxation energy, in emission for the F centers of KBr in Fig. 11.

Weak Coupling For weak coupling the parabola for the excited electron state (Fig. 10b) is not shifted significantly from the ground-state parabola. This causes the most probable transition to be the minimum-to-minimum zero-phonon line with a narrow line width.

In addition, Stokes-shifted lines with phonon emission and anti-Stokes lines with photon absorption occur as phonon replica. The line-shape function here is represented by

$$\begin{aligned} g(\nu - \nu_0) = (1 - S)\,\delta(\nu - \nu_0) + \frac{1}{2}(S + S_i)\,\delta(\nu - \nu_0 - \hbar\omega_{\text{char}}/(2\pi)) \\ + \frac{1}{2}(S - S_i)\,\delta(\nu - \nu_0 + \hbar\omega_{\text{char}}/(2\pi)) + \ldots, \end{aligned} \tag{56}$$

where the dots indicate higher-order replica. Each δ function is lifetime broadened, as discussed in Sect. 2.1. S_i is the Huang-Rhys factor given by Eq. 12, and

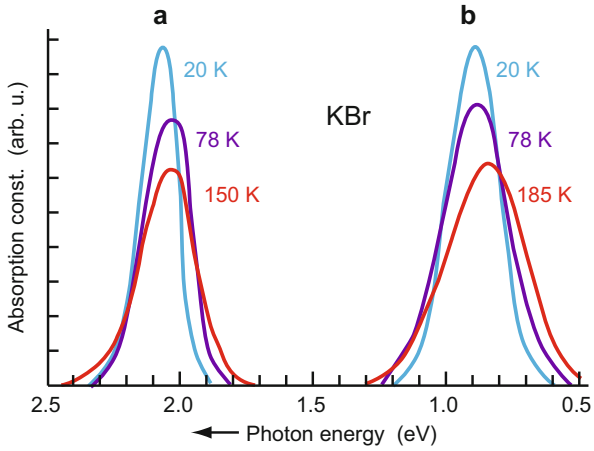


Fig. 11 (**a**) Optical absorption and (**b**) luminescent emission spectra of F centers in KBr for various temperatures (After Gebhardt and Kuhnert 1964)

$$S = S_i \coth\left(\frac{\hbar\omega_{\text{char}}}{kT}\right). \tag{57}$$

For weak coupling, these replica are obtained from Eq. 56 as a series of spikes.

Intermediate Coupling With increasing intermediate coupling, the relative strength of the zero-phonon line gradually decreases, and the phonon replica broaden, which renders them less distinct. Finally, all features combine into a band, which then becomes the broad absorption "line" of the defect for strong coupling. This development is shown in Fig. 12 for the R_2 center – which is an associate of 3 F centers – in different alkali halides. These defect associates behave

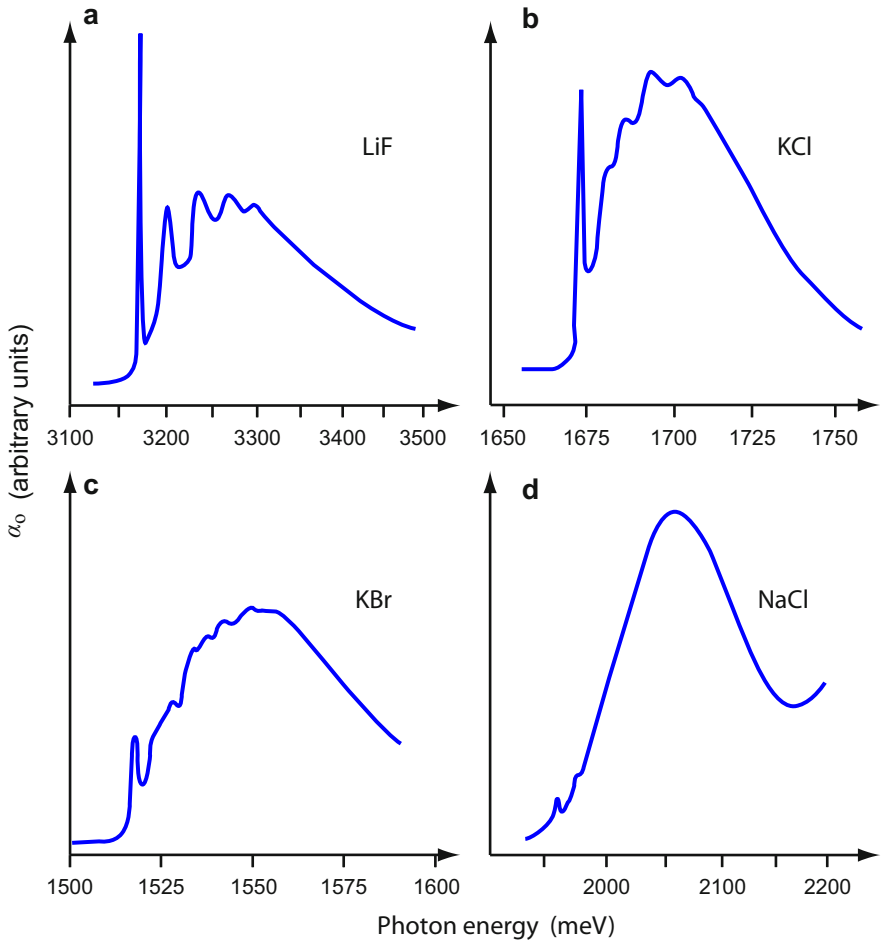


Fig. 12 Optical absorption spectra of R_2 centers at 4 K for LiF, KCl, KBr, and NaCl, showing decreasing structure with increasing lattice coupling (After Fitchen et al. 1963)

as one somewhat larger defect. This results in an intermediate behavior between a deep center and a center with a more extended electron eigenfunction.

With increasing coupling (see Fig. 14), the features in absorption and emission become mirror symmetric (Fig. 13): the zero-phonon line in absorption finally disappears, both maxima become more and more separated (Fig. 11), and the distinction between different phonon replica becomes completely washed out (Fig. 12).

The degree of coupling depends on the elastic and electronic properties of the host. A simple empirical relationship was observed between relaxation energy and the electronegativity χ of the host anions (Baranowski 1979). This dependence is shown for the zinc chalcogenides in Fig. 15.

Quantitative Estimates from the Adiabatic Approximation The first quantitative evaluation of the optical absorption of deep centers with strong coupling was given by Huang and Rhys (1950) for the F center absorption in alkali halides. Following the adiabatic approximation with $\phi(\mathbf{R},\mathbf{r}) = \varphi(\mathbf{R})\psi(\mathbf{R},\mathbf{r})$ (see Eq. 2 of ▶ Chap. 7,

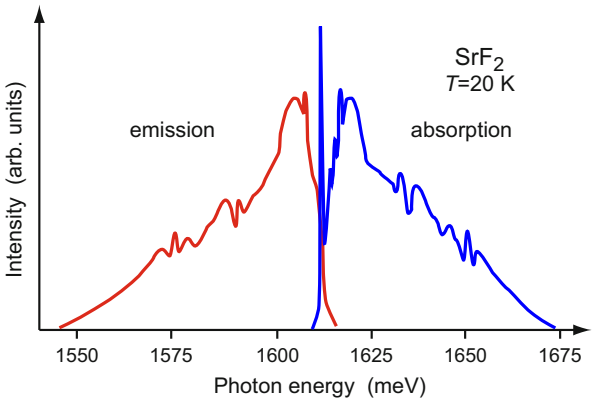


Fig. 13 Mirror-symmetric optical absorption and emission specta for R centers in SrF_2 at 20 K as an example for intermediate lattice coupling (After Beaumont et al. 1972)

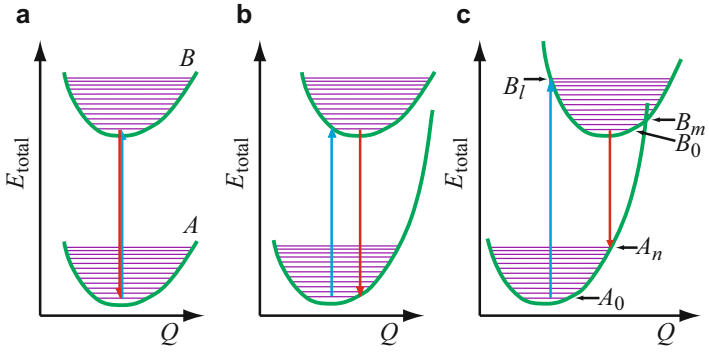


Fig. 14 Ground and excited states in a configuration-coordinate diagram for coupling increasing from subfigs. (**a**) to (**c**)

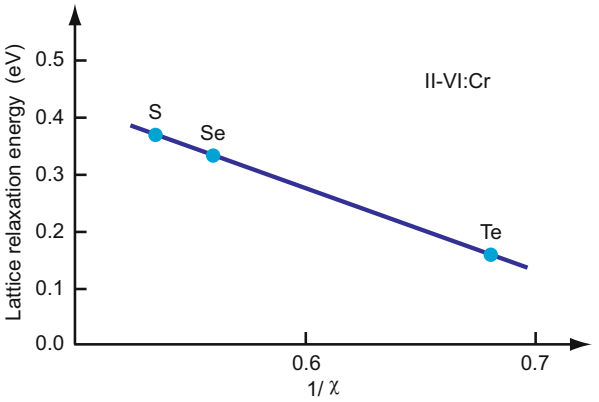


Fig. 15 Lattice-relaxation energy (Franck-Condon shift) for Cr centers as a function of the reciprocal electronegativity of S, Se, and Te in the corresponding zinc chalcogenid host crystals (After Baranowski 1979)

"Quantum Mechanics of Electrons in Crystals"), we separate electron and phonon eigenfunctions. This permits us to express the matrix elements for the optical transition as the product of the matrix elements for the excitation from the electronic ground ($\mu 0$) to the excited state ($\mu 1$) and the overlap integral between the vibrational wavefunctions:

$$\mathbf{M}^{\mu 1\, n1}_{\mu 0\, n0} \cong \int \psi^*_{\mu 0} \left(-\frac{\hbar^2}{2m^*} \frac{\partial}{\partial \mathbf{r}} \right) \psi_{\mu 1} d\mathbf{r} \int \varphi^*_{\mu 0\, n0}(\mathbf{R})\ \varphi_{\mu 1\, n1}(\mathbf{R})\ d\mathbf{R}, \tag{58}$$

where $\psi_{\mu 0}$ and $\psi_{\mu 1}$ are the electronic eigenfunctions, and $\varphi_{\mu 0\ n0}$ and $\varphi_{\mu 1\ n1}$ are the vibrational wavefunctions with the electron in its initial and final states, respectively. The absorption constant within the optical frequency range ν, $\nu + \Delta\nu$ is given by

$$\alpha_o(\nu) = \frac{8\pi^2 N_{\text{deep}}\, e^2}{3h\, m_0^2\, c\, n_r \varepsilon_0 \nu} \left| \mathbf{M}_{\mu 0\, \mu 1} \right|^2 G(\nu), \tag{59}$$

where $\mathbf{M}_{\mu 0 \mu 1}$ is the matrix element for the electronic transition (first integral in Eq. 58), N_{deep} is the density of deep centers, and $G(\nu)$ given as the sum of the overlap integrals

$$G(\nu) = \frac{1}{\Delta\nu} \sum_{n1}^{\nu, \nu+\Delta\nu} \left| \int \varphi^*_{\mu 0\, n0}(X)\ \varphi_{\mu 1\, n1}(X)\ dX \right|^2. \tag{60}$$

The first factor in Eq. 59 determines the amplitude of the optical absorption; the second factor $G(\nu)$ determines the line shape which can be approximated by the line-shape function given in Eq. 54.

There is reasonable agreement between the Huang-Rhys theory and experimental observations of F centers in KBr (see Fig. 16) with three adjustable parameters. These include a very large Huang-Rhys factor ($S \cong 25$), which is characteristic of

such F centers. For a review of the measurements of the related oscillator strengths, see Huber and Sandeman (1986).

2.3.3 Line Shape of Resonant States

Resonant states are those that overlap with a band, thereby substantially reducing the lifetime of an electron in the resonant state. Excitations into resonant states (Fig. 17) can be distinguished by the shape of the absorption peak, which can be approximated by (included in the absorption cross-section)

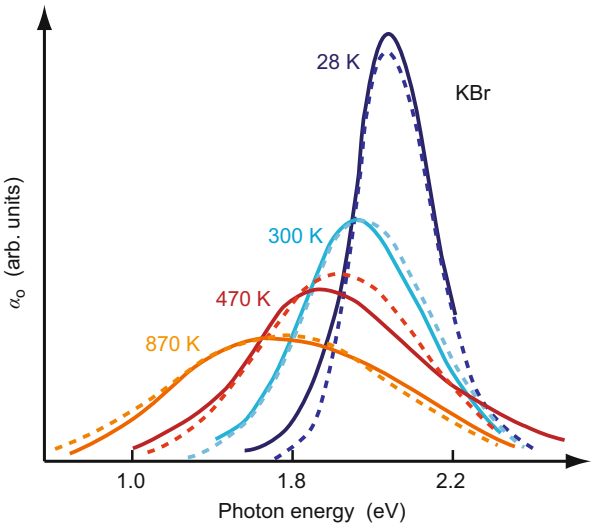


Fig. 16 Theoretical (*dashed*) and experimental (*solid*) optical absorption curves of F centers in KBr at various temperatures (After Huang and Rhys 1950)

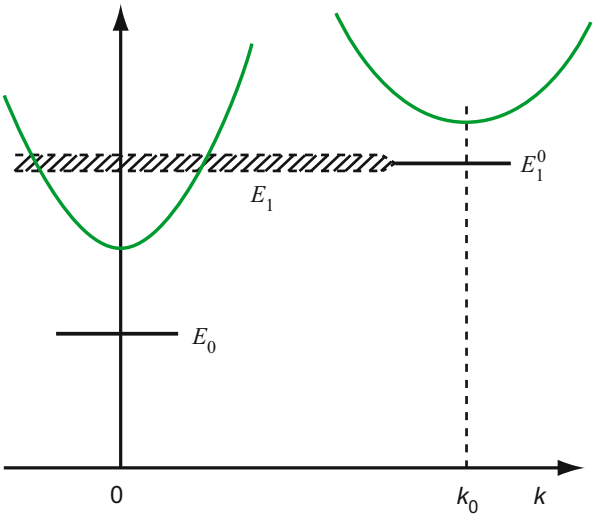


Fig. 17 Conduction bands with minima at $k = 0$ and $k \neq 0$. The state E_0 lies in the bandgap and is a bound state with negligible line width. The state E_1 originates from the unperturbed state E_1^0 and is a resonant state with greatly expanded line width, as the lifetime of an electron in this state is substantially reduced

$$\sigma_{\text{res}}(h\nu) = \frac{a + b(h\nu - E_0)}{(h\nu - E_0)^2 + \gamma^2_{\text{deep, res}}\, h^2}\,, \tag{61}$$

where $1/\gamma_{\text{deep,res}}$ is the lifetime of the electron in the resonant state, and a and b are the empirical parameters (Toyozawa et al. 1967; Velicki and Sak 1966). Resonant peaks are observed which relate to shallow donors, acceptors, or to excitons. See, for example, Onton (1971) for such resonant states in GaP.

2.4 Photoionization Edge Shape

Excitation from a level directly into a band that results in free carriers is referred to as photo ionization; the optical absorption spectrum is edge like – see Sect. 1.1.1 of ▶ Chap. 31, "Photoconductivity". The spectral distribution of the absorption is usually expressed in terms of the photoionization cross-section $\sigma_n(h\nu)$ for electrons or $\sigma_p(h\nu)$ for holes (see Eq. 20).

2.4.1 Photoionization from Deep Centers

Photoionization, i.e., the transition from a trapped to a free-electron state, can be calculated from the deep-level eigenfunction, the Bloch function of free carriers, and the appropriate perturbation operator. The ground state of the deep-level defect can be approximated by

$$\psi_{\text{i}} = \sqrt{\frac{\alpha}{2\pi}}\frac{\exp(-\alpha r)}{r} \quad \text{with} \quad \alpha^2 = \frac{2m_{\text{Trap}}E_{\text{I}}}{\hbar^2} \tag{62}$$

as its initial state, where E_{I} is the ionization energy, and m_{Trap} is a pseudomass of the electron within the deep center, used as an adjustable parameter of the theory. When ionized, the electron in the conduction band is described by a plane wave

$$\psi_{\text{f}} = \sum_{\mathbf{k}} u_n(\mathbf{k})\exp(\text{i}\,\mathbf{k}\cdot\mathbf{r}) \tag{63}$$

as its final state. With these, first ignoring any lattice relaxation (for this, see Sect. 2.4.3), we can obtain (Lucovsky 1965; Grimmeiss and Ledebo 1975) for the ionization cross-section into a parabolic and isotropic band

$$\sigma_n(h\nu) = \frac{8\pi\nu_{\text{D}}}{3}\,\frac{e^2\hbar}{\varepsilon_0 n_{\text{r,c}}}\sqrt{E_{\text{I}}}\,\frac{\sqrt{m_{\text{Trap}}m_n}}{\text{m}_{\text{H}}^2}\,\frac{(h\nu - E_{\text{I}})^{3/2}}{h\nu\left(h\nu + E_{\text{I}}\left\{\frac{m_0}{m_n} - 1\right\}\right)^2}\,. \tag{64}$$

Here ν_{D} is the degeneracy of the occupied impurity state, and m_{H} is an effective carrier mass defined by the perturbation operator. With $m_{\text{Ttrap}} \cong m_{\text{H}} \cong m_0$, Grimmeiss and Ledebo (1975) obtained agreement with photoionization of electrons from O-doped GaAs into the conduction band. See also Zheng and Allen (1994).

2.4.2 Photoionization from Shallow Centers

When considering photoionization from shallow (hydrogen-like) donors, we start with a somewhat better-known eigenfunction of this defect, see Sect. 2. We then evaluate a transition-matrix element with final states that are the eigenstates of the conduction band.

We also have to consider the occupancy of the involved states: it requires an occupied ground state, described by the Fermi distribution, and a free excited state – see Sect. 3.2. Let us look at transitions from filled acceptor levels E_a into the conduction band as an example. We can express for free electrons $k(E)$ from

$$E = \frac{\hbar^2 k^2}{2m_n} = h\nu - \left(E_g - E_a\right). \tag{65}$$

This yields for the matrix element with Eq. 21 of ▶ Chap. 13, "Band-to-Band Transitions", Eqs. 27, and 28

$$\left|\mathbf{M}_{\mathrm{fi,\,qH}}\right|^2 = \frac{3}{2V}\frac{m_0^2}{\mu}\frac{E_g\left(E_g+\Delta_0\right)}{3E_g+2\Delta_0}\frac{32\pi\ a_{\mathrm{qH}}^3}{\left[1+\dfrac{2m_n a_{\mathrm{qH}}\left\{h\nu-\left(E_g-E_a\right)\right\}}{\hbar^2}\right]}. \tag{66}$$

The joint density of states is given by the product of the density of acceptors and the density of band states near the conduction-band edge:

$$J_{a,c} = N_a\frac{V^2(2m_n)^{3/2}\sqrt{h\nu-\left(E_g-E_a\right)}}{\hbar^3}. \tag{67}$$

The optical absorption coefficient (Eqs. 17 and 20) can now be obtained from

$$\alpha_{o,\,qH,\,c} = \frac{e^2}{6\varepsilon_0 n_r c m_0^2 \nu V}\left|\mathbf{M}_{\mathrm{fi,\,qH}}\right|^2 J_{a,c}, \tag{68}$$

yielding a lengthy equation when inserting Eqs. 66 and 67 into Eq. 68. However, for semiconductors with a small effective mass and a large quasi-hydrogen radius, the second term of the sum in square brackets of Eq. 66 can be neglected, and the absorption constant can be approximated as

$$\alpha_{o,\,qH,\,c} = \frac{64\pi^2\nu_D\sqrt{2m_n}e^2a_{\mathrm{qH}}^3N_a}{3\,\varepsilon_0 n_r c\hbar^2}\frac{E_g+\Delta_0}{3E_g+2\Delta_0}\sqrt{h\nu-\left(E_g-E_a\right)}, \tag{69}$$

where ν_D is the degeneracy of the conduction band. The absorption coefficient increases steeply with increasing quasi-hydrogen radius and linearly with the density of acceptors. The optical absorption has an edge-like, square-root shape whenever bands are involved, rather than appearing as lines when the absorption occurs from a

ground to an excited state within the center (see Fig. 1). When the transition proceeds into a band, photoconductivity is usually observed. Therefore, most of the discussion relating to such optical excitation can be found in ▶ Chap. 31, "Photoconductivity".

2.4.3 Phonon Broadening of the Ionization Edge

For many centers, the phonon coupling plays a major role in broadening this edge of the optical absorption cross-section $\sigma_n(h\nu)$. Here σ_n can be expressed approximately (see Noras 1980 or Kopylov and Pikhtin 1975) as

$$\sigma_{n,\,\mathrm{phon}} = \frac{C}{\sqrt{\pi}} \int_0^\infty \left\{ \frac{\left(\sqrt{2}\Gamma x\right)^b}{h\nu\left(\sqrt{2}\Gamma x + E_\mathrm{I}\frac{m_\mathrm{Trap}}{m_n}\right)^a} \exp\left[-\left(x - \frac{h\nu - E_\mathrm{I}}{\sqrt{2}\Gamma} \right) \right] \right\} dx, \tag{70}$$

with Γ as the phonon broadening factor

$$\Gamma = \sqrt{S(\hbar\omega_\mathrm{char})^2 \coth\left(\frac{\hbar\omega_\mathrm{char}}{2kT}\right)}; \tag{71}$$

S is the Huang-Rhys factor. The exponents chosen appropriately for the corresponding model are

$$a = \begin{cases} 2 & \text{for } \delta \text{ function (short range) potential} \\ 4 & \text{for Coulomb (long range) potential} \end{cases} \tag{72}$$

$$b = \begin{cases} 3/2 & \text{for a forbidden transition} \\ 4 & \text{for an allowed transition} \end{cases} \tag{73}$$

This broadened edge-type appearance is shown in Fig. 18 for photoionization of deep centers in CdTe.

Several of these centers are identified, each of them with different coupling to the surrounding lattice. For instance, the center with a capture cross-section of $\sim 10^{-17}$ cm^2 for holes and $\sim 10^{-16}$ cm^2 for electrons, as shown in Figs. 18a, b, indicates a coupling[7] to the lattice with $S \cong 6.7$. Another center, shown in Figs. 18c, d, has a photoionization energy for electrons that is only slightly larger than the first center ($\sim$1.3 eV). It also has a similar ionization cross-section, although it is coupled more strongly to the lattice with $S \cong 30$. After electron capture, the center relaxes by more than 0.62 eV, as indicated in the configuration-coordinate diagram of Fig. 18d, and thereby releases about 30 phonons (see Takebe et al. 1982).

The example shown in Fig. 18 illustrates that various centers – even in the same host material – can have vastly different lattice coupling. This can range from nearly vanishing coupling of shallow centers to very large coupling of some of the deep

[7]This is obtained from the lattice relaxation of 0.14 eV (=1.27–1.13 or = 0.62–0.48 eV) with breathing mode phonons of $\hbar\omega_\mathrm{char} \cong 21$ meV after capture of a hole or an electron, as indicated in Fig. 18a, b.

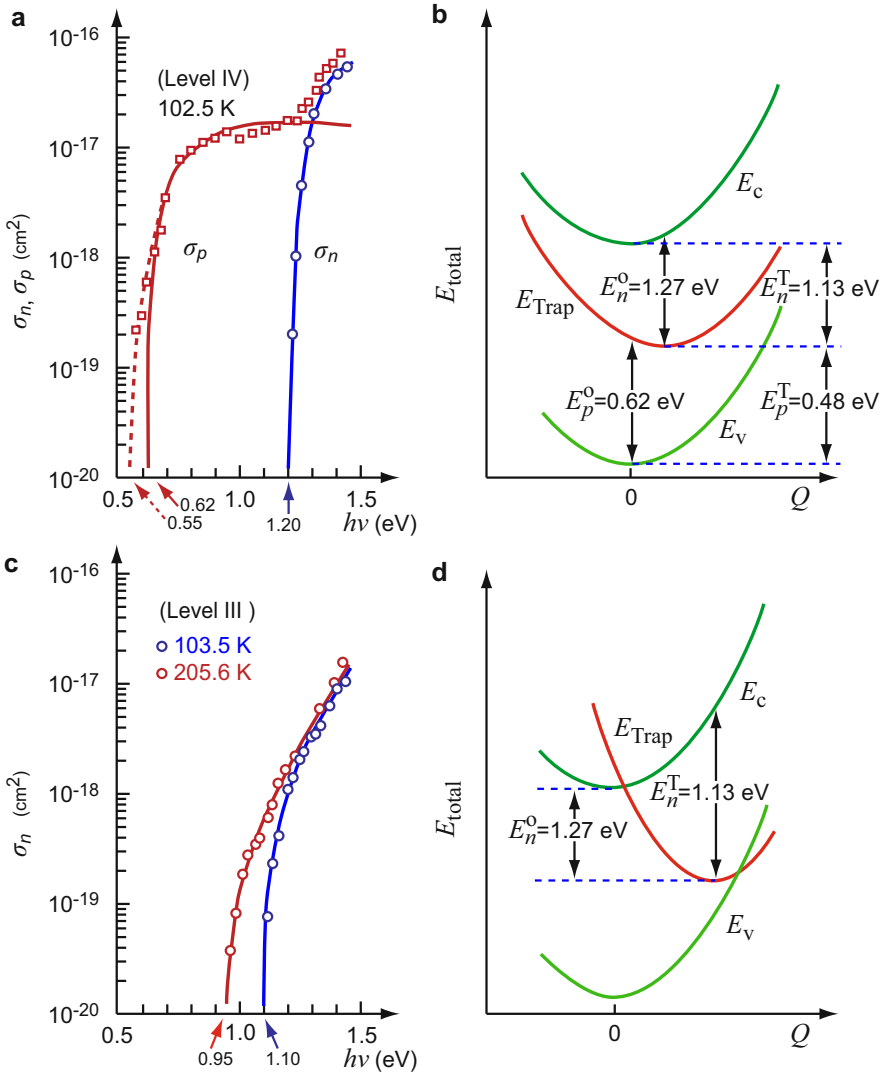


Fig. 18 (**a**) Cross section for photoionization of deep impurity centers in CdTe at 102 K. Solid curves are calculated according to Eq. 70 with $C = 6.4 \cdot 10^{-17}$ and $5.4 \cdot 10^{-15}$ cm^{2}eV for holes and electrons, respectively, $a = 2$, $b = 3/2$, $m_n = 0.11\ m_0$, $m_p = 0.35\ m_0$, $m_{Trap} = 0.35\ m_0$, $E_n^0 = 1.27$ eV, $E_p^0 = 0.62$ eV, and $S = 6.7$ for $\hbar\omega_{char} = 21$ meV. (**b**) Corresponding configuration-coordinate diagram **(c)** and **(d)** the same as for (*a*) and (*b*) but for a different deep center in CdTe (possibly a doubly ionized Cd interstitial) with $E_n^0 = 0.66$ eV, $E_p^0 = 1.23$ eV, and $S \cong 30$ (After Takebe et al. 1982)

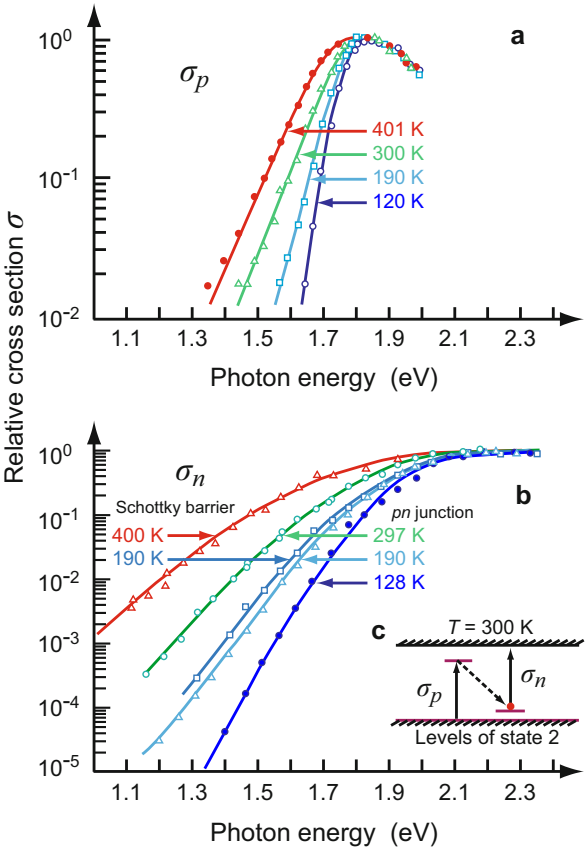


Fig. 19 Optical cross section for the second electron state of oxygen in GaP for (**a**) electrons and (**b**) holes with the temperature as family parameter. (**c**) Band model of GaP with the O level shown (After Henry and Lang 1977)

centers, in which trapped carriers can relax to bridge almost the entire bandgap by emission of many phonons. An example of the temperature dependence of the optical cross section for a center with large lattice relaxation is shown in Fig. 19 for the second electron state of oxygen in GaP. The relaxation is larger for the captured electron than for the hole and amounts to a major fraction of the bandgap.

3 Optical Absorption of Disordered Crystals

Most point defects in an otherwise ideal lattice can be regarded as missing lattice atoms, which are replaced by lattice defects. For every defect that creates one or more levels in the bandgap, the same number of levels that would have been created by the missing lattice atom within the bands are missing. Moreover, the lattice atoms surrounding the lattice defect relax into shifted positions and also create levels which could be shifted into the bandgap. Finally, electron-electron scattering, exchange-correlation effects, and thermal vibration of the lattice atoms give rise to a perturbation of the band edge.

3.1 Band Tailing

At low concentrations, lattice defects cause energy levels within the bandgap with spatially localized wavefunctions. As the defect density N_{defect} increases, these localized wavefunctions start to overlap and form an impurity band (for donors below the conduction band, for acceptors above the valence band). Assuming hydrogen-like defects with a quasi-hydrogen radius a_{qH} the condition for this *Anderson-Mott transition* (Mott and Davis 1979) is

$$N_{\text{defect}}\, a_{\text{qH}}^3 \cong 0.01. \tag{74}$$

At higher defect concentration

$$N_{\text{defect}}\, a_{\text{qH}}^3 \geq 1 \tag{75}$$

the gap between the impurity band and the conduction (or valence) band vanishes, and both bands merge, see Fig. 20. Equation 74 defines the threshold for a heavily doped semiconductor considered below.

Such a tail of band states into the bandgap is also referred to as a Lifshitz tail (Lifshitz 1964). It is well pronounced in heavily doped and amorphous semiconductors and can be experimentally observed from the spectral distribution of the optical absorption (Shklovskii and Efros 1984). For disorder due to dopants, having a correlation length on the order of interatomic spacing, the absorption coefficient shows an exponential decline and can be expressed as

$$\alpha_{\text{o}} = \alpha_{\text{o0}} \exp\left(\frac{h\nu - E_{\text{g}}}{E_0}\right). \tag{76}$$

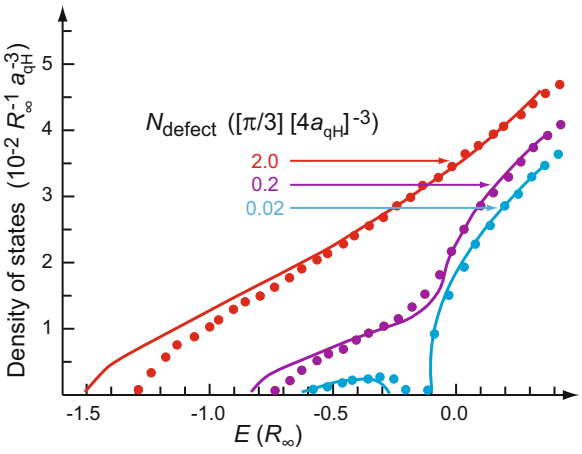


Fig. 20 Density of states as a function of energy for various doping levels given in units of $(\pi/3)(4a_{\text{qH}})^{-3}$. *Circles* and *solid lines* represent solutions obtained from factorized impurity potentials (Haufe et al. 1988) and exact solutions according to Ghazali and Serre (1982), respectively

Here, α_{o0} and E_0 are empirical parameters depending on the semiconductor and its defect structure as it relates to preparation, doping, and treatment of the semiconductor. This dependence of the optical absorption is widely observed and is referred to as the Urbach tail – see Fig. 21 (Urbach 1953). The temperature dependence is usually expressed in terms of the *Urbach rule*

$$\alpha_o(v,T) = \alpha_{o0}\exp\left(\sigma\frac{hv - \hbar\omega_0(T)}{kT}\right), \tag{77}$$

where $\hbar\omega_0(T)$ is a linear function of temperature, which at $T = 0$ K is defined to be E_g, and σ is a fit parameter. The Urbach rule Eq. 77 has been observed universally for crystals as well as for glasses. A detailed analysis of the optical behavior is involved (Casey and Stern 1976). For reviews, see Sritrakool et al. (1986), Jain et al. (1991), Van Mieghem (1992), Abram (1993), and Chakraborty and Biswas (1997).

There are several approaches for estimating the level distribution caused by a statistical distribution of point defects in an otherwise ideal lattice. In a recent work, the Urbach band-tail parameter E_0 is obtained from *spds** tight-binging calculations by modeling electron scattering on polar optical phonons and charged impurities (Sarangapani et al. 2019). E_0 is extracted from the slope of the exponentially decaying density of states below the band edge for GaAs, InAs, GaSb, and GaN, yielding the dependence of E_0 on doping and temperature as well as the related bandgap narrowing in good agreement with experimental data. In GaAs/AlGaAs quantum wells, a similarly observed Urbach tail was assigned to interface roughness on a mesoscopic scale, with a fundamental minimum Urbach energy found for high-quality samples arising from lifetime broadening by phonon scattering (Bhattacharya et al. 2015). Even for the heavily distorted silica glass, spectra

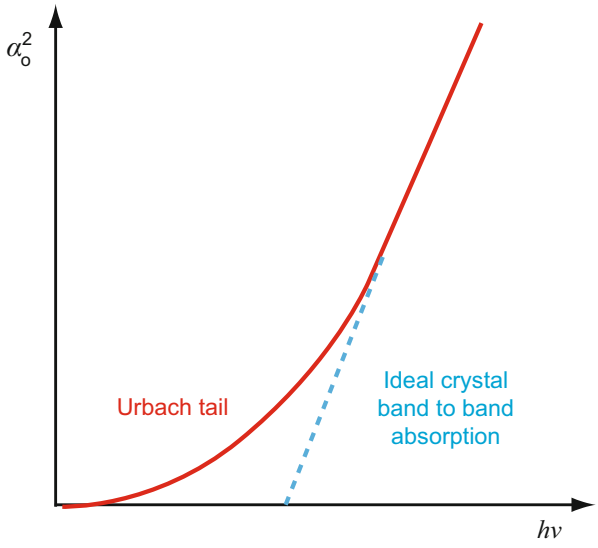


Fig. 21 Optical absorption spectrum of a typical direct-bandgap semiconductor with the absorption constant α_o proportional to the extended density of states in the Urbach tail

calculated from first principles exhibit exponential band-edge tails which follow the Urbach rule in good agreement with experiments (Sadigh et al. 2011).

We consider in the following the more intuitive classical approach. The general behavior can be obtained from a semiclassical model evaluated by Kane (1963) – see also Keldysh and Proshko (1964). Using an independent-electron model, considering the lattice by replacing m_0 with m_n, and ignoring any change in the electron kinetic energy due to the impurities, the density of states is locally perturbed by the potential of lattice defects $V(\mathbf{r})$. When this potential lowers the conduction band, as a Coulomb-attractive center does, there are additional levels accessible to quasi-free electrons at energies below the unperturbed band edge. Near a *repulsive* center, however, there are fewer levels. That is, we assume that the density of states is given by the usual expression for a Fermi gas $g_n \propto \sqrt{E - E_c}$ (Sect. 1.1 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"), however, with an energy scale shifted by the potential energy $eV(\mathbf{r})$ at any point $\mathbf{r}$, yielding

$$g(E, \mathbf{r}) = \frac{1}{2\pi^2} \left(\frac{2m_n}{\hbar^2} \right)^{3/2} \sqrt{E - eV(\mathbf{r})}. \tag{78}$$

Integrating over the entire volume, we obtain the density of states $g(E)$. This integration can be replaced by an integration over an actual potential distribution

$$g(E) = \frac{1}{2\pi^2} \left(\frac{2m_n}{\hbar^2} \right)^{3/2} \int_{-\infty}^{E} \sqrt{E - eV} f(V) dV, \tag{79}$$

with $f(V)$ as the distribution function of the potential. $V(\mathbf{r})$ is assumed to be variable with a Gaussian distribution around a mean potential $\overline{V}$ (Kane approximation), i.e.,

$$f(V) = \frac{1}{\sqrt{\pi \overline{V^2}}} \exp\left[- \frac{\left(V - \overline{V}\right)^2}{\overline{V^2}} \right], \tag{80}$$

where $e^2\overline{V^2}$ is the mean square potential energy. For a density N of charged impurity centers and an assumed Debye screening length λ_D (Sect. 1.3.5 of ▶ Chap. 10, "Interaction of Light with Solids"), we obtain

$$e^2\overline{V^2} = \frac{e^4}{\left(4\pi\varepsilon_{stat}\varepsilon_0 r_0\right)^2} \frac{4\pi}{3} \lambda_D^3 N. \tag{81}$$

Consequently, we obtain from Eqs. 79 and 80

$$g(E) = \frac{1}{2\pi^2} \left(\frac{2m_n}{\hbar^2} \right)^{3/2} \int_{-\infty}^{E} \frac{\sqrt{E - eV}}{\sqrt{\pi \overline{V^2}}} \exp\left[- \frac{\left(V - \overline{V}\right)^2}{\overline{V^2}} \right] dV. \tag{82}$$

At high energy within a band ($E \gg eV$), Eq. 82 yields the unperturbed density of states, e.g., for the conduction band:

$$g(E)\,dE = \frac{1}{2\pi^2}\left(\frac{2m_n}{\hbar^2}\right)^{3/2}\sqrt{E - E_c}\,dE, \tag{83}$$

that is, high within a band, there are no changes in the density of states compared with the ideal lattice. For $E < E_c$, however, the density of states is modified to

$$g(E)\,dE = \frac{1}{4\pi^2}\left(\frac{2m_n}{\hbar^2}\right)^{3/2}\sqrt{\overline{E^2}}(E_c - E)^{-3/2}\exp\left(-\frac{(E - E_c)^2}{\overline{E^2}}\right)dE, \tag{84}$$

with $\overline{E^2} = e^2\overline{V^2}$. Equation 84 indicates that $g(E)$ decreases exponentially ($\propto \exp(-E^2)$) below the edge of the conduction band (see Fig. 22, Kane 1963).

Halperin and Lax (1966, 1967) developed a quantum-mechanical theory for deeper tail states, based on statistical fluctuation of the potential of lattice defects. They applied a variational method that maximizes the density of states with respect to the wavefunction containing statistically fluctuating potentials of Coulomb centers, which are assumed to be of the Yukawa screened Coulomb type.

The Halperin and Lax theory gives the level distribution in the tail as

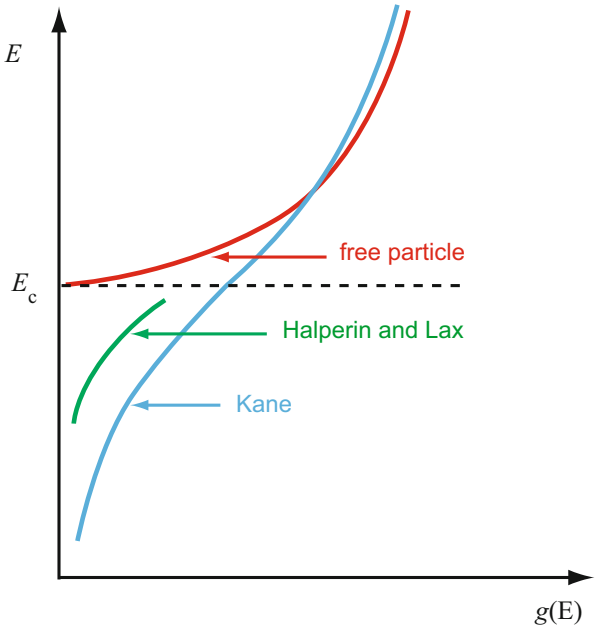


Fig. 22 Density of states for a quasi-free electron near the edge of the conduction band (E_c) in an ideal crystal, and Kane tail of states extending into the bandgap in a crystal with random point defects. The figure also contains the results of the Halperin and Lax approximation (1966, 1967)

$$g(E) = c(\nu)\exp\left(-\frac{a_{\mathrm{qH}}^2}{16\pi N\lambda_{\mathrm{scr}}^5}\,b(\nu)\right), \tag{85}$$

with $\nu = 2m_n|E|\lambda_{\mathrm{scr}}^2/\hbar^2$, a_{qH} as the quasi-Bohr radius, and λ_{scr} as the screening length.[8] The value of $b(\nu)$ is approximately 10 for $\nu = 1$, increases $\propto \nu^{\,2}$ for $\nu \gg 1$, and decreases $\propto \nu^{\,-1/2}$ for $\nu \ll 1$; c(ν) is a proportionality factor that varies only slowly with ν compared to the exponential. The Halperin and Lax distribution is also shown in Fig. 22 and decreases more rapidly than the Kane approximation.

Other more rigorous estimates of the density of state distribution (Sa-yakanit and Glyde 1980; Sa-yakanit et al. 1982) near the band edge also use randomly distributed Coulomb-attractive centers and employ a Feynman's path-integral method (Feynman and Hibbs 1965). They obtain analytic results for $g(E)$, which can be written as

$$g(E) \propto \exp\left(-\frac{E_L^2}{2\overline{E^2}}\,c(\nu)\,\nu^{n(\nu)}\right). \tag{86}$$

Here, $\nu = (E_{\mathrm{c}} - E)/E_L$; $E_L = \hbar^2/\left(2m_nL^2\right)$ is the energy to localize an electron within the correlation distance L, and $c(\nu)$ is a slowly varying function of the order of 1/10. The exponent n depends on this correlation distance (Sa-yakanit 1979) as

$$\begin{aligned} n(\nu) &= \frac{32\nu}{\left(\sqrt{1+16\nu}-1\right)\left(\sqrt{1+16\nu}+7\right)} \\ &= \begin{cases} 2 & \text{for } \nu \to \infty \text{ or } L \to \infty \\ 0.5 & \text{for } \nu \to 0 \text{ or } L \to 0. \end{cases} \end{aligned} \tag{87}$$

With a correlation distance of the defects typically between 1 and 10 Å, n is usually bracketed between 1 and 1.2, yielding with $c(1) \cong 1/(18\sqrt{3})$

$$g(E) \propto \exp\left[-\frac{E_L^2\,c(1)}{2\overline{E^2}}\left(\frac{E_{\mathrm{c}}-E}{E_L}\right)^n\right] \cong \exp\left(-\frac{E-E_1}{E_0}\right), \tag{88}$$

and

$$E_0 \cong \frac{2\overline{E^2}}{E_L\,c(1)} \cong 0.06\frac{\overline{E^2}}{E_L}. \tag{89}$$

[8] For a self-consistent determination of the screening, which depends on the carrier density, which in turn depends on the level density, which again is influenced by the screening length, see Hwang and Brews (1971).

For a review see Shklovskii and Efros (1984). For semiconductors with $m_n \cong m_0$, we have $\overline{E^2}/E_L \cong 1$ eV; hence, $E_0 \cong 0.06$ eV. E_0 is identical with the characteristic energy for the Urbach tail.

These results confirm the tailing nature for deeper tail states, calculated numerically by Halperin and Lax. They approach the Kane approximation in the classical limit and result in an improved agreement with the experiment (Sritrakool et al. 1985, 1986).

For heavily doped semiconductors with lower effective mass ($m_n < 0.1\ m_0$), the correlation length is on the order of the screening length, i.e., typically 20–100 Å. The characteristic energy of such semiconductors is typically one to two orders of magnitude smaller than given above. Here the extent of the Urbach tail is much reduced in agreement with the experiment (Sritrakool et al. 1986).

A comprehensive treatment of light, intermediate, and heavy doping using a Green's-function approach was given by Ghazali and Serre (1982). Their results obtained by exact numerical solution are given in Fig. 20. For a review of more rigorous approaches, see Shklovskii and Efros (1984), Haufe et al. (1988).

3.2 Heavy Doping and Burstein-Moss Effect

In heavily doped semiconductors, the absorption edge of band-to-band transitions is shifted and deformed due to several reasons:

- Tailing of band states into the bandgap
- Shrinking of the bandgap because of many-body effects (Sect. 2.3 of ▶ Chap. 8, "Bands and Bandgaps in Solids")
- Partial filling of the conduction-band states with electrons (Burstein-Moss effect)

Such changes can be detected via excitation spectroscopy (Wagner 1985), with an optical transition probability from filled states in or near the valence band to empty states in the conduction band. Experimental results on heavily doped GaN are reported by Zhao et al. (1999), Feneberg et al. (2014), and results on InN by Davydov et al. (2002). The band-to-band transitions are discussed in Sect. 1 of ▶ Chap. 13, "Band-to-Band Transitions" and the level-to-band transitions in Sect. 2.4. For reasons relating to highly disordered lattices, we will deviate here from this treatment as explained below.

Since the translational symmetry is broken by the random potential of the impurities, **k** is no longer a good quantum number. Consequently, it is more appropriate to use the energy as a label, yielding for the absorption coefficient (Abram et al. 1978)

$$\alpha_o(\nu) = \frac{\pi^2 e^2 \hbar^2}{m_0^2 c\, n_r \nu} \int (f_p(E) - f_n(E + h\nu))\, P(E, E + h\nu) N_v(E) N_c(E + h\nu)\, dE, \quad (90)$$

where the probability $P(E,\ E + h\nu)$ can be expressed by the sum of the matrix elements in a small energy interval around E and around $E + h\nu$:

$$P(E, E + h\nu) = \frac{1}{3}\sum \left| \int \psi_E \frac{\partial}{\partial \mathbf{r}_j} \psi_{E+h\nu} d\mathbf{r} \right|^2 . \tag{91}$$

Examples of some quantitative estimates of such probabilities are given by Lasher and Stern (1964), Casey and Stern (1976), and Berggren and Sernelius (1981). $N_v(E)$ and $N_c(E + h\nu)$ are the density-of-state functions near the respective band edges. These band edges are deformed from the ideal band distribution by the tailing of states, due to band perturbation from the random impurity potential as described in Sect. 3.1.

The occupancy factor, given by the Fermi distributions $f_p(E)$ and $f_n(E + h\nu)$, accounts for the probability of finding occupied states near the valence band with energy E and empty states in the conduction band with energy $E + h\nu$. The tailing of valence-band states and partial filling of conduction-band states with heavy doping result in an asymmetric excitation, e.g., for n-type material from the tail of the valence-band states to states above the Fermi level. These upper states may be shifted to lie well within the conduction band, see Fig. 23.

The Burstein-Moss Effect In semiconductors with a low effective mass, the density of states $N_c \propto m_n^{3/2}$ near the lower edge of the conduction band (Sect. 1.3 of ▶ Chap. 21, "Equilibrium Statistics of Carriers") can be so low that even with moderate donor doping the lower states in the conduction band become filled.

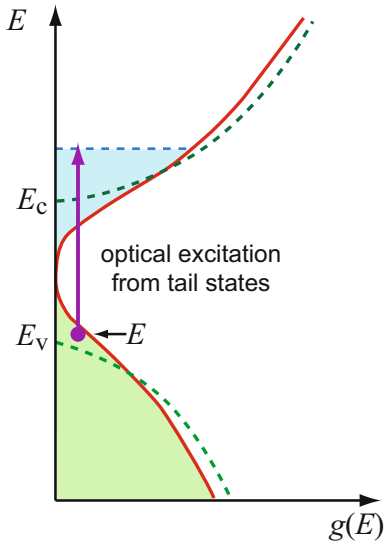


Fig. 23 Optical transition from tailing states of the valence band to free states above the Fermi level within the conduction band

Hence, the Fermi level can be significantly shifted above this band edge (Sect. 2.3.1 of ▶ Chap. 8, "Bands and Bandgaps in Solids").[9] Since an optical excitation can proceed only into unoccupied states, this filling results in a shift of the absorption edge toward higher energies, causing a bandgap widening (see Moss 1961). The shift, referred to as Burstein-Moss shift, can be estimated from the position of the Fermi level and is given by

$$E_{g,\mathrm{BM}} \cong E_{g0} + (E_F - E_c). \tag{92}$$

With this effect, a fine-tuning of the absorption edge can be achieved, which is used to produce optical filters of a rather precisely determined long-wavelength cutoff. Examples for HgTe and InN are given in Fig. 24.

The band-edge narrowing (also termed *band-gap renormalization*), explained in Sect. 2.3 of ▶ Chap. 8, "Bands and Bandgaps in Solids", and the tailing of band states, discussed in Sect. 3.1, complicate this picture, as they have the opposite effect of the Burstein-Moss shift. In GaAs, both effects can be seen. The change of the optical absorption with higher doping densities, measured by Casey et al. (1975), shows the tailing (Fig. 25). In *n*-type GaAs, the Burstein-Moss shift prevails because of the lower effective mass for electrons, causing a significant shift of the Fermi level into the conduction band. For similar doping densities in *p*-type GaAs, however, the heavier hole mass permits only a negligible Burstein-Moss shift. Therefore, the shrinking of the gap and tailing-state transitions provide the major cause for the changes in the absorption spectrum: The absorption edge shifts in the opposite direction and the effective bandgap narrows – see Fig. 25.

Since heavy *n*-type doping moves the Fermi level into the conduction band, optical excitation requires larger energies, labeled E_{g1} in Fig. 26a. Band-to-band luminescence, however, has a lower energy threshold, shown as E_{g2} in Fig. 26a. The width of the emission band is shown as ΔE_{g12}. In elemental indirect-gap semiconductors, this luminescence band involves momentum-conserving TA or TO phonons with their corresponding replica, as shown for Si in Fig. 26b. In compound indirect semiconductors, LO phonons are employed in this process. In Ge, the LA replica are most prominent.

The band filling, obtained by heavy doping and observed by the Burstein-Moss effect, is obtained from the luminescent line shape. The band filling for Si is shown in Fig. 27a. The corresponding line broadening is significant only for a carrier density in excess of 10^{18} cm^{-3}; it amounts to ~60 meV at 10^{20} cm^{-3} and is slightly larger for holes than for electrons, since it relates to the density-of-states mass, with $m_{n\mathrm{ds}}$ ~1.062 m_0 and $m_{p\mathrm{ds}} = 0.55\ m_0$ (see Barber 1967).

The decrease of the bandgap energy with band tailing at higher doping levels is obtained by a square-root fitting of the low-energy tail of the luminescence. It shows

[9]For instance, in InSb with $m_n = 0.0116$, the effective density of states is $N_c \cong 3 \cdot 10^{16}$ cm^{-3}; hence doping with a shallow donor density in excess of 10^{17} cm^{-3} will cause a significant filling of conduction-band states.

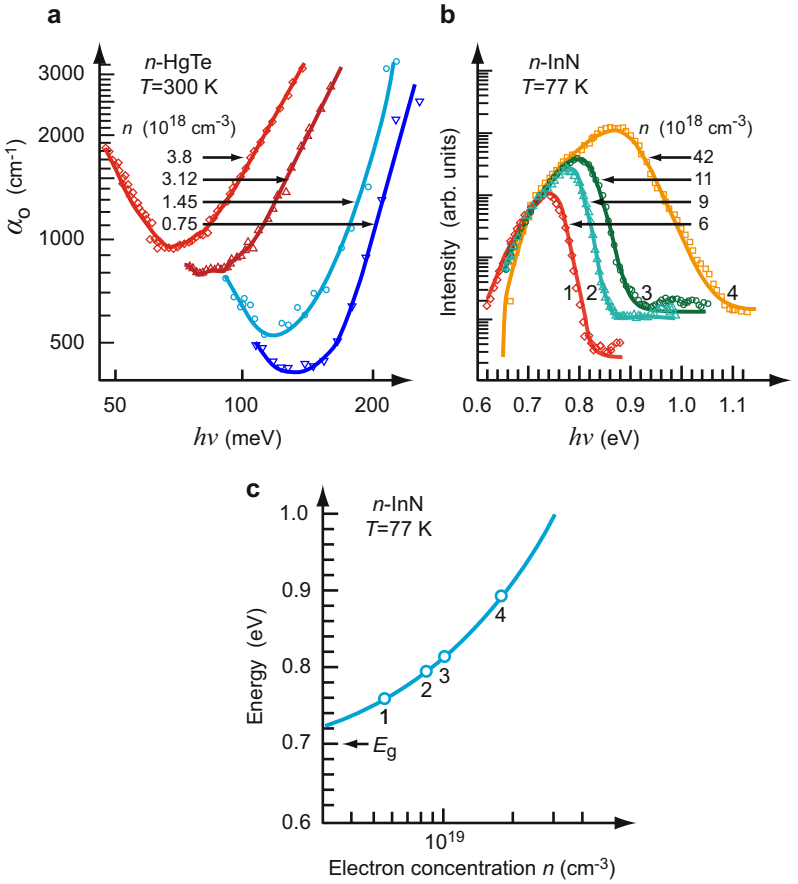


Fig. 24 (**a**) Band-edge shift of HgTe due to the Burstein-Moss effect caused by Al doping at various densities (After Vérié 1967). (**b**) Photoluminescence spectra of band-to-band transitions and (**c**) calculated shift of the fundamental absorption edge in *n*-type InN for increasing electron concentration due to the Burstein-Moss effect; circles mark ($E_g + E_F$) positions estimated from PL spectra in (*b*) (After Davydov et al. 2002)

for Si an essentially constant bandgap energy up to carrier densities of ~10^{18} cm^{-3}, and a significant decrease of E_g starting near the critical Mott density, as shown in Fig. 27b (arrows). For further detail, see Wagner (1985).

The opposing effects of the Burstein-Moss shift and bandgap narrowing was more recently calculated using density-functional theory (Walsh et al. 2008). The modeling showed that the bandgap narrowing of the studied wide-gap semiconductors due to electron doping mainly arises from the nonparabolic shape of the conduction band and may be enhanced by hybridization of dopant states with the host conduction states.

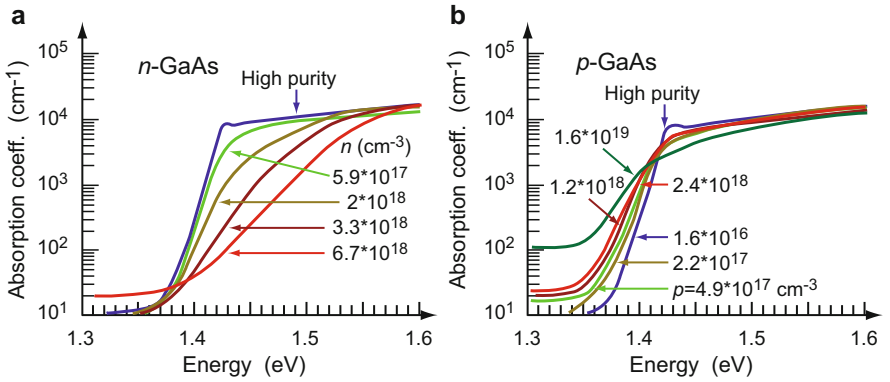


Fig. 25 Absorption coefficient for (**a**) *n*-type and (**b**) *p*-type GaAs measured at room temperature with doping densities as family parameter (After Casey et al. 1975)

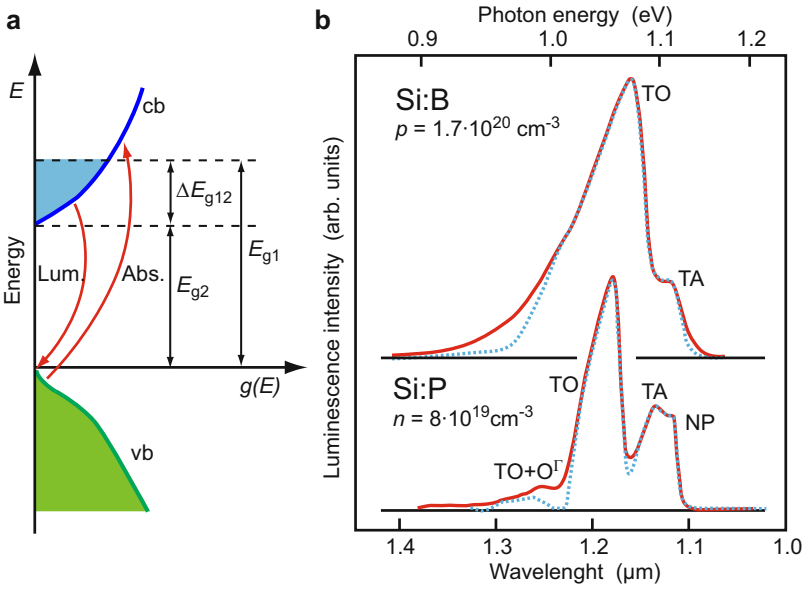


Fig. 26 (**a**) Density-of-state distribution of a heavily doped *n*-type semiconductor with partially filled conduction band (schematic). (**b**) Band-to-band luminescence of *p*-type Si:B (*upper curve*) and *n*-type Si:P (*lower curve*); dotted curve: theoretical fit (After Wagner 1985)

4 Summary

Two major classes of defects result in substantially different types of absorption spectra: shallow-level, hydrogen-like defects and deep-level, tight-bonding defects.

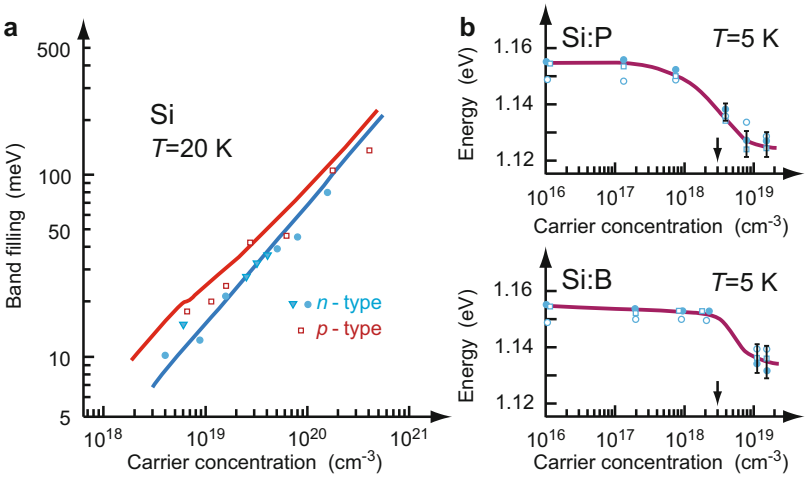


Fig. 27 (**a**) Band filling of Si as a function of carrier density: calculated curves for *n*-type (*blue*) and for *p*-type (*red*) Si at $T = 20$ K. (**b**) Bandgap reduction as a function of carrier density for Si obtained from optical absorption and from the photoluminescence cutoff at $T = 5$ K. The critical Mott density is indicated by the arrows (After Wagner 1985)

Optical absorption at a shallow-level defect results in a line spectrum near the fundamental bandgap; this spectrum is characterized by the energy, amplitude, and width of each line. The line energy provides important information about the electronic structure of the defect center. The line amplitude permits judgment on the density of these centers or joint density of states and their symmetry relation within the lattice, matrix elements, and selection rules. The line shape relates to information on the lifetime of the excited states and center-to-lattice coupling for homogeneously broadened lines, as well as on the center perturbation due to defect-sensitive stress and electric fields of the surrounding lattice for inhomogeneously broadened lines.

Shallow-level defects have electronic eigenfunctions which are substantially influenced by mixing with one band, the conduction band for donors, and the valence band for acceptors. The shallow-level defects show little electron-lattice coupling and consequently little relaxation after a change in excitation or in charging of the center, i.e., the lattice atoms readjust only to a minor degree.

The deep-level defects, in contrast, have eigenfunctions of excited states which mix with *both* valence and conduction bands. Deep-level centers show a much larger lattice relaxation after a change in excitation or in charging of the centers. The degree of relaxation – the number of phonons emitted during relaxation – differs greatly with the type of defect and the specific electronic state. One may picture the difference between such eigenfunctions as those which tend to press surrounding atoms apart and those which extend over surrounding atoms; the latter have much less influence on their equilibrium position.

Both shallow and deep-level defects also show band-to-free electron absorption. Such a photoionization causes an edge-like spectrum, which is characterized by its

onset energy and its slope. These features yield information about the defect density and lattice coupling.

A high density of lattice defects causes their states to overlap; the related lattice perturbation forms an exponential tail of states, extending from the band edges. Such tailing states can be observed as Urbach tails in optical absorption extending from the band edge into the extrinsic range. They are a measure of the degree of lattice disorder, resulting from intrinsic defects, or from a high level of doping. High donor doping influences the position of the Fermi level, and may shift it into the conduction band, thereby causing a shift of the optical absorption edge to higher energies, opposite to the Urbach-tail shift. This Burstein-Moss effect may result in substantial shifts of the absorption edge in semiconductors with a low effective mass.

References

Abram RA (1993) Effects of heavy doping and high excitation on the band structure of gallium arsenide. Semicond Semimetals 39:259

Abram RA, Rees GJ, Wilson BLH (1978) Heavily doped semiconductors and devices. Adv Phys 27:799

Aleksandrov IA, Zhuravlev KS (2020) Luminescence line shapes of band to deep centre and donor–acceptor transitions in AlN. J Phys Condens Matter 32:435501

Alkauskas A, Lyons JL, Steiauf D, Van de Walle CG (2012) First-principles calculations of luminescence spectrum line shapes for defects in semiconductors: the example of GaN and ZnO. Phys Rev Lett 109:267401

Alkauskas A, Yan Q, Van de Walle CG (2014) First-principles theory of nonradiative carrier capture via multiphonon emission. Phys Rev B 90:075202

Alkauskas A, Dreyer CE, Lyons JL, Van de Walle CG (2016) Role of excited states in Shockley-Read-Hall recombination in wide-band-gap semiconductors. Phys Rev B 93:201304

Baranovskii SD, Doerr U, Thomas P, Naumov A, Gebhardt W (1994) Exciton line broadening by compositional disorder in $ZnSe_xTe_{1-x}$ quantum wells. Solid State Commun 89:5

Baranowski JM (1979) unpublished. Work published in: Jaros M (1980) Deep levels in semiconductors. Adv Phys 29:409

Barber HD (1967) Effective mass and intrinsic concentration in silicon. Solid State Electron 10:1039

Beaumont JH, Harmer AL, Hayes W (1972) The F_3 centre in alkaline earth fluorides. J Phys C Solid State Phys 5:257

Berggren K-F, Sernelius BE (1981) Band-gap narrowing in heavily doped many-valley semiconductors. Phys Rev B 24:1971

Bhattacharya R, Mondal R, Khatua P, Rudra A, Kapon E, Malzer S, Döhler G, Pal B, Bansal B (2015) Measurements of the electric field of zero-point optical phonons in GaAs quantum wells support the Urbach rule for zero-temperature lifetime broadening. Phys Rev Lett 114:047402

Callaway J (1963) Transition processes in semiconductor lasers. J Phys Chem Solid 24:1063

Casey HC Jr, Stern F (1976) Concentration-dependent absorption and spontaneous emission of heavily doped GaAs. J Appl Phys 47:631

Casey HC Jr, Sell DD, Wecht KW (1975) Concentration dependence of the absorption coefficient for *n*- and *p*-type GaAs between 1.3 and 1.6 eV. J Appl Phys 46:250

Chakraborty PK, Biswas JC (1997) Conduction-band tailing in parabolic band semiconductors. J Appl Phys 82:3328

Condon EU, Morse PM (1929) Quantum mechanics. McGraw-Hill, New York

Davydov VYu, Klochikhin AA, Emtsev VV, Ivanov SV, Vekshin VV, Bechstedt F, Furthmüller J, Harima H, Mudryi AV, Hashimoto A, Yamamoto A, Aderhold J, Graul J, Haller EE (2002) Band gap of InN and In-rich $In_xGa_{1-x}N$ alloys ($0.36 < x < 1$). Phys Status Solid B 230:R4

Dexter DL (1958) Theory of the optical properties of imperfections in nonmetals. In: Seitz F, Turnbull D (eds) Solid state physics, vol 6. Academic, New York, pp 353–411

Eagles DM (1960) Optical absorption and recombination radiation in semiconductors due to transitions between hydrogen-like acceptor impurity levels and the conduction band. J Phys Chem Solid 16:76

Feneberg M, Osterburg S, Lange K, Lidig C, Garke B, Goldhahn R, Richter E, Netzel C, Neumann MD, Esser N, Fritze S, Witte H, Bläsing J, Dadgar A, Krost A (2014) Band gap renormalization and Burstein-Moss effect in silicon- and germanium-doped wurtzite GaN up to 10^{20} cm^{-3}. Phys Rev B 90:075203

Feynman RP, Hibbs AR (1965) Quantum mechanics and path integrals. McGraw-Hill, New York

Fitchen D, Silsbee RH, Fulton TA, Wolf EL (1963) Zero-phonon transitions of color centers in Alkali Halides. Phys Rev Lett 11:275

Gebhardt W, Kuhnert H (1964) Temperature dependence of F-centre absorption and emission. Phys Lett 11:15

Ghazali A, Serre J (1982) Multiple-scattering approach to the formation of the impurity band in semiconductors. Phys Rev Lett 48:886

Goede O, John L, Hennig D (1978) Compositional disorder-induced broadening for free excitons in II-VI semiconducting mixed crystals. Phys Stat Solid B 89:K183

Grimmeiss HG (1985) Deep energy levels in semiconductors. In: Chadi JD, Harrison WA (eds) Proceeding of the 17th international conference on the physics of semiconductors, San Francisco 1984, pp 589–600. Springer, New York

Grimmeiss HG, Ledebo L-Å (1975) Spectral distribution of photoionization cross sections by photoconductivity measurements. J Appl Phys 46:2155

Haller EE, Navarro H, Keilmann F (1987) Intrinsic linewidth of 1S → nP donor transitions in ultrapure germanium. In: O Engström (ed) Proceedings of the 18th international conferences on the physics of semiconductors, Stockholm 1986, pp 837–840. World Scientific, Singapore

Halperin BI, Lax M (1966) Impurity-band tails in the high-density limit. I. Minimum counting methods. Phys Rev 148:722

Halperin BI, Lax M (1967) Impurity-band tails in the high-density limit. II. Higher order corrections. Phys Rev 153:802

Haufe A, Schwabe R, Feiseler H, Ilegems M (1988) The luminescence lineshape of highly doped direct-gap III-V compounds. J Phys C 21:2951

Hayes W, Stoneham AM (1984) Defects and defect processes in nonmetallic solids. Wiley, New York

Henry CH (1980) Large lattice relaxation processes in semiconductors. In: Kubo R, Hanamura E (eds) Relaxation of elementary excitation. Springer, Berlin, pp 19–33

Henry CH, Lang DV (1977) Nonradiative capture and recombination by multiphonon emission in GaAs and GaP. Phys Rev B 15:989

Huang K, Rhys A (1950) Theory of light absorption and non-radiative transitions in F centres. Proc R Soc London A204:406

Huber MCE, Sandemann RJ (1986) The measurement of oscillator strengths. Rep Prog Phys 49:397

Hwang CJ, Brews JR (1971) Electron activity coefficients in heavily doped semiconductors with small effective mass. J Phys Chem Sol 32:837

Jagannath C, Grabowski ZW, Ramdas AK (1981) Linewidths of the electronic excitation spectra of donors in silicon. Phys Rev B 23:2082

Jain SC, Mertens RP, Van Overstraeten RJ (1991) Bandgap narrowing and its effects on the properties of moderately and heavily doped germanium and silicon. Adv Electronics Electron Phys 82:197

Jin Y, Govoni M, Wolfowicz G, Sullivan SE, Heremans FJ, Awschalom DD, Galli G (2021) Photoluminescence spectra of point defects in semiconductors: validation of first-principles calculations. Phys Rev Mater 5:084603

Kane EO (1963) Thomas-Fermi approach to impure semiconductor band structure. Phys Rev 131:79

Karaiskaj D, Stotz JAH, Meyer T, Thewalt MLW, Cardona M (2003) Impurity absorption spectroscopy in ^{28}Si: the importance of inhomogeneous isotope broadening. Phys Rev Lett 90:186402

Kayanuma Y, Fukuchi S (1984) Nonradiative transitions in deep impurities in semiconductors–study in a semiclassical model. J Phys Soc Jpn 53:1869

Keldysh LV, Proshko GP (1964) Infrared absorption in highly doped germanium. Sov Phys – Solid State 5:2481

Kim S, Hood SN, Walsh A (2019) Anharmonic lattice relaxation during nonradiative carrier capture. Phys Rev B 100:041202

Klauder JR, Anderson PW (1962) Spectral diffusion decay in spin resonance experiments. Phys Rev 125:912

Kopylov AA, Pikhtin AN (1975) Effect of temperature on the optical absorption spectra of deep centers. Sov Phys Sol State 16:1200

Kubo R (1952) Thermal ionization of trapped electrons. Phys Rev 86:929

Larsen DM (1976) Inhomogeneous broadening of the Lyman-series absorption of simple hydrogenic donors. Phys Rev B 13:1681

Lasher G, Stern F (1964) Spontaneous and stimulated recombination radiation in semiconductors. Phys Rev 133:A553

Lax M (1952) The Franck-Condon principle and its application to crystals. J Chem Phys 20:1752

Lifshitz IM (1964) The energy spectrum of disordered systems. Adv Phys 13:483

Lucovsky G (1965) On the photoionization of deep impurity centers in semiconductors. Sol State Commun 3:299

Markham JJ (1956) Electron-nuclear wave functions in multiphonon processes. Phys Rev 103:588

Moss TS (1961) Optical properties of semiconductors. Butterworths Scientific Publications, London

Mott NF (1974) Metal-insulator transitions. Barnes and Noble, New York

Mott NF, Davis EA (1979) Electronic processes in noncrystalline materials. Claredon Press, Oxford, UK

Noras JM (1980) Photoionisation and phonon coupling. J Phys C 13:4779

Onton A (1971) Donor-electron transitions between states associated with the X_{1c} and X_{3c} conduction-band minima in GaP. Phys Rev B 4:4449

O'Rourke RC (1953) Absorption of light by trapped electrons. Phys Rev 91:265

Pekar SI (1953) On the effect of lattice deformations by electrons on optical and electrical properties of crystals. Uspekhi Fiz Nauk 50:197

Peuker K, Enderlein R, Schenk A, Gutsche E (1982) Theory of non-radiative multiphonon capture processes; solution of old controversies. Phys Status Solid B 109:599

Sadigh B, Erhart P, Åberg D, Trave A, Schwegler E, Bude J (2011) First-principles calculations of the urbach tail in the optical absorption spectra of silica glass. Phys Rev Lett 106:027401

Sallen G, Tribu A, Aichele T, André R, Besombes L, Bougerol C, Richard M, Tatarenko S, Kheng K, Poizat J-Ph (2011) Subnanosecond spectral diffusion of a single quantum dot in a nanowire. Phys Rev B 84:041405

Sarangapani P, Chu Y, Charles J, Klimeck G, Kubis T (2019) Band-tail formation and band-gap narrowing driven by polar optical phonons and charged impurities in atomically resolved III–V semiconductors and nanodevices. Phys Rev Appl 12:044045

Sa-yakanit V (1979) Electron density of states in a Gaussian random potential: path-integral approach. Phys Rev B 19:2266

Sa-yakanit V, Glyde HR (1980) Impurity-band density of states in heavily doped semiconductors: a variational calculation. Phys Rev B 22:6222

Sa-yakanit V, Sritrakool W, Glyde HR (1982) Impurity-band density of states in heavily doped semiconductors: numerical results. Phys Rev B 25:2776

Shklovskii BI, Efros AL (1984) Electronic properties of doped semiconductors. Springer, Berlin

Singh J, Bajaj KK (1986) Quantum mechanical theory of linewidths of localized radiative transitions in semiconductor alloys. Appl Phys Let 48:1077

Smakula A (1930) Über Erregung und Entfärbung lichtelektrisch leitender Alkalihalogenide. Z Phys 59:603 (On excitation and decoloration of photoconducting alkali halides, in German)

Sritrakool W, Sa-yakanit V, Glyde HR (1985) Absorption near band edges in heavily doped GaAs. Phys Rev B 32:1090

Sritrakool W, Sa-yakanit V, Glyde HR (1986) Band tails in disordered systems. Phys Rev B 33:1199

Stoneham AM (1969) Shapes of inhomogeneously broadened resonance lines in solids. Rev Mod Phys 41:82

Stradling RA (1984) Studies of the free and bound magneto-polaron and associated transport experiments in n-InSb and other semiconductors. In: Devreese JT, Peeters FM (eds) Polarons and excitons in polar semiconductors and ionic crystals. Plenum Press, New York

Sumi H (1983) Nonradiative multiphonon capture of free carriers by deep-level defects in semiconductors: adiabatic and nonadiabatic limits. Phys Rev B 27:2374

Takebe T, Saraie J, Matsunami H (1982) Detailed characterization of deep centers in CdTe: photoionization and thermal ionization properties. J Appl Phys 53:457

Toyozawa Y, Inoue M, Inui T, Okazaki M, Hanamura E (1967) Coexistence of local and band characters in the absorption spectra of solids I. Formulation. J Phys Soc Jpn 22:1337; Okazaki M, Inoue M, Toyozawa Y, Inui T, Hanamura E (1967) II. Calculations for the simple cubic lattice. J Phys Soc Jpn 22:1349

Türck V, Rodt S, Stier O, Heitz R, Engelhardt R, Pohl UW, Bimberg D, Steingrüber R (2000) Effect of random field fluctuations on excitonic transitions of individual CdSe quantum dots. Phys Rev B 61:9944.

Urbach F (1953) The long-wavelength edge of photographic sensitivity and of the electronic absorption of solids. Phys Rev 92:1324

Van Mieghem P (1992) Theory of band tails in heavily doped semiconductors. Rev Mod Phys 64:755

Velický B, Sak J (1966) Excitonic effects in the interband absorption of semiconductors. Phys Status Solid 16:147

Vérié C (1967) Electronic properties of $Cd_xHg_{1-x}Te$ alloys in the vicinity of the semimetal-semiconductor transition. In: Thomas DC (ed) II-VI semiconductor compounds. Benjamin, New York, p 1124

Wagner J (1985) Heavily doped silicon studied by luminescence and selective absorption. Sol State Electron 28:25

Walsh A, Da Silva JLF, Wei SH (2008) Origins of band-gap renormalization in degenerately doped semiconductors. Phys Rev B 78:075211

Zeiger HJ (1964) Impurity states in semiconducting masers. J Appl Phys 35:1657

Zhao GY, Ishikawa H, Jiang H, Egawa T, Jimbo T, Umeno M (1999) Optical absorption and photoluminescence studies of *n*-type GaN. Jpn J Appl Phys 38:L993

Zheng J, Allen JW (1994) Photoionization of a deep centre in zinc selenide giving information about the conduction band structure. J Cryst Growth 138:504

Shallow-Level Centers

18

Karl W. Böer and Udo W. Pohl

Contents

1 Hydrogen-Like Defects 756
1.1 The Chemical Identity 761
1.2 Hydrogen-Like Donors in Indirect-Bandgap Semiconductors 762
1.3 Hydrogen-Like Ground State and Chemical Shift 765
1.4 Hydrogen-Like Acceptors 771
1.5 Shallow Defects in Compound Semiconductors 772
1.6 Donor-Acceptor Pair and Free-To-Bound Transitions 775
1.7 Higher Charged Coulomb Centers and Metal-Ion Interstitials 779
2 Excitons Bound to Impurity Centers 782
2.1 Excitons Bound to Ionized Donors or Acceptors 782
2.2 Excitons Bound to Neutral Donors or Acceptors 783
2.3 Excitons Bound to Isoelectronic Centers 787
3 Influence of External Fields on Defect Levels 789
3.1 Influence of Hydrostatic Pressure 790
3.2 Influence of Uniaxial Stress 793
3.3 Influence of an Electric Field 796
3.4 Influence of a Magnetic Field 797
4 Summary 802
References 803

Abstract

Shallow defect centers play a dominant role as donors and acceptors in nearly all semiconducting devices. The major features of their spectrum can be described by a quasi-hydrogen model, modified only by the dielectric constant and the effective

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_18

mass of the host semiconductor. This relation yields very good results for higher excited states of a large variety of such defects, while the ground state shows substantial deviations according to the chemical individuality of the defect center. Such an individuality can be explained by considering the core potential and the deformation of the lattice after incorporating the defect.

Band anisotropies and the interaction between bands such as conduction-band valleys cause the lifting of some of the degeneracies of the quasi-hydrogen spectrum. Local stress and electric fields cause additional splitting. The dependence of levels on hydrostatic pressure can be used to identify shallow-level defects which are connected to one band only. The influence of externally applied uniaxial stress and electric or magnetic fields can be used for further identification.

Donors and acceptors can bind excitons. The binding energy sensitively depends on the effective mass ratio of electron and hole and the charge state of the impurity. For the energy of exciton binding at neutral donors and acceptors, a linear dependence from the ionization energy is found. Isoelectronic impurities can also bind excitons by attracting the electron at the core potential.

Keywords

Acceptor · Acceptor-bound exciton · Acceptor energies · Bound excitons · Central-cell correction · Core potential · Donor · Donor-acceptor pairs · Donor-bound exciton · Donor energies · Electron spin resonance · Effective-mass impurity · Exciton binding energy · Free-to-bound transitions · Hayne's rule · Strain-induced splitting · Quasi-hydrogen model · Shallow acceptor · Shallow donor · Stark effect · Two-electron transitions · Two-hole transitions · Rydberg energy · Uniaxial strain · Zeeman effect

1 Hydrogen-Like Defects

In the previous chapter, a general overview of the different defect levels and their optical absorption spectra was presented. We will now discuss in more detail the defect-level spectrum of shallow centers. These centers have eigenfunctions that extend beyond their neighbor atoms and mix only with the nearest band states. We will indicate that the ground state is influenced by the chemistry, how it is to be considered, to what degree the quasi-hydrogen approximation can be used, and what refinements are necessary to obtain a better agreement with the experiment.

In ▶ Chap. 19, "Deep-Level Centers", we will then give a more detailed description of the *deep*-level centers. In contrast, these have highly localized eigenfunctions of their ground states, mix with conduction *and* valence bands, and require a more thorough knowledge of the core potential for the calculation of their eigenvalues.

Hydrogenic Model Shallow-level defects can be described as hydrogen-like defects – see Sect. 1.1 of ▶ Chap. 17, "Optical Properties of Defects". As an example, we will discuss here the electronic states of a substitutional donor, such as a phosphorus atom

on a lattice site in a silicon host crystal. The P atom becomes positively charged after it has given its electron to the host. This electron, now near the bottom of the conduction band, is a quasi-free Bloch electron with an energy

$$E(k) = E_c + \varepsilon_{stat} \frac{\hbar^2 k^2}{2m_n}. \tag{1}$$

Near the P_{Si}^{+} center, the electron can become localized. Its new eigenstate can be calculated by solving the Schrödinger equation $H\psi = E\psi$ with the Hamiltonian

$$H = H_0 - \frac{e^2}{4\pi\varepsilon_{stat}\varepsilon_0 r}, \tag{2}$$

where H_0 is the unperturbed Hamiltonian of the host lattice, to which the attractive Coulomb potential of the defect is added. This potential is modified by the screening action of the host, which is expressed by the static dielectric constant. This is in contrast to a Bloch electron in an ideal lattice, which interacts only with the electronic part of the lattice and therefore involves the optical dielectric constant ε_{opt}. When trapped, the electron becomes localized near the defect and causes a shift of the surrounding ions according to its averaged Coulomb potential. Therefore, the static dielectric constant is used here.

The eigenfunctions to H_0 are Bloch functions. They form a complete orthonormalized set. The solutions to H can be constructed near the defect from a wave packet of Bloch functions:

$$\psi = \sum_{n,\mathbf{k}} c_n(\mathbf{k})\psi_n(\mathbf{k},\mathbf{r}) \cong \sum_{\mathbf{k}} c_c(\mathbf{k})\psi_c(\mathbf{k},\mathbf{r}). \tag{3}$$

The summation over several bands with index n is dropped, since as shallow levels – here for a donor – their eigenfunctions are constructed primarily from eigenfunctions of the nearest band only. $\psi_c(\mathbf{k}, \mathbf{r})$ are the Bloch functions of *conduction-band electrons*:

$$\psi_c(\mathbf{k},\mathbf{r}) = u_c(\mathbf{k},\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r}). \tag{4}$$

As will be verified below, the eigenfunctions of such shallow-level defects extend over several lattice constants, thus restricting $\mathbf{k}$ to values close to the center of the Brillouin zone. Since $u(\mathbf{k})$ changes only slowly with $\mathbf{k}$, we can pull $u_c(\mathbf{k} = \mathbf{k}_0, \mathbf{r})$ as constant from the sum for $k \cong 0$, or near any of the minima of $E(\mathbf{k} \cong \mathbf{k}_0)$, and introduce with[1]

$$\psi(\mathbf{k},\mathbf{r}) = \left[\sum_{\mathbf{k}} c_c(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{r})\right] u_c(\mathbf{k}_0,\mathbf{r}) = F(\mathbf{r})\ \ \psi_c(\mathbf{k}_0,\mathbf{r}) \tag{5}$$

[1]This can easily be seen at the Γ point for $k_0 = 0$: here, we have $\psi(\mathbf{k} = 0, \mathbf{r}) = u(0, \mathbf{r})\exp(i0 \cdot \mathrm{r}) = u(0, \mathbf{r})$.

an *envelope function*

$$F(\mathbf{r}) = \sum_{\mathbf{k}} c_{\mathrm{c}}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{r}), \tag{6}$$

where $\psi_{\mathrm{c}}(\mathbf{k}_0, \mathbf{r})$ is the Bloch function in the minimum[2] of $E(\mathbf{k})$. The envelope function satisfies an appropriately modified Schrödinger equation for the quasi-hydrogen model:

$$\left(-\frac{\hbar^2}{2m_n} \frac{\partial^2}{\partial \mathbf{r}^2} - \frac{e^2}{4\pi\varepsilon_{\mathrm{stat}}\varepsilon_0 r} \right) F(\mathbf{r}) = (E - E_{\mathrm{c}}) F(\mathbf{r}) \tag{7}$$

with the energy normalized to the edge of the conduction band and with an effective mass m_n for Bloch electrons near this band edge. This Schrödinger equation is identical to that for a hydrogen atom, but for an electron of effective mass m_n in a medium of dielectric constant $\varepsilon_{\mathrm{stat}}$. Therefore, the solution can be transcribed directly from that of a hydrogen atom and yields for the envelope eigenfunction of the l*s* ground state:

$$F(r) = \frac{1}{\sqrt{\pi a_{\mathrm{qH}}^3}} \exp\left(-\frac{r}{a_{\mathrm{qH}}} \right), \tag{8}$$

with

$$a_{\mathrm{qH}} = \frac{4\pi\varepsilon_{\mathrm{stat}}\varepsilon_0 \hbar^2}{m_n e^2} = \frac{\varepsilon_{\mathrm{stat}} m_0}{m_n} a_{\mathrm{H}}. \tag{9}$$

a_{qH} is an effective Bohr radius, the *quasi-hydrogen radius*, and $a_{\mathrm{H}} = 0.529$ Å is the Bohr radius of the hydrogen atom. This envelope function is shown as the dashed curve in Fig. 1. The total wavefunction (solid curve) shows the modulation with the rapidly oscillating Bloch function with a period length of the lattice constant *a*.

The resulting eigenstates of the envelope function are bound states below the lowest free states in the conduction band E_{c} and are given by a quasi-hydrogen energy spectrum[3]:

[2] In semiconductors with several equivalent minima (Si, Ge), the wavefunction becomes a sum of contributions from each of the minima: $\sum_j \alpha_j F_{jc}(\mathbf{r})\ u_{jc}(\mathbf{k}_{j0}, \mathbf{r})$.

[3] n is the principal quantum number, describing the entire energy spectrum for a simple hydrogen atom. All other states are degenerate. Therefore, in a pure Coulomb potential, this quantum number is the only one that determines the energy of a hydrogen level. When deviations from this spherical potential appear in a crystal, the $D = \Sigma l(l + 1) = n^2$ degeneracy of each of these levels is removed, and the energy of the *s*, *p*, *d*, ... states is shifted according to $R_\infty/(n + l)^2$. The importance of these transitions is discussed in Sect. 1.2. To further lift the remaining degeneracies of the magnetic quantum number, a magnetic field must act (see Sect. 3.4).

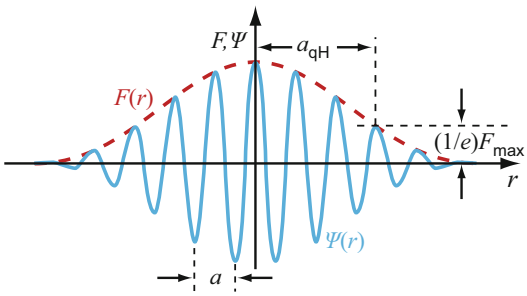


Fig. 1 Relationship between the envelope function $F(r)$ and the wavefunction $\psi_n(r)$ of a Bloch wave packet for an electron localized near a hydrogen-like impurity. Here, a denotes the lattice constant

$$E_{\mathrm{qH},n} = E_c - \frac{m_n e^4}{32\pi^2(\varepsilon_{\mathrm{stat}}\varepsilon_0)^2\hbar^2} \quad \frac{1}{n^2} = E_c - \left(\frac{m_n}{m_0}\frac{1}{\varepsilon_{\mathrm{stat}}^2}\right)R_\infty \frac{1}{n^2} = E_c - E_d \frac{1}{n^2}. \quad (10)$$

R_∞ is the Rydberg energy of the hydrogen atom, and E_d is the *donor ionization energy*, also referred to *donor Rydberg energy*.

The dispersion behavior in k space can be obtained from the Fourier transform of the envelope function:

$$F(\mathbf{r}) = \int F(\mathbf{k})\exp(i\mathbf{k}\cdot\mathbf{r})\ d\mathbf{r}, \quad (11)$$

which yields for the ground state (compare with Eq. 8)

$$F(k) = \frac{8\sqrt{\pi}}{a_{\mathrm{qH}}^{5/2}} \frac{1}{\left[k^2 + a_{\mathrm{qH}}^{-2}\right]^2} \quad (12)$$

and indicates that the wave packet extends in k space approximately to $k \cong 1/a_{\mathrm{qH}}$. That is, $c_c(k)$ (see Eq. 6) is approximately constant for k up to $1/a_{\mathrm{qH}}$ and decreases rapidly ($\propto 1/k^4$) for $k > 1/a_{\mathrm{qH}}$. For higher excited states, the extent in k shrinks proportionally to $1/n^2$ as a_{qH} increases $\propto n^2$ – see Eq. 2 of ▶ Chap. 17, "Optical Properties of Defects" and Fig. 2.

The envelope function yields the charge distribution of such a shallow defect:

$$\rho(\mathbf{r}) = \frac{1}{V}\int F^*(\mathbf{r})F(\mathbf{r})\ d\mathbf{r}, \quad (13)$$

A direct means to check such a charge distribution is by analyzing the electron nuclear double resonance (ENDOR) signal with the assistance of an electron spin resonance (ESR) line width analysis (see Feher 1959, 1998 and Sect. 3.4.2).

Bound and Resonant States The bound electron of a donor can be described as a wave packet of Bloch states with vanishing group velocity, which is localized at the

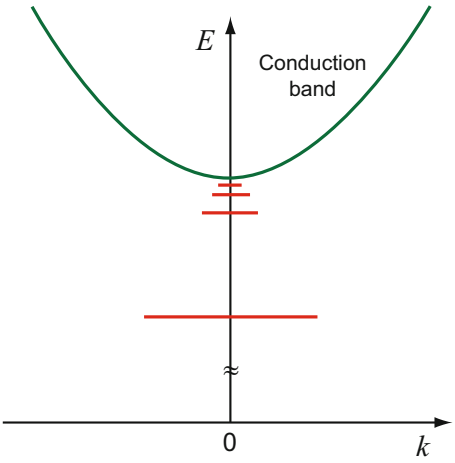


Fig. 2 Ground and excited states of the donor level in the $E(k)$ diagram (*red lines*), indicating the extent of these levels in k space. The ground state has the smallest radius; hence, its extension in k space is the largest

lattice defect (Fig. 1). These Bloch functions must be centered around critical points where $\partial E(\mathbf{k})/\partial \mathbf{k}$ vanishes – see Sect. 4.2 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals", and Callaway (1976).

At higher energies and multiple bands, one obtains a permitted excited state of the defect center within the bandgap with similar features as described above. Alternately, one finds states which are related to higher bands (e.g., the X band in GaP or the spin-orbit split-off valence band in Si) and observes an overlap of these states with lower bands (Fig. 3). Here, the eigenfunctions form a *resonant state* within such a band composed of running Bloch waves (Bassani et al. 1969, 1974). The resonant states have a width that depends on the exchange integral between localized and nonlocalized states. The broadening of the resonant state occurs because the electron has a much reduced lifetime in the quasi-hydrogen state relating to the upper band before it tunnels to the lower band and relaxes to its $E(k)$ minimum. This is easier for higher energies; here, the lifetime is shorter, and therefore the broadening is larger – see Fig. 3a.

Similar features are obtained for localized and resonant states of an acceptor relating to the spin-orbit split-off band and tunneling into the light- and heavy-hole bands (Bassani and Pastori Parravicini 1975); see Fig. 3b. Resonant states with substantial broadening have been observed experimentally – see Onton et al. (1967, 1972), Onton (1971), and Bassani et al. (1969); see also Sect. 3.1.

Prominent examples for the formation of resonant states are observed for dilute nitride III-V-N alloys such as $GaAs_{1-x}N_x$. Replacing a small fraction of the group V atoms by N strongly perturbs the conduction-band structure due to the large difference in electronegativity and atomic size. As a consequence, N complexes are formed, introducing resonant defect levels above the conduction-band edge; for a review, see O'Reilly et al. (2009). The general trend is well described by an empirical two-level band-anticrossing (BAC) model (Shan et al. 1999). It accounts for the strong bandgap bowing at low N composition observed in $GaAs_{1-x}N_x$ in

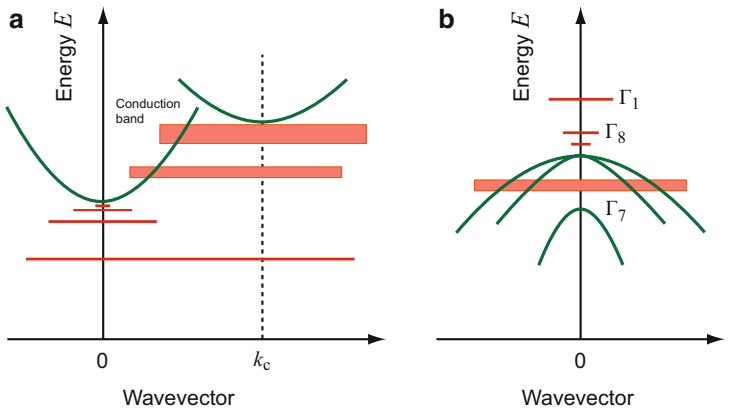


Fig. 3 Schematic of (**a**) a donor or (**b**) an acceptor with excited localized states within the gap and resonant states within the band

terms of an interaction between the GaAs conduction band-edge and a higher lying band of localized N resonant states, and the dependence of the energy gap on composition, pressure, and temperature. It provides also a good qualitative explanation of the electronic properties for other dilute nitride alloys. A quantitative description requires, however, an explicit inclusion of the distribution of N-related defect levels by, e.g., sp^3s^* tight-binding modeling (O'Reilly et al. 2009). Analogous results were found for dilute bismide alloys $GaAs_{1-x}Bi_x$, where the resonant levels are strongly broadened due to their interaction with *valence*-band states (Usman et al. 2011).

1.1 The Chemical Identity

In the simple hydrogen-like approximation, the chemical identity of the donor is totally lost. The identity of the host is provided by $\varepsilon_{\text{stat}}$ and m_n. In contrast to this theory, the ground state is observed to depend significantly on the chemical identity of the donor, such as the different elements of group V impurities in Si (P, As, Sb, or Bi) or monovalent metal interstitials such as Li (Sect. 1.7). The approximation used before is too coarse to show such a dependency. In Sect. 1.3, a number of refinements will be discussed, which will address the complexity of the ground state.

It is remarkable, however, that *excited states* of these shallow donors lose their chemical identity and are rather well explained by the simple theory given above, provided the correct effective mass and dielectric constant (see Sect. 1.3.3 of ▶ Chap. 10, "Interaction of Light with Solids") are used, as will be explained in Sect. 1.2. There are several reasons why an improved agreement is obtained for higher states:

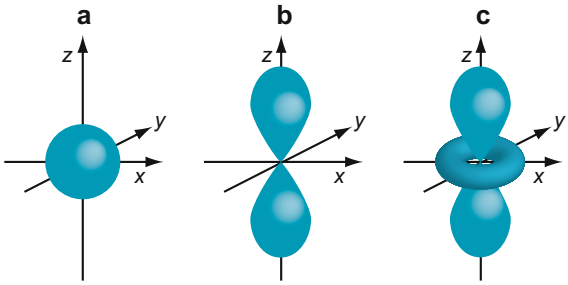


Fig. 4 Shape of the eigenfunction of atomic orbitals: (**a**) *s* type, (**b**) *p* type, and (**c**) *d* type

1. Higher *s* states extend to much larger diameters[4] ($\propto n^2$).
2. *p* states show a node of the wavefunction near the core (see Fig. 4), making the wavefunction less sensitive to the actual potential near the core region.

One needs extremely pure crystals, however, to avoid significant overlap of higher-state eigenfunctions with wavefunctions of other impurities which would cause a perturbation of these excited states. The eigenfunctions of excited states of such shallow impurities often extend beyond 1000 Å or more, interacting with each other when the distance between them is less than 1000 Å, equivalent to a density of $(1/1000\ \text{Å})^3 \cong 10^{15}\ \text{cm}^{-3}$. This requires ultrapure crystals and controlled doping in the < 100 ppb range. In addition, native defects, which are usually frozen-in at densities much in excess of 1 ppm, interfere by influencing the surrounding lattice. Finally, line defects and surface defects (▶ Chap. 15, "Crystal Defects"), which produce internal stresses and electric fields, interfere by perturbing the electronic eigenvalues of defect states. Careful preparation of near-perfect crystals is therefore essential to yield unambiguous results. Very sharp spectral lines of excited states are indeed measured in ultrapure silicon and germanium; see, e.g., Fig. 1 of ▶ Chap. 17, "Optical Properties of Defects", and Steger et al. (2009) for isotopically pure silicon.

1.2 Hydrogen-Like Donors in Indirect-Bandgap Semiconductors

Most free electrons in indirect semiconductors are in one of the sidevalleys rather than at $k = 0$. The energy of such electrons is given by

[4]For instance, when $a_{qH} \cong 50$ Å for the 1*s* state, it is 200 Å for the 2*s* and 450 Å for the 3*s* states, making the hydrogenic effective mass approximation a much improved approximation. In addition, in semiconductors where ε/m^* is already very large, e.g., in GaAs with $\varepsilon_{stat} m_0/m_n = 192.5$, resulting in $a_{qH} = 101.9$ Å $\cong 18a$, this approximation is quite good for the 1*s* state. In GaAs, it results in $E_{qH} = 5.83$ meV, while the experimental values vary from 5.81 to 6.1 meV for GaAs/Si and GaAs/Ge. For more comparisons between theory and experiment, see Bassani et al. (1974).

$$E(\mathbf{k}) = E_{\mathrm{c}}^{(i)} + \frac{\hbar^2}{2}\left(\frac{k_x^2 + k_y^2}{m_{\mathrm{t}}} + \frac{k_z^2}{m_{\mathrm{l}}}\right), \tag{14}$$

where m_{t} and m_{l} are the transverse and longitudinal effective masses due to the ellipsoidal shape of these valleys – see Figs. 10 and 16 of ▶ Chap. 8, "Bands and Bandgaps in Solids". The effective mass Schrödinger equation for the envelope function in one of these valleys, here in the k_z direction, is

$$\frac{-\hbar^2}{2}\left[\frac{1}{m_{\mathrm{t}}}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + \frac{1}{m_{\mathrm{l}}}\frac{\partial^2}{\partial z^2}\right] F^{(z)}(\mathbf{r}) - \frac{e^2}{4\pi\,\varepsilon_{\mathrm{stat}}\varepsilon_0\, r}F^{(z)}(\mathbf{r}) = (E - E_{\mathrm{c}})F^{(z)}(\mathbf{r}). \tag{15}$$

The anisotropic envelope function for a hydrogen-like donor electron in Si within the satellite valley (Kohn 1957) is given by

$$F^{(x)}(\mathbf{r}_l) = F(\mathbf{r}_l)_{\mathrm{isotropic}}\sqrt{\frac{\tilde{a}^3}{a^2 b}}\,\frac{\exp\left[-\left(\frac{x_l^2}{(nb)^2} + \frac{y_l^2 + z_l^2}{(na)^2}\right)\right]}{\exp\left(-\frac{r_l}{n\tilde{a}}\right)}, \tag{16}$$

with $a = 25$ Å, $b = 14.2$ Å, $\tilde{a} = 21$ Å, $n = \sqrt{0.029/E_d(\mathrm{eV})}$, and E_d as the ground-state ionization energy of the donor. The x axis is aligned with the direction of $\mathbf{k}$. Similar expressions are given for $F^{(y)}$ and $F^{(z)}$.

In addition, in noncubic semiconductors the dielectric constant is anisotropic (*anisotropic shielding*). The corresponding Schrödinger equation for the envelope function is then given by

$$\frac{-\hbar^2}{2}\left[\frac{1}{m_{\mathrm{t}}}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + \frac{1}{m_{\mathrm{l}}}\frac{\partial^2}{\partial z^2}\right] F_j(\mathbf{r}) - \frac{e^2}{4\pi\sqrt{\varepsilon_{||}\varepsilon_{\perp}}\varepsilon_0\sqrt{x^2 + y^2 + \left(\varepsilon_{\perp}/\varepsilon_{||}\right)z^2}}F_j(\mathbf{r}) = (E - E_{\mathrm{c}})F_j(\mathbf{r}), \tag{17}$$

with $\|$ and $\perp$ denoting directions parallel or perpendicular to the z direction and aligned with the crystallographic c direction. It shows substantial anisotropy of the envelope function in its ls state as an ellipsoid rather than a sphere – see Eq. 14. The nomenclature ls, 2s, etc., is usually retained for these states, to which the eigenfunctions would converge in the limit of vanishing anisotropy. Quasi-Bohr radius and effective Rydberg energy are defined by using m_{t} and $\sqrt{\varepsilon_{||}\varepsilon_{\perp}}$ rather than m_n and $\varepsilon_{\mathrm{stat}}$.

The degree of changes of various excited states as a function of the degree of sidevalley anisotropy γ can be seen from Fig. 5; often the anisotropy is expressed in terms of the factors γ, γ_a, or α, with $\gamma = m_{\mathrm{t}}/m_{\mathrm{l}}$, $\gamma_a = m_{\perp}\varepsilon_{\perp}/\left(m_{||}\varepsilon_{||}\right) = 1 - \alpha$. The

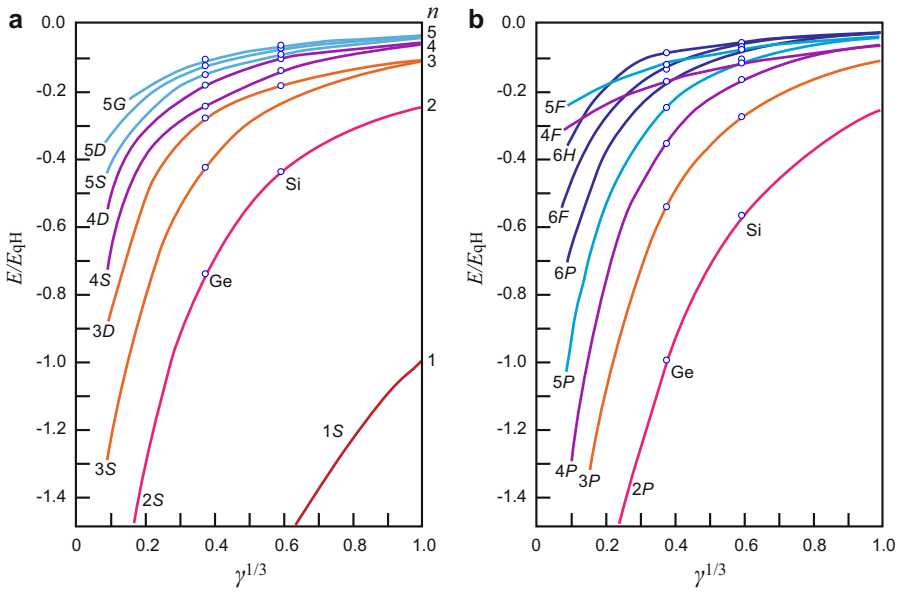


Fig. 5 Energy levels of donor states calculated with hydrogenic effective-mass approximation as a function of the anisotropy parameter γ (see text) for (**a**) *s*-like and (**b**) *p*-like states. The limits, $\gamma = 1.0$ and $\gamma = 0$, indicate isotropic and two-dimensional semiconductors, respectively, and $E = 0$ corresponds to E_c. The dots show the levels for Ge and Si (After Faulkner 1969)

values of γ for Ge and Si are marked by dots in the figure. Higher anisotropy compresses the eigenfunction in one direction and thereby increases the binding energy as the electron is forced closer to the nucleus ($m_t > m_n$; $\varepsilon_\perp > \varepsilon_{stat}$). Changes of the eigenstates can be significant as a function of the anisotropy parameter (up to a factor of 4 for the ground-state energy; Shinada and Sugano 1966), as shown for the ground state in Fig. 6 (Pollmann 1976).

The ground-state energies, although improved, do not yet agree with the experiment. Excited states, however, show rather impressive agreement, as indicated in Fig. 7. In this figure, all experimental values are shifted so that the $2p_0$ level agrees with the theory. This figure also includes an alkali interstitial and a group VI element. The S^+-defect yields an energy four times higher because of the double charge of the center. This example shows that the agreement goes beyond the original quasi-hydrogen list of shallow donors in group IV semiconductors. The radius of substitutional, interstitial, and double-donor wavefunctions in Si was more recently determined from the ratio of the coefficients of the linear and quadratic Zeeman effects, independently confirming the validity of the effective-mass approach for excited states (Li et al. 2018).

For donors, the *chemical shift* of the ground state is largest in Si, namely, 111 meV between B and In. The shift is much less severe in semiconductors with smaller effective mass and larger dielectric constant. In Ge this shift is smaller by a

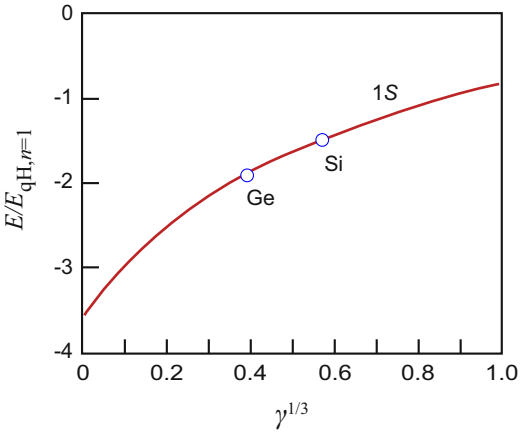


Fig. 6 Ground-state energy (normalized to the l*s* state energy of the quasi-hydrogen donor in an isotropic crystal) as a function of the anisotropy factor

factor of ~100. Several other approximations, discussed in the following section, are used to obtain improved donor-state energies [Luttinger and Kohn (1955), Faulkner (1969), Torres et al. (1997); see also reviews by Baldereschi and Lipari (1973) and Bassani et al. (1974)].

1.3 Hydrogen-Like Ground State and Chemical Shift

The disagreement of the ground-state calculation in the hydrogenic effective mass theory with the experiment is substantial. Immediately apparent from Fig. 7 is the fact that the donor 1*s* state is split into three (or two) lines for Si (or Ge). This is caused by coupling between the states relating to the six (or four) equivalent sidevalleys of the conduction band in Si (or Ge), which was first analyzed by Morita and Nara (1966). The splitting varies for different donors and is by far the largest contribution to the observed deviation of the ground state from the simple effective mass hydrogen-like model. We will discuss this contribution in Sect. 1.3.1.

Other contributions to the chemical shift deal with the short-range actual potential of the impurity. Such *central cell-corrections* to the effective-mass theory included various attempts to:

- Use modified dielectric screening for estimating the short-range potential.
- Use short-range model potentials with adjustable parameters.
- Introduce local pseudopotentials.
- Consider the influence of lattice relaxation surrounding the impurity.

Lattice Relaxation The lattice surrounding a substitutional impurity relaxes. This was computed using normconserving pseudopotentials to describe the interaction between core and valence electrons and the local-density approximation for exchange and correlation interaction with an ab initio calculation with no adjustable

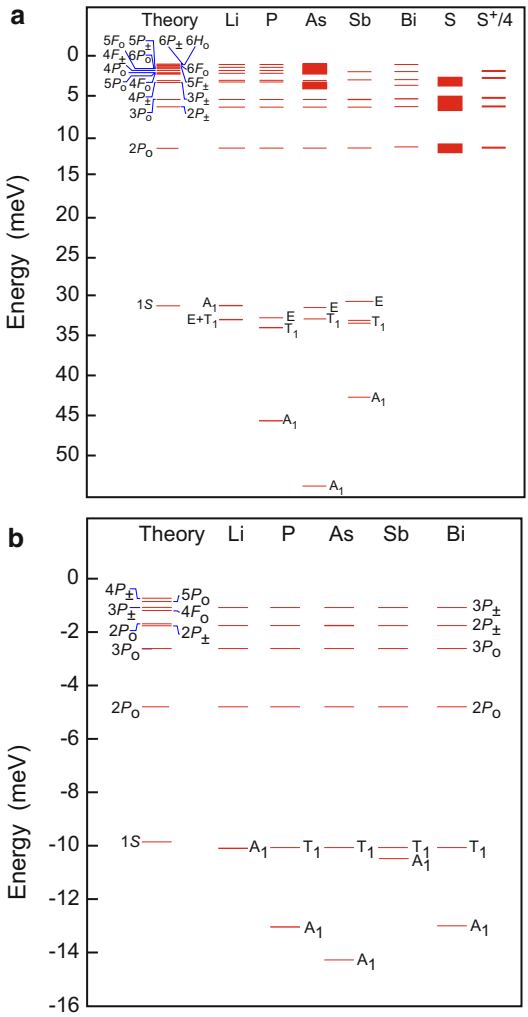


Fig. 7 (**a**) Energy levels of shallow donors in silicon. The theoretical level distribution is obtained with $\varepsilon = 11.4$. The experimental values are shifted so that the $2p_0$ levels line up with the theoretical level. The experimental uncertainty is indicated as the width of the levels. (**b**) Same as (**a**) for Ge (After Faulkner 1969)

parameters. For instance, incorporating a substitutional B atom relaxes the four surrounding Si atoms inward (like a breathing mode) by 0.21 Å (9%) and moves the B atom slightly (0.1 Å) off center toward the plane with three Si atoms, for a slight tendency of a threefold coordination. The total energy gain by this relaxation is 0.9 e V (Denteneer et al. 1989). A small fraction of it contributes to the chemical shift of the electronic energy of the hydrogen-like ground state.

There are many models, which are reviewed by Stoneham (1975). We will deal with the more important aspects of the central-cell correction to the hydrogenic potential in the following subsections (see also Pantelides 1978).

1.3.1 Band Mixing

In element semiconductors, the mixing of hydrogen-like donor states with different subbands results in the splitting of the ground state.

Mixing through *intervalley interaction* in Si (or Ge) (reviewed by Pantelides 1978) removes the degeneracy of the sixfold degenerate ground state (E_0) in the neighborhood of the defect, splitting it into three (two) levels of symmetry Γ_1, Γ_{12}, and Γ_{15} (or Γ_1 and Γ_{15}). The corresponding energies for Si:P (experimental values from Aggarwal and Ramdas 1965a, b) are

$$\begin{aligned} E_c - E_{\Gamma_1} &= E_0 - \lambda - 4\mu \quad (A_1 \text{ singlet}) = 45.3 \text{ meV} \\ E_c - E_{\Gamma_{12}} &= E_0 - \lambda + 2\mu \quad (T_2 \text{ triplet}) = 33.7 \text{ meV} \\ E_c - E_{\Gamma_{15}} &= E_0 + \lambda \qquad (E \text{ doublet}) = 32.3 \text{ meV} \end{aligned} \tag{18}$$

The A_1 singlet is an *s*-like state, and the T_2 triplet is a *p*-like state, with A_1 and T_2 being irreducible representations of the tetrahedral group T_d. For Ge, we have (experimental values from Reuszer and Fisher 1964)

$$\begin{aligned} E_c - E_{\Gamma_1} &= E_0 - 3\lambda' \quad (A_1 \text{ singlet}) = 14.2 \text{ meV} \\ E_c - E_{\Gamma_{15}} &= E_0 + \lambda' \quad (T_2 \text{ triplet}) = 10.0 \text{ meV} \end{aligned} \tag{19}$$

where λ and μ are the matrix elements for transitions from (k_0, 0, 0) to (0, k_0, 0) and ($-k_0$, 0, 0), respectively, in Si and where $\pm\ \lambda'$ is the matrix element between the impurity and its mirrored position at $(\pm\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ in Ge. These matrix elements are computed by Baldereschi (1970); see also Bassani et al. (1974).

Mixing of hydrogen-like states with several bands is considered when the bands have their edges at similar energies:

- Different equivalent subbands (as discussed above)
- Different nonequivalent subbands (e.g., in GaAs or GaSb)
- Different degenerate bands at $k = 0$ (valence bands for most semiconductors)

Mixing with *nonequivalent valleys* is important when the valleys are closely spaced in energy to E_c (Vul' et al. 1971). A detailed analysis was performed by Altarelli and Iadonisi (1971). The coupling increases when, e.g., with hydrostatic pressure, the energy separation between the Γ and X minima in GaAs decreases (Costato et al. 1971). Castner (1970) presents indications that higher valleys *and* saddle points contribute to the split ground-state energies in Si.

The progressive improvement of the theoretical description of hydrogen-like defects is indicated in Fig. 8. The corresponding energies are listed in Table 1.

There are optical transitions possible between these split ground states. Three absorption lines due to transitions from the $1s(A_1)$ to the $1s(T_2)$ state are observed, while the transitions to $1s(E)$ are forbidden; such transitions can be observed, however, under uniaxial stress (Bergman et al. 1987). In addition, one needs to use the proper dielectric constant $\varepsilon(k)$ at the position k of the minima of the sidevalleys (Baldereschi 1970).

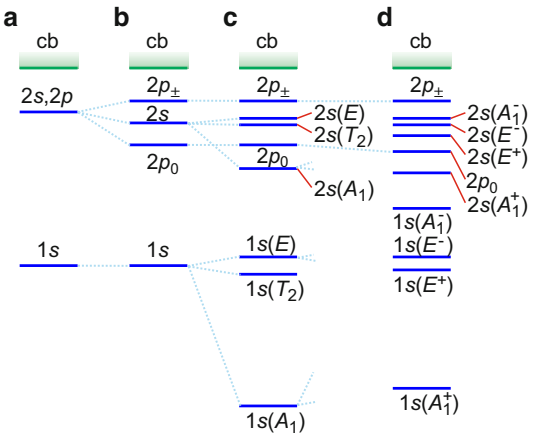


Fig. 8 Lower energy states of an electron bound to a donor: (**a**) one-valley treatment with isotropic effective mass (e.g., GaAs); (**b**) one-valley treatment, anisotropic effective mass; (**c**) multivalley treatment in T_d symmetry; (**d**) multivalley treatment with appropriate D_{3d} symmetry (e.g., Si) (After Grimmeiss and Janzén 1986). $p_{\pm}$ denotes $m_l = \pm 1$ states of the p level

Table 1 Binding energy of donors in Si and Ge (meV); 1s values in brackets are additional levels of the split 1s ground state (After Bassani et al. 1974)

	State						
Semiconductor	1s	$2p^0$	2s	$2p^{\pm}$	$3p^0$	3s	$3p^{\pm}$
Si (theo.)	31.27 (10.6, 1.1)	11.51	8.82	6.40	5.48	4.75	3.12
Si:P	45.5 (33.9, 32.6)	11.45		6.39	5.46		3.12
Si:As	53.7 (32.6, 31.2)	11.49		6.37	5.51		3.12
Si:Sb	42.7 (32.9, 30.6)	11.52		6.46	5.51		3.12
Ge (theo.)	9.81 (0.6)	4.74	3.52	1.73	2.56	2.01	1.03
Ge:P	12.9 (9.9)	4.75		1.73	2.56		1.05
Ge:As	14.17 (10.0)	4.75		1.73	2.56		1.04
Ge:Sb	10.32 (10.0)	4.74		1.73	2.57		1.04

1.3.2 Short-Range Potential Corrections

The energy of the impurity ground-state also depends on the net impurity potential. In its most simple form, it is given as the difference between the Coulomb potential of the impurity and the lattice atom that is replaced and the central-cell potential:

$$V(\mathbf{r}) = \left(Z_{\text{host}} - Z_{\text{impurity}}\right) \frac{e^2}{4\pi\,\varepsilon_{\text{stat}}\varepsilon_0\, r} + V_{\text{cc}}(\mathbf{r}). \tag{20}$$

The central-cell potential V_{cc} extends over the range of the unit cell. It contains the short-range components due to the chemical individuality of the defect and the nonlinearity of the polarization in the direct neighborhood of the impurity. In the first approximation, one connects the central-cell potential asymptotically with the Coulomb potential, however, using a k-dependent ε near the impurity.

For the central-cell potential, various approximations are made (Csavinszky 1965; Bebb 1969). When fulfilling proper orthogonality requirements on the core orbitals of the impurity, such impurity potentials can be used for isocoric impurities only in a generalized effective-mass approximation to obtain better fitting ground-state energies of shallow defect centers (Sah and Pantelides 1972; Pantelides and Sah 1974). For nonisocoric impurities, see next Sect. 1.3.3. A central-cell correction of the potential near the core is reviewed by Stoneham (1975). It is included in an effective impurity potential

$$V_{\text{pseudo}}(\mathbf{r}) = \begin{cases} \dfrac{e^2 Z(r)}{4\pi\,\varepsilon_{\text{stat}}\varepsilon_0\, r} & \text{for } r > r_0 \\ \displaystyle\int \frac{d\mathbf{k}}{(2\pi)^3} \frac{V'(\mathbf{k})}{4\pi\,\varepsilon(\mathbf{k})\,\varepsilon_0} \exp(i\mathbf{k}\cdot\mathbf{r}) & \text{for } r < r_0, \end{cases} \tag{21}$$

where $V'(\mathbf{k})$ is the Fourier transform of the unscreened impurity potential and $\varepsilon(\mathbf{k})$ is the wavevector-dependent dielectric function – see Fig. 12 of ▶ Chap. 10, "Interaction of Light with Solids". $Z(r) = Z_0$ is used here for $r > r_0$, with Z_0 as the point charge of the impurity with respect to the replaced lattice atom (1 for As, 2 for Se, etc.), and r_0 as an adjustable arbitrary boundary between the inner and outer regions of the approximation. This is a simple example for an often used treatment. Namely, it is the splitting of the region of interest into an inner region which is treated in more detail and an outer region, with a Coulomb potential asymptotically approaching the unperturbed lattice potential. Boundary conditions between these two regions are of importance. For better approximations, the inner region should be made as large as possible and should contain more than just the neighbor atom of the defect center (the Keating potential extends to the fifth neighbor; see Stoneham 1986).

1.3.3 Local Pseudopotentials and Model Potentials

Local Pseudopotentials Instead of computing the true impurity potential, which could yield the set of quasi-hydrogen levels of interest and also deeper core levels of limited interest, one usually proceeds to estimate a local pseudopotential[5] by which only outer-shell electrons are influenced.

The *pseudopotential formalism* as applied to impurities can best be explained (Phillips and Kleinman 1959; Austin et al. 1962) by comparing the Schrödinger equation containing the true impurity potential $V(\mathbf{r})$

$$\left[-\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial \mathbf{r}^2} + V(\mathbf{r})\right] \psi_n = E_n \psi_n \tag{22}$$

[5]Such a local pseudopotential is used near an impurity as opposed to the nonlocal pseudopotential used for band structure analysis (see Sect. 1.3 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals").

with the Schrödinger equation containing the pseudopotential

$$\left[-\frac{\hbar^2}{2m_0}\frac{\partial^2}{\partial \mathbf{r}^2}+V_{\text{pseudo}}(\mathbf{r})\right]\phi_n=E_n'\phi_n. \tag{23}$$

Bassani and Celli (1961) have shown that the eigenvalues E_n and E_n' for Eqs. 22 and 23 are nearly identical for wavefunctions $\psi_n=\psi_c$ and $\phi_n=\phi_c$ for the conduction band, i.e., if only the behavior of outer electrons is of interest. Therefore, it is sufficient to obtain the shallow energy states of an impurity when composed of conduction-band eigenfunctions by using the pseudopotential, rather than the true impurity potential. A short review on how to obtain the proper pseudopotential is given by Pantelides (1978).

An example of the potential distribution for several mono- and divalent hydrogen-like impurities in Si, as computed by Pantelides (1975), is given in Fig. 9a. It can be seen that isocoric impurities (P and S) have nearly point-charge Coulomb behavior, since the contribution from the core potential is similar to that of the replaced host atom and therefore cancels. Other elements with a different atomic core, however, show a major variation of the atomic potential in the core region indicated for As and Se. For these impurities, $V_{\text{pseudo}}(r)$ remains closer to the behavior of a point charge, as shown for As_{Si} in Fig. 9b.

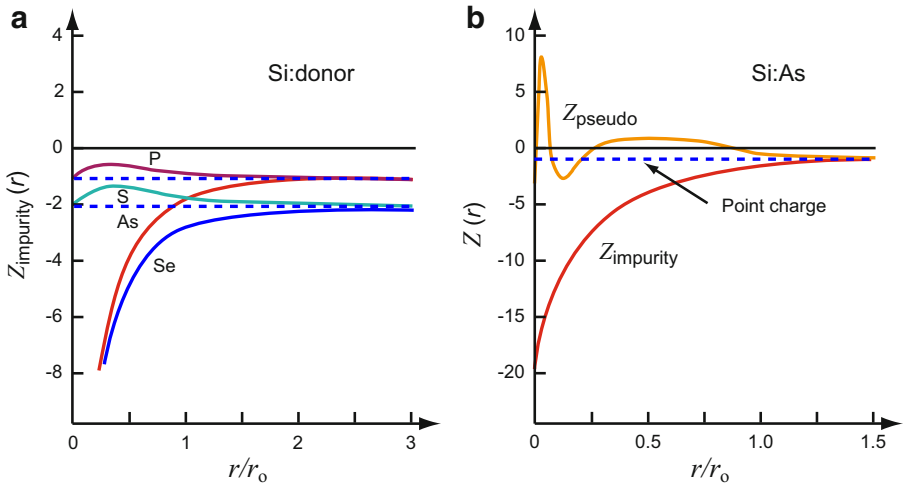


Fig. 9 (**a**) Effective impurity charge $Z(r)$ modifying the Coulomb potential $V(r)=Z(r)e^2/(4\pi\varepsilon_{\text{stat}}\varepsilon_0 r)$ to present the "true potential" for substitutional impurities in Si, as contrasted with $Z(r)=Z=1$ or $Z=2$ for a single or double point charge, and shown as *horizontal dashed lines*; (**b**) Z_{impurity} as in subfigure (**a**) for As in an extended scale. Z_{pseudo} is the corresponding pseudopotential charge with cancellation of the strong core part that is only relevant for core levels (After Pantelides 1975)

Pseudopotentials were used to calculate the eigenvalues for a number of shallow-level defects in Si (Pantelides and Sah 1974) and compared to ENDOR data (Schechter 1975).

Model Potentials Simple model potentials have been used with moderate success (Abarenkov and Heine 1965; Appapillai and Heine 1972). These potentials contain the essential features of the pseudopotential, namely, a rather smooth core potential that gives the chemical identity of the impurity and a Coulomb tail that provides good agreement with higher states of quasi-hydrogen impurities. An example is the Abarenkov-Heine potential

$$V_{\text{pseudo}}(\mathbf{r}) = \begin{cases} -\dfrac{e^2}{4\pi\,\varepsilon_{\text{stat}}\varepsilon_0\, r} & \text{for } r > r_0 \\ \displaystyle\sum_l A_l P_l & \text{for } r < r_0, \end{cases} \tag{24}$$

where P_l is the angular projection operator and A_l are the empirical energy constants. This model potential is shown in Fig. 10 and yields reasonable agreement with the experiment if the summation in Eq. 24 includes higher terms in l (Baldereschi and Lipari 1976).

1.4 Hydrogen-Like Acceptors

In principle, the hydrogenic effective-mass theory for a shallow acceptor is much like that for a donor, except that three bands must be considered: the light- and heavy-hole valence bands and the split-off spin-orbit band. With larger split-off energies Δ compared to kT, the contribution from the split-off band to the ground-state energy can be neglected. In addition, band warping causes an anisotropy of the effective masses, which needs to be considered. For example, the characteristic decay length of the eigenfunction in the [111] direction in Ge:Ga is 92 Å and in the [100] direction is 87 Å.

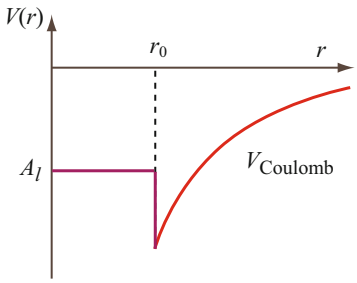


Fig. 10 Abarenkov-Heine model potential of a shallow impurity

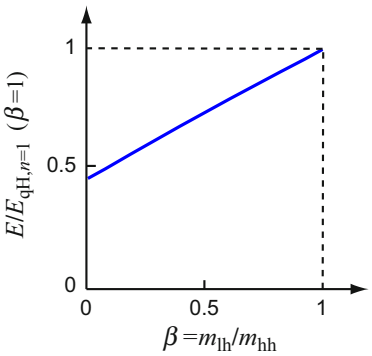


Fig. 11 Energy of the acceptor ground state as a function of the effective-mass ratio β between light and heavy holes

The ground-state energy for acceptors depends on the ratio between light- and heavy-hole masses ($\beta = m_{lh}/m_{hh}$), as computed by Gel'mont and D'yakonov (1972) and shown in Fig. 11. It decreases nearly linearly with decreasing β.

Each of these bands contributes to the *excited state level spectrum* of the acceptors and thereby produces a greater wealth of levels. This is shown in Fig. 12, with levels identified as corresponding to the light- and heavy-hole bands of symmetry Γ_8 and spin-orbit split band of symmetry Γ_7. The quantum numbers preceding the band notation in Fig. 12 indicate successively higher excited states. In addition, the valence bands are anisotropic, requiring the use of the appropriate Luttinger parameters, also called the inverse effective-mass constants – see Luttinger and Kohn (1955), Luttinger (1956), and Sects. 1.2.2 and 1.2.3 of ▶ Chap. 8, "Bands and Bandgaps in Solids". For a review, see Bassani et al. (1974) and Pantelides (1978).

Baldereschi and Lipari (1973) have computed a number of acceptor levels (Table 2) in various semiconductors with point-charge screened potential. These computations provide the correct trend, but the so-obtained levels still show some differences to the experimental values given in Table 3.

1.5 Shallow Defects in Compound Semiconductors

The shallow donor or acceptor in a compound semiconductor is more complex than in an elemental semiconductor for two reasons:

1. The interaction of the electron or hole with the alternatingly charged ions of the lattice
2. The differentiation between incorporating the defect on an anion and cation site

Although the degree of ionicity of most compound semiconductors is small, the above-stated effects are not negligible. For example, isocoric acceptors in GaP should be well described by point-charge quasi-hydrogenic models – see Sect. 1.3. However, the two isocoric acceptors, GaP:Zn_{Ga} and GaP:Si_P, have substantially different ground-state energies: 64 and 204 meV, respectively. This difference cannot

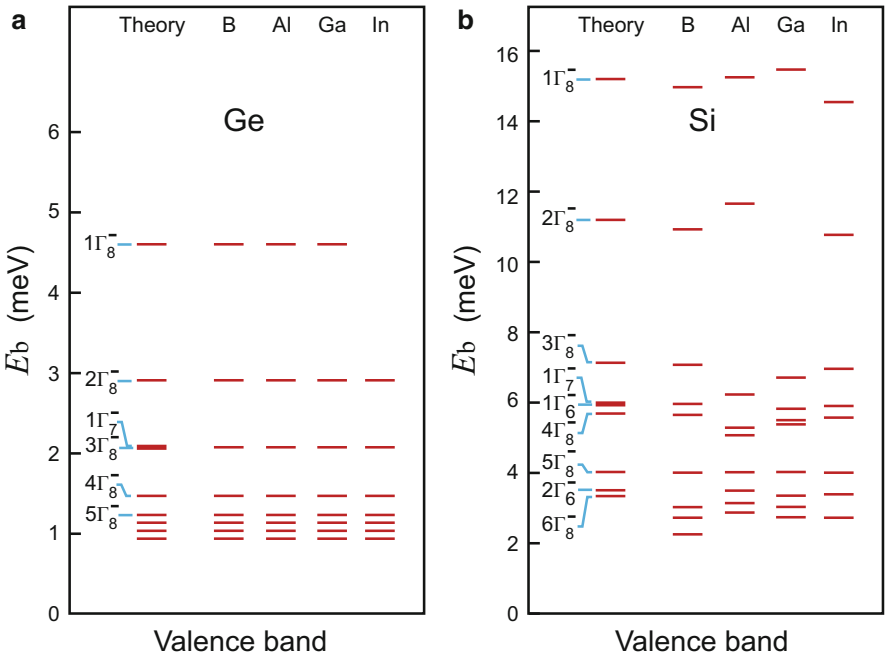


Fig. 12 Ground and excited states of acceptors in (**a**) Ge and (**b**) Si (After Lipari and Baldereschi 1978; experimental data from Haller and Hansen 1974)

Table 2 Acceptor levels in (meV) computed from point-charge screened potentials (After Baldereschi and Lipari 1973)

Material	$1S_{3/2}$ (Γ_8)	$2S_{3/2}$ (Γ_8)	$P_{1/2}$ (Γ_6)	$P_{3/2}$ (Γ_8)	$P_{5/2}$ (Γ_8)	$P_{5/2}$ (Γ_7)
Si	31.56	8.65	4.18	12.13	8.51	5.86
Ge	9.73	2.89	0.61	4.30	2.71	2.04
AlSb	42.45	12.40	3.35	18.46	12.00	8.22
GaP	47.40	13.69	4.21	19.17	13.04	9.42
GaAs	25.67	7.63	1.60	11.38	7.20	5.33
GaSb	12.55	3.77	0.650	5.74	3.59	2.61
InP	35.20	10.53	1.97	15.89	9.98	7.32
InAs	16.31	5.00	0.420	7.91	4.76	3.63
ZnS	175.6	51.98	11.65	77.62	49.55	35.37
ZnSe	110.2	32.98	6.07	50.04	31.47	22.68
ZnTe	77.84	23.07	5.09	34.72	22.32	15.36
CdTe	87.26	26.42	3.70	41.43	25.85	17.68

Table 3 Observed shallow acceptor levels in meV (After Bassani et al. 1974)

Semiconductor	Acceptor	Transitions								Ionization energy
		1st	2nd	3rd	4th	5th	6th	7th	8th	
Ge	B	6.24	7.57	7.94	8.66	9.06	9.32	9.65	9.81	10.47
	Al	6.59		8.27	9.02		9.67	10.02	10.15	10.80
	Ga	6.74	8.02	8.44	9.19		9.84	10.17	10.31	10.97
	In	7.39	8.42	9.08	9.86	10.20	10.48	10.81	10.96	11.61
	Tl	8.87	9.83	10.57	11.32	11.65	11.92	12.26	12.43	13.10
Si	B	30.38	34.53	38.35	39.64	41.52	42.50	42.79	43.27	44.5
					39.91					
	Al				64.08					
		54.88	58.49		64.96	66.28	66.75	(67.1)	67.39	68.5
					65.16					
	Ga				67.95					
		58.23		67.12	68.25	69.85	70.49	(70.8)	71.11	72
					68.43					
	In	141.99	145.79	149.74	150.80	(152.8)	153.27		153.97	155
					151.08					
AlSb				27.17						
		24.66	26.21	27.53	29.41	30.9	32.2			33
				27.87						
InSb	Zn, Cd, Ag	5.7	6.94	7.31	7.93					
		25.5	26.5	27.3	27.9					

be explained by site-dependent screening, even though neighboring anions are expected to screen more effectively since they are surrounded by more electrons. It needs a more sophisticated analysis, beyond that of the effective-mass approximation (Bernholc and Pantelides 1977).

The degree of ionicity in compound semiconductors also determines the coupling of electrons with the lattice, i.e., with phonons; it is described by Fröhlich's coupling constant α_c (Sect. 1.2.1 of ▶ Chap. 22, "Carrier-Transport Equations"). Such an interaction can be included by considering, instead of a Bloch electron, a polaron to interact with the defect center, with an effective mass $m_{pol} = m_n/(1 - \alpha_c/6)$. The eigenstates of a hydrogen-like defect in compound semiconductors can be estimated as

$$E_{qH}^{(compound)} = \left(1 + \frac{\alpha_c}{6} + \frac{\alpha_c}{24}\frac{E_{qH}}{\hbar\omega_{LO}}\right)E_{qH}; \tag{25}$$

(see Sak 1971) where E_{qH} is the quasi-hydrogen energy (Eq. 1 of ▶ Chap. 17, "Optical Properties of Defects"). The modified hydrogenic effective-mass approximation describes reasonably well the level spectrum of excited states of shallow donors and acceptors in compound semiconductors with a sufficiently large ε/m_p ratio, i.e., for many III–V and II–VI compounds. However, this approximation is not sufficient to explain the observed variations of the ground-state energies indicated above.

The binding energy of substitutional shallow-level impurities is mainly obtained from photoluminescence and IR spectroscopy. Some experimental values are listed in Table 4 (Kirkman et al. 1978; Dean 1973a; Kopylov and Pikhtin 1978).

1.6 Donor-Acceptor Pair and Free-To-Bound Transitions

Donor-Acceptor Pairs When donors and acceptors are both present in a semiconductor, there is some probability that a given donor has an acceptor nearby. Since ionized donors and acceptors are Coulomb attractive to each other, they occupy preferably nearest lattice sites at low temperatures. In addition to electrostatic

Table 4 Binding energy in meV of hydrogenic donors (E_d) and acceptors (E_a) in GaAs and GaP measured at low temperatures (typ. 1.8–4.2 K); EMT denotes energy calculated by effective-mass theory

Donor	E_d in GaAs	E_d in GaP	Acceptor	E_a in GaAs	E_a in GaP
EMT	5.72	59	EMT	27.0	56.3
C	5.91		Be	28.0	56.6
Si	5.84	85	Mg	28.7	59.9
Ge	5.88	204	Zn	30.6	69.7
S	5.87	107	Cd	34.7	102.2
Se	5.79	105	C	26.9	54.3
			Si	34.8	210
			Ge	40.4	265

attraction discussed below, space-filling aspects also create attractive forces. For instance, incorporation of an As donor in Si causes a slight shrinkage of the lattice $\Delta a/(a \cdot N_{\mathrm{As}}) \cong -0.4 \cdot 10^{-24}$ cm^3, while inclusion of a Ga acceptor dilates the lattice by $\Delta a/(a \cdot N_{\mathrm{Ga}}) \cong 0.8 \cdot 10^{-24}$ cm^3, resulting in a strain-induced attraction of Ga and As within the Si lattice. Normally, a distribution of these pairs is found with various distances between them.

Such donor-acceptor pairs may be composed of cation vacancies or impurities that act as acceptors, as well as substitutionals, cation interstitials, anion vacancies, or other impurities that act as donors. See Williams (1968), Dean (1973b), and Taguchi and Ray (1983) for II-VI compounds.

Donor-acceptor pairs in GaP are well investigated. The Coulomb interaction between the donor and acceptor is "quantized", since the relative positions are dictated by the discrete lattice sites. This gives rise to a sequence of discrete lines in the recombination luminescence spectrum. One distinguishes substitutionals of the same sublattice, such as a C_P-O_P (carbon or oxygen on a phosphorus site) pair acting as acceptor and donor, respectively, which is referred to as type I pair. An example of a type II pair is Zn_{Ga}-O_P in GaP, in which the two sublattices are involved. The energy of donor-acceptor pairs is influenced by their relative distance $r_{d\text{-}a}$ and can be expressed as

$$E_{d-a} = E_g - (E_d + E_a) - \frac{e^2}{4\pi\,\varepsilon_{stat}\varepsilon_0\, r_{d-a}} - E^*(r_{d-a}), \tag{26}$$

where E_d and E_a are the distance of donor or acceptor levels from the respective band edges. The pair energy is substantially influenced by the Coulomb attraction and by a correction energy term $E^*(r_{d-a})$, caused by overlap of donor and acceptor eigenfunctions. The correction includes a Van der Waals term and two multipole terms, which are discussed for zincblende GaP and SiC by Sun et al. (2011). The correction term is much reduced at larger distances.

At low temperatures ($T < 10$ K), the absorption lines are sharp enough to distinguish a large number of pairs with different pair distance. This distance can be identified in crystallographic notation for a cubic lattice by the position of both atoms – the first assumed at 000 and the second at $n_1\, n_2\, n_3$, with m the *shell number* and n_i integers starting at 0:

$$2m = n_1^2 + n_2^2 + n_3^2. \tag{27}$$

In an *fcc* sublattice of zincblende structure, the sum $(n_1 + n_2 + n_3)$ is even. The separation distance in the cubic zincblende lattice with lattice constant a is then given for type I pairs or for type II pairs by

$$(r_{d-a})_{\mathrm{I}} = \sqrt{\frac{m}{2}}a, \quad \text{or} \quad (r_{d-a})_{\mathrm{II}} = \sqrt{\left(\frac{m}{2} - \frac{5}{16}\right)}\,a. \tag{28}$$

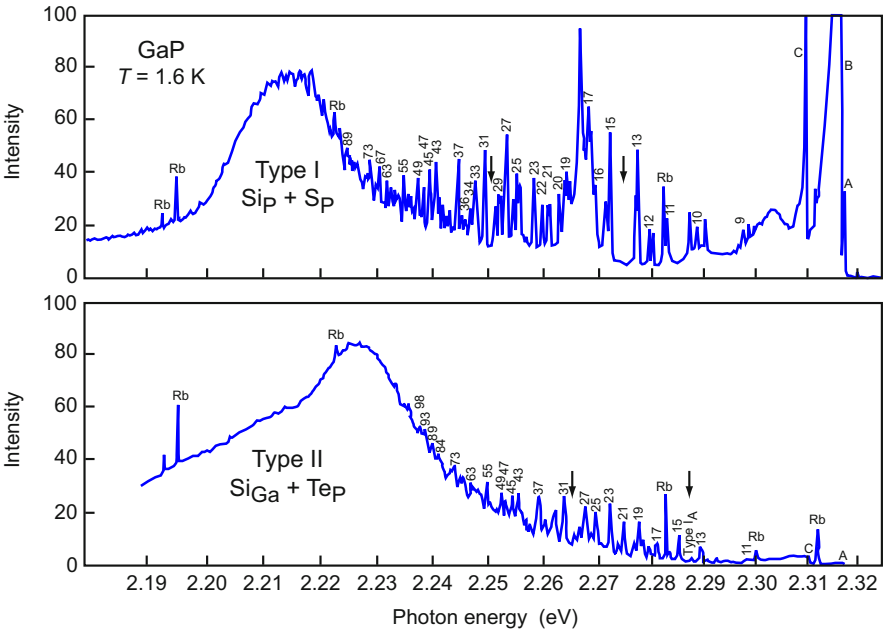


Fig. 13 Luminescence spectrum of GaP type I, doped with S_P and Si_P ($E_a + E_d = 0.140$ eV) in the *upper* diagram, measured at 1.6 K, and type II, doped with Si_{Ga} and Te_P ($E_a + E_d = 0.124$ eV) in the lower diagram. The numbers in the figure refer to the shell number m in Eq. 28, and *arrows* in the type I spectrum indicate empty shells. Rb lines are used for calibration (After Thomas et al. 1964)

The absorption lines of the donor-acceptor pair are identified sequentially according to m. For some m condition Eq. 28 predicts distances where no lattice site occurs in type I pairs; these distances are referred to as empty shells.[6] A typical spectrum showing the wealth of distinguished lines is given in Fig. 13. The energy as a function of the relative pair distance is shown in Fig. 14. We see that for higher pairs ($m > 15$), the fit neglecting the perturbation term E^* is excellent. For less distant pairs, however, this term is necessary to provide a better agreement between theory and experiment. In the highly structured luminescence spectrum of pairs in GaP peaks up to $m \cong 90$ are identified (in pair spectra of ZnSe up to $m = 117$, see Merz et al. 1973), indicating donor-acceptor interaction over at least seven lattice constants. At still higher separation, the peaks are more closely spaced and can no longer be resolved; in GaP, they produce a broad band near $h\nu \cong 2.22$ eV of rather intense luminescence.

In addition to donor-acceptor pairs, we distinguish isoelectronic center pairs, e.g., N-N pairs of substitutional N_P in GaP shown in Fig. 24. Other pairs which

[6] An empty shell in an *fcc* sublattice of zincblende structure occurs if $2m = 4^i\,(8j + 7)$ with integers $i, j \geq 0$ (Pohl and Busse 1989); see arrows at $m = 14$ and 30 in the upper spectrum of Fig. 13.

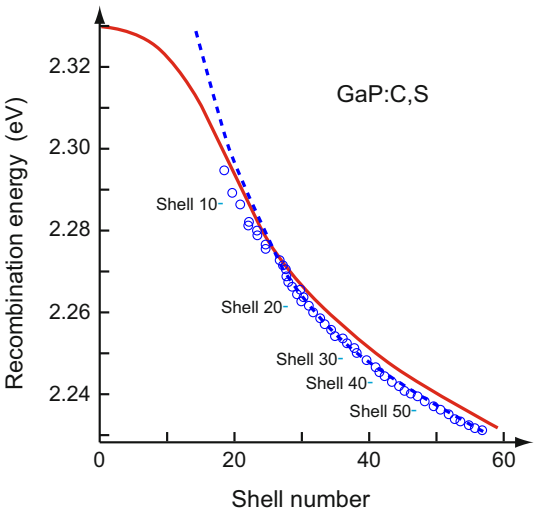


Fig. 14 Energy of donor-acceptor pairs C_{Ga}-S_P in GaP as a function of the shell number and fit with Eq. 26; dashed curve without and solid curve with $E^*(r_{d-a})$ consideration (After Hayes and Stoneham 1984)

have been the subject of extensive research are the chalcogen pairs in Si (Wagner et al. 1984).

Free-To-Bound Transitions Luminescence transitions involving shallow hydrogen-like defects occur from the band to the acceptor or donor (free-to-bound transition) or between donor-acceptor pairs described above. With increasing temperature, more recombination transitions involving at least one of the bands are favored since the lifetime of carriers in shallow donors is decreased.

Free-to-bound transitions from a band into a hydrogen-like donor or acceptor reflect the optical ionization energy. Because of a Franck-Condon shift (Markham 1966), this optical ionization energy usually is different from the thermal ionization energy used for determining the semiconductivity. For hydrogen-like defects, however, such a shift is small because of a very small electron-lattice coupling, except for transitions to the ground state.

The line shape reflects the carrier distribution within the band. As an example, Fig. 15a shows the transition of an electron from the conduction band into an acceptor in GaAs at 1.9 K. This emission line shows a sharp cutoff at low energies corresponding to the energy from the band edge to the ground state of the acceptor (Fig. 15b). Toward higher energies, a long exponential tail is observed, which reflects the electron-energy distribution (f_n) within the band:

$$I(h\nu) = A g_c(h\nu) N_a f_n\big((h\nu - E_g + E_a)/(kT)\big), \tag{29}$$

where $g_c(h\nu)$ is the density of states near the conduction-band edge, N_a is the density of the acceptors, and A is a proportionality constant. From the exponential slope of this tail, one estimates an electron temperature of 14.4 K. This increased electron temperature is due to the optical excitation, thermalized with phonons. Figure 15a

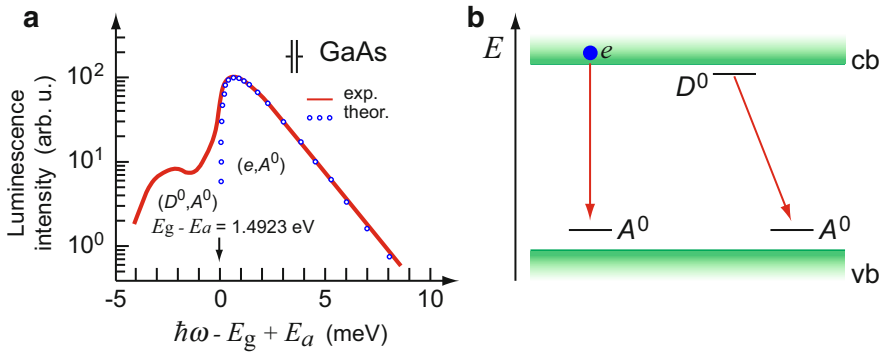


Fig. 15 (**a**) Luminescence spectrum for transitions from the conduction band (e, A^0) and from a donor (D^0, A^0) into a hydrogen-like acceptor in GaAs at 1.9 K. *Circles:* theoretical line shape for $T_e = 14.4$ K (After Ulbrich 1978). (**b**) Band model with corresponding transitions

also contains a lower-energy luminescent feature labeled (D^0, A^0), which is caused by the recombination of electrons from a hydrogen-like donor with a nearby acceptor.

1.7 Higher Charged Coulomb Centers and Metal-Ion Interstitials

Higher Charged Coulomb Centers With higher charges, a Coulomb-attractive center has its eigenfunctions closer to the core and requires more attention to core correction. Therefore, its ground state is more akin to deep-level centers – see ▶ Chap. 19, "Deep-Level Centers".

Higher charged centers can be created by substitutional impurities. These are further away from the group of the replaced element, and act similarly to a hydrogen-like donor or acceptor, except that more electrons or holes are donated. For instance, if in an Si lattice one of its atoms is replaced by a sulfur, selenium, or tellurium atom, *two* electrons can be donated. Since the first electron is bound to a doubly charged center, it is bound at a level four times as deep (charge $Z = 2$); the ionization energy is given by

$$E_{\mathrm{I}} = E_{\mathrm{c}} - \frac{Z^2\, e^4}{2(4\pi\varepsilon_0\hbar)^2}\ \frac{m_n}{\varepsilon_{\mathrm{stat}}^2} = R_\infty \frac{Z^2}{\varepsilon_{\mathrm{stat}}^2}\ \frac{m_n}{m_0} \tag{30}$$

The second electron behaves like an electron attached to an ordinary hydrogen-like donor. Hence the substitutional sulfur can be represented by a double donor with two levels (Fig. 16). Occupancy, however, determines which of the two levels is active: when filled with two electrons, only the shallow level is active; ionized once, the other electron becomes more strongly bound and the deeper level is active. From

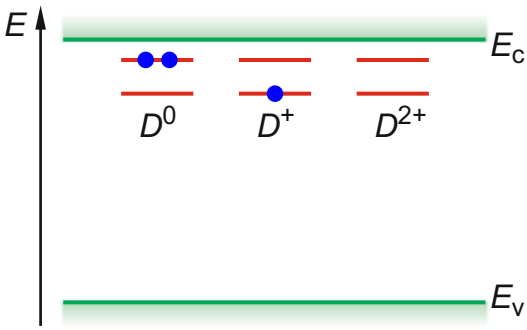


Fig. 16 Double donor (e.g., S in Si) with corresponding charge character relative to the neutral lattice

the quasi-hydrogen model, we estimate ionization energies of ~50 meV and ~120 meV for the second donor level in Ge and Si, respectively, compared to 10 or 32 meV for the first level. The actual energies for S in Si are 302 and 587 meV for the first and second ionization levels (Grimmeiss et al. 1980). The larger energies indicate tight binding, which makes the hydrogen-like approximation less accurate and requires central-cell potential consideration. We will therefore return to these centers in ▶ Chap. 19, "Deep-Level Centers".

A similar behavior is expected and observed for two-level acceptors from substitutional group II elements, e.g., Zn or Cd in Ge or Si with a second acceptor level at ~25 or ~210 meV, respectively, compared to 10 or 53 meV for the first level. Elements from groups with larger valence difference are more difficult to implant as substitutionals. Anions are often too large for the host lattice; cations tend to become more easily incorporated as interstitials (see below).

In an *AB compound*, the situation is more complex, as illustrated by incorporating a group IV element such as Sn (which is two groups away from Cd) as a substitutional into CdS. Replacing the Cd ion, it acts as a two-level donor; replacing an S ion, it acts as a two-level acceptor. This ambiguity makes it difficult to predict the behavior of such a type of dopants without additional information. In ZnO, the elements Si, Ge, and C form double donors in most stable 2+ charge state on Zn site with large inward relaxations, while on O or interstitial sites unfavorably high formation energies were found (Lyons et al. 2009, 2014).

Overcharged Donors or Acceptors One observes doubly charged normal donors (or acceptors) when the neutral donor (or acceptor) can trap an additional electron (or hole). These are known as D^- or A^+ centers and can be compared to an H^- ion (Faulkner 1969).

The binding energy of the additional carrier is very small (0.54 and 1.7 meV for hydrogen-like impurities in Ge and Si, respectively; see Lampert 1958) and results in a very large radius of the quasi-hydrogen eigenfunction, requiring high purity to avoid complications due to overlap.

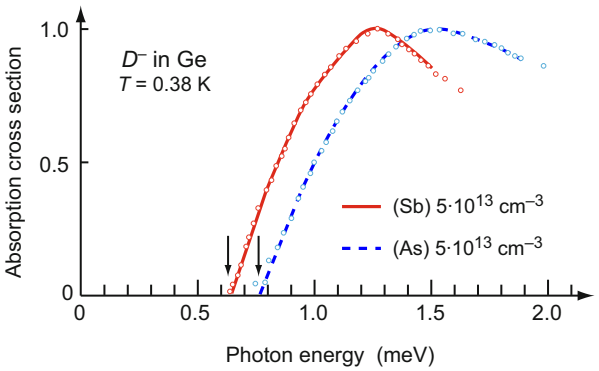


Fig. 17 Photoionization cross section for D^- states in ultrapure Ge with As or Sb doping at 0.38 K. The low-energy threshold indicates the binding energy of the D^- center (After Narita 1985)

The ionization energy can be measured from the photoconduction threshold (see Fig. 17) and shows a chemical shift for As_{Ge}^- and Sb_{Ge}^-. Anisotropy and multivalley effects are responsible for the deviations from the simple H^--ion model estimates.

Another type of center develops when a doubly charged donor (or acceptor), as described in the previous section, traps another electron (or hole); it becomes overcharged and binds three carriers. Because of the Pauli principle, this center does not have an isolated atom (such as He^-) as an analogue, which exists only in the metastable $(1s)(2s)(2p)$ state. On the other hand, the overcharged $Z = 2$ center can exist in a $(1s)^3$ binding state (McMurray 1985). The ground-state energy of such a center requires central-cell consideration as well as radial and angular correlation between the trapped carriers.

Metal-Ion Interstitials Cation interstitials, either extrinsic or intrinsic, usually behave as donors. An example was given in Sect. 1.2 (Fig. 7) with Li as an interstitial in Si – see Reiss et al. (1956) and Haller et al. (1981). The metal atom on an interstitial position prefers to donate its valence electron(s). When it has no counterpart to form a charge-compensating bond, it can be described as a hydrogen-like donor. Since ionized cations are usually much smaller than anions, they are more easily incorporated on interstitial sites. Therefore, interstitial donors are more readily observed than interstitial acceptors.

The ground-state level of the interstitial depends on its site of incorporation, see, e.g., Fig. 5a of ▶ Chap. 15, "Crystal Defects", and Jansen and Sankey (1987); for interstitials in II-VI compounds, see Watkins (1977). Depending on the valency of the incorporated metal ion, they can act as single- or multilevel donors.

In a simple hydrogen-like model, higher excited states of these impurities cannot be distinguished from the classical substitutional donor. The chemical shift of the ground state, however, is substantial. In a pseudopotential approximation (Pantelides 1975), a complication arises since there is no cancellation, such as for the potential of a replaced host atom in a substitutional impurity. Self-interstitials in elemental semiconductors are rather deep centers and will be discussed in Sect. 2.3 of ▶ Chap. 19, "Deep-Level Centers".

2 Excitons Bound to Impurity Centers

When a neutral impurity is incorporated, it becomes charged when it traps a carrier. This now-charged defect acts as a Coulomb-attractive center, which can in turn trap a carrier of the opposite sign and thereby form a quasi-hydrogen state (Hopfield et al. 1966). This state can be described as an exciton (see Sect. 1.2 of ▶ Chap. 14, "Excitons") bound to a neutral defect center.

The binding energy of such a center (also referred to as localization energy) is the difference between the free-exciton energy and the bonding energy of the exciton to this center, which is typically on the order of 10 meV (Faulkner 1968); in wide-bandgap semiconductors such as GaN, the binding energy may be significantly larger. There are numerous neutral centers to which excitons can be bound, such as deep centers (e.g., isoelectronic defects) or shallow centers like neutral hydrogen-like donors or acceptors. We will briefly discuss both excitons bound to neutral or to ionized donors or acceptors in the following sections.

2.1 Excitons Bound to Ionized Donors or Acceptors

Binding to neutral centers at first view shows some similarities to the binding at deep-level centers, although the bonding mechanism is different. While the first carrier binds to a deep center entirely through short-range interaction, consequently attracting an oppositely charged carrier into a quasi-hydrogen orbit, here both carriers (the exciton) are attracted simultaneously into an H_2-like quasi-hydrogen state. In some respects, this state may be compared to a trion, i.e., two electrons and a hole attracted to an ionized donor or two holes and an electron attracted to an ionized acceptor – see Sect. 1.4 of ▶ Chap. 14, "Excitons". For a review, see Dean (1983) and Pajot (2009).

It is relatively easy to calculate the bonding of an exciton to an ionized donor. The Hamiltonian for such a (D^+, X) center can be written as

$$E_{\mathrm{b,exc}} = -\frac{\hbar^2}{2m_p}\frac{\partial^2}{\partial \mathbf{r}_p^2} - \frac{\hbar^2}{2m_n}\frac{\partial^2}{\partial \mathbf{r}_n^2} + \frac{e^2}{4\pi\,\varepsilon_{\mathrm{stat}}\varepsilon_0\,r_p} - \frac{e^2}{4\pi\,\varepsilon_{\mathrm{stat}}\varepsilon_0\,r_n} - \frac{e^2}{4\pi\,\varepsilon_{\mathrm{stat}}\varepsilon_0\,r_{np}}, \quad (31)$$

where r_p and r_n represent the distances of the trapped hole and electron from the donor and r_{np} is their relative distance. The energy of the bound exciton is sensitive to the effective mass ratio m_n/m_p, see Fig. 18. Above a critical value of ~0.45, no bound exciton states are observed: the kinetic energy of the hole then becomes too large to be bound to the neutral donor. Such (D^+, X) states therefore exist in CdS but are absent in Si and GaP, where the mass ratio is larger than 0.45 (Rotenberg and Stein 1969; see also Dean and Herbert 1979). Luminescence lines due to the recombination of (D^+, X) complexes are found in various II-VI semiconductors and commonly labeled I_3, and those of (D^0, X) complexes are labeled I_2.

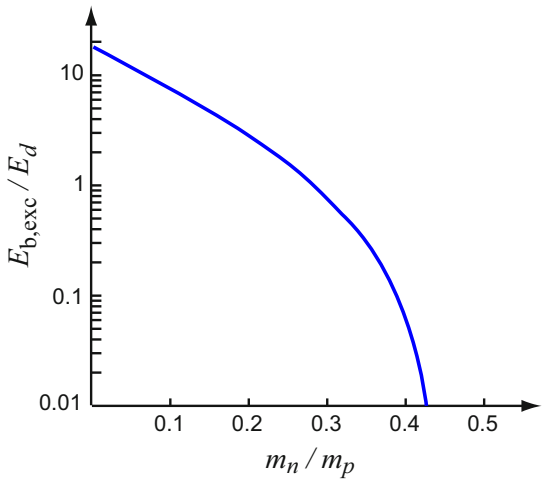


Fig. 18 Calculated binding energy $E_{b,exc}$ for an exciton bound to an ionized donor, normalized to the ionization energy E_d of the donor, as a function of the mass ratio m_n/m_p (After Skettrup et al. 1971)

For an ionized acceptor, the mass-ratio condition is inverted and therefore is not fulfilled for most semiconductors. Consequently, excitons are usually not bound to ionized acceptors.

2.2 Excitons Bound to Neutral Donors or Acceptors

The bound exciton states on neutral acceptors (A^0, X) or on neutral donors (D^0, X) are not limited by the effective-mass ratio. Their binding energy is proportional to the ionization energy of donors or acceptors, as shown in Fig. 19 – see Dean 1973b. This empirical relation is referred to as Hayne's rule (see, e.g., Halsted and Aven 1965) and is given by (Halsted 1967)

$$E_{b,exc} = a + bE_a \quad \text{or} \quad E_{b,exc} = c + dE_d. \tag{32}$$

In GaP the value of a is positive, while c is negative with $b \cong 0.1$ and $d \cong 0.2$. The intersects are related to the increase of the relative charge in the central cell when the exciton is bound; consequently, a and b are close to zero in Si (Dean 1983; Haynes 1960).

When calculated in an H_2-molecule approximation, the binding energies of (D^0, X) or (A^0, X) complexes are about 30% of the ionization energy of the corresponding donors or acceptors. These energies are listed in Table 5 for a number of II-IV compounds. It should be noted that also shallow donor *complexes* can localize excitons obeying Hayne's rule, as shown, e.g., for complexes of the double donor Mg with group-III acceptors Ga, In, and Al in silicon (Abraham et al. 2019).

In addition to the binding energy of such excitons, lines corresponding to the excited states of the donors or acceptors, as suggested by Thomas and Hopfield (1962), can be observed (Thewalt et al. 1985).

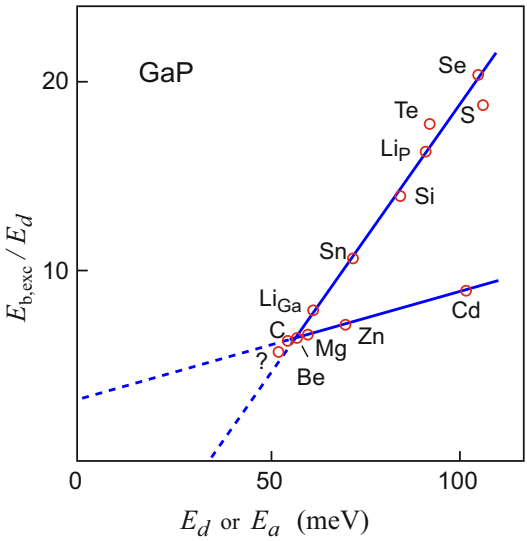


Fig. 19 Binding energy of excitons to neutral donors or acceptors in GaP as a function of the ionization energy of single donors or acceptors (After Dean 1983)

Table 5 Bound-exciton lines with transition energies E_{BE} and binding energies $E_{b,exc}$ of bound excitons in II-IV compounds measured at low temperatures (typ. 1.7–4.2 K) (After Taguchi and Ray 1983)

Semiconductor	E_{BE} (eV)	$E_{b,exc}$ (meV)	Assignment	Energy level
ZnS	3.758	34.4	(A^0, X)	$E_v + 1.22$ eV
	3.724	68.3		
ZnSe	2.799	2–3	(D^0, X); V^0_{Se}	$E_c - 0.02$ eV
	2.780	19–22	(A^0, X); V^0_{Zn}	$E_v + 0.28$ eV
			V^{2-}_{Zn}	$E_v + 1.1$ eV
ZnTe	2.377	4	(D^0, X)	
	2.375	6	(A^0, X); V^0_{Zn}	$E_v + 0.06$ eV
	2.362	19	(D^+, X); V^+_{Te}	$E_c - 0.03$ eV
CdS	2.547	5	(D^0, X)	
	2.536	16	(A^0, X); V^0_{Cd}	$E_v + 0.8$ eV
CdSe	1.822	4	(D^0, X); Cd^0_i	
	1.817	9	(A^0, X); Se^0_i	
CdTe	1.594	1	(D^0, X); V^0_{Te}	$E_c - 0.02$ eV
	1.590	4	(A^0, X); V^0_{Cd}	$E_v + 0.06$ eV

Excitons bound to hydrogen-like donors or acceptors are known to have high luminescence efficiencies. They show extremely sharp luminescence lines since the kinetic energy contribution, which causes the broadening for the free exciton, vanishes. There is a rich spectrum in high-quality single crystals involving transitions from the ground or excited state of the bound exciton leaving the state of the donor or acceptor unchanged or with a final state of these impurities at $n = 2, 3....$ There is an additional structure to the luminescence due to the chemical

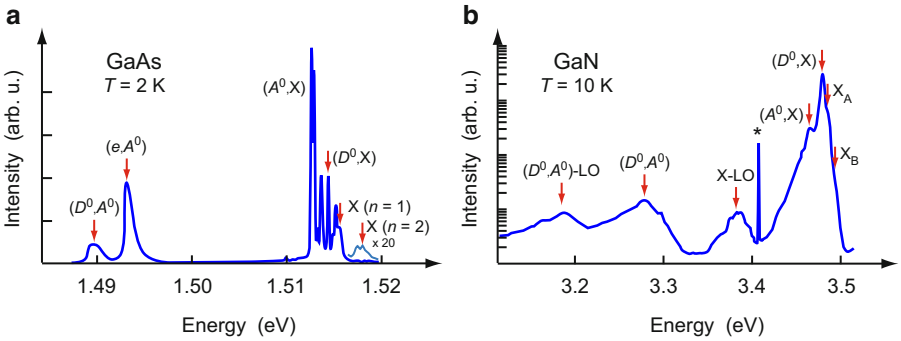


Fig. 20 Photoluminescence spectra for (**a**) GaAs and (**b**) GaN, recorded at low temperature. Labels denote typical emission lines near the band edge: free-exciton recombination X, donor- and acceptor-bound exciton (D^0, X) and (A^0, X), conduction-band-to-acceptor (free-to-bound) transition (e, A^0), and donor-acceptor pair recombination (D^0, A^0); the asterisk in (**b**) labels a plasma line of the excitation laser (Spectra adapted from Ulbrich (1978) and Cojocari et al. (2004))

individuality, i.e., the central-cell potential that shifts the 1s state. This effect is more pronounced for the acceptor than for the donor, because of the larger effective mass of the hole, which renders the acceptor a deeper center.

The intensity of bound-exciton lines is usually very large, since their line width is narrow and they have very large oscillator strength (Fig. 20). The oscillatory strength is proportional to the volume of the ground-state eigenfunction, typically to a_{qH}^3. Consequently, a giant oscillatory strength is found for materials of low m^* and high $\varepsilon_{\mathrm{stat}}$ (Rashba and Gurgenishvilli 1962).

2.2.1 Two-Electron and Two-Hole Transitions

In many semiconductors like GaAs or InP, the energy of shallow donors and acceptors does not vary much; identification of the chemical nature of an impurity may then not be inferred from the bound-exciton recombination energy, since, e.g., strain in the samples leads to comparable energy differences. More reliable assignments are obtained from energy differences of levels within a given bound-exciton complex; such (usually weak) transitions are two-electron transitions (TET) of (D^0, X) complexes and correspondingly two-hole transitions (THT) of (A^0, X) complexes. Radiative recombination of two-electron transitions leaves the donor in an excited state D^{0*} (Dean et al. 1967), see Fig. 21. The energy difference between the (D^0, X) 1S ground state of the donor-bound exciton complex and the final state of the two-electron transition yields the difference between the D^0 1s ground state and the D^{0*} 2s excited state of the donor. This difference is sensitive to the central-cell potential of the donor. The term two-electron transition indicates that the bound exciton recombines simultaneously producing an excitation of the second electron in the bound-exciton complex. Analogous conditions apply for acceptors. In the particle models of Fig. 21, the large charge represents the impurity and the small charges electrons and holes.

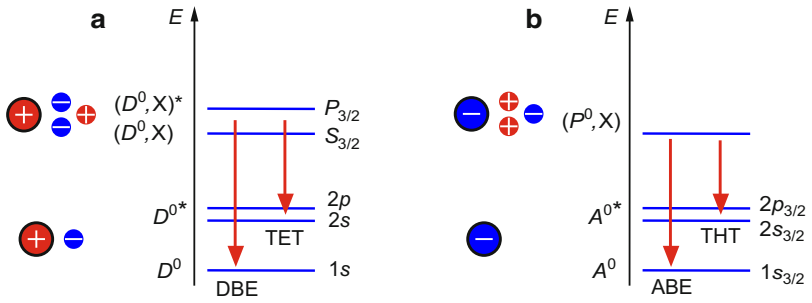


Fig. 21 Particle model and energy-level scheme for (**a**) two-electron transitions (TET) in the recombination of a donor-bound exciton (DBE) complex (D^0, X) and (**b**) two-hole transitions (THT) of an acceptor-bound exciton (ABE) complex (A^0, X). The stars indicate the excited state of the impurity center

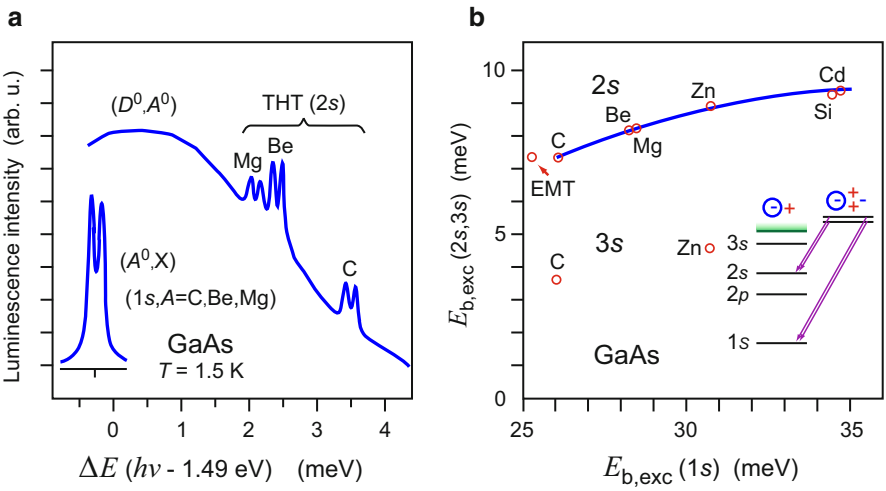


Fig. 22 Two-hole transitions of acceptor-bound excitons in GaAs. (**a**) Photoluminescence of the principal bound-exciton lines due to C, Be, and Mg acceptors appearing at the same energy (*inset*, 1*s*) and corresponding well-separated two-hole transitions (THT) that leave the acceptor in the 2*s* state, superimposed on the broad (D^0, A^0) donor-acceptor pair emission. (**b**) 2*s* and 3*s* excited state-exciton binding energy depending on the 1*s* acceptor ground state for various acceptors in GaAs, determined from two-hole transitions; EMT denotes the effective-mass binding energy (After Ashen et al. 1975)

The luminescence of GaAs shown in Fig. 22 illustrates unresolved differences in the principal (A^0, X) emission of excitons bound to different acceptors, despite their significantly different acceptor energies E_a. In contrast, the two-hole transitions appearing at the high-energy side of the broad donor-acceptor pair luminescence are well separated.

2.3 Excitons Bound to Isoelectronic Centers

These centers have a central core potential term but no Coulomb field for binding of the exciton. There is a large variety of such centers, for instance, GaP:N_P.

The exciton bound to such centers is referred to as an isoelectronically bound exciton (Hopfield et al. 1966): an attractive core potential of the center tightly binds an electron which, in turn, can bind a hole into a quasi-hydrogen orbital. This electron-hole pair is equivalent to an exciton bound to the center, while it acts as an acceptor because of its asymmetry in carrier binding. Similar defect centers can be expected when first attracting a hole into a deep neutral center and then an electron into the resulting Coulomb-attractive defect, which would act as a donor-like defect center with the bound exciton. For a bound-exciton model of isoelectronic centers in Si, see Davis and Nazaré (1994).

Isoelectronic centers are of interest because they show a high luminescence efficiency at room temperature. This is caused by the fact that these centers, produced by neutral atoms of the same valency as the replaced lattice atoms, extend to rather large k values due to their short-range potential. This permits electron transitions from $k \neq 0$ without phonon assistance, an important feature for indirect bandgap materials.

Examples of isoelectronic centers are N, As, or Bi in GaP (Thomas et al. 1965; Trumbore et al. 1966; Dean et al. 1969), N pairs in GaAs (Kita and Wada 2006; St-Jean et al. 2015), O in ZnTe (Dietz et al. 1962; Hopfield et al. 1966 and Merz 1968) and Te pairs in ZnSe (St-Jean et al. 2016), or Te in CdS (Roessler 1970). For reviews, see Bergh and Dean (1976) or Craford and Holonyak (1976). At first view, the isoelectronic impurity, which replaces a lattice atom, seems to be a relatively simple defect; in its incorporation into the lattice, it resembles a shallow donor or acceptor. An estimate of the binding energy, however, is more difficult (Baldereschi and Hopfield 1972). The high luminescence efficiency is suggested by an efficient capture of an exciton at such a center and a high radiative recombination probability of this exciton. The exciton capture may be a two-step process by which the impurity first captures an electron (or hole) if it is a donor (or acceptor)-like defect and consequently, now charged, captures the oppositely charged carrier (Hopfield 1967).

The energy of radiative recombination from isoelectronic centers lies close to the band edge and, at low temperatures, can be resolved as originating from two different exciton states, e.g., with $J = 1$ and $J = 2$ for emission lines A and B, respectively, in GaP:N_P as shown in Fig. 23.

In Fig. 23 pronounced replica of the primary luminescence (A, B) of excitons trapped at isoelectronic centers in GaP are visible; they are identified as A-TO, A-LO, and B-LO. These strong lines indicate that radiative transitions occur with phonon assistance, i.e., after LO or TO phonon emission, induced by strong lattice coupling with a large Huang-Rhys factor (see Sect. 2.3.2 of ▶ Chap. 17, "Optical Properties of Defects"). The phonon interaction with such isoelectronic centers is so strong because of a severe lattice deformation that surrounds the impurity and provides

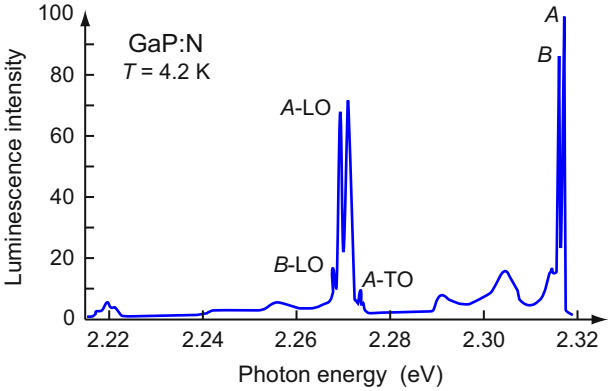


Fig. 23 Luminescence spectrum from GaP:N with $5 \cdot 10^{16}$ cm^{-3} N_P at 4.2 K, identified as an exciton transition bound at the isoelectronic center. The *lines* identified with LO and TO are phonon replicas (After Thomas and Hopfield 1966)

an efficient coupling. The excitation spectrum shows similar phonon replica and is almost mirror symmetric to the luminescence spectrum except for slightly different phonon energies.

At higher nitrogen concentration ($> 10^{18}$ cm^{-3}), many lines labeled NN_i appear at lower energy, the deepest (NN_1) being 0.143 eV below the intrinsic exciton, see Fig. 24. The concentration of the *NN* centers varies as the square of the *A* center concentration,[7] indicating that the *NN* centers consist of two nitrogen atoms; the NN_1 center arises from two substitutional N atoms on the closest possible P sites, the NN_2 center from two nitrogen on the second nearest sites, and so on (Thomas and Hopfield 1966; Cohen and Sturge 1977): the eigenstates of the exciton, bound to such isoelectronic pairs, depends on the intrapair distance. Such spectral series due to pairs are also observed for donor-acceptor pairs discussed in more detail in Sect. 1.6.

Another example of a defect pair that can bind an exciton is the Be-Be center in Si. Here, two Be atoms replace one Si atom and produce an axial field in ⟨111⟩ direction that causes a characteristic splitting of the ground state of the center, helping in identifying states of the same center. The basic line structure of excited states of this bound exciton behaves like that of an acceptor (e.g., B in Si), but it is redshifted by the binding energy of the exciton (3 meV); for more details, see Lipari and Baldereschi (1978), Thewalt et al. (1985), and Ishikawa et al. (2010).

[7]The concentration of the *A* center and the *NN* centers is proportional to their absorption strengths.

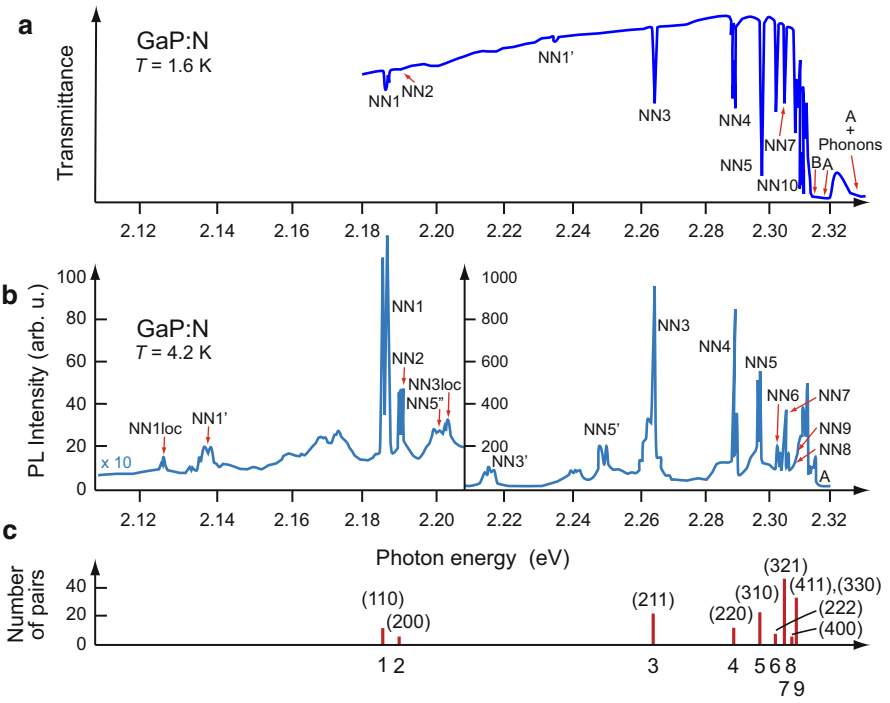


Fig. 24 (**a**) Absorption and (**b**) luminescence spectrum of GaP doped with ~19 cm^{-3} nitrogen, forming N-N pairs, and measured at low temperature. The prime symbols identify phonon replicas, and LOC identifies local modes. (**c**) Expected numbers of pairs for a random distribution of N in GaP and next-neighbor distances corresponding to the respective pair lines above in the spectra; *labels* indicate the orientation of the pairs (After Thomas and Hopfield 1966)

3 Influence of External Fields on Defect Levels

Different kinds of fields may act on a considered defect. As to mechanical stress fields, we distinguish between hydrostatic stress, which leaves the symmetry of the lattice unchanged, and uniaxial and shear stresses, which induce changes of the lattice symmetry; stress occurs also in epitaxial layers and may have gradients in case of partial relaxation. Electric fields can be external, usually homogeneous fields, and built-in fields which are space-charge induced and show gradients. External magnetic fields again are homogeneous, while such fields exerted from spin or orbit coupling contain inhomogeneous components on the defect atomic scale. We will discuss the influence of these fields on shallow-level centers in the following sections.

Although the influence of electric and magnetic fields is much larger in shallow levels than in deep levels due to the large volume occupied by their eigenfunctions, there are also substantial influences on some deep-level centers. These, however, are

more complex in nature. Furthermore, some deep centers can also show rather extended wavefunctions, e.g., the Al center in Si which extends 70% up to the seventh neighbor shell.

3.1 Influence of Hydrostatic Pressure

Hydrostatic pressure does not influence the impurity potential appreciably, nor does it change the position of deeper bands. There is a much smaller shift of the valence band than for the conduction band for most semiconductors.

With increasing pressure a decreasing lattice constant causes characteristic changes, which are most pronounced for the conduction bands and are observable by a change in the bandgap energies. In cubic semiconductors, the energy separation from valence to conduction band increases at Γ and *L* (see Fig. 25 for GaAs) but decreases at *X*. These shifts occur with a similar pressure coefficient of approximately 10 μeV/bar in column IV and III–V semiconductors. For cubic GaN and AlN, pressure coefficients of ~4 μeV/bar for the Γ conduction band and significantly smaller values for defects were calculated using Green's functions (Gorczyca et al. 1997); respective tight-binding calculations are reported in Jancu et al. (1998). For a general review on pressure effects, see Goñi and Syassen (1998), and for recent first-principles calculations, see Freysoldt et al. (2014).

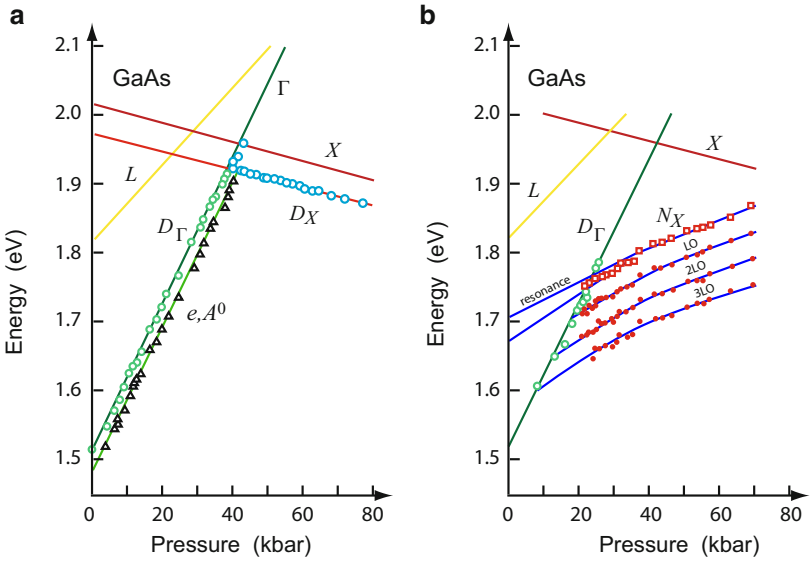


Fig. 25 Luminescence of GaAs at 5 K. (**a**) Shallow defect levels in GaAs are connected to the lowest band (Γ below and *X* above 41 kbar); D_Γ and D_X are excitons bound to donors, and eA^0 is the band-to-acceptor luminescence emission. (**b**) Deep N levels in GaAs labeled N_X are not connected with the lowest conduction band. Shown also are phonon replicas (LO … 3LO) and resonance states of the N levels below 22 kbar within the Γ continuum (After Wolford 1987)

Shallow donor states shift together with the conduction bands and can be identified as such by their connection to respective bands with an identical shift. Such a shift can be observed by the recombination luminescence of excitons, which are bound to such donors and shown in Fig. 25a. Changes of the connection from the Γ band to the *X* band are observed near the crossover pressure (~41 kbar for GaAs). These bound-exciton lines are given here to exemplify shallow defects. Free excitons and conduction-band-to-acceptor luminescence behave similarly.

Near the crossover pressure, the Γ-related shallow levels can become resonant with the X_1 minima. This can be observed by a large increase in carrier scattering to resonance scattering when the X_1 minimum is shifted to coincide with the donor level. At this pressure a minimum in the carrier mobility is observed (Kosicki and Paul 1966). Other effects occur when, for example, in Ge the minimum goes from L_1 to Δ_1 at $p \cong 50$ kbar. The degeneracy is increased from four- to tenfold and is decreased to sixfold when the Δ_1 minimum is lowered below L_1. The consequence is a splitting from two sublevels (E_{Γ_1} and $E_{\Gamma_{15}}$) into five and then back to three sublevels – E_Γ, $E_{\Gamma_{12}}$, and $E_{\Gamma_{15}}$.

Deeper donor levels show a lesser influence from hydrostatic pressure since they are more dependent on the core potential, which is little changed. Deep levels also are connected to several bands which tend to reduce and partially cancel their individual contributions. This can be seen in nitrogen-doped GaAs at pressures in excess of 20 kbar, as shown in Fig. 25b (Wolford et al. 1979). Below ~20 kbar the N levels become resonant within the Γ band and can no longer be observed by photoluminescence. A theoretical estimate of the shift of deep levels in GaAs was given by Ren et al. (1982) and is shown in Fig. 26. The pressure derivatives *dE/dP* of the levels are significantly smaller than that of the fundamental bandgap dE_g/dP =12.6 meV/kbar.

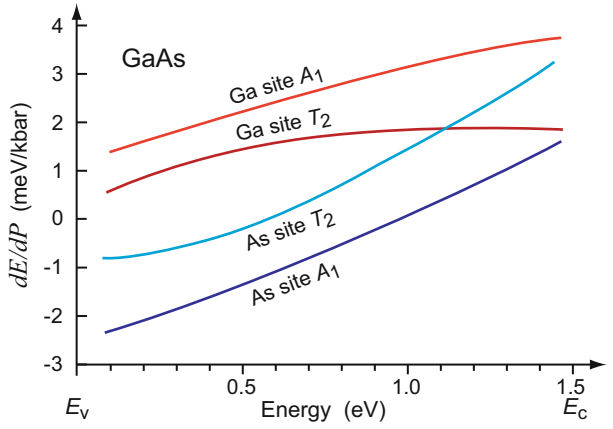


Fig. 26 Calculated pressure dependence of deep levels in GaAs as a function of their energies in the bandgap. A_1- and T_2-type levels are *s*-like pulled down from the conduction band (antibonding) and *p*-type pushed up from the valence band (bonding)

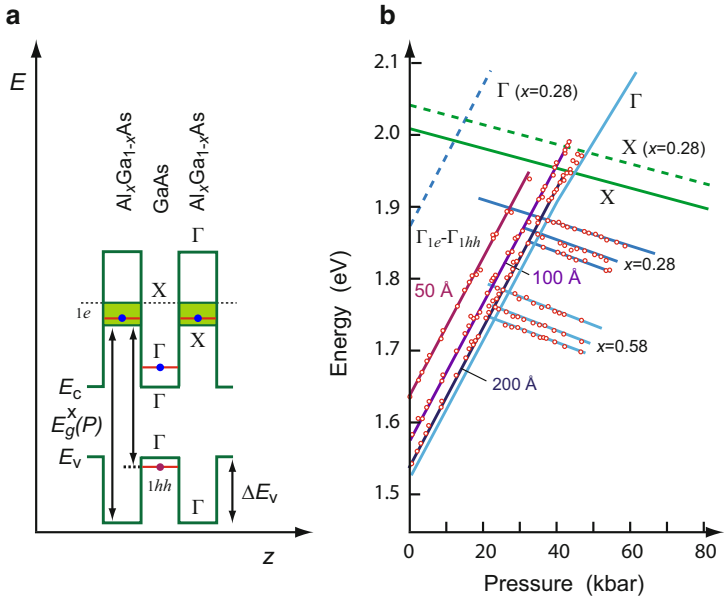


Fig. 27 (**a**) Schematics of a GaAs/Al$_x$Ga$_{1-x}$As superlattice structure with transitions from Γ and *X* bands indicated. (**b**) The photoluminescence follows the pressure dependence of the direct bandgap for the Γ transition, although shifted toward higher energies due to confinement. Above the crossover pressure Γ → *X*, indirect transitions are shown with phonon replica for the superlattice structure. These transitions occur from *X* electrons inside the barrier to Γ holes within the well (see subfigure **a**) (After Wolford 1987)

Acceptors are insignificantly influenced by hydrostatic pressure, as it has a lesser influence on the valence-band structure near $k = 0$. A more involved effect was recently reported for Be in GaN that gives rise to a yellow luminescence, which is replaced by an UV emission at high pressure. Hybrid functional calculations indicate that the yellow emission originates from the transition of a deep (Be_{Ga}–O_N) donor complex with a strong anisotropic relaxation, which disappears a high pressure; the emerging UV emission is explained by the resulting shallow donor complex (Teisseyre et al. 2017).

Quantum-well states and **superlattice states** show the connection of direct transitions with the Γ band in GaAs/Al$_x$Ga$_{1-x}$As heterostructures (Fig. 27). Above the crossover pressure, spatially indirect transitions are observed in a superlattice (Fig. 27a), including phonon replica, which now shift parallel to the *X* band with pressure and composition (Fig. 27b); in a GaAs/Al$_x$Ga$_{1-x}$As quantum well confined conduction-band states exist only in GaAs and consequently only direct Γ → Γ transitions. For more information, see Wolford (1987).

Table 6 Removal of degeneracies between equivalent critical points and donor states (s and p_0 envelope functions) under uniaxial stress, yielding new symmetries of the resulting deformed valleys (After Bassani et al. 1974)

Conduction-band minima at	Cubic crystal (T_d group)	[100] Stress (D_{2d} group)	[111] Stress (C_{3v} group)	[110] Stress (C_{2v} group)
(100)	$\Gamma_1 + \Gamma_{12} + \Gamma_{15}$(Si)	$X_1 + X_3$	$2\Lambda_1 + 2\Lambda_3$	$\Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4$
$(\bar{1}00)$				
(010)		$X_1 + X_2 + X_5$		
$(0\bar{1}0)$				
(001)				$2\Sigma_1$
$(00\bar{1})$				
$\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$	$\Gamma_1 + \Gamma_{15}$(Ge)	$X_1 + X_3 + X_5$	Λ_1	$\Sigma_1 + \Sigma_3$
$\left(\frac{\bar{1}}{2}, \frac{1}{2}, \frac{1}{2}\right)$			$\Lambda_1 + \Lambda_3$	
$\left(\frac{1}{2}, \frac{\bar{1}}{2}, \frac{1}{2}\right)$				$\Sigma_1 + \Sigma_4$
$\left(\frac{1}{2}, \frac{1}{2}, \frac{\bar{1}}{2}\right)$				

3.2 Influence of Uniaxial Stress

Uniaxial stress lowers the symmetry of the semiconductor, depending on the relative direction of the stress with respect to the crystal axes. Consequently, the band structure becomes more complicated. The band degeneracies are removed, producing a splitting and a shift of the bands, as indicated for the six, without stress, equivalent valleys in Si and for the four valleys in Ge in Table 6. The shift is proportional to the stress, with the deformation potentials Ξ_{dil} and Ξ_{shear} for dilatation (subscript dil) in the direction normal to the main axes of the sidevalley ellipsoids and for uniaxial shear (subscript shear). This results in stretching along the main axes and compression in the two normal directions – see Fritzsche (1962), Pollak (1965), and Cardona (1969a, b). The deformation potentials are proportionality factors for the energy shift which, in cubic materials, is given by

$$\Delta E_{(j)} = \sum_{\alpha,\beta} \left(\Xi_{\text{dil}}\delta_{\alpha\beta} + \Xi_{\text{shear}}\tilde{k}_\alpha\tilde{k}_\beta\right)\ u_{\alpha\beta}, \tag{33}$$

with the strain components $u_{\alpha\beta}$ (Herring and Vogt 1956). Here $\delta_{\alpha\beta}$ is the Kronecker δ symbol, and $\tilde{k}_{\alpha\beta}$ are the components of the unit $\mathbf{k}$ vector on the α and β axes. Uniaxial stress in the [111] direction for Si and in the [100] direction for Ge, however, results in no splitting, since the symmetry of the valleys remains unchanged under such a stress – see Table 6.

Shallow impurities relating to these bands split accordingly. In Si, for instance, the $1s$ donor state, which is split into A_1, T_2, and E (Γ_1, Γ_{12}, and Γ_{15}) because of intervalley interaction, splits further when uniaxial stress is applied in [100] direction. The Γ_{12} and Γ_{15} levels split and the Γ_1 level shifts. The splitting is proportional to the stress, and the

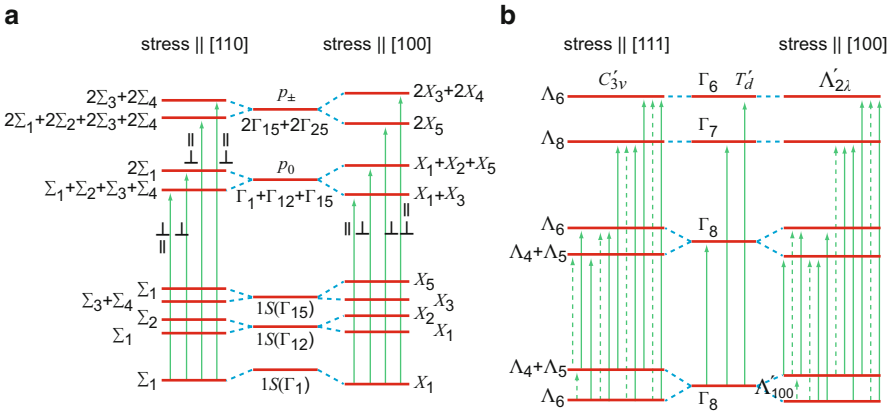


Fig. 28 Schematic representation of shifts and splitting of (**a**) donor and (**b**) acceptor states in Si under uniaxial stress. *Dashed lines* are transitions for **E** parallel to the stress axis, and *solid lines* are for **E** perpendicular to the stress axis (After Bassani et al. 1974)

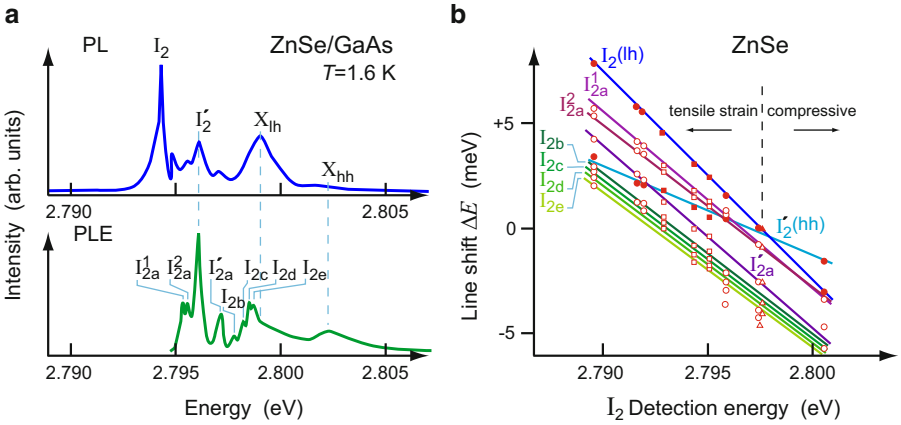


Fig. 29 (**a**) Photoluminescence (*top*) and I_2 excitation spectrum (*bottom*) of the donor-bound exciton (lines I_2 and I_2') in a strained ZnSe/GaAs layer; X_{lh} and X_{hh} denote light- and heavy-hole excitons; lines labeled I_{2a}, I_{2b}, etc., refer to excited states of the donor-bound excitons. (**b**) Strain-induced energy shifts of the I_2 lines in differently strained ZnSe/GaAs layers measured at 1.6 K; the I_2 detection energy is proportional to the lateral strain. *Squares* and *circles* refer to layers on substrate and free-standing layers with removed substrate, respectively (After Kudlek et al. 1992)

shifting, as a second-order effect, is proportional to the square of the stress. Splittings and shifts are shown in Fig. 28a; see also Wilson and Feher (1961).

The effect of uniaxial stress is similar to that of biaxial stress exerted on epitaxial layers with some lattice mismatch to the substrate. Figure 29 shows the splitting and shift of the (D^0, X) recombination (lines I_2, I_2') in ZnSe/GaAs layers (Kudlek

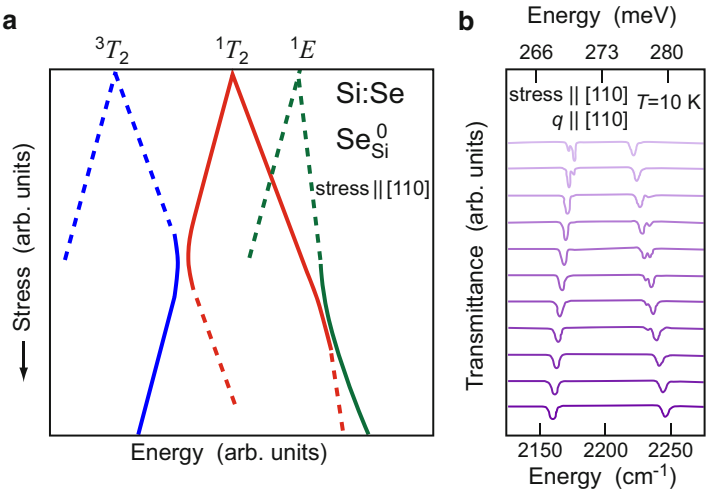


Fig. 30 (**a**) Schematic energy diagram of the Se donor in Si (solid curves: observed). (**b**) Transmission spectra for increasing (from *top* to *bottom*) uniaxial [110] stress as family parameter with lower- and upper-branch $^1T_2$ transitions allowed at low stress, changing to $^3T_2$ and ^1E transitions, respectively, at higher stress (After Bergman et al. 1987)

et al. 1992). ZnSe is biaxially strained in the interface plane to GaAs due to a 0.3% larger lattice constant, and layer growth at various temperatures leads to a different degree of partial plastic strain relaxation in the layers. The free exciton X consequently splits with different spacing into a light- and heavy-hole excitons and related (D^0, X_{lh}) and (D^0, X_{hh}) lines I_2 and I_2', which have different pressure coefficients. At the cryogenic measuring temperature, the ZnSe layers are tensely strained, since the thermal expansion coefficient of ZnSe is larger than that of the GaAs substrate.

The influence of uniaxial stress on **acceptor states** is determined by the degeneracy of the valence bands at $\mathbf{k} = 0$. The states relating to Γ_6 and Γ_7 are shifted, and the Γ_8 state is shifted and split. The higher excited states are influenced accordingly (Rodriguez et al. 1972). The transitions are polarization dependent and are shown in Fig. 28b. Such changes of the level spectrum can be observed in optical absorption or luminescence and are helpful in identifying the levels with respect to their symmetry. For a review, see Rodriguez et al. (1972).

An example of the shifting and splitting of levels of the Se donor in Si (see also Sect. 1.7) is shown in Fig. 30 for the transmission spectrum. The allowed $^1T_2$ level splits with increasing stress and avoids crossing with the ^1E and $^3T_2$ levels, which are forbidden at low stress and become allowed with increasing linear stress in the [110] direction (Bergman et al. 1987).

3.3 Influence of an Electric Field

An electric field acts in a manner similar to uniaxial stress by a shifting and splitting of the defect levels due to a lowering of the crystal symmetry. Although the changes are similar, they are smaller than those described in Sect. 3.2 for usual experimental conditions. The ground state typically shifts by 100 μeV for fields of 10 kV/cm; the shift of the first excited state for the same field is on the order of 1 meV. The effect is similar to the atomic *Stark effect*.

A splitting of degenerate eigenstates proportional to the applied homogeneous field was observed by Stark (1914) in hydrogen (*linear Stark effect*). This effect is due to the superposition of a perturbation term $e\mathbf{F} \cdot \mathbf{r}$ in addition to the Coulomb potential $\propto e^2/r$. A comprehensive description of the classical effect can be found in Herzberg (1937) or Sommerfeld (1950). Because of the larger quasi-hydrogen radius of shallow defect levels, the corresponding splitting within a semiconductor takes place at much lower electric fields. For an example, a shift of 0.1 meV requires ~10^4 V/cm for the ground state and only ~10^3 V/cm for the 2*s* excited state of shallow donors. An interpretation of the Stark effect for excited states, however, is complicated when the extended wavefunction starts to overlap with band states – see Franz-Keldysh effect, Sect. 2.3.3 of ▶ Chap. 29, "Carrier Generation". A Stark-effect splitting is shown in Fig. 31 (Blossey 1970). The normalizing field is the so-called ionization field of such defects, given by

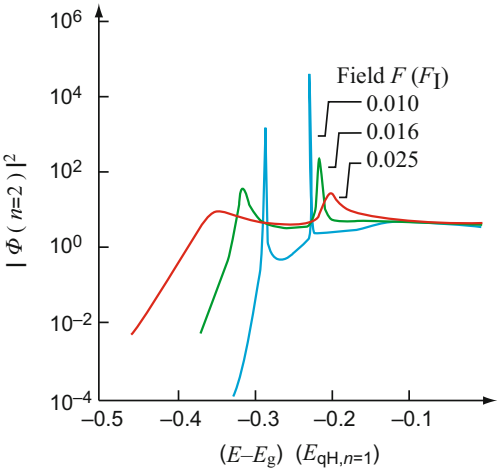


Fig. 31 Calculated Stark effect splitting of the $n = 2$ hydrogen-like level at normalized fields F as indicated in units of the ionization field F_I at the respective curves (After Blossey 1970)

Table 7 Ionization fields given in (kV/cm) by Eq. 34 for quasi-hydrogen defects in some compound semiconductors (After Blossey 1970)

AlSb	12	InP	7.8	ZnS	200	CdS	140
GaAs	5.7	InAs	0.70	ZnSe	75	CdSe	60
GaSb	1.0	InSb	0.08	ZnTe	47	CdTe	31

$$F_{\mathrm{I}} = \frac{E_{\mathrm{qH}}}{e\, a_{\mathrm{qH}}}, \tag{34}$$

is listed for some compound semiconductors in Table 7.

If the field is produced by other defects or electrons, the Stark effect is more complicated. These defects will already cause some random splitting because of the statistical distribution of their relative distance from each other, except for very low doping densities (Guichar et al. 1972). Therefore, the external field has only a minor influence, and requires a larger field strength, resulting in a dependence $\propto F^2$, the *quadratic Stark effect*. In addition to the described splitting and shift, the perturbation through the field permits optical transitions which, without a field, are forbidden by selection rules. For more information, see Blossey (1970, 1971).

For acceptors in cubic semiconductors, one observes that the splitting of the Γ_8-related acceptor states is much larger than the shifts of the Γ_6- and Γ_7-related levels (Kohn 1957; White 1967; Blossey 1970).

3.4 Influence of a Magnetic Field

The influence of the magnetic field on shallow, hydrogen-like defects can be divided into an effect on the conduction-band states and the effect on the eigenstates of the defect. As discussed earlier (Sect. 1.2.4 of ▶ Chap. 8, "Bands and Bandgaps in Solids"), the magnetic field forces electrons into orbits, circling with cyclotron frequency (Eq. 24 of ▶ Chap. 8, "Bands and Bandgaps in Solids"). When the magnetic field is small enough so that

$$\hbar\omega_{\mathrm{c}} \ll E_{\mathrm{qH},n}, \tag{35}$$

where $E_{\mathrm{qH},n}$ is the quasi-hydrogen bonding energy of the n^{th} level, the influence of this magnetic field can be regarded as a small perturbation. This case is discussed first. When the field is larger, the band states split into distinct Landau levels. Here, additional effects need to be considered that will be discussed later.

3.4.1 The Zeeman Effect

The splitting of spectral lines of atoms in a longitudinal or transverse magnetic field was first observed by Zeeman (1897) and is termed the *Zeeman effect*. The Zeeman effect is traditionally related to single atoms – see Herzberg (1937). Within a semiconductor, the influence of the surrounding lattice must be taken into consideration. In hydrogen-like defects, the electron orbits are very large. Consequently, a much smaller magnetic induction causes a major splitting.

Defect levels in semiconductors show a Zeeman splitting, which relates to the defect quantum number and the appropriate Zeeman levels of the impurity (Bassani et al. 1974). At higher magnetic induction, which can easily be reached for excited states, the wavefunction becomes severely compressed normal to **B**, which causes an increase in the binding energy of the electron at the defect (Baldereschi and Bassani 1970). In addition, the band states split into Landau levels, causing further complication – see below and Sect. 2 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients". The influence of the magnetic field can be included in the Schrödinger equation by replacing the momentum operator $-i\hbar\partial/\partial\mathbf{r}$ with the operator $-i\hbar\partial/\partial\mathbf{r} - e\,\mathbf{A}$, with the vector potential

$$\mathbf{A} = \frac{1}{2}(\mathbf{B} \times \mathbf{r}) \tag{36}$$

and **B** the magnetic induction. This yields in the effective-mass approximation

$$\left[-\frac{\hbar^2}{2m_n}\frac{\partial^2}{\partial\mathbf{r}^2} - \frac{e\hbar}{m_n}\left(\frac{\partial}{\partial\mathbf{r}}\cdot\mathbf{A}\right) + \frac{e^2}{2m_n}(\mathbf{A}\cdot\mathbf{A}) - \frac{e^2}{4\pi\,\varepsilon_{\text{stat}}\varepsilon_0\,r}\right] F(\mathbf{r}) = (E - E_c)F(\mathbf{r}). \tag{37}$$

The second term gives *normal Zeeman splitting*, which is linear in B for all $l \neq 0$ levels:

$$E_l = \frac{e}{2m_n}\mathbf{B}\cdot\mathbf{L}, \tag{38}$$

where **L** is the angular momentum operator; $\mathbf{L}^2$ has the eigenvalues $\hbar^2 l(l+1)$. The splitting occurs into $2l + 1$ equidistant levels with spacing $\pm\,\hbar\omega_c$.

When including the spin with spin operator **S**, an additional paramagnetic interaction due to the orientation of **S** in the magnetic field must be considered (see Sect. 1.1 of ▶ Chap. 9, "Magnetic Semiconductors"). Eq. 38 is then replaced by

$$E_J = g\mu_B\mathbf{B}\cdot\mathbf{J}, \tag{39}$$

where $\mu_B = e\hbar/(2m_0)$ is the Bohr magneton and $\mathbf{J} = \mathbf{L} + \mathbf{S}$ is the total angular momentum operator; $\mathbf{J}^2$ has the eigenvalues $\hbar^2 j(j+1)$. g is the Landé factor, given for isolated atoms by

$$g = 1 + \frac{j(j+1) + s(s+1) - l(l+1)}{2j(j+1)}. \tag{40}$$

For electrons in a semiconductor, the g factor is affected by the interaction with adjacent atoms, resulting in substantially different values. The paramagnetic interaction described by Eq. 39 is the *anomaleous Zeeman splitting* which is also linear in B. The splitting results in $2j + 1$ equidistant levels with spacing $g\mu_B B$.

In addition to the linear effects, a diamagnetic term, stemming from the third term in Eq. 37, produces a quadratic correction

$$\Delta E_{\text{dia}} = \frac{e^2}{8m_n}(\mathbf{B}\times\mathbf{r})^2 \cong \left(\frac{(\hbar\omega_c)^2}{8E_{\text{qH}}}\right)_s, \tag{41}$$

which is given here explicitly for the *s* states (van Vleck 1932). The quadratic term is the only magnetic field dependence for *s* states which show no normal Zeeman splitting ($l = 0$). It produces a compression of the wavefunction and thereby results in an increase in the binding energy.

For the small-field approximation, we obtain a Zeeman splitting of the degenerate levels of the quasi-hydrogen defect and a diamagnetic shift, similar to the observation in isolated atoms. Within a semiconductor, the magnetic field removes all degeneracies, including the Kramer's degeneracy due to time reversal, and consequently causes the most extensive splitting of defect levels (Condon and Shortley 1959; Haug 1972).

At high magnetic field ($\hbar\omega_c \gg E_{\text{qH}}$), the fourth term in Eq. 37, can be disregarded, and we obtain directly the Landau levels in the conduction band as solution of Eq. 37.

There is a large body of experimental results showing the influence of the magnetic induction on defect levels. For example, in Ge a magnetic field in the [111] direction will split the p_0 donor level in Voigt geometry (Sect. 3.1 of ▶ Chap. 12, "Photon–Free-Electron Interaction") into two levels; only one is allowed in Faraday geometry. The $p_\pm$ levels split into four levels, all of which are allowed in

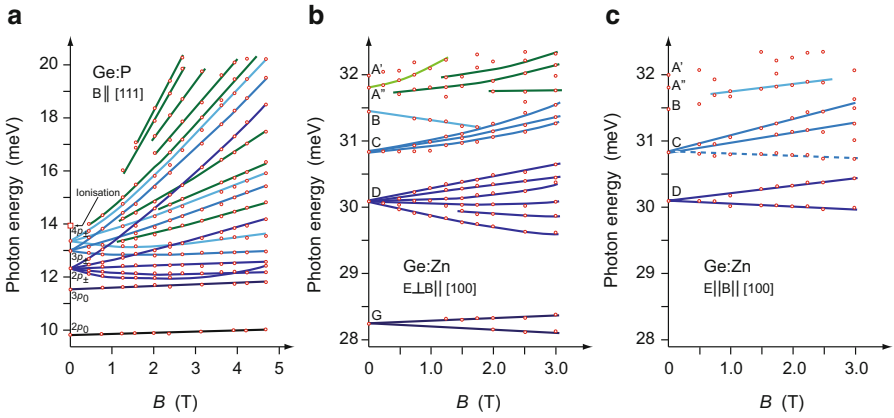


Fig. 32 Energy of absorption lines as a function of the magnetic induction in Ge: (**a**) for a P donor; (**b**) for a divalent Zn acceptor in Faraday geometry; (**c**) the same in Voigt geometry ($\mathbf{E}\perp\mathbf{H}\parallel[100]$) with conventional notation of G, D, C, B, and A for the excited states of the lower excitation lines of this acceptor (After Horii and Nisida 1970; Moore 1971)

Faraday geometry. This splitting has helped to identify these and higher excited states (Horii and Nisida 1970), as is shown for P donors in Ge in Fig. 32a.

The splitting of acceptor impurity-states is more complicated because of the band degeneracy and spin-orbit coupling; with a magnetic field, all degeneracies are consequently removed. For instance, a $\Gamma_8 \rightarrow \Gamma_8$ transition shows a splitting of the D^- term which can be well separated from the others into eight lines – six of which are observed in Faraday configuration and two in Voigt configuration. As many as eight Faraday and four Voigt lines are possible. This is shown in Fig. 32b and c for the Zn acceptor in Ge – see also Soepangkat et al. (1972) and Carter et al. (1976). For the Zeeman effect on donor-bound excitons in CdTe:In, see Zimmermann et al. (1994), and for acceptor-bound excitons in ZnSe:Li, see Pohl et al. (1996). For reviews, see Kaplan (1970) and Hasegawa (1969).

3.4.2 Magnetic Resonances at Lattice Defects

Electron Spin Resonance Electron spin-flip resonances can be measured in point defects with unpaired spin. For an introduction, see, e.g., Lancester (1966) and Lund et al. (2011). Spin-flip resonance absorption is related to the Zeeman splitting discussed in the previous section. Spin-flipping resonance of defect levels must be distinguished from the spin resonance of conduction electrons between different Landau levels. In a typical spin-doublet splitting, the upper branch relates to the electronic state with the spin parallel to the magnetic induction; for the lower state, the spin is antiparallel. In thermal equilibrium, more lower energy states are filled; a transition from the lower to the upper state can be initiated by supplying the necessary energy difference between these two states. This energy usually lies in the microwave range, where the corresponding resonance absorption can be observed (Poole 1983; Abragam and Bleaney 1976). The absorption permits a rather sensitive determination of the magnitude of the Zeeman splitting. Examples of typical spin-flip absorptions are shown in Fig. 33. For a review of such resonances

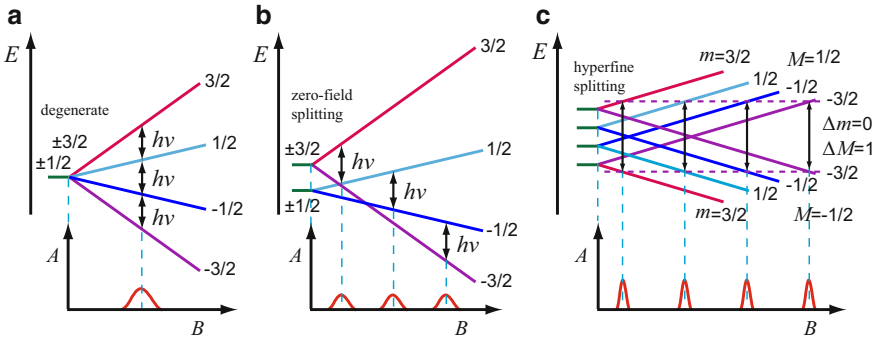


Fig. 33 Typical types of spin-flip resonances for (**a**) a degenerate line; (**b**) a line with spin-orbit splitting; (**c**) a line including interaction with unpaired nuclear spin

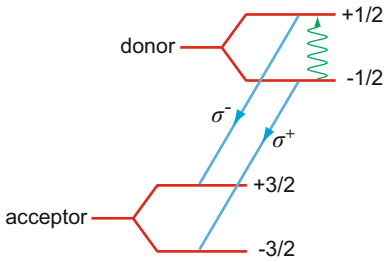


Fig. 34 Magnetic resonance transition for a donor-acceptor pair in CdS, causing an increase of the σ^- transition from the $S = 1/2$ donor to the $J = 3/2$ acceptor state (mixing with the uppermost valence band, e.g., in CdS)

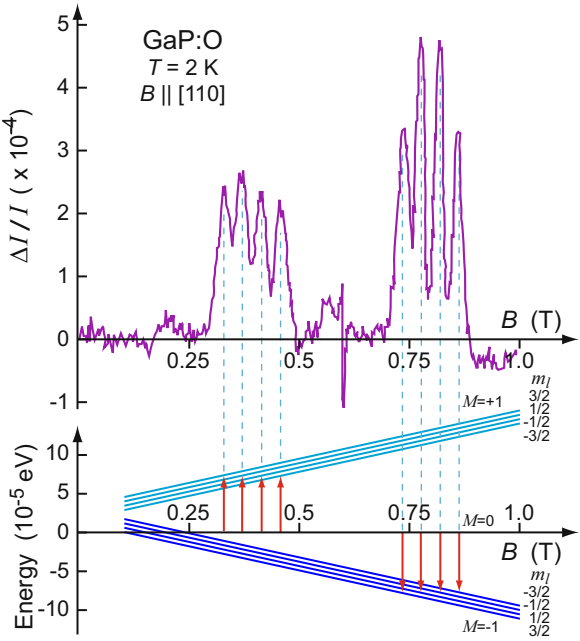


Fig. 35 Optically detected magnetic resonance (ODMR) emission spectrum of the O^- center in GaP with its level scheme due to splitting by hyperfine interaction with the Ga nucleus shown in the lower part (After Gal et al. 1979)

in III-V compounds, see Schneider (1982), and in II-VI compounds, see Schneider (1967).

Spin-flip resonances can be detected optically when the microwave-induced transition occurs from the lower energy spin-up to the higher energy spin-down states. This also results in corresponding changes of the polarization of absorption or emission (luminescence) in the presence of the magnetic field. For normal Zeeman transitions in longitudinal observation (Faraday geometry), one sees two components which are left and right circularly polarized. In transverse observation (Voigt geometry), one sees three components with linear polarization, of which one is parallel and two are perpendicularly polarized to the field axis. These are known

as π and σ components, for parallel and perpendicular (the latter for German senkrecht) polarization. Such changes occur in single defects or in defect associates (Cavenett 1981) as indicated in Fig. 34 for a donor-acceptor pair.

Hyperfine Splitting When the electrons in a defect center interact with the magnetic momentum of the nucleus, one observes hyperfine splitting of these levels, which provides information on the chemical identity of the center. As an example, the optically detected hyperfine magnetoresonance signal of oxygen-doped GaP is shown in Fig. 35.

Super-hyperfine Splitting When the electrons of a defect center interact also with the nuclei of the surrounding lattice atoms, one observes a super-hyperfine or ligand hyperfine splitting. The interaction causes a splitting of electronic levels of the unpaired electron: each level is split into $2I + 1$ sublevels with I the spin of the nucleus. The energy difference between the sublevels is proportional to the charge density of the wavefunction of the defect state.

This type of splitting provides valuable information on the structure of the immediate neighborhood of the defect center (Spaeth 1986) and on the chemical identity of the defect. Super-hyperfine splitting is used to analyze the eigenfunction, i.e., the electron distribution of defect centers using the electron nuclear double resonance (ENDOR) signal combined with the electron spin resonance (ESR)[8] signal; see Feher (1959). The ENDOR technique is based on changes of the ESR signal caused by the spin flip of appropriate nuclei [e.g., Si^{29} in a silicon crystal with natural Si [Si^{28} (92.2%), Si^{29} (4.7%), Si^{30} (3.1%)]] within the reach of the electron cloud of the defect and induced by an external electromagnetic field.

4 Summary

Major features of shallow-level defects are easily described by a quasi-hydrogen spectrum, modified by the dielectric constant and the effective mass of the host semiconductor. While this relatively simple relationship holds surprisingly well for higher excited states of many shallow defects, substantial deviations occur in the ground state due to the core potential and lattice relaxation.

Some of the degeneracies of the quasi-hydrogen spectrum are lifted by band anisotropies and interaction between band valleys. Local stress and electric fields cause additional splitting. Because of far-reaching interaction between defects through such fields, the line spectrum is usually substantially inhomogeneously broadened. Narrow lines can only be observed in ultrapure, strain-free materials. The dependence of levels on hydrostatic pressure can be used to identify those which are connected to one band only. Shallow-level defects are distinguished therefore from so-called deep-level centers which originate from tightly bound centers and

[8]Electron spin resonance (ESR) is also referred to as electron paramagnetic resonance (EPR).

connect with both bands; only accidentally may they lie close to one of them. The effect of uniaxial stress and electric or magnetic fields applied along specific crystallographic directions and selection rules related to field-distorted symmetries can be used for an identification of the impurity center. Further analysis is obtained from magnetic resonance techniques in point defects with unpaired spin by applying resonances in the electron spin-flip typically in the GHz regime at common magnetic fields.

Donors and acceptors can form pairs even at relatively low defect concentration. Their discrete distances in the crystal lattice give rise to optical transitions with discrete energies, given by the bandgap energy reduced by the donor and acceptor ionization energies and a Coulomb attraction term; an additional correction accounts for the overlap of donor and acceptor eigenfunctions becoming significant for close pairs.

Donors and acceptors can bind excitons with a binding energy typically on the order of 10 meV. For neutral defects a linear dependence of the binding energy from the ionization energy of the impurity is observed in experiment. For ionized donors, the binding energy depends on the ratio of electron-to-hole effective mass, and no bound exciton state exists above a critical ratio of ~0.45. Ionized acceptors usually do not bind excitons. Isoelectronic impurities can bind excitons by attracting the electron at the core potential, that in turn binds the hole on a quasi-hydrogen orbital.

References

Abarenkov IV, Heine V (1965) The model potential for positive ions. Philos Mag 12:529

Abragam A, Bleaney B (1976) Electron paramagnetic resonance of transition ions. Claredon, Oxford, UK

Abraham RJS, Shuman VB, Portsel LM, Lodygin AN, Astrov YuA, Abrosimov NV, Pavlov SG, Hübers H-W, Simmons S, Thewalt MLW (2019) Shallow donor complexes formed by pairing of double-donor magnesium with group-III acceptors in silicon. Phys Rev B 99:195207

Aggarwal RL, Ramdas AK (1965a) Effect of uniaxial stress on the excitation spectra of donors in silicon. Phys Rev 137:A602

Aggarwal RL, Ramdas AK (1965b) Optical determination of the symmetry of the ground states of group-V donors in silicon. Phys Rev 140:A1246

Altarelli M, Iadonisi G (1971) Donor ground states of group IV and III-V semiconductors. Nuovo Cimento B 5:21

Appapillai M, Heine V (1972) Solid state group tech rep 5. Cavendish Lab, Cambridge, UK

Ashen DJ, Dean PJ, Hurle DTJ, Mullin JB, White AM (1975) The incorporation and characterization of acceptors in epitaxial GaAs. J Phys Chem Solid 36:1041

Austin BJ, Heine V, Sham LJ (1962) General theory of pseudopotentials. Phys Rev 127:276

Baldereschi A (1970) Valley-orbit interaction in semiconductors. Phys Rev B 1:4673

Baldereschi A, Bassani F (1970) Optical absorption by excitons in a uniform magnetic field. In: Keller SP, Hensel JC, Stern F (eds) Proceedings of the 10th international conference on the physics of semiconductors, Cambridge MA, USA. National Bureau of Standards, Technical Information Service, Springfield, pp 191–196

Baldereschi A, Hopfield JJ (1972) Binding to isoelectronic impurities in semiconductors. Phys Rev Lett 28:171

Baldereschi A, Lipari NO (1973) Spherical model of shallow acceptor states in semiconductors. Phys Rev B 8:2697

Baldereschi A, Lipari NO (1976) Interpretation of acceptor spectra in Si and Ge. In: Fumi FG (ed) Proceedings of the 13th international conference on the physics of semiconductors, Rome 1976. Tipografia Marves, Rome, pp 595–598

Bassani GF, Celli V (1961) Energy-band structure of solids from a perturbation on the "empty lattice". J Phys Chem Solid 20:64

Bassani GF, Pastori Parravicini G (1975) Electronic states and optical transitions in solids. Pergamon, Oxford/New York

Bassani GF, Iadosini G, Preziosi B (1969) Band structure and impurity states. Phys Rev 186:735

Bassani GF, Iadosini G, Preziosi B (1974) Electronic impurity levels in semiconductors. Rep Prog Phys 37:1099

Bebb HB (1969) Application of the quantum-defect method to optical transitions involving deep effective-mass-like impurities in semiconductors. Phys Rev 185:1116

Bergh AA, Dean PJ (1976) Light emitting diodes. Claredon, Oxford

Bergman K, Grossmann G, Grimmeiss HG, Stavola M, Holm C, Wagner P (1987) Spin triplet states of double donors in silicon. In: Engström O (ed) Proceedings of the 18th international conference on the physics of semiconductors, Stockholm 1986. World Scientific, Singapore, pp 883–886

Bernholc J, Pantelides ST (1977) Theory of binding energies of acceptors in semiconductors. Phys Rev B 15:4935

Blossey DF (1970) Wannier exciton in an electric field. I. Optical absorption by bound and continuum states. Phys Rev B 2:3976

Blossey DF (1971) Wannier exciton in an electric field. II. Electroabsorption in direct-band-gap solids. Phys Rev B 3:1382

Callaway J (1976) Quantum theory of solid state. Academic Press, New York

Cardona M (1969a) Modulation spectroscopy. In: Seitz F, Turnbull D, Ehrenreich H (eds) Solid state physics, suppl 11. Academic, New York

Cardona M (1969b) Optical constants of insulators: dispersion relations. In: Nudelman S, Mitra SS (eds) Optical properties of solids. Plenum Press New York, pp 137–151

Carter AC, Skolnick MS, Stradling RA, Leotin JP, Askenazy S (1976) The Zeeman splitting of Si, S, and Te donors in GaP. In: Fumi FG (ed) Proceedings of the 13th international conference on the physics of semiconductors, Rome 1976. Tipografia Marves, Rome, pp 619–622

Castner TG Jr (1970) Configuration mixing of subsidiary minima: corrections to the ground-state wave function for donors in silicon. Phys Rev B 2:4911

Cavenett BC (1981) Optically detected magnetic resonance (O.D.M.R.) investigations of recombination processes in semiconductors. Adv Phys 30:475

Cohen E, Sturge MD (1977) Excited states of excitons bound to nitrogen pairs in GaP. Phys Rev B 15:1039

Cojocari O, Popa V, Ursaki VV, Tiginyanu IM, Hartnagel HL, Daumiller I (2004) GaN Schottky multiplier diodes prepared by electroplating: a study of passivation technology. Semicond Sci Technol 19:1273

Condon EU, Shortley GH (1959) The theory of atomic spectra. Cambridge University Press, Cambridge, UK

Costato M, Manchinelli F, Reggiani L (1971) Anomalous behavior of shallow donor ground state levels in Ge under pressure. Solid State Commun 9:1335

Craford MG, Holonyak N Jr (1976) The optical properties of the nitrogen isoelectronic trap in $GaAs_{1-x}P_x$. In: Seraphin BO (ed) Optical properties of solids – new developments. North Holland, Amsterdam, pp 187–253

Csavinszky P (1965) Corrections to the effective mass theory of shallow impurity states in Si and Ge. J Phys Soc Jpn 20:2027

Davis G, Nazaré MH (1994) The bound exciton model for isoelectronic centres in silicon. In: Heinrich H, Jantsch W (eds) Proceedings of the 17th international conference on defects of semiconductors. Mater Sci Forum 143–147. pp 105–109
Dean PJ (1973a) Inter-impurity recombinations in semiconductors. Prog Solid State Chem 8:1
Dean PJ (1973b) Lithium donors and the binding of excitons at neutral donors and acceptors in gallium phosphide. In: Williams F (ed) Luminescence of crystals, molecules, and solutions. Plenum Press, New York, pp 538–552
Dean PJ (1983) Excitons in semiconductors. In: Di Bartolo B (ed) Collective excitation in semiconductors. Plenum Press, New York, pp 247–315
Dean PJ, Herbert DC (1979) Bound excitons in semiconductors. In: Cho K (ed) Excitons. Springer, Berlin, pp 55–182
Dean PJ, Cuthbert JD, Thomas DG, Lynch RT (1967) Two-electron transitions in the luminescence of excitons bound to neutral donors in gallium phosphide. Phys Rev Lett 18:122
Dean PJ, Cuthbert JD, Lynch RT (1969) Interimpurity recombinations involving the isoelectronic trap bismuth in gallium phosphide. Phys Rev 179:754
Denteneer PJH, Van de Walle CG, Pantelides ST (1989) Microscopic structure of the hydrogen-boron complex in crystalline silicon. Phys Rev B 39:10809
Dietz RE, Thomas DG, Hopfield JJ (1962) "Mirror" absorption and fluorescence in ZnTe. Phys Rev Lett 8:391
Faulkner RA (1968) Toward a theory of isoelectronic impurities in semiconductors. Phys Rev 175:991
Faulkner RA (1969) Higher donor excited states for prolate-spheroid conduction bands: a reevaluation of silicon and germanium. Phys Rev 184:713
Feher G (1959) Electron spin resonance experiments on donors in silicon. I. Electronic structure of donors by the electron nuclear double resonance technique. Phys Rev 114:1219
Feher G (1998) The development of ENDOR and other reminescences of the 1950's. In: Eaton GR, Eaton SS, Salikov KM (eds) Foundation of modern EPR. Worlds Scientific, Singapore, pp 548–556
Freysoldt C, Grabowski BJ, Hickel T, Neugebauer J, Kresse G, Janotti A, Van de Walle CG (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86:253
Fritzsche H (1962) Effect of stress on the donor wave functions in germanium. Phys Rev 125:1560
Gal M, Cavenett BC, Smith P (1979) New evidence for the two-electron O^- state in GaP. Phys Rev Lett 43:1611
Gel'mont BL, D'yakonov MI (1972) Acceptor levels in diamond-type semiconductors. Sov Phys Semicond 5:1905
Goñi AR, Syassen K (1998) Optical properties of semiconductors under pressure. In: Suski T, Paul W (eds) High pressure in semiconductor physics I. Semiconductors and semimetals, vol 54. Academic Press, San Diego, pp 247–425
Gorczyca I, Svane A, Christensen NE (1997) Calculated defect levels in GaN and AlN and their pressure coefficients. Solid State Commun 101:747
Grimmeiss HG (1985) Deep energy levels in semiconductors. In: Chadi JD, Harrison WA (eds) Proceedings of the 17th international conference on the physics of semiconductors, San Francisco 1984. Springer, New York, pp 589–600
Grimmeiss HG, Janzén E (1986) Chalcogen impurities in silicon. In: Pantelides ST (ed) Deep centers in semiconductors. Gordon and Breach, New York, pp 87–146
Grimmeiss HG, Janzén E, Skarstam B (1980) Deep sulfur-related centers in silicon. J Appl Phys 51:4212
Guichar GM, Sébenne C, Proix F, Balkanski M (1972) Lowering the extrinsic photoconductivity threshold of Si:P. In: Proceedings of the11th international conference on the physics of semiconductors. PWN Polish Scientific Publishers, Warsaw, pp 877–881
Haller EE, Hansen WL (1974) High resolution Fourier transform spectroscopy of shallow acceptors in ultra-pure germanium. Solid State Commun 15:687

Haller EE, Hansen WL, Goulding FS (1981) Physics of ultra-pure germanium. Adv Phys 30:93
Halsted RE (1967) Radiative recombination in the band edge region. In: Aven M, Prener JS (eds) Physics and chemistry of II-VI compounds. North Holland, Amsterdam, pp 383–431
Halsted RE, Aven M (1965) Photoluminescence of defect-exciton complexes in II-VI compounds. Phys Rev Lett 14:64
Hasegawa H (1969) Effects of high magnetic fields on electronic states in semiconductors – the Rydberg series and the Landau levels. In: Haidemenakis ED (ed) Physics of solids in intense magnetic fields. Plenum Press, New York, pp 246–270
Haug A (1972) Theoretical solid state physics. Pergamon, Oxford
Hayes W, Stoneham AM (1984) Defects and defect processes in nonmetallic solids. Wiley, New York
Haynes JR (1960) Experimental proof of the existence of a new electronic complex in silicon. Phys Rev Lett 4:361
Herzberg G (1937) Atomic spectra and atomic structure. Dover, New York
Herring C, Vogt E (1956) Transport and deformation-potential theory for many-valley semiconductors with anisotropic scattering. Phys Rev 101:944
Hopfield JJ (1967) Radiative recombination at shallow centers. In: Thomas DG (ed) Proceedings of international conference on II-VI semiconducting compounds. Benjamin, New York, p 786
Hopfield JJ, Thomas DG, Lynch RT (1966) Isoelectronic donors and acceptors. Phys Rev Lett 17:312
Horii K, Nisida Y (1970) Identification of $^4p_0$ and $^4p_{\pm 1}$ of germanium donor from the Zeeman spectra. J Physical Soc Jpn 29:798
Ishikawa T, Sekiguchi T, Yoshizawa K, Naito K, Thewalt MLW, Itoh KM (2010) Zeeman photoluminescence spectroscopy of isoelectronic beryllium pairs in silicon. Sol State Commun 150:1827
Jancu J-M, Scholz R, Beltram F, Bassani F (1998) Empirical spds* tight-binding calculation for cubic semiconductors: general method and material parameters. Phys Rev B 57:6493
Jansen RW, Sankey OF (1987) Trends in the energy levels and total energies of s p – valence interstitials in compound semiconductors – an ab initio tight-binding study for GaAs. In: Engström O (ed) Proceedings of the 18th international conference on the physics of semiconductors, Stockholm 1986. World Scientific, Singapore, pp 813–816
Kaplan R (1970) Far-infrared magnetooptical studies of semiconductors using Fourier spectroscopy and photoconductivity techniques. In: Haidemenakis D (ed) Optical properties of solids. Gordon & Breach, London, pp 301–330
Kirkman RF, Stradling RA, Lin-Chung PJ (1978) An infrared study of the shallow acceptor states in GaAs. J Phys C 11:419
Kita T, Wada O (2006) Bound exciton states of isoelectronic centers in GaAs:N grown by an atomically controlled doping technique. Phys Rev B 74:035213
Kohn W (1957) Shallow impurity states in silicon and germanium. In: Seitz F, Turnbull D (eds) Solid state physics, vol 5. Academic Press, New York, pp 257–320
Kopylov AA, Pikhtin AN (1978) Shallow impurity states and the free exciton binding energy in gallium phosphide. Solid State Commun 26:735
Kosicki BB, Paul W (1966) Evidence for quasilocalized states associated with high-energy conduction-band minima in semiconductors, particularly Se-doped GaSb. Phys Rev Lett 17:246
Kudlek G, Presser N, Pohl UW, Gutowski J, Lilja J, Kuusisto E, Imai K, Pessa M, Hingerl K, Sitter A (1992) Exciton complexes in ZnSe layers: a tool for probing the strain distribution. J Cryst Growth 117:309
Lampert MA (1958) Mobile and immobile effective-mass-particle complexes in nonmetallic solids. Phys Rev Lett 1:450
Lancester G (1966) Electron spin resonance in semiconductors. Hilger & Watts, London
Li J, Le NH, Litvinenko KL, Clowes SK, Engelkamp H, Pavlov SG, Hübers H-W, Shuman VB, Portsel LM, Lodygin AN, Astrov YuA, Abrosimov NV, Pidgeon CR, Fisher A, Zeng Z, Niquet

Y-M, Murdin BN (2018) Radii of Rydberg states of isolated silicon donors. Phys Rev B 98:085423
Lipari NO, Baldereschi A (1978) Interpretation of acceptor spectra in semiconductors. Solid State Commun 25:665
Lund A, Shiotani M, Shimada S (2011) Principles and applications of ESR spectroscopy. Springer, Heidelberg/New York
Luttinger JM (1956) Quantum theory of cyclotron resonance in semiconductors: general theory. Phys Rev 102:1030
Luttinger JM, Kohn W (1955) Motion of electrons and holes in perturbed periodic fields. Phys Rev 97:869
Lyons JL, Janotti A, Van de Walle CG (2009) Role of Si and Ge as impurities in ZnO. Phys Rev B 80:205113
Lyons JL, Steiauf D, Janotti A, Van de Walle CG (2014) Carbon as a shallow donor in transparent conducting oxides. Phys Rev Applied 2:064005
Markham JJ (1966) F-centers in alkali halides. Academic Press, New York
McMurray RE Jr (1985) Spectroscopy of positively charged multiple acceptors. Solid State Commun 53:1127
Merz JL (1968) Isoelectronic oxygen trap in ZnTe. Phys Rev 176:961
Merz JL, Nassau K, Shiever JW (1973) Pair spectra and shallow acceptors in ZnSe. Phys Rev B 8:1444
Moore WJ (1971) Magnetic field effects on the excitation spectra of neutral group II double acceptors in germanium. J Phys Chem Solid 32:93
Narita S (1985) Effects of uniaxial stress and magnetic field of D-center in germanium and silicon. Solid State Commun 53:1115
Onton A (1971) Donor-electron transitions between states associated with the X_{1c} and X_{3c} conduction-band minima in GaP. Phys Rev B 4:4449
Onton A, Fisher P, Ramdas AK (1967) Spectroscopic investigation of group-III acceptor states in silicon. Phys Rev 163:686
Onton A, Yacoby Y, Chicotka RJ (1972) Direct optical observation of the subsidiary X_{1c} conduction band and its donor levels in InP. Phys Rev Lett 28:966
O'Reilly EP, Lindsay A, Klar PJ, Polimeni A, Capizzi M (2009) Trends in the electronic structure of dilute nitride alloys. Semicond Sci Technol 24:033001
Pajot B (2009) Optical absorption of impurities and defects in semiconducting crystals, I. Hydrogen-like Centres. Springer, Berlin
Pantelides ST (1975) Theory of impurities in semiconductors. In: Queisser HJ (ed) Festkörperprobleme/ Advances in solid state physics, vol 15. Vieweg, Braunschweig, pp 149–181
Pantelides ST (1978) The electronic structure of impurities and other point defects in semiconductors. Rev Mod Phys 50:797
Pantelides ST (ed) (1986) Deep centers in semiconductors. Gordon and Breach, New York
Pantelides ST (1987) The effect of hydrogen on shallow dopants in crystalline silicon. In: Engström O (ed) Proceedings of the 18th international conference on the physics of semiconductors, Stockholm 1986. World Scientific, Singapore, pp 987–990
Pantelides ST, Sah CT (1974) Theory of localized states in semiconductors. II. The pseudo impurity theory application to shallow and deep donors in silicon. Phys Rev B 10:638
Phillips JC, Kleinman L (1959) New method for calculating wave functions in crystals and molecules. Phys Rev 116:287
Pohl UW, Busse W (1989) Probability tables for small clusters of impurity atoms in sc, bcc and fcc lattices assuming long range interaction. J Chem Phys 90:6877
Pohl UW, Wiesmann D, Kudlek GH, Litzenburger B, Hoffmann A (1996) Magneto-optical investigation of the shallow lithium acceptor in zinc selenide. J Cryst Growth 159:414
Pollak FH (1965) Effect of uniaxial compression on impurity conduction in p-germanium. Phys Rev 138:A618
Pollmann J (1976) Exciton ground-state in strongly anisotropic crystals. Solid State Commun 19:361

Poole CP Jr (1983) Electron spin resonance – a comprehensive treatise on experimental techniques, 2nd edn. Wiley, New York
Rashba EI, Gurgenishvilli GE (1962) To the theory of the edge absorption in semiconductors. Sov Phys Solid State 4:759
Reiss H, Fuller CS, Morin FJ (1956) Chemical interactions among defects in germanium and silicon. Bell Sys Tech J 35:535
Ren SY, Dow JD, Wolford DJ (1982) Pressure dependence of deep levels in GaAs. Phys Rev B 25:7661
Reuszer JH, Fisher P (1964) An optical determination of the ground-state splittings of group V impurities in germanium. Phys Rev 135:A1125
Rodriguez S, Fisher P, Barra F (1972) Spectroscopic study of the symmetries and deformation-potential constants of singly ionized zinc in germanium, theory. Phys Rev B 5:2219
Roessler DM (1970) Luminescence in tellurium-doped cadmium sulfide. J Appl Phys 41:4589
Rotenberg M, Stein J (1969) Use of asymptotically correct wave function for three-body Rayleigh-Ritz calculations. Phys Rev 182:1
Sah CT, Pantelides ST (1972) Theory of impurity states in semiconductors. Solid State Commun 11:1713
Sak J (1971) Perturbation theory for a bound polaron. Phys Rev B 3:3356
Schechter D (1975) Pseudopotential theory of shallow-donor ground states II. Phys Rev B 11:5043
Schneider J (1967) Electron spin resonance of defect centers in II-VI cemiconductors. In: Thomas DG (ed) II-VI semiconducting compounds. Benjamin, New York, p 40
Schneider J (1982) ESR of defects in III–V compounds. MRS Proc 14:225
Shan W, Walukiewicz W, Ager III JW, Haller EE, Geisz JF, Friedman DJ, Olson JM, Kurtz SR (1999) Band anticrossing in GaInNAs alloys. Phys Rev Lett 82:1221
Shinada M, Sugano S (1966) Interband optical transitions in extremely anisotropic semiconductors. I. Bound and unbound exciton absorption. J Physical Soc Jpn 21:1936
Skettrup T, Suffczynski M, Gorzkowski W (1971) Properties of excitons bound to ionized donors. Phys Rev B 4:512
Soepangkat HP, Fisher P, Rodriguez S (1972) g-factors of boron in germanium. Phys Lett A 39:379
Sommerfeld A (1950) Atombau und Spektrallinien, vol I. Vieweg, Braunschweig (Atomic structure and spectral lines, in German)
Spaeth J-M (1986) Application of optically detected magnetic resonance to the characterization of point defects in semiconductors. In: von Bardeleben HJ (ed) Proceedings of the 14th international conference on defects in semiconductors. Mater Sci Forum 10–12. Trans Tech Publications, Switzerland, pp 505–514
Stark J (1914) Beobachtungen über den Effekt des elektrischen Feldes auf Spektrallinien, I Quereffekt. Ann d Phys 43:965 (Observations on the effect of an electric field on spectral lines, I transversal effect, in German)
Stark J, Wendt G (1914) Beobachtungen über den Effekt des elektrischen Feldes auf Spektrallinien, II Längseffekt. Ann d Phys 43:983 (Observations on the effect of an electric field on spectral lines, II longitudinal effect, in German)
Steger M, Yang A, Karaiskaj D, Thewalt MLW, Haller EE, Ager III JW, Cardona M, Riemann H, Abrosimov NV, Gusev AV, Bulanov AD, Kaliteevskii AK, Godisov ON, Becker P, Pohl H-J (2009) Shallow impurity absorption spectroscopy in isotopically enriched silicon. Phys Rev B 79:205210
St-Jean P, Éthier-Majcher G, Francoeur S (2015) Dynamics of excitons bound to nitrogen isoelectronic centers in GaAs. Phys Rev B 91:115201
St-Jean P, Éthier-Majcher G, André R, Francoeur S (2016) High-fidelity and ultrafast initialization of a hole spin bound to a Te isoelectronic center in ZnSe. Phys Rev Lett 117:167401
Stoneham AM (1975) Theory of defects in solids. Claredon, Oxford, UK
Stoneham AM (1986) Hot topics: theory. In: von Bardeleben HJ (ed) Proceedings of the 14th international conference on defects in semiconductors. Mater Sci Forum 10–12. Trans Tech Publications, Switzerland, pp 9–19

Sun JW, Ivanov IG, Juillaguet S, Camassel J (2011) Splitting of type-I (N-B, P-Al) and type-II (N-Al, N-Ga) donor-acceptor pair spectra in 3C-SiC. Phys Rev B 83:195201
Taguchi T, Ray B (1983) Point defects in II–VI compounds. Progr Cryst Growth Charact 6:103
Teisseyre H, Lyons JL, Kaminska A, Jankowski D, Jarosz D, Boćkowski M, SuchockiA, Van de Walle CG (2017) Identification of yellow luminescence centers in Be-doped GaN through pressure-dependent studies. J Phys D Appl Phys 50:22LT03
Thewalt MLW, Labrie D, Timusk T (1985) The far infrared absorption spectra of bound excitons in silicon. Solid State Commun 53:1049
Thomas DG, Hopfield JJ (1962) Optical properties of bound exciton complexes in cadmium sulfide. Phys Rev 128:2135
Thomas DG, Hopfield JJ (1966) Isoelectronic traps due to nitrogen in gallium phosphide. Phys Rev 150:680
Thomas DG, Gershenzon M, Trumbore FA (1964) Pair spectra and "edge" emission in gallium phosphide. Phys Rev 133:A269
Thomas DG, Hopfield JJ, Frosch CJ (1965) Isoelectronic traps due to nitrogen in gallium phosphide. Phys Rev Lett 15:857
Torres VJB, Oberg S, Jones R (1997) Theory of nitrogen-hydrogen complexes in GaP. In: Davies G, Nazaré MH (eds) Proceedings of the 19th international conference on defect in Semiconductors. Mater Sci Forum 258–263. Trans Tech Publications, Switzerland, pp 1063–1067
Trumbore FA, Gershenzon M, Thomas DG (1966) Luminescence due to the isoelectronic substitution of bismuth for phosphorus in gallium phosphide. Appl Phys Lett 9:4
Ulbrich RG (1978) Low density photoexcitation phenomena in semiconductors: aspects of theory and experiment. Solid State Electron 21:51
Usman M, Broderick CA, Lindsay A, O'Reilly EP (2011) Tight-binding analysis of the electronic structure of dilute bismide alloys of GaP and GaAs. Phys Rev B 84:245202
van Vleck JH (1932) The theory of electric and magnetic susceptibilities. Claredon, Oxford, UK
Vul' AY, Bir GL, Shmartsev YV (1971) Donor states of sulfur in gallium antimonide. Sov Phys Semicond 4:2005
Wagner P, Holm C, Sirtl E, Oeder R, Zulehner W (1984) Chalcogens as point defects in silicon. In: Grosse P (ed) Festkörperprobleme/Advances in solid state physics, vol 24. Vieweg, Braunschweig, pp 191–228
Watkins GD (1977) Lattice defects in II-VI compounds. In: Urli NB, Corbett JW (eds) Proceedings of international conference on radiation effects in semiconductors, vol 31, Institute of physics conference series. Institute of Physics, Bristol, p 95
White JJ (1967) Effects of external and internal electric fields on the boron acceptor states in silicon. Can J Phys 45:2695
Williams F (1968) Donor-acceptor pairs in semiconductors. Phys Status Solidi 25:493
Wilson DK, Feher G (1961) Electron spin resonance experiments on donors in silicon. III. Investigation of excited states by the application of uniaxial stress and their importance in relaxation processes. Phys Rev 124:1068
Wolford DJ (1987) Electronic states in semiconductors under high pressures. In: Engström O (ed) Proceedings of the 18th international conference on the physics of semiconductors, Stockholm 1986. World Scientific, Singapore, pp 1115–1123
Wolford DJ, Hsu WY, Dow JD, Streetman BG (1979) Nitrogen trap in the semiconductor alloys $GaAs_{1-x}P_x$ and $Al_xGa_{1-x}As$. J Lumin 18–19:863
Zeeman P (1897) Doublets and triplets in the spectrum produced by external magnetic forces. Philos Mag 44:55; and: Philos Mag 44:255
Zimmermann H, Boyn R, Lehr MU, Schulz H-J, Rudolph P, Kornac J-T (1994) The Zeeman effect on bound-exciton states of indium-related complex centres in CdTe. Semicond Sci Technol 9:1598

Deep-Level Centers

19

Karl W. Böer and Udo W. Pohl

Contents

1 Modeling of Deep-Level Centers 812
1.1 General Properties of Deep-Level Centers 813
1.2 Theoretical Methods to Analyze Defect Centers 816
1.3 The Jahn-Teller Effect 819
1.4 Crystal-Field Theory 820
2 Deep Centers in Semiconductors 824
2.1 Defects in Alkali Halides 824
2.2 Vacancies in Semiconductors 826
2.3 Self-Interstitials and Antisite Defects 829
2.4 Hydrogen in Semiconductors 831
2.5 Substitutional Defects 832
2.6 Chalcogens in Semiconductors 835
2.7 The DX Center 837
2.8 Negative-*U* Centers 838
2.9 Instabilities of Shallow and Deep Centers 840
3 Transition-Metal Impurities 841
3.1 Effect of Site, Charge, Spin, and Excitation 841
3.2 The Energy of Levels 845
3.3 Optical Transitions of Transition-Metal Dopants 850
4 Summary 853
References 854

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_19

Abstract

A wide variety of deep-level centers exists of both intrinsic and extrinsic origin. They often provide a preferred path for carrier recombination or act as deep traps. Deep-level defects have tightly bound electrons in small orbits which, for the ground state, often do not extend beyond the distance to the next neighbors of the semiconductors. Their electronic eigenfunctions can mix with both conduction and valence bands. Many of the deep defect levels relax substantially after defect recharging or excitation, causing a significant change in the equilibrium position of the surrounding atoms. In optical spectra, the strong lattice coupling leads to a broadened spectrum with a weak zero-phonon transition and hardly identifiable individual phonon lines.

The description of deep-level centers requires a tight-binding analysis. Such centers may, however, have higher excited states which have quasi-hydrogen character with corresponding orbits extending well into the surrounding lattice. For extrinsic deep-level centers, the atomic electronegativity is an indicator for the depth of the central-cell potential. The deep centers act as deep traps for either electrons or holes. Many of them have various charge states in the bandgap with corresponding levels depending on the Fermi energy, leading to effective compensation of doping. Deep-level centers may also act as recombination centers when their relaxation is significant and thereby form a bridge between conduction and valence band for nonradiative recombination. Transition-metal impurities are a prominent group with vastly different defect-level behaviors, showing both intracenter excitations and ionization with charge transfer.

Keywords

Deep trap · Central-cell potential · Color center · Compensation · Crystal-field splitting · Crystal-field theory · DX center · F center · Ham effect · Intraionic transitions · Jahn-Teller distortion · Negative-U center · Passivation · Recombination centers · Tanabe-Sugano diagram · Transition-metal impurities · Vacancy

1 Modeling of Deep-Level Centers

Deep-level centers generally require a tight-binding analysis in which, at least for the ground state, the wavefunction remains localized close to the core of the defect. Such centers cannot be described by a hydrogenic effective-mass approximation. Deep levels, however, do not necessarily require a large binding energy.[1] They are not connected to a specific conduction-band minimum or valence-band maximum, that is, these deep trap levels do not follow one specific band edge

[1]Deep levels also appear in narrow-bandgap materials (see Lischka 1986); deep levels may appear as well close to the conduction- or valence-band edge (see Hjalmarson et al. 1980).

when perturbed by alloying or the application of hydrostatic pressure. Instead, the wavefunction of deep levels is derived from conduction- and/or valence-band states throughout the Brillouin zone. Their central core potential dominates their behavior at the ground state, or they have unsaturated inner shells in transition-metal impurities, permitting electronic transitions here. Specifically, such defects may or may not be charged relative to the lattice; they may be isoelectronic or isovalent.

The levels of these centers are described by a short-range potential. The pseudopotential method is an advantageous tool for determining $V(\mathbf{r})$. In addition, the deformed lattice environment must be considered – see Sects. 2.1.1 and 2.8, and for Jahn-Teller distortion Sect. 1.3. Since the deep states of the center extend throughout the entire Brillouin zone, states from both bands and all near-bandgap valleys are necessary to construct the ground-state electron eigenfunction. The resulting deep defect levels communicate with both bands and act as deep traps for electrons or holes or as recombination centers (Sect. 1.1.2 of ▶ Chap. 30, "Carrier Recombination and Noise"). A review of such deep centers can be found by Queisser (1971), Stoneham (1975), Lannoo and Bourgoin (1981), Pantelides (1978, 1992), Madelung and Schulz (1989), Watkins (2000), Weber (1993), and Stavola (1998, 1999a).

In contrast to shallow-level centers, which are easily identified chemically in their specific lattice environment and are rather well understood in their electronic level structure, it is much more difficult to identify a specific deep-level lattice defect and to describe theoretically its electronic behavior. Except for a large variety of centers in ionic crystals, which were identified earlier (the well-known class of F centers in alkali halides), many deep centers in semiconductors are still described by a combination of letters and numbers given to them by the authors who started their analysis; these centers are identified by their spectral signature.

By a concerted effort of various experimental methods reviewed by Stavola (1998, 1999a), it is possible to identify some deep-level centers unambiguously. These methods include optical absorption, luminescence emission and excitation spectroscopy, electron-spin resonances (ESR, ENDOR), optical detection of magnetic resonances (ODMR), x-ray studies (EXAFS), and deep-level transient spectroscopy (DLTS). An important contribution was the improved growth techniques of ultrapure and stress-free crystals, which eliminate disturbing influences of the defect environment. In spite of new results, other deep level centers seem to escape generally accepted identification. An important problem with deep centers is their tendency to form associates or to incorporate into their structure major lattice deformation of decisive consequences.

In such a complex situation, we will introduce stepwise certain elements characteristic of deep centers.

1.1 General Properties of Deep-Level Centers

First, we will show that the eigenvalue spectrum of deep-level defects is substantially different from hydrogen-like centers. For its instructional value, the very simple, one-dimensional example of a deep potential well connected to *one band only* is presented first.

1.1.1 Model of Square-Well Potential

A means of introducing the chemical individuality of a center is provided by assuming a rectangular one-dimensional well of depth $-V_0$ and width $2a$. A very similar approach is used to evaluate a level spectrum in two-dimensional quantum wells and in superlattices (Sect. 3.1 of ▶ Chap. 8, "Bands and Bandgaps in Solids"). The steady-state electron behavior is described by a solution of the Schrödinger equation

$$\frac{d^2\psi}{dx^2} - k^2\psi = 0 \quad \text{with} \quad k^2 = \frac{2m[E - V(x)]}{\hbar^2}. \tag{1}$$

Here V is used again as potential *energy* (eV), following the conventional use. With arguments similar to those in Sect. 1.2.1 of ▶ Chap. 6, "The Origin of Band Structure", we can show that solutions of Eq. 1 exist for k values that are solutions of the transcendental equations:

$$\sqrt{\mu^2 + k^2} = k\tan(ka) \quad \text{or} \quad \sqrt{\mu^2 + k^2} = -k\cot(ka), \tag{2}$$

where

$$\mu^2 = \frac{2m|V_0|}{\hbar^2}. \tag{3}$$

These solutions can be obtained graphically from the intersection of the left- and right-hand sides of Eq. 2, shown in Fig. 1.

With discrete values of k obtained as solutions of Eq. 2, the permitted values of E inside the well are discrete and real. These are given by

$$E_n = \frac{\hbar^2}{2m_0}k_n^2 + V_0, \tag{4}$$

with $k_n \cong n\pi/a$ for lower values of n (see Fig. 1). The electron rest-mass m_0 is used here since the electron remains close to the center and does not move through the lattice. The eigenstates of such wells are approximately

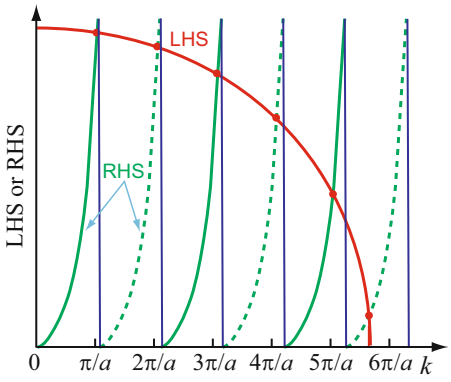


Fig. 1 Left-hand (*LHS*) and right-hand sides (*RHS*) of Eqs. 2 (first equation: *solid curve*; second equation: *dashed curve*) with solutions indicated by *dots* (intersections of LHS and RHS)

$$E_n \cong \frac{\hbar^2 \pi^2}{2m_0 a^2} n^2 + V_0, \quad \text{where } n = 1, 2, \ldots \tag{5}$$

and increase *quadratically* with n. The individuality of each center is given by a different a and V_0.

In contrast, the eigenstates of a simple *hydrogen-like defect* are given by

$$E_n = \frac{m_n e^4}{2\hbar^2 (4\pi \varepsilon_{\text{stat}} \varepsilon_0)^2} \frac{1}{n^2}, \quad \text{where } n = 1, 2, \ldots. \tag{6}$$

and decrease $\propto 1/n^2$, with eigenstates *converging toward the continuum* of free states at the edge of the conduction band (see Fig. 2b).

The use of a better central-cell potential $V(r)$ of the defect in the Schrödinger equation would yield more realistic results for the deep-defect level spectrum. However, the inclusion of additional bands into the model of deep centers is more important. In fact, it is this connection to more than one band extremum that permits the distinction between deep and shallow levels, as will be discussed in the following section. This will replace the linear relation between well depth (V_0) and the depth of the ground state with a much compressed relationship, as indicated in the example shown for the square well in Figs. 3 and 16; see Sects. 1.1.2 and 2.5.

1.1.2 Coulomb Tail and Deep-Center Potential

In addition to the central-cell potential, we have to consider the long-range Coulomb potential of charged deep centers. The Coulomb tail determines higher excited states of these centers and renders them hydrogen-like, similar to shallow centers. Consequently, one or several deep levels are observed, followed by a series of hydrogen-like shallow levels close to the respective bands (Pantelides and Sah 1974).

The model-potential combination of a square well and Coulomb potential describes the ground state of centers with the Coulomb potential predominating

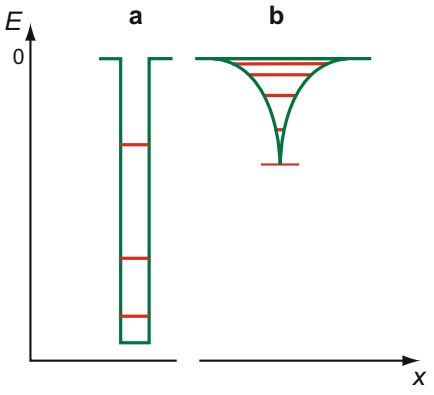


Fig. 2 Electron eigenstates (**a**) in rectangular well and (**b**) in a Coulomb-attractive well of a quasi-hydrogen defect

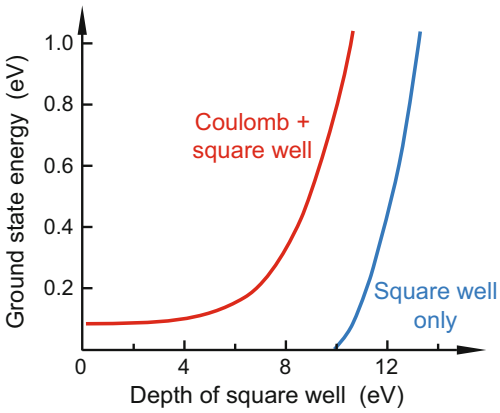


Fig. 3 Influence of an increasingly attractive square-well potential, added to a screened Coulomb potential for $\varepsilon_{stat} = 10$, $m_n/m_0 = 0.1$, well diameter 5 Å (After Vogl 1981)

until, with increasing well depth, the short-range part of the potential becomes very large (>10 eV in Fig. 3). The ground state of the center then shows typical deep-level behavior, here calculated properly with interaction of valence *and* conduction bands.

The *atomic electronegativity*[2] can be used as an indicator for the depth of the square well representing the core potential. Figure 4 shows the experimentally observed chemical trend: namely, a flat branch for Coulomb-dominated centers and a steeply decreasing branch for core-dominated centers, which have a deeper well potential. The reason for this empirical relationship will become clear in the discussion of Sect. 2.5.

Some important results obtained for deep-level defects can be summarized as follows:

- Deep centers are connected to more than one valence- or conduction-band edge and may be connected to both valence *and* conduction bands.
- The ground-state energy of the deep level varies at least an order of magnitude less than the impurity potential.
- Although the short-range potential is dominating, the eigenfunctions of some of the deep-level impurity centers, such as substitutional chalcogens in Si or III-V compounds, extend well beyond nearest neighbors and do not change much with the chemistry of the impurity if incorporated at the same site (Ren et al. 1982).

1.2 Theoretical Methods to Analyze Defect Centers

The theoretical methods deal with approximations for solving the Schrödinger equation of the defect within the lattice environment.

[2]The atomic electronegativity is defined as the difference between the *s* energies of host and impurity atoms for donors and the respective *p* energies for acceptors.

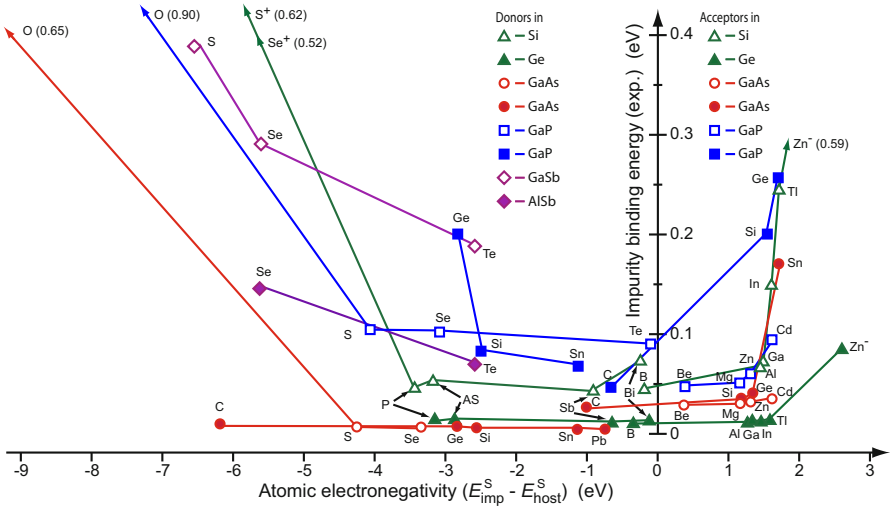


Fig. 4 Ionization energy of donors and acceptors in some host semiconductors, plotted versus the atomic electronegativity given by the difference of *s* or *p* energies, respectively, of impurity and host atoms (After Vogl 1981)

1.2.1 Perturbative Methods and Green's Function Technique

Perturbative Methods Perturbative methods use a defect potential that can be written as

$$V = V_0 + U \quad \text{with the Hamiltonian} \quad H = H_0 + U, \tag{7}$$

where H_0 is the one-electron Hamiltonian of the unperturbed lattice and U is the defect perturbation. The eigenfunctions of the corresponding Schrödinger equation $H\psi_\nu = E_\nu\psi_\nu$ are determined by expanding ψ_ν in terms of a complete set of functions ϕ_ν:

$$\psi_\nu = \sum_\lambda F_\lambda \phi_\lambda. \tag{8}$$

As such a set of functions, one may choose Bloch functions (as done in Sect. 1 of ▶ Chap. 18, "Shallow-Level Centers"), Wannier functions, or other orthonormalized functions – for instance, simple exponentials or Gaussian orbitals. The eigenvalues are then obtained from the secular matrix.

When the range of the perturbation potential is shorter, one advantageously uses Wannier or other *localized* functions for the expansion of ψ_ν. Thus, one obtains the eigenvalues from the corresponding Koster and Slater (1954a, b) determinant (see also Bassani et al. 1969; Jaros and Brand 1976).

Green's Function Technique Self-consistent Green's-function calculations powerfully determine the differences between an ideal crystal and the changes introduced

by the defect center, recognizing their *localization in space.* Although more complicated than cluster calculations outlined below, they supply more accurate solutions (Bernholc and Pantelides 1978; Baraff and Schlüter 1980).

This method has been used extensively, providing very valuable information on deep-level defect centers (Hjalmarson et al. 1980). For a brief review, see also Pantelides (1986). In more recent years, the computational complexity of Green's-function calculations has led to a decrease in their use.

1.2.2 Cluster Calculation and Supercell Technique

The eigenstates of a deep center can be estimated by considering only the atoms in its neighborhood, that is, in an atomic cluster (Messmer and Watkins 1973). Cluster calculations are carried out by calculating the eigenfunctions of such a group of atoms, treating it as a large molecule. When initially calculating it with atoms from the ideal crystal, and then inserting the impurity into its center, one obtains information on its energy-level structure. Although these cluster calculations are easily implemented, they converge slowly with cluster size and the results are very sensitive to conditions at the cluster boundary. Moreover, the defect-level energy is not very accurate, and corresponding changes of band states are difficult to obtain.

An example for a diamond crystal is given in Fig. 5. A level splits into the bandgap when a nitrogen atom is incorporated into this cluster. In addition, a large shift to lower energies is seen in Fig. 5, when lattice relaxation in the neighborhood of the N atom is permitted (Jahn-Teller distortion-see Sect. 1.3 and Sturge 1967). The replacement of a carbon atom by a nitrogen atom causes a substantial trigonal distortion of the four neighbor atoms (see Kajihara et al. 1991; Briddon and Jones 1993).

Supercell Technique Related to a cluster calculation is the supercell technique, in which the defect center is placed periodically in an otherwise perfect crystal. This technique replaces the questionable boundary condition for a cluster with less problematic *periodic* boundary conditions. The method, however, produces artificially broadened defect levels caused by defect interaction (Louie et al. 1976; Kauffer et al. 1977). With large computers, this effect can be minimized by using larger clusters for each cell. Currently, the supercell technique is a valuable tool. For a recent calculation of point defects in group III nitrides from first-principles, see, e.g., Gao et al. (2019); for a review, see Freysoldt et al. (2014).

1.2.3 Semiempirical Tight-Binding Approximation

The chemical trend of deep impurities can be obtained rather well from a semiempirical pseudo-Hamiltonian based on a small number of pseudo-orbitals (one *s-,* three *p-,* and one excited *s** orbital, sometimes complemented by five *d* orbitals; see Sect. 2.1.4 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals") for substitutional impurities in tetrahedrally bonded semiconductors. The approximation is based on the band-orbital model of Harrison (1973), and adopted by Vogl et al. (1983), to reproduce valence- and conduction-band structures of semiconductors. Hjalmarson et al. (1980) employed the same model to obtain information on the chemical trend of deep levels of substitutional impurities-see Sect. 2.5.

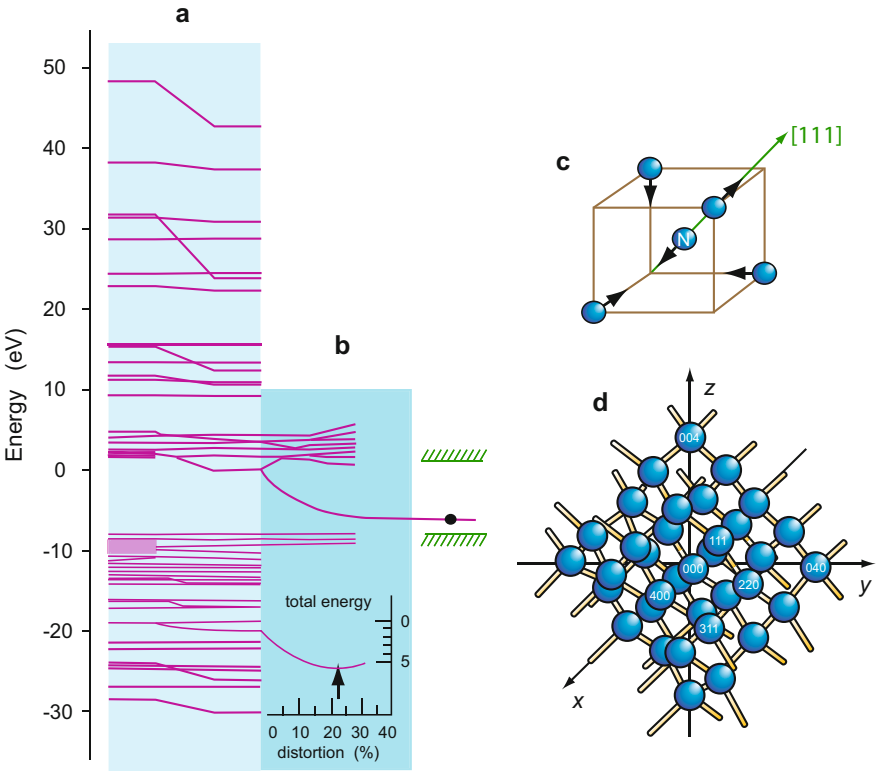


Fig. 5 (**a**) Electron energy-level distribution for the cluster of 35 carbon atoms (left half of **a**) depicted in (**d**) and including a nitrogen atom in its center (right half of **a**). (**b**) Energy shifts in the spectrum when the carbon atoms surrounding the center nitrogen atom relax by a Jahn-Teller distortion; the shift is shown as a function of the lattice relaxation (minimum total energy at 26% of the nearest-neighbor distance). (**c**) Unit cell and (**d**) 35 atom cluster (After Watkins and Messmer 1970)

1.3 The Jahn-Teller Effect

The Jahn-Teller effect is the intrinsic instability of a complex with electronic degenerate states against distortions which can remove this degeneracy. The effect was predicted by Jahn and Teller (1937) as a general phenomenon in any symmetric nonlinear configuration of atoms. Besides isolated molecules, there are many defects in solids affected by the Jahn-Teller effect, such as the F center in halides, the vacancy in silicon, and transition-metal impurities in elemental and compound semiconductors. For reviews, see Sturge (1967), Bersuker (1984), and for paramagnetic crystals Bates (1978). A first-principles density functional calculation of the Jahn-Teller effect in the tetrahedral T_d symmetry of diamond is reported by Prentice et al. (2017) and Thiering and Gali (2017).

When occupied states of the defect split, at least one of the states will be lower than the degenerate state of the undisturbed defect, and the energy of the defect can

be lowered by a Jahn-Teller stabilization energy δE. The local symmetry of the distorted defect is lower than the point symmetry of the crystal. The distortion that removes the electronic degeneracy is itself degenerate with respect to spatially equivalent configurations[3]: there is more than one equivalent position of the distorted center with equal energy. The energy of these equivalent configurations is separated by a potential barrier.

One distinguishes between *static* and *dynamic Jahn-Teller effects.* In the static effect, the lowering of the site symmetry is detectable, for example, in magnetic resonance studies; the symmetry of the defects is frozen-in in the equivalent configurations. In the dynamic Jahn-Teller effect, the energy barrier between the configurations is small compared to the thermal energy (or allows for quantum-mechanical tunneling). The distortion then averages out to zero during the characteristic time of the experiment, yielding the point symmetry of the undisturbed system.

Ham Effect In the presence of a Jahn-Teller effect, the orbital momentum **L** and the spin-orbit coupling λ**LS** are reduced by a factor γ. This effect was described by Ham (1965). The reduction γ depends on the Jahn-Teller energy δE and the energy of an effective phonon $\hbar\omega$, yielding for a strong Jahn-Teller effect $\gamma \cong \exp(-3\delta E/2\hbar\omega)$. A strong effect given by a large stabilization energy δE hence leads to a strongly reduced spin-orbit splitting of energy levels. This is often observed in the fine structure of zero-phonon transitions in optical spectra of transition-metal impurities in semiconductors (see Sect. 3.3), albeit a *quantitative* interpretation usually requires complex calculations; for such an analysis on d^5 ions in cubic symmetry, see Parrot et al. 1996. The reduction of magnetic moments of impurities in cubic environment was also calculated more recently by first-principles cluster calculations (Iwahara et al. 2017).

1.4 Crystal-Field Theory

Some qualitative information about the electronic behavior of deep centers can be obtained by starting from the electron eigenvalue spectrum of the isolated impurity atom in vacuum. Then one can determine to what extent this spectrum is influenced after the atom is exposed within the crystal to the field of the surrounding atoms, the *crystal field.* This crystal field is used then as a perturbation.

Such a description is relatively simple when the symmetry of the surrounding lattice environment is known. When introduced as a substitutional impurity without lattice relaxation, the symmetry of the lattice environment is that of the undisturbed crystal.[4]

[3]The electronic degeneracy is hence replaced by a spatial degeneracy. It should be noted that lifting of the degeneracy does not apply for the two-fold Kramers degeneracy, but for both spin and orbital degeneracy. A level with only spin degeneracy (e.g., the ^6S ground state of Mn^{2+}) leads, however, only to a minor effect on the order below 0.1 meV, while the effect of orbitally degenerate levels may be well in the 100 meV range.

[4]However, in actuality, deformations of the surrounding lattice result, with consequent lowering of the symmetry.

This causes splitting of degenerated energy levels of the free atom: the eigenfunctions of any free atom *in vacuum* must be invariant against rotation and reflection, resulting in a large degeneracy of the eigenvalues. However, this is no longer true *within a crystal*, where the point group of the lattice determines the remaining degeneracies with lesser symmetry.

The *crystal-field theory* deals only with the symmetry-related influence of the surrounding atoms but neglects the effects of the neighboring valence electrons. Therefore, it specifically addresses electrons in deeper shells that are partially filled and are shielded from the influence of other valence electrons. Such impurities are transition-metal atoms. The splitting of their d orbitals in a crystal field may be described in terms of an Orgel diagram, where the energy levels are plotted as a function of the crystal-field strength Dq, where D is a symmetry-dependent parameter of the potential and q is the point charge of the ligands (Orgel 1955, see also McClure 1959, and Schläfer and Gliemann 1967). For Orgel diagrams of various symmetries, see König and Kremer (1977).

For an illustrating example, the level splitting is discussed for an atom with two d electrons (for instance, Ti or Zr) substituting for an atom of a host with O_h (cubic) symmetry, such as Si or Ge. Each of these electrons has 10 states available with $l = 2$, $m = -2, -1, 0, 1, 2$, and spin $s = \pm 1/2$, resulting in 45 different states for the two electrons, distinguished by their total quantum numbers L and S:

One	^1S-state with $L = 0$	$S = 0$
Nine	^3P states with $L = 1$	$S = 1$
Five	^1D states with $L = 2$	$S = 0$
Twenty-one	^3F states with $L = 3$	$S = 1$
Nine	^1G states with $L = 4$	$S = 0$

The splitting of these five levels ^{2S+1}L becomes transparent after sequentially introducing electron–electron interaction and crystal field. This is shown in Fig. 6. Electron–electron interaction results in a splitting of the d^2 level into the five levels ^{2S+1}L, distinguished by L as shown in the second column of Fig. 6 (the "isolated impurity atom in vacuum"). The addition of the crystal field results in a further splitting into levels $^{2S+1}\Gamma$ according to their transformation behavior in the crystal symmetry (here O_h): ^1D into 2, ^3F into 3, and ^1G into 4 levels, as shown in the third column of Fig. 6. When the interaction among levels $^{2S+1}\Gamma$ of equal symmetry is finally taken into consideration, many levels are shifted substantially (without further splitting), presenting a reordered level arrangement as given in the fourth column of Fig. 6.

It is instructive to reverse the sequence of a hypothetical interaction by first neglecting the electron–electron interaction[5] (strong-field scheme, Tanabe and Sugano 1954a, b): one obtains the d^2 level splitting by crystal-field interaction first. This results in a split into one e_g level and one t_{2g} level for each d electron, that is, for the two d electrons into three levels with two electrons in e_g or t_{2g}, or one

[5]The use of lower- and upper-case letters to describe the states gives a good example to distinguish between one- and multielectron states. The latter includes electron–electron interaction.

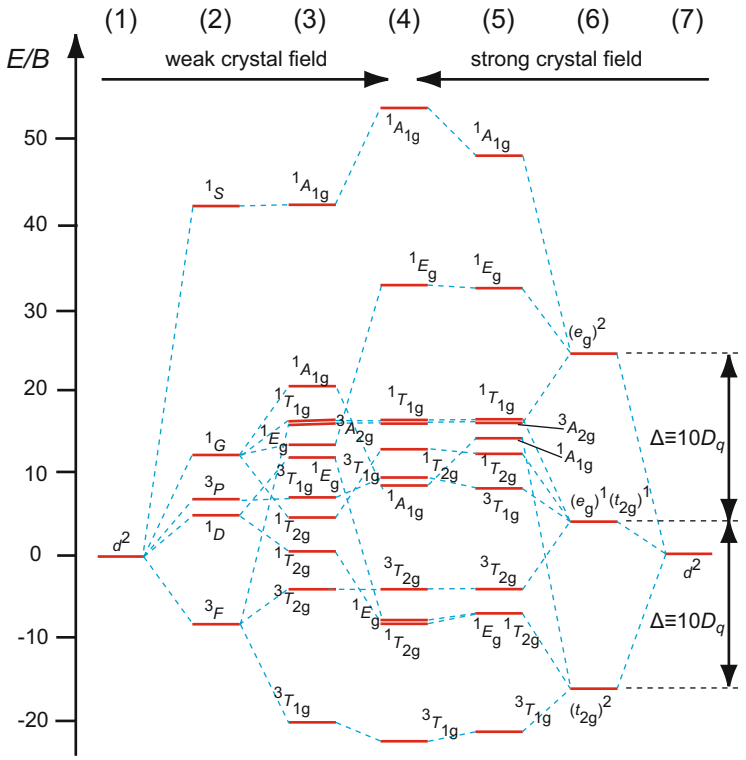


Fig. 6 Splitting of d^2 levels in cubic lattices. Weak-field scheme: (*1*) unperturbed atoms, (*2*) electron–electron interaction only, (*3*) crystal-field interaction added, (*4*) term interaction added; strong-field scheme: (*7*) unperturbed atoms, (*6*) crystal field only, (*5*) electron–electron interaction added, (*4*) configuration interaction added. Energies are scaled in units of Racah's parameter *B*. (After Schläfer and Gliemann 1967)

electron in each e_g and t_{2g} states, as shown in the sixth column of Fig. 6. The splitting Δ between an e_g and a t_{2g} state equals 10 Dq; the index g (for German *gerade*) indicates even parity. Adding the electron–electron interaction further splits $(e_g)^2$ into three, $(e_g)^1$ $(t_{2g})^1$ into four, and $(t_{2g})^2$ into three levels, as shown in the fifth column of Fig. 6. Finally, the level interaction causes level shifting and yields the same results as for the previously discussed center column.[6]

The weak- and strong-field schemes discussed above refer only to the orbital momentum and do not account for the spin-orbit coupling; this causes an additional (small) splitting. The corresponding optical spectra of some of these transition metal

[6]An Orgel diagram is therefore also referred to as Tanabe-Sugano diagram.

impurities in wide-bandgap material are therefore quite complicated – see McClure (1959), and Schläfer and Gliemann (1967). In semiconductors of smaller bandgap, only a few of these levels fall within the bandgap and are easily identifiable-see Ludwig and Woodbury (1962), Milnes (1983), and Bates and Stevens (1986).

The inverse field effect on holes and electrons leads to the same multiplet structure of the ground-state splitting for d^N and d^{10-N} atoms ($N = 1 \ldots 4$), with an inversed order of the crystal-field levels (Fig. 7): a d shell filled with N electrons corresponds to a completely occupied d shell filled with 10-N holes, and the electrical field of the crystalline environment has an inverse effect on electrons and holes.

A corresponding relation exists between the ground-state levels of a d^N atom in crystal fields of either octahedral (O_h) or tetrahedral (T_d) symmetry. Both derive from cubic symmetry, but atoms in a tetrahedral environment are placed off-axis with respect to cubic coordinates; as a consequence the term sequence is reversed. The ground state of d^2 shown in Fig. 7 for an octahedral crystal field is a ${}^3T_{1g}({}^3F)$ level; in tetrahedral environment, this changes to the ${}^3A_2({}^3F)$ level. The index g is omitted here, since tetrahedral symmetry has no center of inversion in contrast to octahedral symmetry. The change of the crystal field for a change from octahedral to tetrahedral crystal field can generally be written

$$\Delta_{\text{tetrahedron}} = -\frac{4}{9}\Delta_{\text{octahedron}}. \tag{9}$$

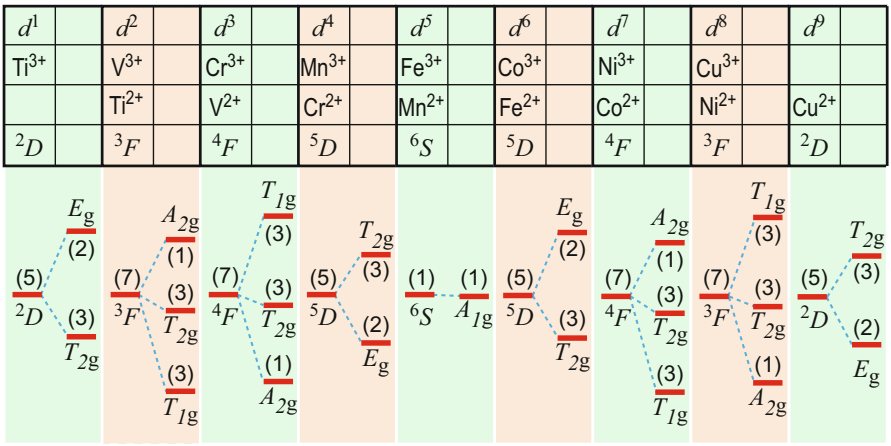


Fig. 7 Schematic level splitting of the ground states of d^n atoms in a crystal field with octahedral (O_h) symmetry

2 Deep Centers in Semiconductors

A great variety of deep centers exist in semiconductors; however, not many of them can be unambiguously identified (Stavola 1998, 1999a; Madelung and Schulz 1989). Such identification was done very early for a variety of deep defect centers in alkali halides. For the purpose of illustration, these will be given first as examples. We will then discuss some of the more important deep-level defects in typical semiconductors.

2.1 Defects in Alkali Halides

2.1.1 Anion Vacancy: The F Center

Anion vacancies act like donors, and *cation vacancies* act like acceptors; however, both are deep centers. They are easy to identify by their optical absorption spectrum of broad isolated lines within wide bandgaps, and their unambiguous response to specific treatments, which stimulate unique defect reactions. These centers show a large amount of lattice relaxation (large Huang-Rhys factor-see Sects. 1.2.2 and 2.3.2 of ▶ Chap. 17, "Optical Properties of Defects") when recharged, providing excellent examples for electron-lattice interaction.

The classical example of an anion vacancy is an F center in an alkali-halide crystal[7] (Fig. 8). The missing negative charge of the anion is replaced by an electron in order to restore local neutrality. That electron is not as tightly bound to the vacancy as it was to the Cl^- anion in Fig. 8, which is now missing. The level associated with this defect therefore lies in the bandgap; in NaCl with a gap of 7.5 eV, the F center lies 2.7 eV below the conduction band. It becomes deeper with increasing lattice-binding strength. This is indicated in Fig. 8b as a function of the lattice constant that decreases monotonically with increasing binding strength. The vacancy changes its charge character with ionization from neutral to positive, relative to the lattice

$$V_{Cl}^{0} \longleftrightarrow V_{Cl}^{+} + e^{-} \tag{10}$$

and thus acts as a donor. *Excited states of the F center* have been observed by Lüty (1960) and are referred to by the symbols *K, L_1, L_2,* and L_3 (Chiarotti and Grassano 1966). These levels are probably resonant states with *X* minima of the conduction band. They are located with large energy spacings of ~0.6 eV for each consecutive L_i level in KCl, which are typical for deep centers.

Replacing the missing ion with an electron does not completely restore the ideal lattice periodicity. The resulting lattice perturbation produces another level which lies

[7] F centers ("Farb" centers: German for color centers) were the first lattice defects correctly identified and described by their electronic structure by Pohl and coworkers (see the review by Pohl 1938). Later associates of two, three, or four F centers were observed and referred to as M, R, and N centers (N_1 in planar and N_2 in tetrahedral arrangement) – see Schulman and Compton (1962).

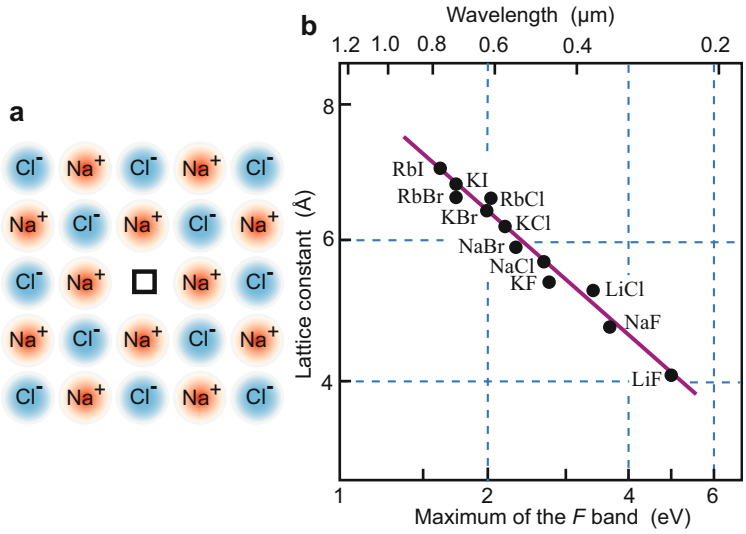


Fig. 8 (**a**) F center in a sodium chloride crystal. (**b**) Position of the maximum of the F center absorption as a function of the lattice constant (After Mollwo 1931)

closer to the respective band: a second carrier can be trapped in this level. This center, derived from an F center, is called **F' center** and returns to an F center when ionized:

$$F^- \rightleftarrows F^0 + e^- \tag{11}$$

In the example of KBr, the ionization energy of the F^- center (traditionally called F' center) is 1.4 eV, which is substantially less then the ionization energy of the F center: 2.05 eV.

In addition, the surrounding lattice near the vacancy is also perturbed; its eigenstates split off into levels in the bandgap. When the vacancy is empty, the resulting perturbation is larger than when this vacancy has an electron trapped in it. These levels are shown schematically in Fig. 9 and are called ***α*** or ***β*** **bands**. Such levels are observed in alkali halide crystals. They result in a decrease of the bandgap energy in the immediate neighborhood of the defect for KBr from 6.55 to 6.44 and 6.15 eV, respectively, at 90 K. These can be interpreted as excitons trapped at the empty vacancy or at the F center.

2.1.2 Other Centers in Alkali Halides

A cation vacancy in an alkali halide is called a **V center** and acts as an acceptor. It is deeper than the F center in the same crystal and less sharp, probably because of stronger lattice relaxation or of distant pair formation. Nearest-neighbor associates of such centers are termed V_2 and V_3 centers (Seitz 1954).

There are more intrinsic centers which can be formed. These include the $\mathbf{V_k}$ **center,** the **I center,** and the **H center,** all of which are related to halogen defects (see

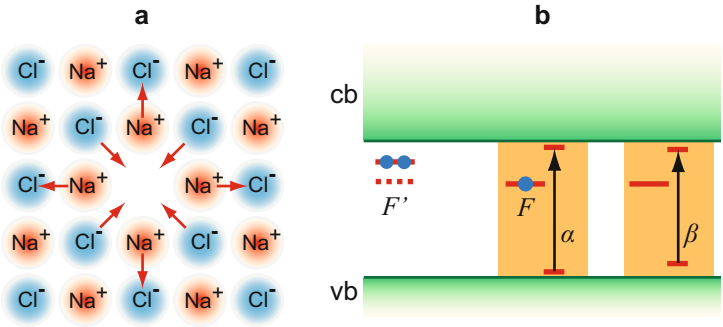


Fig. 9 (**a**) Disturbance of the anion and cation lattice around an anion vacancy in NaCl. The displacement is indicated by *arrows*. (**b**) Resulting split-off levels near valence band (*vb*) and conduction band (*cb*) and corresponding transition β. A smaller perturbation is observed if an electron is trapped, i.e., near an F *center*, resulting in a split-off with transition α. These perturbations represent the corresponding energies of the trapped excitons (Sect. 2 of ▶ Chap. 18, "Shallow-Level Centers")

e.g., Castner et al. 1958) and are, respectively, a self-trapped hole, a halogen ion interstitial, and a halogen molecule on a single halogen lattice site (Itoh 1982). A few of these *color centers* are shown in Fig. 10. Each of these defects has a characteristic signature in optical absorption, luminescence, or in spin resonance. The wide bandgap and tight-binding of defects in alkali halides permit separate identification and give convincing evidence of a wealth of intrinsic defects, which are much more difficult to identify in common semiconductors. For a review, see Schulman and Compton (1962), Fowler (1968), Lüty (1973), Williams (1978), Itoh (1982), and Hayes and Stoneham (1984).

2.2 Vacancies in Semiconductors

2.2.1 Vacancies in Covalent Crystals

The vacancy is one of the most important centers in covalent crystals. The change in the charge distribution when introducing a vacancy into the Si lattice is shown in Fig. 11. We will return to this topic in Sect. 2.5.

Vacancies in covalent crystals result in deep levels which can have several occupation states. Typically, one distinguishes five charge states of the vacancy V^{2+}, V^{+}, V^{0}, V^{-}, and V^{2-}; two of them, V^{+} and V^{-}, are observed in Si in spin resonance and two by diffusion experiments (Watkins 2000). The breaking of covalent bonds results in *dangling bonds*, which combine within the vacancy to form molecular orbitals – a singlet state with a_1 and a triplet state with t_2 symmetry. These yield the ground states of the differently charged vacancies by population of the states with electrons of appropriate spin (Fig. 12). For Si, the corresponding states are $V^{2+}\left(a_1^2\right)$, $V^{+}\left(a_1^2 t_2\right)$, $V^{0}\left(a_1^2 t_2^2\right)$, $V^{-}\left(a_1^2 t_2^3\right)$, and $V^{2-}\left(a_1^2 t_2^4\right)$ –

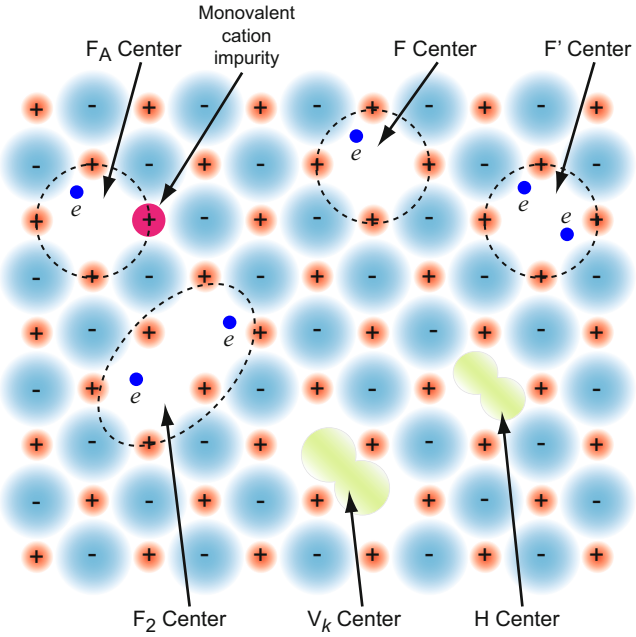


Fig. 10 Some of the color centers in alkali halides (After Hayes and Stoneham 1984)

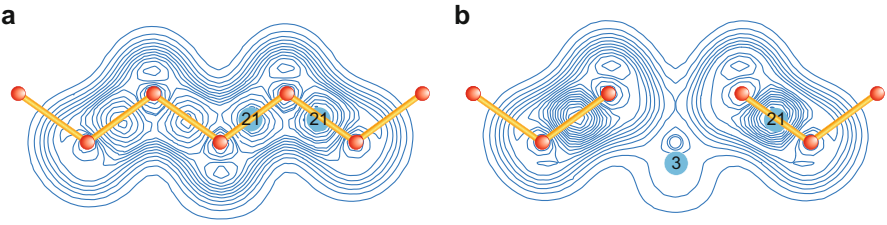


Fig. 11 Electron-density profile in the (110) plane of Si. (**a**) Ideal atomic array, (**b**) array perturbed by center vacancy; values are in units of electrons per bulk unit cell (After Baraff and Schlüter 1979)

see Watkins (2000) or Hayes and Stoneham (1984). A strong tetragonal Jahn-Teller distortion lowers the symmetry and significantly shifts and splits these levels as shown in Fig 12a. The unpaired electron in the V_{Si}^{+} and V_{Si}^{-} states is spread equally over four or two of the surrounding atoms, respectively. The entire electronic behavior of the vacancy can be explained in a one-electron model (Pantelides and Harrison 1976).

The states V^{+} and V^{-} are observed by ESR resonances of unpaired electrons (Fig. 12a), and V^{0} and V^{2-} are indirectly observed after photoexcitation. The $V^{+} \rightarrow V^{0}$ transition is seen to shift below the $V^{2+} \rightarrow V^{+}$ transition, which is typical for a negative-U behavior – see Sect. 2.8. This behavior was predicted by Baraff

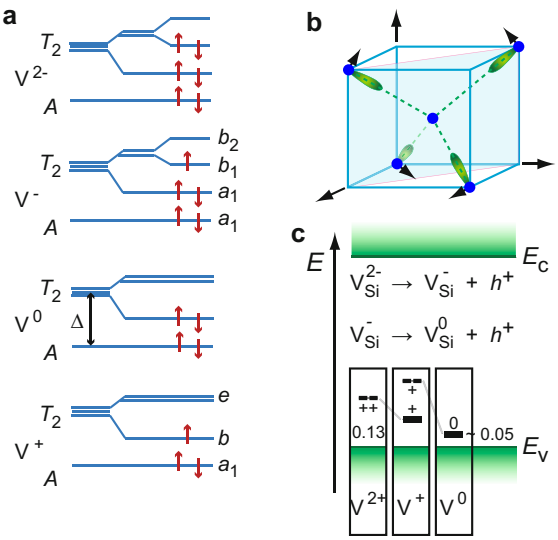


Fig. 12 (**a**) Electronic structure of Si vacancies with symmetry and occupation of one-electron orbitals indicated. (**b**) Jahn-Teller distortion (*arrows*) in the neutral vacancy. (**c**) Shift of levels before (*left*) and after lattice relaxation (*right*) with charge character indicated (After Watkins 2000)

et al. (1980a, b) and later experimentally confirmed by Watkins (1984). See also Stoneham (1975), Lipari et al. (1979), and Jaros et al. (1979).

The V_{Si}^{+} state is metastable and disproportionates to V_{Si}^{0} and the stable V_{Si}^{2+} (Fig. 12c), which lies ~0.13 eV above E_v. This center releases *two* holes when excited and dissociates immediately after creation, since the V_{Si}^{0} center is shallower (0.05 eV).

Much progress has been made with the identification of intrinsic defects in diamond, in part due to the availability of high-quality synthetic diamonds that can be enriched with the ^{13}C isotope (nuclear spin $I = 1/2$) which yields a hyperfine satellite structure in ESR studies. The diamond vacancy has been studied by optical spectroscopy (Davies and Manson 1994) and ESR (Isoya et al. 1992).

A convenient method[8] for producing such vacancies is by bombardment with relatively fast electrons (Loferski and Rappoport 1958) at cryogenic temperatures, and trapping the cogenerated, highly mobile interstitial (Sect. 3.2 of ▶ Chap. 15, "Crystal Defects") at other lattice defects, e.g., group III atoms (Watkins 2000). The threshold energy for producing vacancies with electrons in Ge and Si is 14.5 and 12.9 eV, respectively. Vacancies in *binary semiconductors* can be analyzed in a similar fashion, taking into account the distinction between cation and anion vacancies.

2.2.2 Vacancies in Compound Semiconductors

Vacancies can play an important role in compound semiconductors. They may, for example, limit the achievable doping level: the Ga vacancy in GaAs acts under As-rich conditions as a compensating center at high *n*-type doping (Zhang and Northrup 1991; Baraff and Schlüter 1985). Calculated energy for the formation of

[8]Quenching from high temperature is not fast enough to freeze-in measurable densities of vacancies (Watkins 1986).

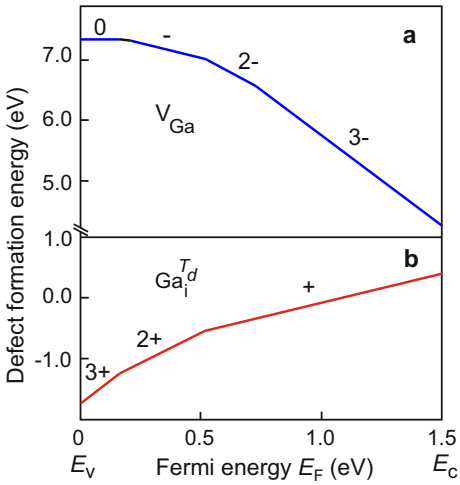


Fig. 13 Calculated formation energies of defects in GaAs. (**a**) Isolated vacancy at the Ga site V_{Ga}, and (**b**) isolated interstitial Ga atom Ga_i at a tetrahedral site with four nearest As atoms. The origin of the Fermi-energy scale is set to the maximum of the valence band E_v (After Baraff and Schlüter 1985)

the Ga vacancy V_{Ga} is given in Fig. 13; it depends on the position of the Fermi energy E_F, and the charge-transfer energies are in agreement with more recent calculations (El-Mellouhi and Mousseau 2005, Komsa and Pasquarello 2012). For a Fermi level near the conduction-band edge in *n*-type GaAs, the interstitial V_{Ga} defect with threefold negative charge has lowest formation energy. The defect hence traps electrons, thereby compensating the extrinsic doping; if the defect traps Z electrons, an energy of $-Z(E_F - E_d)$ is gained, where E_d is the defect energy-level. As the Fermi level is lowered, less electrons are trapped by V_{Ga} due to an increased formation energy. Simultaneously, the formation energy of another intrinsic defect decreases, that of interstitial gallium Ga_i (Fig. 13b). Hence, gradually more holes are trapped by Ga_i as E_F approaches the valence-band edge E_v. Since both native defects can be created by pair formation described by the reaction $0 \leftrightarrow Ga_i + V_{Ga}$, *either* high doping level, *p*-type *and n*-type, is compensated.

Vacancies also mediate impurity diffusion (Northrup and Zhang 1993) and can cause deep-level luminescence that degrades the operation of optoelectronic devices. An example for this is the yellow luminescence in GaN, a broad luminescence band centered around 2.2 eV that commonly appears in *n*-type GaN. It is commonly accepted that this luminescence is caused by Ga vacancies, which introduce energy levels ~1.1 eV above the valence-band edge, and which forms preferentially in *n*-type GaN (Neugebauer and Van de Walle 1996a).

2.3 Self-Interstitials and Antisite Defects

Self-Interstitials Elemental semiconductors such as Si and Ge have a rather loosely packed lattice with coordination number 4 and sufficient space to accommodate interstitials of the host lattice, that is, *self-interstitials.* There are several possibilities for incorporating an additional Si atom in a Si lattice, as shown in Fig. 5a of ▶ Chap. 15,

"Crystal Defects". Three of them are indicated as *T*, *H*, and *B* and have tetragonal, hexagonal, and bond-centered geometry. After incorporation, the surrounding atoms relax to shifted positions.

The formation energy of the interstitials depends on their specific site as well as on their charge state. This is shown in Fig. 14, with the formation energy plotted as a function of the position of the Fermi level (Car et al. 1984); more recent DFT calculations basically confirm the trend, but with lower formation enthalpies and including negative charge states (Rinke et al. 2009; Centoni et al. 2005). Such dependency indicates that the recharging of an interstitial makes it more stable at a different position. Alternating recharging, for example, by recombination with excess carriers followed by thermal ionization at the changed position, can therefore stimulate diffusion-see Sect. 3.2 of ▶ Chap. 15, "Crystal Defects".

The Si_i^+ center is metastable; when recharging and relaxing into the Si_i^{2+} center, it shows a negative-*U* character. Si interstitials are deep-level centers with levels 0.6 and 0.8 eV below the conduction-band edge.

In compound semiconductors, the formation energy of native defects such as the self-interstitial also depends on the stoichiometry of anions and cations. Ga-rich conditions in the growth of GaAs, for example, favor the formation of Ga self interstitals (Baraff and Schlüter 1985). Their formation also depends on the Fermi level as shown in Fig. 13.

In the II-VI semiconductors ZnSe and ZnTe, families of Frenkel pairs have been studied (Watkins 2000). In ZnSe, for example, two distinct configurations for isolated Zn interstitial centers Zn_i^+ were identified by optically detected magnetic resonance, one at the tetrahedral site surrounded by Zn atoms and the other at the tetrahedral site surrounded by Se atoms. In this case, the migration of interstitials at cryogenic temperatures could be directly explored.

Antisite Defects Antisite defects are identified in some of the III-V compounds, for example, As_{Ga} in GaAs. This center can efficiently compensate both shallow

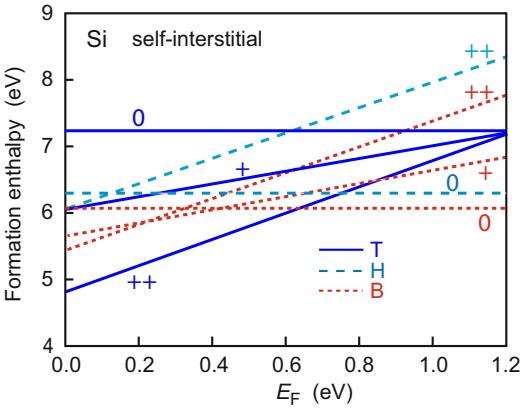


Fig. 14 Si self-interstitial formation enthalpy as a function of the position of the Fermi level for the different interstitial positions *T, H, B* and charging states (After Car et al. 1984). More recently a slightly lower formation enthalpy was calculated

acceptors and shallow donors in GaAs, consequently causing a reduction in semiconductivity, and allows for fabricating semi-insulating GaAs (Trautman et al. 1998). The defect by itself is the so-called EL2 center in GaAs (Chadi 2003; Kaminska and Weber 1993; Baraff 1992; see also Centoni et al. 2005). In excited metastable states, this antisite defect undergoes a large displacement away from T_d symmetry and likely forms associates related to As_i such as the $(V_{Ga} - As_i)$ complex. A comparable defect is the P_{In} antisite in InP (Sun et al. 1993).

In compound AB semiconductors, one distinguishes A_B and B_A as possible antisite defects. In higher compounds, such antisite formation is often more probable and presents a larger variety of defects: six in ABC compounds, although only a few of them are energetically preferred. For a discussion on the relation between stoichiometry and deep-level centers, see Lin et al. (1998).

2.4 Hydrogen in Semiconductors

Hydrogen in semiconductors requires special attention. It can be introduced into the semiconductor in several ways, either intentionally or unintentionally, where it diffuses rapidly and interacts with other impurities and defects. Hydrogen is known to passivate many deep centers and is therefore often used with great benefit for device fabrication, see Pearton et al. (1992), Pankove and Johnson (1991), Estreicher (1995), and Nickel (1999).

Hydrogen is sometimes introduced into semiconductors intentionally to passivate defects and improve the electrical properties of the materials. For example, the efficiency of solar cells is increased by the introduction of H due to the passivation of deep-level defects and grain boundaries. On the other hand, the unintentional introduction of H, for example, during epitaxial growth, can lead to the passivation of the desired shallow dopants. A prominent example is the unintentional passivation of Mg acceptors in GaN during epitaxial layer growth; the solution of this problem enabled the rapid development of the GaN-based device technology for lasers and solid-state lighting (Nakamura et al. 2000).

Experimental and theoretical studies have revealed a variety of interesting properties of hydrogen in elemental and compound semiconductors. In p-type silicon, for example, isolated H atoms occupy a bond-centered site where it acts as a donor which passivates shallow acceptors. In n-type Si, isolated H sits at a thetrahedral site where it acts as an acceptor which can passivate shallow donors. This amphoteric behavior originates from the ability of H to adopt different configurations with different electrical properties (Van de Walle 1991; Johnson et al. 1994). The position of the Fermi energy where the stable charge state of interstitial hydrogen changes from the H^+ donor state to the H^- acceptor state was computed for a wide range of host zincblende and wurtzite elemental as well as III-V and II-VI compound semiconductors by ab initio methods; this study also allowed for predictions of band aligments in heterojunctions (Van de Walle and Neugebauer 2003).

Many hydrogen-containing defects have been studied; for a survey of defects in Si see Stavola (1998, 1999a, b). The properties of hydrogenated shallow impurities in

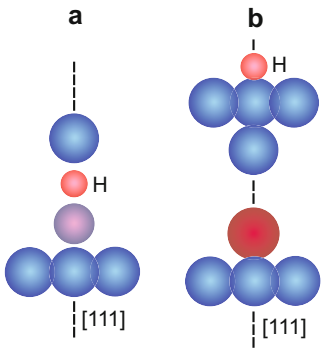


Fig. 15 Structure of (**a**) the acceptor-H and (**b**) the donor-H complex in Si (After Stavola 1999b)

elemental and compound semiconductors are fairly well understood. The structures of acceptor-H and donor-H complexes in Si are illustrated in Fig. 15. Many hydrogenated deep-level centers have also been studied. Examples are hydrogenated vacancies and interstitials in Si and also hydrogenated transition-metal impurities in Si.

2.5 Substitutional Defects

Substitutionals can be considered as impurities replacing a vacancy. Inserting a Si atom into a vacant site of a Si host can be considered as having its s and p orbitals interacting with the A_1 and T_2 states of the vacancy, producing bonding and antibonding states which merge with valence and conduction bands, respectively. Inserting an impurity with substantially lower energy orbitals will produce only a small shift in the levels, causing (a) an impurity-like hyper-deep bonding level in or below the valence band and (b) a vacancy-like antibonding level slightly lower than the vacancy states usually within the bandgap.[9] For impurities with energy levels higher than those of Si, the corresponding levels lie within the conduction band and slightly above the vacancy states for the bonding-antibonding pair.

The vacancy states are limiting cases for donor- or acceptor-like states to which both approach asymptotically with increasing depth of the binding potential of the electron or hole in the impurity. This indicates how the influence of two bands compresses the spectrum of ground states of a deep level from $\partial E/\partial V \cong V_0$ in the simple one-band model (Sect. 1.1.1) to the actual $|\partial E/\partial V|$ that is rapidly decreasing with increasing $|V_0|$, where V_0 is the depth of the potential well. This is shown schematically in Fig. 16.

This behavior is deduced from Green's function calculations using semiempirical tight-binding Hamiltonians (Hjalmarson et al. 1980). It has been used to calculate the chemical trend of numerous deep-level centers as a function of the impurity potential. These levels are shown in Fig. 17. For the impurities to the right of the intersect

[9]The A_1 state of the acceptor is always strongly bound and lies within or below the valence band, while the T_2 states may emerge from the conduction band into the bandgap.

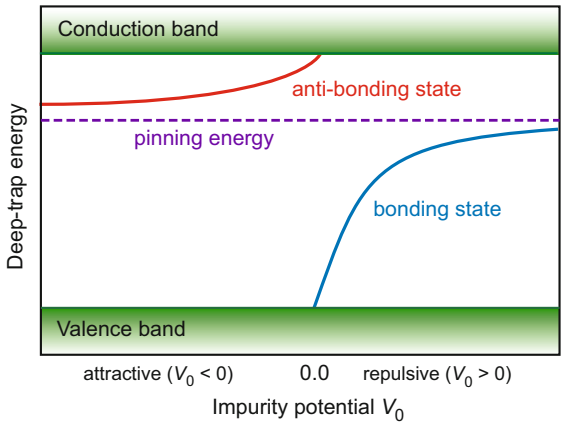


Fig. 16 Depth of deep donor and acceptor states as a function of their impurity binding-potential V_0. The asymptotic "pinning energy" is that of the electron state of the vacancy (After Vogl 1981)

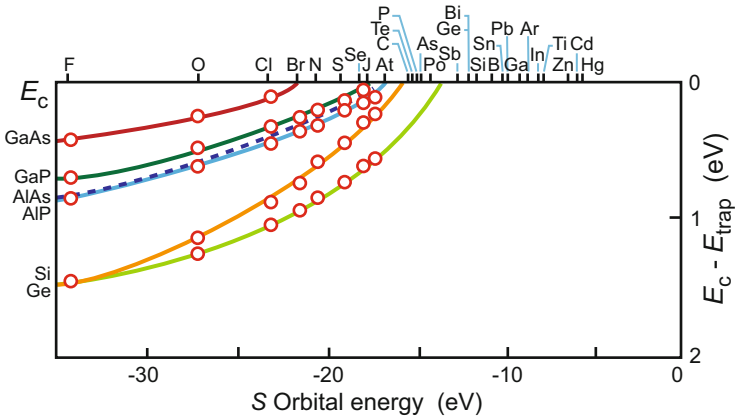


Fig. 17 Calculated ionization energies $E_c - E_{trap}$ for substitutional deep-level impurities on anion sites in the semiconductors listed at the left margin. Impurities with a smaller binding potential than shown by the intersection of the curves with E_c lie as resonant "deep" states within the conduction band (After Hjalmarson et al. 1980)

of the curves with the conduction-band edge in the figure, the impurity potential is not large enough to offer a bond state. Here, the effective-mass approximation for hydrogen-like states yields shallow levels connected to one band only (the conduction band in this example). For a short review, see Dow (1985).

When using an impurity potential comprised of a deep well and a Coulomb tail (e.g., like the Abarenkov-Heine potential) with

$$V(r) = \lambda V_0(r) + \frac{e^2}{4\pi\varepsilon_{\text{stat}}\varepsilon_0 r}, \quad \text{where} \quad V_0(r) = \begin{cases} V_0 & \text{for } r \le r_0 \\ 0 & \text{for } r > r_0 \end{cases}, \tag{12}$$

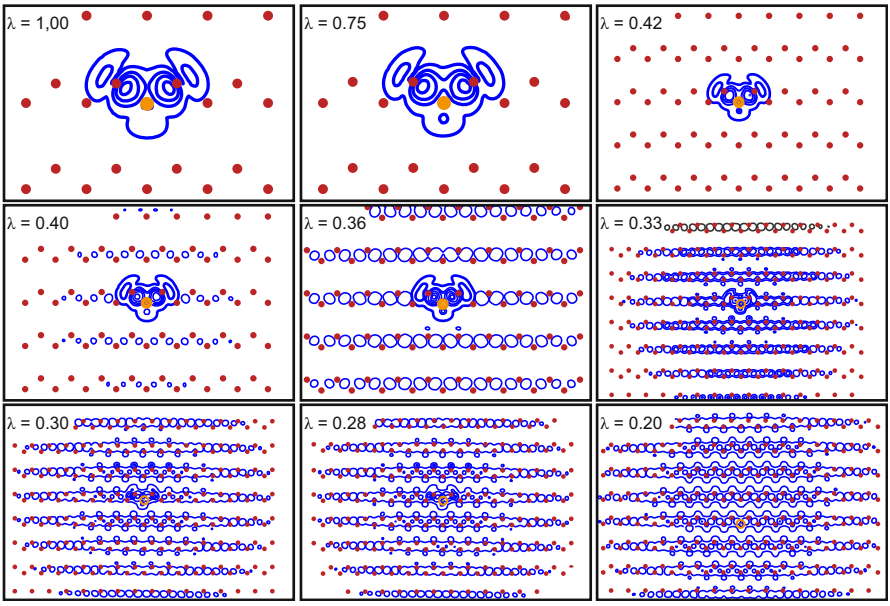


Fig. 18 Charge-density distribution of a substitutional impurity (*orange circle*) in Si, computed from a potential similar to Eq. 12 for decreasing strength of the short-range potential from $\lambda = 1.00$ to $\lambda = 0.20$ in the nine subfigures. $\lambda = 1$ corresponds to the T_2 state of a Si vacancy (After Pantelides 1986)

and a well deep enough to create a bound ground state in the bandgap, one can follow the transition from a deep-level center with essentially a vacancy-like charge-density distribution (dangling-bond-like) to a hydrogen-like (effective-mass-like) spread-out distribution with decreasing impurity potential (Pantelides 1986) as shown in Fig. 18.

The chemical identity of an impurity (i.e., its size, bonding type, and valency) is responsible for changes of the local symmetry after the incorporation of an impurity. Some of these changes are related to the symmetry-breaking lattice relaxation due to the Jahn-Teller effect discussed in Sect. 1.3. Others deal with the strength and angle relation of the bonding forces of the impurity in relation to the tetrahedrally arranged available dangling bonds of the impurity.

The arrangement of atoms nearest to an impurity has an important influence on the density of state (DOS) distribution of the resulting levels. This influence is strong enough to make the DOS of a substitutional sulfur donor in Si look rather similar to an interstitial Si atom, both in a tetrahedral environment (Vigneron et al. 1983).

Isoelectronic Defects Isoelectronic defects are formed by substitutionals from the same column of elements as the replaced host atom. Different homologous elements have a different energy spectrum, since the long-range potential – the Coulomb term – cancels: the replacing atom has the same valency. The remaining central-cell potential

reflects the chemical identity of the center. Replacing a middle-row host atom (e.g., P in GaP) with a highest-row atom (here N), one obtains a defect that acts as an electron trap with binding energy of 10 meV – see Sect. 2.3 in ▶ Chap. 18, "Shallow-Level Centers". Replacing it with a lower-row atom (here Bi), the defect acts as a hole trap with a binding energy of 38 meV (Dean et al. 1969). Replacement with mid-range atoms usually does not produce electronic defect centers, but instead produces *alloys* with the respective sublattice: As forms a $GaAs_{1-x}P_x$ mixed crystal, as does Sb, which forms $GaSb_{1-x}P_x$. There are no corresponding levels in the bandgap.

When traps are created, some of them can be rather deep; for instance, ZnTe:O, with $E_c - E_{trap} = 0.4$ eV, and CdS:Te, with $E_{trap} - E_v = 0.19$ eV (Cuthbert and Thomas 1967, 1968). An extensive review of the experimental observations of isoelectronic traps is given by Dean (1973). Theoretical models are quite sensitive to central-cell approximations (Faulkner 1968; Baldereschi and Hopfield 1972; Jaros and Brand 1979). First-principles calculations are necessary to predict the binding energy with reasonable accuracy.

In contrast to the hydrogenic effective-mass treatment of shallow donors or acceptors, where the central-cell corrections are often minor compared to the far-reaching Coulomb contributions, such short-range potentials are dominant. Furthermore, minor deviations (by 1%) in the estimated potential can result in major changes (by a factor of 2) of the electronic eigenstates of the defect.

Extensive theoretical and experimental work has been done with GaP. In the next Sect. 2.6, $GaP:O_P$ is discussed as one example. For reviews, see Stoneham (1975) and Pantelides (1978).

2.6 Chalcogens in Semiconductors

Chalcogens in Si Chalcogens (0, S, Se, and Te), incorporated in Si, act as deeper donors. They easily form associates, especially oxygen, as described by Wagner et al. (1984). Single substitutional donors are observed for neutral and singly charged S, Se, or Te: they act as double donors. The charge-density distribution of these donors has been calculated by Ren et al. (1982).

The ground state of neutral and single ionized donors is rather deep (Table 1) and shows a substantial chemical shift. In addition, the 1*s* ground state is split into *A, E,* and T_2 states, which are nondegenerate, doubly, and threefold degenerate, respectively; see Fig. 7a of ▶ Chap. 18, "Shallow-Level Centers". The *p* states are shallow and follow rather well the effective mass (hydrogen-like) approximation, as can be seen by comparison with the last column of Table 1.

The small difference of the ionization energy of 1*s* (A_1) states between S and Se is similar to the minor variance of the ionization energy of the free atoms, compared to a more substantial difference to Te. The larger hydrostatic pressure coefficient of the ground state compared to the hydrogen-like centers indicates the connection to conduction bands (in addition to valence bands). This shift is a result of a substantial shift in respect to the *X* valley. For a review, see Grimmeiss and Janzén (1986).

Table 1 Binding energy (in meV) and pressure coefficient (in meV/Pa) of ground and excited states of neutral and singly ionized chalcogen impurities in silicon (After Wagner et al. 1984)

	S^0	Se^0	Te^0	$S^+/4$	$Se^+/4$	$Te^+/4$	E_{qH}
$1s(A_1)$	318.2	306.5	198.7	153.3	148.3	102.8	31.27
$1s(T_2)$	34.6	34.5	39.2	46.05	41.5	44.3	
$1s(E)$	31.6	31.2	31.6				
$2p_0$	11.4	11.5	11.5	11.4	11.5	11.8	11.51
$2s$	9.37	9.3	9.7				8.83
$2p_\pm$	6.4	6.4	6.3	6.43	6.4	6.4	6.40
$3p_0$	5.46	5.47	5.5				5.48
$3p_\pm$	3.12	3.12	3.12				3.12
$3d_0$		3.8	4.0				3.75
$4p_\pm$	2.2	2.2	2.1				2.19
$5p_\pm$		1.5					1.44
$\partial E(1s)/\partial p$ ($\times 10^{-8}$)	−1.7	−1.8	−0.9	−2.05	−2.1	−1.2	

Oxygen-related centers are part of the family of so-called *thermal donors,*[10] some of which may be isolated oxygen centers, incorporated as interstitials, or combined with Si as an interstitialcy or a substitutional. For original literature on the different types of incorporation, see Wagner et al. (1984), Jones (1996).

Oxygen in GaP and GaAs Oxygen in GaP is one of the more extensively studied defect centers, since it was of practical interest for light-emitting diodes. Incorporated as a substitutional of phosphorus, it yields two deep centers stemming from the same defect:

$$O_P^+ + e^- \rightarrow O_P^0 \quad \text{and} \quad O_P^0 + e^- \rightarrow O_P^-. \tag{13}$$

The first center with $(E_c - E_d)^+ \cong 0.8$ eV has a rather small lattice coupling with a Huang-Rhys factor of $S \cong 3$: it relaxes to $(E_c - E_d)^0 = 0.96$ eV. The second center with $(E_c - E_d)^0 \cong 0.6$ eV relaxes to $(E_c - E_d)^- \cong 2.03$ eV; that is, it has a very large Huang-Rhys factor of $S \cong 30$. The corresponding band diagram is shown in Fig. 19. For a review, see Dean et al. (1983).

The strong lattice relaxation (see in ▶ Chap. 17, "Optical Properties of Defects", Sects. 1.2, 2.3 and Fig. 4) that occurs when a second electron is captured produces a lower level than that for the captured first electron. This is indicative of a *negative U center* discussed in Sect. 2.8. Further observations indicate a much more complex behavior of the different oxygen-related centers, a discussion of which is beyond the scope of this book. For a review, see Dean (1986).

Oxygen has long time been supposed to form a deep-level defect in GaAs, until its role was clarified by local-vibrational-mode studies and deep-level transient

[10]Due to the thermal nature of incorporation, that is, during a heat treatment between 350 °C and 555 °C in Czochralski-grown Si.

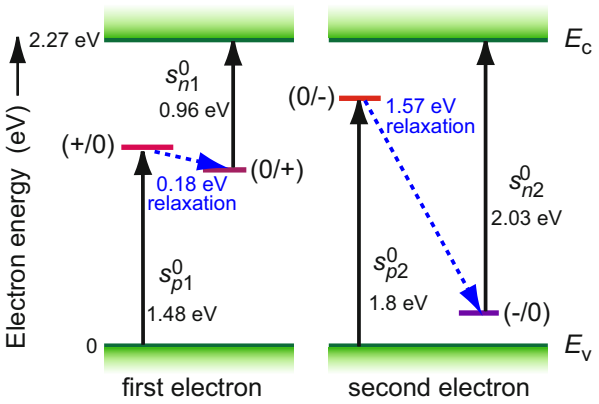


Fig. 19 Band diagram of GaP with one- and two-electron O_P states at 300 K, obtained from photocapacity measurements (After Dean et al. 1983)

spectroscopy. O has been found to form two distinct centers: an interstitial species in which oxygen is near the bond center between neighboring Ga and As atoms, and a second center in which O substitutes for As and is displaced off-center to be bonded to two of its Ga neighbors. The local-mode spectra and structures of these oxygen centers are shown in Fig. 20. The substitutional O_{As} defect is a negative-*U* center. For a review, see Skowronski (1992).

In GaN substitutional oxygen with its small covalent radius is found to be a stable shallow donor, while it forms a deep DX center in AlN (Park and Chadi 1997; Gordon et al. 2014).

2.7 The DX Center

There is a large variety of impurity centers which cause deep levels in the bandgap. An example is the EL2 defect related to an anion-antisite defect discussed in Sect. 2.3, which presents a level near the center of the bandgap and plays a key role in creating semi-insulating GaAs. Another defect in III-V compounds is often referred to as the *DX center*: *D,* since it acts like a donor, and *X,* because it was associated with an unknown defect *X* for a long time. These centers show major lattice relaxation and have deep-center characteristics. They are involved in carrier trapping and recombination traffic and are responsible for persistent photoconductivity and large Stokes shifts (Lang and Logan 1977).

DX centers are most pronounced in heavily doped *n*-type III-V compounds and alloys (e.g., AlGaAs or GaAsP, Lang 1986) and are related to the chemical nature and concentration of dopants, specifically involving S, Se, Te, Si, and Sn donors. In $Al_xGa_{1-x}As$ alloys with $x > 0.2$, the donors show a metastability at ambient pressure and it can be induced also in GaAs by the application of hydrostatic pressure. Recent calculations find a similar behavior for $Al_xGa_{1-x}N$ alloys (Gordon et al. 2014). O, Si, and Ge impurities form donors being shallow in GaN; in AlN, however, these impurities exhibit DX-center behavior; by linearly interpolating the (+/−) transition

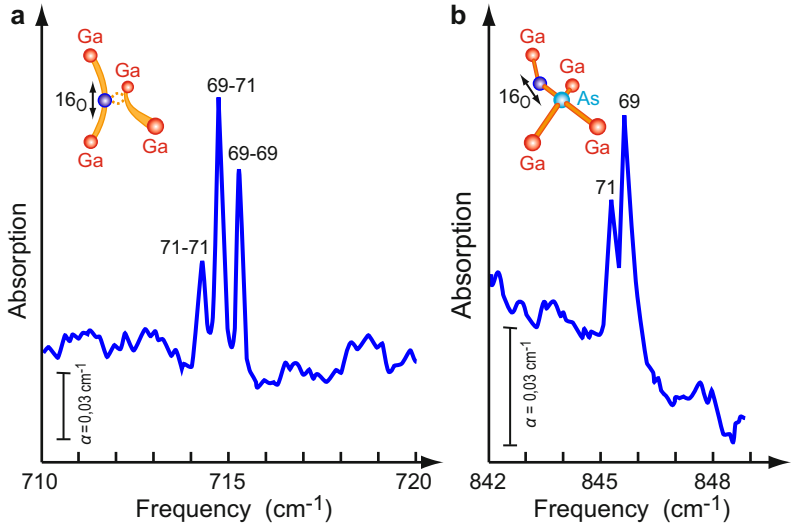


Fig. 20 Local-mode spectra assigned to (**a**) substitutional oxygen O_{As} in GaAs and (**b**) interstitial oxygen in GaAs. The structure of these defects shown in the insets are obtained from the vibrational line structure which results from the naturally abundant isotopes of the oxygen's next-neighbor Ga atoms ^{69}Ga (60%) and ^{71}Ga (40%) (After Schneider et al. 1989)

level energy (relative to the conduction-band maximum) between pure AlN and GaN for varied Al concentration, the onset for the occurrence of the DX instability is found at 61% Al for oxygen, 94% for silicon, and 52% for germanium. A comparable behavior is found for carbon in GaN and InN, where it acts as a shallow donor on cation site, but as a DX center in AlN on cation site (Lyons et al. 2014).

Both theory and experiment find that the DX center is related to the isolated donor. On its metastabile regular lattice site, it is a shallow defect. However, either the donor (for a donor on the group III sublattice) or its first neighbor (for a donor on the group V sublattice) can move off its lattice site to create a deep state. The structure of a Si donor on these two sites is shown in Fig. 21. The DX center traps two electrons in its deep configuration and is a negative-*U* defect. For reviews, see Bhattacharya (1988), Mooney (1990, 1992), Malloy and Khachaturyan (1993), Kaminska and Weber (1993).

2.8 Negative-*U* Centers

In Sects. 2.2.1, 2.3, and 2.6, examples are given for a negative-*U* center. *U* is the *Hubbard correlation energy* that was introduced by Hubbard (1963) as an energy penalty when two electrons with opposite spin occupy the same site. For a free atom, *U* is the difference between the ionization energy and the electron affinity; typically, it is on the order of 10 eV. Embedded in a crystal lattice, *U* is greatly reduced by

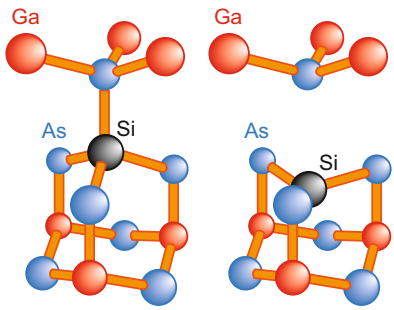


Fig. 21 Structure of the substitutional Si donor in GaAs on its regular lattice site (*left*) and in the distorted DX configuration (*right*) (After Mooney 1992)

lattice shielding and interaction to generally 0.1 . . . 0.5 eV. For most crystal defects, U is positive; this means that a defect, which has several charge states, has the higher charged state closer to the related band. This ordering of levels is easily understood once it is recognized that a second electron is less bound to a donor defect than the first because the two electrons repulse each other.

Consequently, a negative-U center indicates that with one electron already trapped, the second one is even more attracted. This can only happen when the first electron has polarized the defect configuration sufficiently so that the second one is trapped into a substantially different defect environment: for this, a substantial lattice interaction (relaxation) is required. The electron-lattice interaction can be expressed as

$$V(u) = -\lambda u \left(n_\uparrow + n_\downarrow\right) + \beta u^2/2, \tag{14}$$

with u as the atomic displacement. For equilibrium ($\partial V/\partial u = 0$), we obtain for the energy of single and double occupancies $-\lambda^2/2\beta$ and $-\lambda^2/\beta$, respectively, with $n_\uparrow$ and $n_\downarrow = 0$ or 1 as spin occupancy numbers, λ as the electronic lattice coupling, similar to Fröhlich's lattice coupling constant for free electrons α_c (but not dimensionless), and β the elastic restoring term. After adding the always-positive normal Hubbard correlation energy, we have

$$U = U_0 - \lambda^2/\beta, \tag{15}$$

which is a defect property and can become negative for large electron-lattice coupling and small lattice restoring forces. In effect, the lattice near the defect site now harbors a bipolaron (see Sect. 1.2.2 of ▶ Chap. 22, "Carrier-Transport Equations"), as suggested by Anderson (1975). This can be described as forming an *extrinsic Cooper pair,* similar to the Cooper-pair formation in a superconductor (Sect. 1.2 of ▶ Chap. 26, "Superconductivity"). Here, however, the *defect center* assists such pair formation, consequently yielding a substantially increased binding energy (U). For calculation of λ and U, see Baraff et al. (1980b).

As a result, the negative-U center is not stable in its singly occupied state, since energy (U) can be gained by trapping another carrier. Consequently, the singly occupied shallow state is not observed after the filling process (e.g., via optical excitation) is switched off and sufficient time has passed for relaxation. With

continued excitation, however, both the shallow and deep states can be observed (Watkins 1984).

In addition to the GaP:O_P^0 center, the V_{Si}^+ and the Si_i^+ are known to have negative-*U* character and many more deep centers are probably of a similar type (Watkins and Troxell 1980). Hydrogen in Si has been identified as a negative-*U* center, both theoretically (Van de Walle et al. 1989) and experimentally (Johnson et al. 1994). A general explanation for the negative-*U* behavior of hydrogen in semiconductors was discussed by Neugebauer and Van de Walle (1996b).

Negative-*U* in Chalcogenide Glasses Some optical and electron-spin resonance behavior of chalcogenide amorphous semiconductors can be explained by assuming negative-*U* centers (Street and Mott 1975). Such centers were identified by Kastner et al. (1976) as valence-alternation pairs. For instance, dangling bonds in broken chains can be bound to an adjacent unbroken chain by a valence alternation, shown here as a charge disproportionation

$$2\,Se_{(1)}^0 \rightarrow Se_{(1)}^- + Se_{(3)}^+; \tag{16}$$

the two dangling Se atoms in such a chain disproportionate to form a negatively charged dangling bond in onefold coordination, indicated as subscript, and one positively charged Se atom attached to the neighboring chain with threefold coordination (see Fig. 22), in an exothermic reaction: the charged pair has a *lower* energy than the neutral defects. The Coulomb energy between the two electrons in the negatively charged defect is more than compensated by the lattice-relaxation energy. Therefore, this center is a negative-*U* center.

The existence of negative-*U* centers in amorphous semiconductors can explain the pinning of the Fermi level by such defects without showing a high density of uncompensated spins, which would otherwise be expected when compensation occurs between ordinary donors and acceptors (Fritzsche 1976).

2.9 Instabilities of Shallow and Deep Centers

There are centers which show metastability as a shallow or as a deep center. Examples are metal-acceptor pairs in covalent crystals or excitons bound at a donor (see Sect. 2 of ▶ Chap. 18, "Shallow-Level Centers") in AlGaAs. These centers

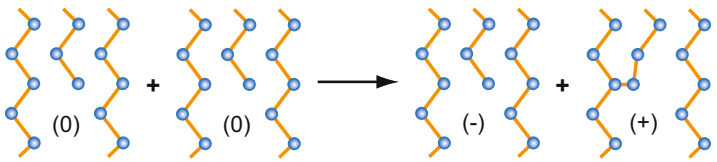


Fig. 22 Negative-*U* model by disproportionation at a dangling bond with charge character indicated (After Street and Mott 1975)

can be explained as having a negative-*U* character. When occupied with one electron, they behave as a typical shallow center, but turn into a deep center when a second electron is trapped.

When a carrier is captured by a deep center, it relaxes often to a substantially lower energy. Such lattice relaxation can be sufficiently large to change the configuration of the defect. Metastable configurations can also exist for which an activation energy is required to return to the ground-state configuration. The vacancy in Si, the EL2 defect in GaAs, the DX center in AlGaAs, and H in Si all exhibit large lattice relaxation and have bistability or metastability. For a discussion of the effects of large lattice relaxation, see Scherz and Scheffler (1993), Watkins (1991).

3 Transition-Metal Impurities

A very large group of deep centers is that involving transition metals. Many of them act as recombination centers. Others are efficient activators for luminescence or for sensitizing the semiconductor for high-gain photoconductivity. *Transition-metal impurities* include 3*d,* 4*d,* and 5*d* transition metals, 4*f* rare earth, and 5*f* actinides. All of these elements have unsaturated inner shells. Their outer-shell electrons shield part of the interaction with the host lattice. The center can be described by a *tight-binding approximation,* that is, like an individual atom, but perturbed by the surrounding lattice field - see Sect. 1.4.

3.1 Effect of Site, Charge, Spin, and Excitation

The properties of the center are determined by its chemistry and its character relating to the host lattice, such as its

- *Site character*, as interstitial or substitutional within a host lattice of given symmetry
- *Charge character* in relation to the oxidation state, conventionally identified by its remaining valence electrons, e.g., Cr^{3+} when replacing Ga^{3+} in a GaAs lattice, and its *charge state relative to the host lattice,* for example, $(Cr^{3+})^{0}_{Ga}$ neutral, $(Cr^{2+})^{-}_{Ga}$, or $(Cr^{4+})^{+}_{Ga}$ as negatively or positively charged when incorporated in a GaAs lattice on a Ga site
- *Spin character*, as high-spin, when the unsaturated *d*-shell electrons have preferably parallel spin (Hund's rule), resulting in a high electron–electron and weak crystal-field interaction, and as low-spin, when electrons have more antiparallel spin,[11] resulting in weak electron–electron and stronger crystal-field interactions (see Table 2)

[11] With parallel spins there are less alternatives to populate states of equal energy (less degeneracy). With antiparallel spin orientation, there is more degeneracy, giving the opportunity to the Jahn-Teller splitting to produce an even lower level. For transition-metal impurities, Hund's rule, and for Si vacancies, the Jahn-Teller effect produces the lower ground state (Zunger 1983).

Table 2 Occupancies of the *d* orbitals (Fig. 24) of a transition-metal impurity in states of either t_2 or *e* symmetry in an octahedral coordination in a semiconductor host crystal

	High-spin		Low-spin	
Number of *d* electrons	t_2	*e*	t_2	*e*
1	↑		↑	
2	↑↑		↑↑	
3	↑↑↑		↑↑↑	
4	↑↑↑	↑	↑↓↑↑	
5	↑↑↑	↑↑	↑↓↑↓↑	
6	↑↓↑↑	↑↑	↑↓↑↓↑↓	
7	↑↓↑↓↑	↑↑	↑↓↑↓↑↓	↑
8	↑↓↑↓↑↓	↑↑	↑↓↑↓↑↓	↑↑
9	↑↓↑↓↑↓	↑↓↑	↑↓↑↓↑↓	↑↓↑

- *Transition character* with respect to an optical excitation, for an *intracenter transition* with charge conservation at each center, or an *ionization* with charge transfer to the conduction or valence band or to another center.

The solubility of 3*d* impurities in Si is extremely low, typically with a maximum concentration of 10^{14} cm^{-3}. This makes positive identification difficult since the solubility limit lies below the threshold of conventional analytic chemistry. The 3*d* elements are preferably incorporated in interstitial sites of Si, while 5*d* elements are mostly substitutionals. The solubility of Cu and Ni in Si, however, is much larger, up to $5 \cdot 10^{17}$ cm^{-3}. In III-V and II-VI compounds, the solubility of the 3*d* elements is higher, typically 10^{17} cm^{-3}. An exception is Mn, which forms continuous solid solutions with II-VI compounds: it produces a dilute magnetic semiconductor (see Sect. 2 of ▶ Chap. 9, "Magnetic Semiconductors").

Site Character Incorporated as a substitutional, the transition-metal impurity entails interaction with the states of the replaced vacancy. This means that certain states hybridize strongly with the corresponding states of the vacancy, for example, the $d(T_2)$ state of Cr_{Si} with the T_2 state of the vacancy (see Zunger and Lindefelt 1983). Interstitials of the Ti to Zn 3*d* elements have little interaction of their *d* orbital with the crystal state.

Charge Character It is characteristic of the 3*d* impurities that many of them have several stable charge characters (Fig. 23), yielding levels in the bandgap with a relatively small energy difference between them, typically 0.5 eV rather than 2–3 eV. Each of these levels split due to the crystal field and the Jahn-Teller effect (dynamic and cooperative-see Bates and Stevens 1986). The strain induced by the defect causes a change in symmetry surrounding each defect center. The tetrahedral symmetry of GaAs becomes orthorhombic with incorporation of Cr^{3+} and becomes tetragonal with Cr^{2+} (Bates and Stevens 1986). In Si, interstitial 3*d* impurities cause an increase in distance of the four nearest neighbors and a decrease in distance of the six next-nearest neighbors. This makes the lattice surrounding the 3*d* impurity nearly tenfold coordinate (Lindefelt and Zunger 1984a, b).

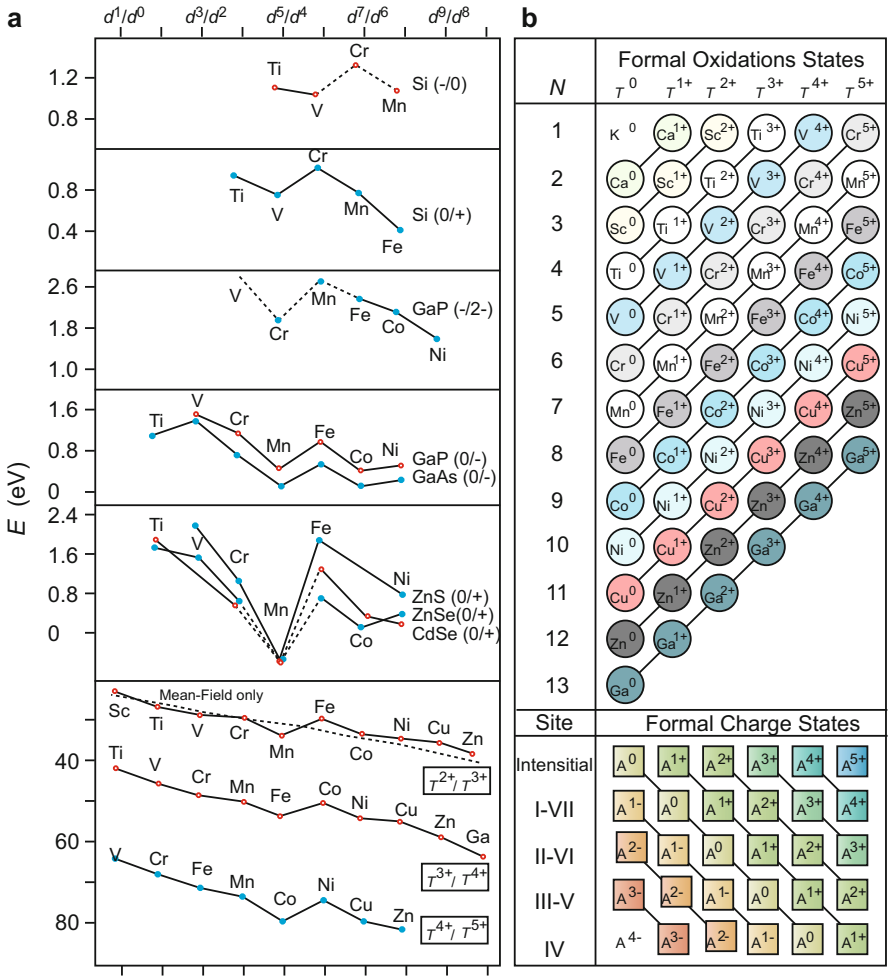


Fig. 23 (**a**) Ionization energy for 3*d* transition elements, free ions and as impurities in various compounds; *bottom*: $E_{ionization}$ for free ion. (**b**) Formal oxidation states with total number *N* of electrons in 3*d* and 4 *s* shells and free ion charges (*upper diagram*). Corresponding charges relative to the semiconductor lattice for cation substitutional incorporation (*lower diagram*) (After Zunger 1986)

Significant concerns include the identification of the ground state and the proper sequencing of higher excited states. A number of empirical rules are used in the classical discussions (Kaufmann and Schneider 1982). Important factors are the *Hund rules,* which require that the ground state has maximum multiplicity $(2S + 1)$ and maximum *L,* and that $J = L \pm S$ when the shell is less (+) or more (−) than half full (see Sect. 1.1 of ▶ Chap. 9, "Magnetic Semiconductors"; Ashcroft and Mermin 1976). For instance, when a Cr atom replaces a Ga atom in *n*-type GaAs and traps an additional electron, it may change from a $3d^5\ 4s^1$ configuration to a $3d^5\ 4s^2$ or

$3d^3\,4s^2\,4p^2$ configuration; the latter is more Ga-like and more probable. The remaining three electrons in the d shell determine the defect-level spectrum and distribute themselves between e orbitals, transforming according to $(2z^2 - x^2 - y^2)/\sqrt{6}$ and $(x^2 - y^2)/\sqrt{2}$ (fourfold degenerate), and t^2 orbitals, transforming according to xy, yz, and zx (sixfold degenerate); lower-case letters indicate one-electron states, see Fig. 24.

The ground states of the differently charged Cr substitutional centers in GaAs are

$$(\mathrm{Cr}^{4+})^{+}\ (3d^2)\ E - E_\mathrm{v} = 0.45\ \mathrm{eV},\ s_n = 9 \cdot 10^{-17}\ \mathrm{cm}^2$$
$$(\mathrm{Cr}^{3+})^{0}\ (3d^3)\ E - E_\mathrm{v} = 0.74\ \mathrm{eV},\ s_n = 10^{-17}\ \mathrm{cm}^2$$
$$(\mathrm{Cr}^{2+})^{-}\ (3d^4)\ E - E_\mathrm{c} = 0.12\ \mathrm{eV},\ \text{inside conduction band}$$

Many other centers in GaAs as well as other semiconductors are identified and reviewed by Clerjaud (1985) and Zunger (1986). A list of the different oxidation states and charges relative to the semiconductor is shown in Fig. 23b. The ionization energy of $3d$ transition metal ions is given in Fig. 23a.

Transition Character The transitions at the transition-metal dopant can be described as intracenter (also referred to as *intraionic*) transitions; for instance,

$$(e)^2 t_2 + h\nu \rightarrow e\ \ (t_2)^2. \tag{17}$$

When, on the other hand, carriers from one of the bands are involved, these transitions can be described as ionization (i.e., as charge-transfer transitions), such as

$$(e)^3 + h\nu \rightarrow (e)^4 + h^+. \tag{18}$$

The corresponding energies for intracenter transitions are listed in Table 3 and for ionization in Table 4 – see also Pantelides and Grimmeiss (1980).

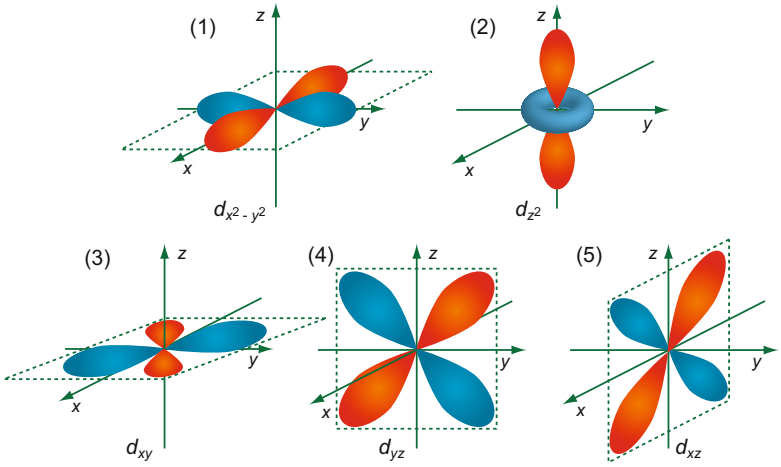


Fig. 24 d orbitals of wavefunctions in e symmetry (*1*) and (*2*) and in t_2 symmetry (*3*)–(*5*)

Table 3 Observed excitation energies (eV) of 3*d*-transition metals as cation substitutionals in doped II-VI and III-V semiconductors; ZPL denotes a zero-phonon line (After Zunger 1986)

Ground state impurity	$^3A_2$ Ti^{2+}	$^4T_1$ V^{2+}	$^5T_2$ Cr^{2+}	$^6A_1$ Mn^{2+}	^5E Fe^{2+}	$^4A_2$ Co^{2+}	$^3T_1$ Ni^{2+}
Host							
ZnS		0.53($^4T_2$)	0.64(^5E)	2.34($^4T_1$)	0.44($^5T_1$)	0.46($^4T_2$)	0.54($^3T_2$)
	1.21 ($^3T_1$)	1.14($^4A_2$)	1.36($^3T_2$)	2.53($^4T_2$)	2.07($^3A_2$)	0.77($^4T_1$)	1.13($^3T_2$)
		1.39($^4T_1$)	1.75($^3T_1$)	2.67(^4E)	2.14($^3A_1$)	1.76($^4T_1$)	1.52($^3T_2$)
ZnSe	0.74 ($^3T_1$)	0.50($^4T_2$)	0.68(^5E)	2.31($^4T_1$)	0.34($^5T_2$)	0.43($^4T_2$)	0.50($^3T_2$)
	1.22 ($^3T_1$)	1.08($^4A_2$)	1.61($^3T_2$)	2.47($^4T_2$)	1.26($^3T_1$)	0.78($^4T_1$)	1.10($^3A_2$)
		1.24($^4T_1$)	1.85($^3T_1$)	2.67(^4E)		1.67($^4T_1$)	1.46($^3T_1$)
ZnTe	–	–	0.68(^3E)	2.3 ($^4T_1$)	0.31($^5T_2$)	0.72($^4T_1$)	–
				2.4 ($^4T_2$)		1.44($^4T_1$)	
				2.6 (^4E)			
CdS	0.40 ($^3T_2$)	0.61	0.66(^5E)	–	0.32($^5T_2$)	6.8($^4T_1$)	0.51($^3T_2$)
	0.71 ($^3T_1$)	1.08				1.73($^4T_1$)	1.01($^3A_2$)
	1.22 ($^3T_1$)	1.8($^4T_1$)					1.58($^3T_1$)
CdSe	0.38 ($^3T_2$)	0.43($^4T_2$)	0.62(^5E)	–	0.37($^5T_2$)	0.37($^4T_2$)	0.52($^3T_2$)
	0.62 ($^3T_1$)	0.86($^4A_2$)			0.29(ZPL)	0.35(ZPL)	(0.45, ZPL)
	1.18 ($^3T_1$)	1.31($^4T_1$)				0.68($^4T_1$)	0.99($^3A_2$)
						1.61($^4T_1$)	1.42($^3T_1$)
							(1.35, ZPL)
CdTe	0.35 ($^3T_2$)	0.85($^4A_2$)	0.63(^5E)	2.2 ($^4T_1$)	0.28($^5T_2$)	0.37($^4T_2$)	–
	0.62 ($^3T_1$)	1.11($^4T_2$)		2.5 ($^4T_2$)		0.72($^4T_1$)	
	1.14 ($^3T_1$)					1.44($^4T_1$)	
GaP	0.60 ($^3T_2$)		0.87(^5E)	1.34($^4T_1$)	0.41($^5T_2$)	0.56($^4T_2$)	0.82($^3T_2$)
	–	1.07		1.53(ZPL)		0.87–1.24 ($^4T_1$)	1.24($^3T_1$)
						1.50($^4T_1$)	1.43($^3A_2$)
GaAs	0.565($^3T_1$)		0.9	–	0.37($^5T_2$)	0.5 (ZPL) ($^4T_2$)	1.15($^3T_1$)
	0.66 ($^3T_1$,$^3T_2$)	1.03	0.84(ZPL, ^5E)			0.87–1.05 ($^4T_1$)	
	1.04 ($^3T_1$)	0.69				1.40(ZPL) ($^4T_1$)	
InP	–	–	0.76(ZPL, ^5E)	–	0.35($^5T_2$)	0.47(ZPL) ($^4T_2$)	–
						0.92($^4T_1$)	–
						0.79(ZPL)	
						0.78–0.92	

3.2 The Energy of Levels

The crystal-field splitting of levels of a transition-metal impurity can be estimated in tetrahedral symmetry as

$$\Delta = |E(e) - E(t_2)| = -\frac{4}{15}\frac{Ze^2}{R}\frac{\langle r^4\rangle}{R^4}, \tag{19}$$

Table 4 Ionization energies (eV) of substitutional 3*d*-transition metal dopants in compound semiconductors and interstitial 3*d* dopants in Si. Where levels are not yet observed, but predicted in valence or conduction bands, or in the gap, this is indicated (vb, cb, or gap) (After Zunger 1986)

Material	Ti	V	Cr	Mn	Fe	Co	Ni	Transition type
ZnS	(gap)	2.11	1.0, 1.74	(vb)	2.01	(gap)	0.75	(0/+)
	–		2.78, 2.48	–	(gap)		2.46	(−/0)
ZnSe	1.75	(gap)	0.46	(vb)	1.1, 1.25	~0.3	0.15	(0/+)
	–		2.26, 2.1	–			1.85	(−/0)
ZnTe	(gap)	(gap)		(vb)		0.86	(vb)	(0/+)
	–	–	1.4	–	(gap)	–	~1	(−/0)
CdS	~2.21	≳1.8	1.47	(vb)	(gap)	(gap)	≥0.27	(0/+)
CdSe	1.69	~1.5	0.64	(vb)	0.64	0.22	0.32	(0/+)
	–	–		–	–	–	1.81	(−/0)
CdTe	~0.55	0.74		(vb)	0.13	–	(vb)	(0/+)
	–	–	1.34	–	–	–	0.92	(−/+)
GaP	–	–	0.5	–	–	–	–	(0/+)
	(gap)	1.55	1.12	0.4	0.86	0.41	0.5	(−/0)
	–	–	1.85	–	2.09	2.02	1.55	(−/2−)
GaAs	–	–	0.32 − 0.45	–	–	–	–	(0/+)
	1.07,1.29	1.38	0.74	0.11	0.49	0.16	0.22, 0.35	(−/0)
	–	–	1.57	–	(cb)	1.67	1.13	(−/2−)
InP	–	(cb)	0.94–1.03	0.22	0.8	0.24–0.32	–	(−/0)
	–	0.2		–	–	–	–	(0/+)
Si	0.89	0.72	0.95	0.75	0.385	–	–	(0/+)
			0.38					
	1.09	1.01	–	1.06	–	–	–	(−/0)
	0.25	0.30	–	0.25	–	–	–	(+/2+)

where R is the distance to the nearest neighbor (ligand) and $\langle r^4 \rangle$ is the expectation value of r^4 for the $3d$ wavefunction (Hayes and Stoneham 1984). Here Δ increases with increasing covalency, higher charge of the transition metal, and higher transition series. The effective crystal-field splitting is given in Table 5 for some of the $3d$ impurities in a few binary compounds.

The multielectron levels in the *irreducible representation* of the tetrahedral T_d point group are a_1, a_2, e, t_1, and t_2 – see Bassani et al. (1974) or Madelung (1981). In Fig. 25, an example for the transition from the lower $^5T_2$ ($t^2\, e^2$) to the excited ^5E ($t^3\, e^1$) state is given for GaAs:$(Cr^{2+})^-$ (Clerjaud 1985). A list of possible splittings of the free transition-metal ion states after incorporation into a semiconductor of T_d symmetry is given in Table 6. In addition, one has to consider splitting due to spin-orbit, lattice-phonon, and the Jahn-Teller effect interactions.

The relationship of the level spectrum to energies of the host lattice and the chemical identity of the transition metal has been analyzed using a self-consistent

Table 5 Effective crystal-field splitting Δ_{eff} (in eV) of $3d$ impurities in binary semiconductors (After Zunger 1986)

		Host crystal				
		ZnS	ZnSe	InP	GaAs	GaP
Impurity	Cr	0.540	0.540	0.64	0.65	0.67
	Mn	0.402	0.400			0.52
	Fe	0.430	0.41	0.43	0.44	0.45
	Co	0.453	0.459	0.575	0.590	0.608
	Ni	0.520	0.510		0.91	0.97

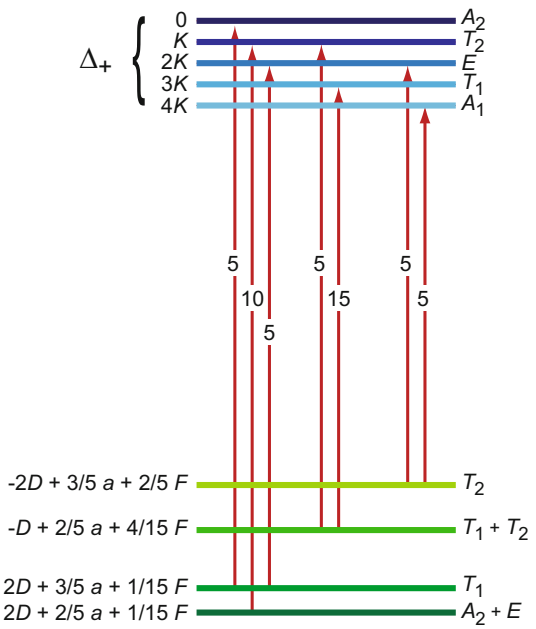


Fig. 25 Level scheme for intra d-shell transitions of GaAs:$(Cr^{2+})^-$ with weight factors for competing transitions (not to scale); see Fig. 28 (After Clerjaud 1985)

Table 6 Splitting of free transition-metal ion terms into many-electron terms after incorporation into semiconductors of T_d symmetry at lattice site (solid underlining) or interstitial site (dashed underlining)

Free-ion many-electron terms	T_d crystalline many-electron terms	Ground state	
		S	J
S Terms			
Fe^{3+},Mn^{2+},Cr^{+};d^5,^6S	$\underline{^6A_1, (t^3e^2)}$	$\frac{5}{2}$	$\frac{5}{2}$
D Terms			
Sc^{2+};d^1,^2D	$\underline{^2T_2(t^1e^0)} + \underline{^2E(e^1t^0)}$	$\frac{1}{2}$	
Mn^{3+},Cr^{2+};d^4,^5D	$\underline{^5T_2(t^2e^2)} + \underline{^5E(t^3e^1)}$	2	1, 2, $\underline{3}$
Co^{3+},Fe^{2+};d^6,^5D	$\underline{^5T_2(t^4e^2)} + \underline{^5E(t^3e^3)}$	2	0, $\underline{\frac{1}{2}}$, 1
Cu^{2+},Ni^{+}; d^9,^2D	$\underline{^2T_2(t^5e^4)} + \underline{^2E(t^6e^3)}$	$\frac{1}{2}$	$\underline{\frac{1}{2}}$, $\frac{3}{2}$
F Terms			
V^{3+},Ti^{2+};d^2,^3F	$\underline{^3T_1(t^2e^0)} + {}^3T_2(t^1e^1) + \underline{^3A_2(e^2t^0)}$	1	
Cr^{3+},V^{2+}; d^3,^4F	$\underline{^4T_1(t^1e^2)} + {}^4T_2(t^2e^1) + \underline{^4A_2(t^3e^0)}$	$\frac{3}{2}$	$\underline{\frac{1}{2}}$, $\frac{3}{2}$, $\underline{\frac{5}{2}}$
Ni^{3+},Co^{2+},Fe^{+};d^7,^4F	$\underline{^4T_1(t^5e^2)} + {}^4T_2(t^4e^3) + \underline{^4A_2(t^3e^4)}$	$\frac{3}{2}$	
Cu^{3+},Ni^{2+}, d^8;^3F	$\underline{^3T_1(t^4e^4)} + {}^3T_2(t^5e^3) + \underline{^3A_2(t^6e^2)}$	1	$\underline{0}$, 1, 2

quasiband crystal-field method and employing a density-functional Green's function approach, introduced by Lindefelt and Zunger (1982). This method was first applied to 3*d* transition-metal impurities in an Si host lattice (Zunger and Lindefelt 1983) and yields results in agreement with the experiment when spin polarization is included (Beeler et al. 1985; for comparative remarks, see Zunger 1986). Typical examples are shown in Fig. 26, identifying t_2 and e states, which are contained within the bands for some of the transition-metal impurities. Their energy sequence is inverted when comparing interstitial and substitutional incorporation into the host lattice. The charge character of these levels is given in Fig. 26b and identifies donors ($- \rightarrow 0$) and acceptors ($+ \rightarrow 0$). The level spectrum for substitutional defects is less structured, and no deep levels are obtained for Ti and Fe. For 4*d* impurities, see Beeler et al. (1986); for other transition metals in III-V compounds, see Clerjaud (1985). For configuration interaction in II-VI and III-V compounds, see Dreyhsig (1998). For charge-transfer spectra in III-V compounds, see Wolf et al. (1993). For intracenter transitions of transition-metal impurities in II-VI compounds, see Bouhelal and Albert (1993); see also Dahan and Fleurov (1994) and Fleurov and Dahan (1995). Defect levels in Si are reviewed by Chen and Milnes (1980).

The investigations of Zunger (1985) have shown some universality of binding energies within the bandgap. When normalized to the vacuum level using the intrinsic work function of the host, all binding energies of the same 3*d* impurity in different host crystals are approximately the same - see Fig. 27. This reflects the screening of the localized 3*d* electrons by outer-shell electrons and was

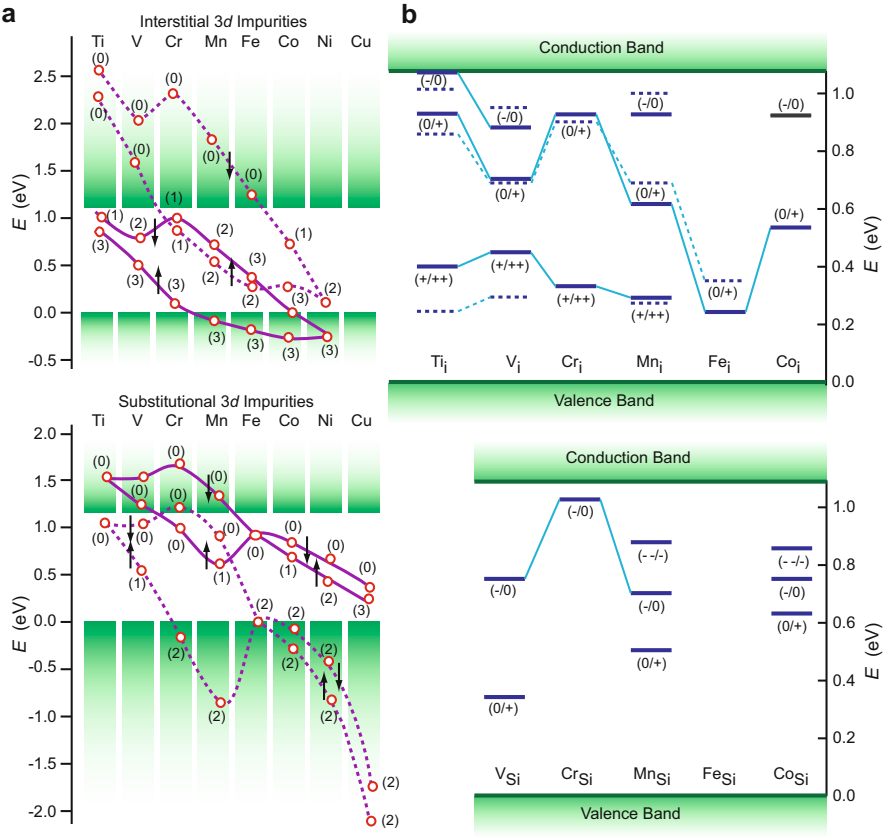


Fig. 26 (**a**) Ground-state energies of neutral 3*d* impurities in a Si host with level occupancy in parentheses. (**b**) Calculated (*solid curves*) and observed (*dashed*) 3*d* impurity levels in Si at interstitial and lattice (substitutional) sites (After Beeler et al. 1985)

employed to obtain valence-band offsets at semiconductor interfaces, see Sect. 2.2.4 of ▶ Chap. 16, "Crystal Interfaces".

Transition-metal atoms in covalent crystals show only minor net charges, that is, a minor fraction of ionic bonding, and tend to approach a noble-metal configuration. In Ni this is best achieved. Its extremely high diffusivity in Si ($D = 10^{-4}\,\mathrm{cm^2\,s^{-1}}$) may be an indicator of this fact as it renders the Ni atom small and inert.

The ability of transition-metal impurities to trap carriers is used for compensation, that is, to produce less-conductive (semi-insulating) semiconductors like, for example, InP:Fe. For changing charge states in II-VI semiconductors see Kreissl and Schulz (1996). Much of the earlier work to identify the level spectrum was done by electron-paramagnetic resonance (EPR), see Ludwig and Woodbury (1962); for a review, see Weber (1983).

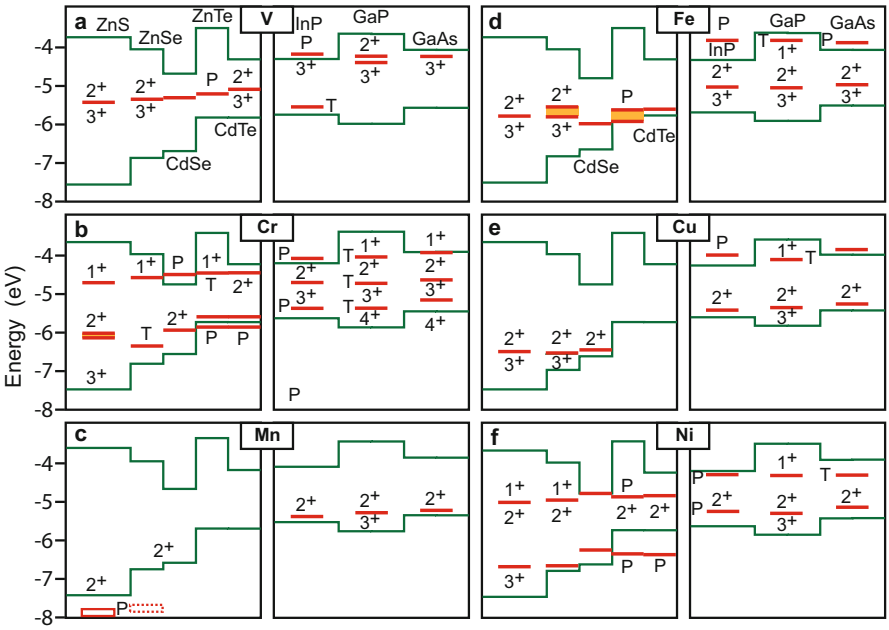


Fig. 27 Binding energies of 3*d* impurities (such as V and Cr) in different host semiconductors [only identified in the *upper diagrams* (**a**) and (**d**)]. Energies are related to the vacuum level. The numbers 1+, 2+, . . . indicate stable oxidation states. *P* predicted, *T* tentative (After Caldas et al. 1985)

3.3 Optical Transitions of Transition-Metal Dopants

In Fig. 28, a typical optical absorption spectrum of Cr^{2+} in GaAs at 6 K is shown which can be understood as intraionic $^5T_2(^5D) \rightarrow {}^5E(^5D)$ transitions of the d^4 electrons, split by the effects of Jahn-Teller and spin-orbit interactions (Williams et al. 1982). The corresponding zero-phonon transitions are identified in Fig. 25. The shape of the Cr^{2+} absorption band is explained by phonon coupling and lattice strains in a tetragonal site; the Jahn-Teller energy of the $^5T_2$ state is 75 meV and that of the ^5E state is 6.2 meV (see Deveaud et al. 1984). For a review of other absorption spectra of transition metals, see Clerjaud (1985) and Pajot and Clerjaud (2013) . In semi-insulating GaAs, ionizing transitions from the Cr^{3+} center are observed in addition to those listed above:

$$\left(Cr^{3+}\right)^0_{Ga} + e^-_{vb} + h\nu \rightarrow \left(Cr^{2+}\right)^-_{Ga}. \tag{20}$$

Most of the deep-center optical transitions have been studied for transition metals and rare earths in II-VI compounds, which show well-localized states due to the strong interaction of 3*d* electrons. Some of these are described by crystal-field theory (Griffith 1964), as a result of the splitting of *d* states by the field from nearest neighbors. The transitions are also sensitive to variations of the crystal field

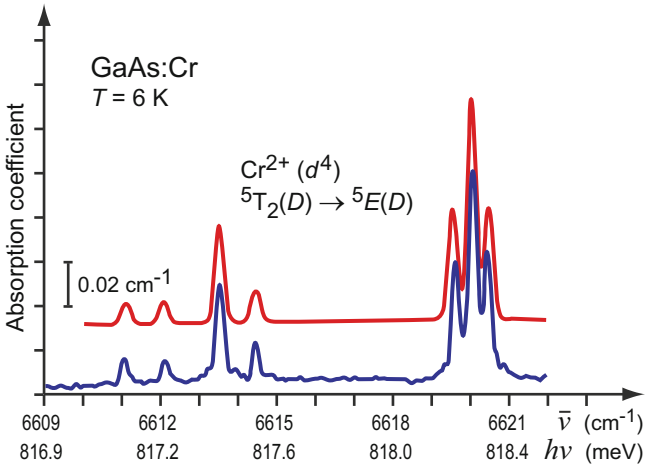


Fig. 28 Optical absorption spectrum of Cr^{2+} in GaAs at 6 K. The red curve is a theoretical estimation using spin-Hamiltonian parameters from EPR measurements (After Clerjaud 1985)

beyond the nearest naighbors, induced, for example, by stacking faults (Pohl et al. 1990) or other impurities (Pohl and Gumlich 1989). The Jahn-Teller shift (see Uba and Baranowski 1978) has been used to correlate some of the rather complex features of the optical absorption of ZnSe:Co (Fig. 29).

The optical spectra of transition-metal impurities in wide-gap crystals is rich in absorption and emission lines due to several charge states that are available within the bandgap for a variety of these elements and various crystal-field splittings. Vanadium doping of the II-VI semiconductors ZnS (Biernacki et al. 1988), ZnSe (Goetz et al. 1992), and ZnTe (Peka et al. 1996) leads to three pronounced luminescence bands related to intraionic multiplett transitions $^5E(D) \rightarrow {}^5T_2(D)$ of V^+ (d^4), $^4T_2(F) \rightarrow {}^4T_1(F)$ of V^{2+} (d^3), and $^3T_2(F) \rightarrow {}^3A_2(F)$ of V^{3+} (d^2), as shown for ZnTe:V in Fig. 30. In ZnTe, the respective acceptor level for the (V^+/V^{2+}) charge-transfer transition lies 9400 cm^{-1} (1.17 eV) above the valence-band edge, and the donor level for the (V^{2+}/V^{3+}) charge-transfer transition lies 12,500 cm^{-1} (1.55 eV) below the conduction-band edge. The broad emission bands with weak zero-phonon lines at their high-energy onset illustrate the strong lattice coupling leading to the pronounced phonon-assisted emissions.

The absorption of some transition-metal impurities is responsible for the color of the host crystal. For example, the red color of ruby is due to Cr^{3+} in Al_2O_3, and the blue color of sapphire is due to charge transfer from Fe^{2+} to Ti^{4+} in Al_2O_3.

Transition Metal Impurities in Alloys The transition energies of 3*d* impurities track with the vacuum level of the host semiconductor. In an alloy $A_xB_{1-x}C$, the transition energy is given by

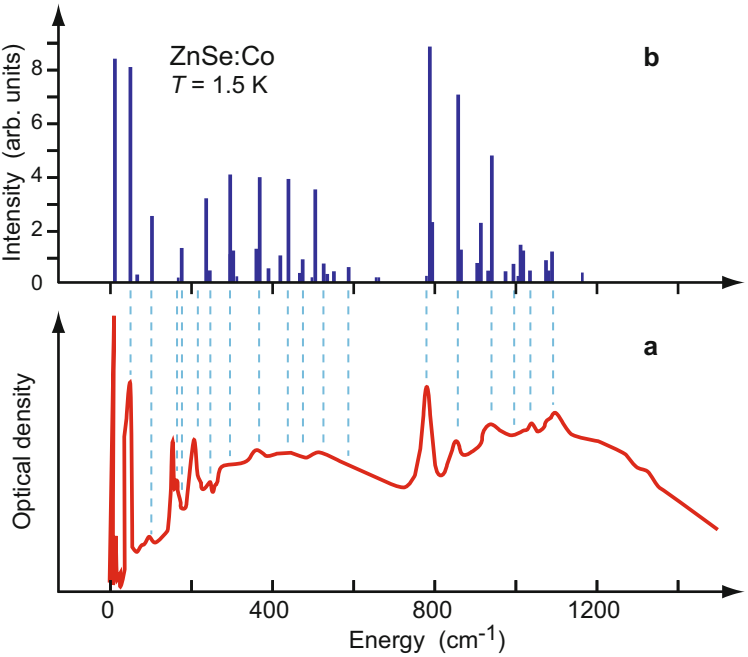
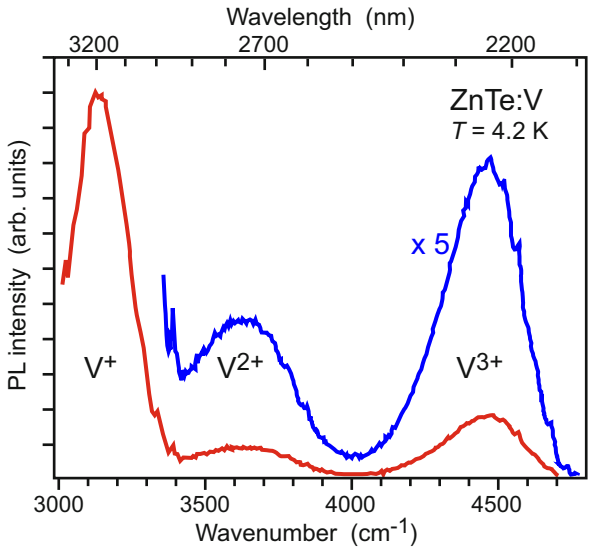


Fig. 29 (**a**) Measured optical absorption of ZnSe:Co at 1.5 K. (**b**) Prediction of the Jahn-Teller model for transitions within the $3d^7$ configuration of Co^{2+} (After Uba and Baranowski 1978)

Fig. 30 Photoluminescence spectrum of ZnTe:V at 4.2 K, showing intraionic 3*d* multiplett emissions of V^{+} (d^4), V^{2+} (d^3), and V^{3+} (d^2) (After Peka et al. 1996)



$$E(x) = E(0) + \overline{\alpha}\ x, \quad \text{where } \overline{\alpha} = \varphi(x = 1) - \varphi(x = 0), \tag{21}$$

using a linear variation of the slope $\overline{\alpha}$ with the intrinsic semiconductor workfunction φ on composition x; see also Sect. 2.2.4 of ▶ Chap. 16, "Crystal Interfaces".

4 Summary

Deep-level defect centers have tightly bound electrons which, for the ground state, are localized near the core by a central-cell potential. The potential depth for substitutional defects is indicated by the atomic electronegativity, expressed by the difference between the *s* energies of host and impurity for donors and the respective *p* energies for acceptors. The electronic eigenfunctions can mix with *both* conduction *and* valence bands, in contrast to shallow-level defects. The potential has usually long-range Coulomb tails, rendering higher excited states of a deep-level defect hydrogen-like with carrier orbits extending well into the surrounding lattice.

Many deep defects relax substantially after excitation or charging; consequently the equilibrium positions of the surrounding atoms can significantly change. Such deep-level centers may act as centers for nonradiative recombination by forming a bridge between conduction and valence band. Furthermore, strong lattice coupling leads to optical spectra with hardly discernible zero-phonon transition lines and broad phonon-assisted transitions, which are substantially shifted between optical absorption and emission.

Deep centers act as deep traps for either electrons or holes, yielding donor character for a (0/+) charge-transfer transition and acceptor behavior for a (0/–) transition. Many such centers have positive *and* negative charge states in the bandgap with energy levels depending on the position of the Fermi energy, providing the ability to effectively compensate extrinsic doping.

Prominent *intrinsic* deep-level defects are vacancies, self-intersitials, and antisites. They may be metastable in some charge state with substantial lattice relaxation and also be affected by Jahn-Teller distortion. A prominent group of *extrinsic* deep defect centers are transition-metal impurities with their uncompleted inner *d* (or *f*) shell. Excitation leads either to an intracenter transition within the partially filled inner shell, leaving the charge state unaffected, or to a charge-transfer transition changing the charge state.

As traps and recombination centers, these deep-level defects are usually detrimental for semi- and photoconducting devices, by reducing their response and rendering them slow. However, specific types of deep traps can be beneficial by storing carriers for a long time. This is important for sensitizing photoconductors or by accumulating the effect of radiation in solid-state dosimeters. Other deep centers are essential for many applications in luminescence, for example, in LEDs or solid-state lasers, or to render semiconductor layers semi-insulating for field-effect transistors.

References

Anderson PW (1975) Model for the electronic structure of amorphous semiconductors. Phys Rev Lett 34:953

Ashcroft NW, Mermin ND (1976) Solid state physics. Harcourt Brace College Publishers, New York

Baldereschi A, Hopfield JJ (1972) Binding to isoelectronic impurities in semiconductors. Phys Rev Lett 28:171

Baraff GA (1992) The midgap donor level EL2 in gallium arsenide: recent developments. In: Pantelides ST (ed) Deep centers in semiconductors, 2nd edn. Gordon and Breach, New York, pp 547–590

Baraff GA, Schlüter M (1979) Electronic structure of localized defects in covalent semiconductors. In: Treusch J (ed) Festkörperprobleme/advances in solid state physics, vol 19. Vieweg, Braunschweig, pp 303–329

Baraff GA, Schlüter M (1980) New self-consistent approach to the electronic structure of localized defects in solids. Phys Rev B 19:4965

Baraff GA, Schlüter M (1985) Electronic structure, total energies, and abundances of the elementary point defects in GaAs. Phys Rev Lett 55:1327

Baraff GA, Kane EO, Schlüter M (1980a) Simple parametrized model for Jahn-Teller systems: vacancy in *p*-type silicon. Phys Rev B 21:3563

Baraff GA, Kane EO, Schlüter M (1980b) Theory of the silicon vacancy: an Anderson negative-*U* system. Phys Rev B 21:5662

Bassani GF, Iadosini G, Preziosi B (1969) Band structure and impurity states. Phys Rev 186:735

Bassani GF, Iadosini G, Preziosi B (1974) Electronic impurity levels in semiconductors. Rep Prog Phys 37:1099

Bates CA (1978) Jahn-Teller effects in paramagnetic crystals. Phys Rep 35:187

Bates CA, Stevens KWH (1986) Localised electron states in semiconductors. Rep Prog Phys 49:783

Beeler F, Andersen OK, Scheffler M (1985) Theoretical evidence for low-spin ground states of early interstitial and late substitutional 3*d* transition-metal ions in silicon. Phys Rev Lett 55:1498

Beeler F, Andersen OK, Scheffler M (1986) Electronic structure calculation of 3d and 4d transition metal point defects in silicon. In: Engström O (ed) Proceedings of 18th international conference physics of semiconductors, Stockholm 1986. World Scientific, Singapore, pp 875–878

Bernholc J, Pantelides ST (1978) Scattering-theoretic method for defects in semiconductors. I. Tight-binding description of vacancies in Si, Ge, and GaAs. Phys Rev B 18:1780

Bersuker IB (1984) The Jahn-Teller effect and vibronic interaction in modern chemistry. Plenum Press, New York

Bhattacharya P (1988) The relationship of the D-X centre in $Al_xGa_{1-x}As$ and other III-V alloys with the conduction band structure. Semicond Sci Technol 3:1145

Biernacki SW, Roussos G, Schulz H-J (1988) The luminescence of $V^{2+}(d^3)$ and $V^{3+}(d^2)$ ions in ZnS and an advanced interpretation of their excitation levels. J Phys C 21:5615

Bouhelal A, Albert JP (1993) Intracenter transitions of transition metal impurities in II-VI semiconductors. Physica B 185:255

Briddon PR, Jones R (1993) Theory of impurities in diamond. Physica B 185:179

Caldas M, Fazzio A, Zunger A (1985) Chemical trends and universalities in the the spectra of transition metal impurities in semiconductors. In: Kimerling LC, Parsey JM Jr (eds) Proceedings of 13th international conference on defects in semiconductors. The Metallurgical Society of AIME, Warrendale, pp 1035–1041

Car R, Kelly PJ, Oshiyama A, Pantelides ST (1984) Microscopic theory of atomic diffusion mechanisms in silicon. Phys Rev Lett 52:1814

Castner G, Känzig W, Woodruff TO (1958) The electronic structure of a *V*-Center. Nuovo Cimento 7(Suppl):612

Centoni SA, Sadigh B, Gilmer GH, Lenosky TJ, Díaz de la Rubia T, Musgrave CB (2005) First-principles calculation of intrinsic defect formation volumes in silicon. Phys Rev B 72:195206

Chadi DJ (2003) Arsenic-antisite defect in GaAs: multiplicity of charge and spin state. Phys Rev B 68:193204
Chen JW, Milnes AG (1980) Energy levels in silicon. Annu Rev Mater Sci 10:157
Chiarotti G, Grassano UM (1966) Modulated F-center absorption in KCl. Phys Rev Lett 16:124
Clerjaud B (1985) Transition-metal impurities in III-V compounds. J Phys C Sol State Phys 18:3615
Cuthbert JD, Thomas DG (1967) Fluorescent decay times of excitons bound to isoelectronic traps in GaP and ZnTe. Phys Rev 154:763
Cuthbert JD, Thomas DG (1968) Optical properties of tellurium as an isoelectronic trap in cadmium sulfide. J Appl Phys 39:1573
Dahan P, Fleurov V (1994) The sum rule for the luminescence of semiconductors doped with transition metal impurities. J Phys Condens Matter 6:101
Davies G, Manson NB (1994) Properties and growth of diamond, vol 9, EMIS data review series. INSPEC IEE, London, p 159
Dean PJ (1973) Inter-impurity recombinations in semiconductors. Prog Solid State Chem 8:1
Dean PJ (1986) Oxygen and oxygen associates in gallium phosphide and related semiconductors. In: Pantelides ST (ed) Deep centers in semiconductors. Gordon and Breach, New York, pp 185–347
Dean PJ, Cuthbert JD, Lynch RT (1969) Interimpurity recombinations involving the isoelectronic trap bismuth in gallium phosphide. Phys Rev 179:754
Dean PJ, Skolnick MS, Uihlein C, Herbert DC (1983) New aspects of the oxygen donor in gallium phosphide. J Phys C 16:2017
Deveaud B, Picoli G, Lambert B, Martinez G (1984) Luminescence processes at chromium in GaAs. Phys Rev B 29:5749
Dow JD (1985) Localized perturbations in semiconductors. In: Bassani F, Fumi F, Tosi MP (eds) Highlights Condens Matter Theory. North Holland Publ, pp 465–494
Dreyhsig J (1998) Configuration-interaction energy level diagrams for d^7 and d^8 charge-state impurities in II–VI and III–V compound semiconductors. J Phys Chem Sol 59:31
El-Mellouhi F, Mousseau N (2005) Self-vacancies in gallium aresenide: an ab initio calculation. Phys Rev B 71:125207
Estreicher SK (1995) Hydrogen-related defects in crystalline semiconductors: a theorist's perspective. Mater Sci Eng 14:319
Faulkner RA (1968) Toward a theory of isoelectronic impurities in semiconductors. Phys Rev 175:991
Fleurov V, Dahan P (1995) Covalent effects and optical spectra of transition metal impurities in semiconductors. Proc SPIE 2706:296
Fowler WB (1968) The physics of color centers. Academic Press, New York
Freysoldt C, Grabowski B, Hickel T, Neugebauer J, Kresse G, Janotti A, Van de Walle CG (2014) First-principles calculations for point defects in solids. Rev Mod Phys 86:253
Fritzsche H (1976) Summary remarks. In: Kolomiets BT (ed) Electronic phenomena in non-crystalline semiconductors. Nauka, Leningrad, p 65
Gao Y, Sun D, Jiang X, Zhao J (2019) Point defects in group III nitrides: a comparative first-principles study. J Appl Phys 125:215705
Goetz G, Pohl UW, Schulz H-J (1992) Optical properties of vanadium ions in ZnSe. J Phys C 4:8253
Gordon L, Lyons JL, Janotti A, Van de Walle CG (2014) Hybrid functional calculations of *DX* centers in AlN and GaN. Phys Rev B 89:085204
Griffith JS (1964) The theory of transition metal ions. Cambridge University Press, Cambridge, UK
Grimmeiss HG, Janzén E (1986) Chalcogen impurities in silicon. In: Pantelides ST (ed) Deep centers in semiconductors. Gordon and Breach, New York, pp 87–146
Ham FS (1965) Dynamical Jahn-Teller effect in paramagnetic resonance spectra: orbital reduction factors and partial quenching of spin-orbit interaction. Phys Rev 138:A1727

Harrison WA (1973) Bond-orbital model and the properties of tetrahedrally coordinated solids. Phys Rev B 8:4487
Hayes W, Stoneham AM (1984) Defects and defect processes in nonmetallic solids. Wiley, New York
Hjalmarson HP, Vogl P, Wolford DJ, Dow JD (1980) Theory of substitutional deep traps in covalent semiconductors. Phys Rev Lett 44:810
Hubbard J (1963) Electron correlations in narrow energy bands. Proc Roy Soc (London) A276:238
Isoya J, Kanda H, Uchida Y, Lawson SC, Yamasaki S, Itoh H, Morita Y (1992) EPR identification of the negatively charged vacancy in diamond. Phys Rev B 45:1436
Itoh N (1982) Creation of lattice defects by electronic excitation in alkali halides. Adv Phys 31:491
Iwahara N, Vieru V, Ungur L, Chibotaru LF (2017) Zeeman interaction and Jahn-Teller effect in the Γ_8 multiplet. Phys Rev B 96:064416
Jahn HA, Teller E (1937) Stability of polyatomic molecules in degenerate electronic states, I. Orbital degeneracy. Proc R Soc Lond A Math Phys Sci A161:220
Jaros M, Brand S (1976) Localized defects in III-V semiconductors. Phys Rev B 14:4494
Jaros M, Brand S (1979) Electronic states associated with the substitutional nitrogen impurity in GaP_xAs_{1-x}. J Phys C 12:525
Jaros M, Rodriguez CO, Brand S (1979) Self-consistent pseudopotential calculation of electronic states associated with a reconstructed silicon vacancy. Phys Rev B 19:3137
Johnson NM, Herring C, van de Walle CG (1994) Inverted order of acceptor and donor levels of monatomic hydrogen in silicon. Phys Rev Lett 73:130
Jones R (ed) (1996) Early stages of oxygen precipitation in silicon. Kluwer, Dordrecht
Kajihara SA, Antonelli A, Bernholc J, Car R (1991) Nitrogen and potential *n*-type dopants in diamond. Phys Rev Lett 66:2010
Kaminska M, Weber ER (1993) EL2 defect in GaAs. In: Weber ER (ed) Imperfections in III/V materials: semiconductors and semimetals, vol 38. Academic Press, Boston, pp 59–89
Kastner M, Adler D, Fritzsche H (1976) Valence-alternation model for localized gap states in lone-pair semiconductors. Phys Rev Lett 37:1504
Kauffer E, Pecheur P, Gerl M (1977) Comment on the electronic structure of the neutral vacancy in silicon. Phys Rev B 15:4107
König E, Kremer S (1977) Ligand field energy diagrams. Plenum Press, New York
Kaufmann U, Schneider J (1982) Point defects in GaP, GaAs, and InP. Adv Electronics Electron Phys 58:81
Komsa H-P, Pasquarello A (2012) Intrinsic defects in GaAs and InGaAs through hybrid functional calculations. Physica B 407:2833
Koster GF, Slater JC (1954a) Wave functions for impurity levels. Phys Rev 95:1167
Koster GF, Slater JC (1954b) Simplified impurity calculation. Phys Rev 96:1208
Kreissl J, Schulz H-J (1996) Transition-metal impurities in II–VI semiconductors: characterization and switching of charge states. J Cryst Growth 161:239
Lang DV (1986) DX centers in III-V alloys. In: Pantelides ST (ed) Deep centers in semiconductors. Gordon and Breach, New York, pp 489–539
Lang DV, Logan RA (1977) Large-lattice-relaxation model for persistent photoconductivity in compound semiconductors. Phys Rev Lett 39:635
Lannoo M, Bourgoin J (1981) Poind defects in semiconductors. Springer, Berlin
Lin L, Chen NF, Zhong X, He H, Li C (1998) Relationship between deep-level centers and stoichiometry in semi-insulating gallium arsenide. J Appl Phys 84:5826
Lindefelt U, Zunger A (1982) Quasiband crystal-field method for calculating the electronic structure of localized defects in solids. Phys Rev B 26:846
Lindefelt U, Zunger A (1984a) Breathing-mode relaxation around tetrahedral interstitial 3*d* impurities in silicon. Phys Rev B 30:1102
Lindefelt U, Zunger A (1984b) Interstitial transition atom impurities in silicon: electronic structure and lattice relaxation. J Phys C Solid State Phys 17:6047
Lipari NO, Bernholc J, Pantelides ST (1979) Electronic structure of the Jahn-Teller distorted vacancy in silicon. Phys Rev Lett 43:1354

Lischka K (1986) Deep level defects in narrow gap semiconductors. Phys Stat Sol 133:17
Loferski JJ, Rappoport P (1958) Radiation damage in Ge and Si detected by carrier lifetime changes: damage thresholds. Phys Rev 111:432
Louie SG, Schlüter M, Chelikowsky JR, Cohen ML (1976) Self-consistent electronic states for reconstructed Si vacancy models. Phys Rev B 13:1654
Ludwig GW, Woodbury HH (1962) Electron spin resonance in semiconductors. In: Seitz F (ed) Solid state physics, vol 13. Academic Press, New York, pp 223–304
Lüty F (1960) Höhere Anregungszustände von Farbzentren. Z Phys 160:1 (Higher excited states of color centers, in German)
Lüty F (1973) Electromodulation spectroscopy of localized excitations in crystals. Surf Sci 37:120
Lyons JL, Janotti A, Van de Walle CG (2014) Effects of carbon on the electrical and optical properties of InN, GaN, and AlN. Phys Rev B 89:035204
Madelung O (1981) Introduction to solid state theory. Springer, Berlin
Madelung O, Schulz M (eds) (1989) Impurities and defects in group IV elements and III-V compounds. Landolt-Börnstein, New Series Vol III/22b. Springer, Berlin
Malloy KJ, Khachaturyan K (1993) DX and related defects in semiconductors. In: Weber ER (ed) Imperfections in III/V materials: semiconductors and semimetals, vol 38. Academic Press, Boston, pp 235–291
McClure DS (1959) Electronic spectra of molecules and ions in crystals part II. In: Seitz F, Turnbull D (eds) Solid state physics, vol 9. Academic Press, New York, pp 399–525
Messmer RP, Watkins GD (1973) Molecular-orbital treatment for deep levels in semiconductors: substitutional nitrogen and the lattice vacancy in diamond. Phys Rev B 7:2568
Milnes AG (1983) Impurity and defect levels (experimental) in gallium arsenide. Adv Electronics Electron Phys 61:63
Mollwo E (1931) Ueber die Absorptionsspektren photochemisch verfärbter Alkalihalogenid-Kristalle. Nachr Ges Wiss, Göttingen, pp 97–100 (On the absortion spectra of photochemically colored alkali halide crystals, in German)
Mooney PM (1990) Deep donor levels (DX centers) in III-V semiconductors. J Appl Phys 67:R1
Mooney PM (1992) DX centers in III-V alloys: recent developments. In: Pantelides ST (ed) Deep centers in semiconductors, 2nd edn. Gordon and Breach, New York, p 643
Nakamura S, Pearton S, Fasol G (2000) The blue laser diode. Springer, Berlin
Neugebauer J, Van de Walle CG (1996a) Gallium vacancies and the yellow luminescence in GaN. Appl Phys Lett 69:503
Neugebauer J, Van de Walle CG (1996b) Role of hydrogen in doping of GaN. Appl Phys Lett 68:1829
Nickel NH (ed) (1999) Hydrogen in semiconductors II: semiconductors and semimetals, vol 61. Academic, San Diego
Northrup JE, Zhang SB (1993) Dopant and defect energetics: Si in GaAs. Phys Rev B 47:6791
Orgel LE (1955) Spectra of transition-metal complexes. J Chem Phys 23:1004
Pajot B, Clerjaud B (2013) Optical absorption of impurities and defects in semiconducting crystals. Springer, Berlin
Pankove JI, Johnson NM (eds) (1991) Hydrogen in semiconductors: semiconductors and semimetals, vol 34. Academic Press, Boston
Pantelides ST (1978) The electronic structure of impurities and other point defects in semiconductors. Rev Mod Phys 50:797
Pantelides ST (1986) Perspectives in the past, present, and future of deep centers in semiconductors. In: Pantelides ST (ed) Deep centers in semiconductors. Gordon and Breach, New York, pp 1–86
Pantelides ST (ed) (1992) Deep centers in semiconductors, 2nd edn. Gordon and Breach, New York
Pantelides ST, Grimmeiss HG (1980) Final-state effects in the excitation spectra of deep impurities in semiconductors. Sol State Commun 35:653
Pantelides ST, Harrison WA (1976) Electronic structure, spectra, and properties of 4:2-coordinated materials. I. Crystalline and amorphous SiO_2 and GeO_2. Phys Rev B 13:2667
Pantelides ST, Sah CT (1974) Theory of localized states in semiconductors. I. New results using an old method. Phys Rev B 10:621

Park CH, Chadi DJ (1997) Stability of deep donor and acceptor centers in GaN, AlN, and BN. Phys Rev B 55:12995

Parrot R, Boulanger D, Diarra MN, Pohl UW, Litzenburger B, Gumlich HE (1996) Model for the electronic and vibronic structure of $^4T_1$ levels of d^5 ions coupled to vibrational E-modes: case of the fluorescent level of Mn^{2+} in ZnS. Phys Rev B 54:1662

Pearton SJ, Corbet JW, Stavola M (1992) Hydrogen in crystalline semiconductors. Springer, Heidelberg

Peka P, Lehr MU, Schulz H-J, Pohl UW, Kreissl J, Irmscher K (1996) Vanadium centers in ZnTe crystals Part I: optical properties. Phys Rev B 53:1907

Pohl RW (1938) Zusammenfassender Bericht über Elektronenleitung und photochemische Vorgänge in Alkalihalogenidkristallen. Phys Z 39:36 (Summarizing report on electron conductivity and photochemical processes in alkali halide crystals, in German)

Pohl UW, Gumlich H-E (1989) Optical transitions of different Mn-ion pairs in ZnS. Phys Rev B 40:1194

Pohl UW, Ostermeier A, Busse W, Gumlich H-E (1990) Influence of stacking faults in polymorphic ZnS on the d^5 crystal field states of Mn^{2+}. Phys Rev B 42:5751

Prentice JCA, Monserrat B, Needs RJ (2017) First-principles study of the dynamic Jahn-Teller distortion of the neutral vacancy in diamond. Phys Rev B 95:014108

Queisser HJ (1971) Deep impurities. In: Madelung O (ed) Festkörperprobleme/advances in solid state physics, vol 11. Vieweg, Braunschweig, pp 45–64

Ren SY, Dow JD, Wolford DJ (1982) Pressure dependence of deep levels in GaAs. Phys Rev B 25:7661

Rinke P, Janotti A, Scheffler M, Van de Walle CG (2009) Defect formation energies without the band-gap problem: combining density-functional theory and the GW approach for the silicon self-interstitial. Phys Rev Lett 102:026402

Scherz U, Scheffler M (1993) Density-functional theory of *sp*-bonded defects in III/V semiconductors. In: Weber ER (ed) Semiconductors and semimetals, vol 38. Academic Press, Boston, pp 1–58

Schläfer HL, Gliemann G (1967) Einführung in die Ligandenfeldtheory. Akademische Verlagsgesellschaft, Frankfurt am Main (Introduction to ligand field theory, in German)

Schneider J, Dischler B, Seelewind H, Mooney PM, Lagowski J, Matsui M, Beard DR, Newman RC (1989) Assessment of oxygen in gallium arsenide by infrared local vibrational mode spectroscopy. Appl Phys Lett 54:1442

Schulman JH, Compton WD (1962) Color centers in solids. Pergamon Press, New York

Seitz F (1954) Color centers in alkali halide crystals II. Rev Mod Phys 26:7

Skowronski M (1992) Oxygen in gallium arsenide. In: Pantelides ST (ed) Deep centers in semiconductors, 2nd edn. Gordon and Breach, New York, pp 379–406

Stavola M (ed) (1998) Identification of defects in semiconductors: semiconductors and semimetals, vol 51A. Academic Press, San Diego

Stavola M (ed) (1999a) Identification of defects in semiconductors: Semiconductors and semimetals, vol 51B. Academic Press, San Diego

Stavola M (1999b) Hydrogen in silicon and germanium. In: Proceedings of 5th international symposium advanced science and technology of silicon materials, pp 337–343; reproduced in Hull R (ed) Properties of crystalline silicon, EMIS datareviews series vol 20, p 511

Stoneham AM (1975) Theory of defects in solids. Claredon Press, Oxford, UK

Street RA, Mott NR (1975) States in the gap in glassy semiconductors. Phys Rev Lett 35:1293

Sturge MD (1967) The Jahn-Teller effect in solids. In: Seitz F, Turnbull D, Ehrenreich H (eds) Solid state physics, vol 20. Academic Press, New York, pp 91–211

Sun HJ, Gislason HP, Rong CF, Watkins GD (1993) Different P_{In} antisites in *n*- and *p*-type InP. Phys Rev B 48:17092

Tanabe Y, Sugano S (1954a) On the absorption spectra of complex ions I. J Phys Soc Jpn 9:753 and 9:766

Tanabe Y, Sugano S (1954b) On the absorption spectra of complex ions II. J Phys Soc Jpn 9:766

Trautman P, Baj M, Baranowski JM (1998) Hydrostatic pressure and uniaxial stress in investigations of the EL2 defect in GaAs. In: Suski T, Paul W (eds) High pressure in semiconductor physics I. Semiconductors and semimetals, vol 54. Academic Press, San Diego, pp 427–455

Thiering G, Gali A (2017) *Ab initio* calculation of spin-orbit coupling for an NV center in diamond exhibiting dynamic Jahn-Teller effect. Phys Rev B 96:081115

Uba SM, Baranowski JM (1978) Depression of vibronic levels and transition from the dynamic to static Jahn-Teller effect in the $^4T_1$ multiplet: the case of Co^{2+} in ZnSe. Phys Rev B 17:69

Van de Walle CG (1991) Theory of isolated interstitial hydrogen and muonium in crystalline semiconductors. In: Pankove JI, Johnson NM (eds) Semiconductors and semimetals, vol 34. Academic Press, Boston, pp 585–622

Van de Walle CG, Denteneer PJH, Bar-Yam Y, Pantelides ST (1989) Theory of hydrogen diffusion and reactions in crystalline silicon. Phys Rev B 39:10791

Van de Walle CG, Neugebauer J (2003) Universal alignment of hydrogen levels in semiconductors, insulators and solutions. Nature 423:626

Vigneron JP, Scheffler M, Pantelides ST (1983) Electronic structure of self-interstitials and *sp*-bonded interstitial impurities in semiconductors. Physica B + C 117–118:137

Vogl P (1981) Chemical trends of deep impurity levels in covalent semiconductors. In: Treusch J (ed) Festkörperprobleme/advances in solid state physics, vol 21. Vieweg, Braunschweig, pp 191–219

Vogl P, Hjalmarson HP, Dow JD (1983) A Semi-empirical tight-binding theory of the electronic structure of semiconductors. J Phys Chem Sol 44:365

Wagner P, Holm C, Sirtl E, Oeder R, Zulehner W (1984) Chalcogens as point defects in silicon. In: Grosse P (ed) Festkörperprobleme/advances in solid state physics, vol 24. Vieweg, Braunschweig, pp 191–228

Watkins GD (1984) Negative-*U* properties for defects in solids. In: Grosse P (ed) Festkörperprobleme/advances in solid state physics, vol 24. Vieweg, Braunschweig, pp 163–189

Watkins GD (1986) The lattice vacancy in silicon. In: Pantelides ST (ed) Deep centers in semiconductors. Gordon and Breach, New York, pp 147–183

Watkins GD (1991) Metastable defects in silicon: hints for DX and EL2? Semicond Sci Technol 6:B111

Watkins GD (2000) Intrinsic point defects in semiconductors. In: Jackson KA, Schröter W (eds) Handbook of semiconductor technology, vol 1. Wiley-VCH, Weinheim, pp 121–165

Watkins GD, Messmer RP (1970) An LCAO-MO treatment for a deep defect level in a semiconductor. In: Keller SP, Hensel JC, Stern F (eds) Proceedings of 10th international conference on physics semiconductors, Cambridge MA, USA. NTL Bureau of Standards, Tech Inf Service, Springfield, pp 623–629

Watkins GD, Troxell JR (1980) Negative-*U* properties for point defects in silicon. Phys Rev Lett 44:593

Weber ER (1983) Transition metals in silicon. Appl Phys A 30:1

Weber ER (ed) (1993) Semiconductors and semimetals, vol 38. Academic Press, Boston

Williams RT (1978) Photochemistry of F-center formation in halide crystals. Semicond and Insul 3:251

Williams PJ, Eaves L, Simmonds PE, Henry MO, Lightowlers EC, Uihlein C (1982) High-resolution optical absorption spectroscopy on Cr-related defects in GaAs and GaP. J Phys C 15:1337

Wolf T, Ulrici W, Côte D, Clerjaud B, Bimberg D (1993) New evidence for bound states in the charge transfer spectra of transition-metal-doped III-V semiconductors. Mater Sci Forum 143:317

Zhang SB, Northrup JE (1991) Chemical potential dependence of defect formation energies in GaAs: application to Ga self-diffusion. Phys Rev Lett 67:2339

Zunger A (1983) One-electron broken-symmetry approach to the core-hole spectra of semiconductors. Phys Rev Lett 50:1215

Zunger A (1985) Theory of 3*d* transition atom impurities in semiconductors. Annu Rev Mater Sci 15:411

Zunger A (1986) Electronic Structure of 3*d* transition-atom impurities in semiconductors. In: Ehrenreich H, Turnbull D (eds) Solid state physics, vol 39. Acedemic Press, Orlando, pp 275–464

Zunger A, Lindefelt U (1983) Electronic structure of transition-atom impurities in semiconductors: substitutional 3*d* impurities in silicon. Phys Rev B 27:1191

Defects in Amorphous and Organic Semiconductors

20

Karl W. Böer and Udo W. Pohl

Contents

1 Tailing States in Disordered Semiconductors 862
1.1 The Anderson Model 863
1.2 Anderson and Anderson-Mott Localizations 865
1.3 Band Tails, Localization, and Mobility Edge 867
1.4 Simulation of Defect States in Amorphous Semiconductors 868
2 Defects in Amorphous Semiconductors 869
2.1 Classes of Amorphous Semiconductors 869
2.2 Defect Types in Amorphous Semiconductors 871
3 Defects in Organic Semiconductors 876
3.1 Classes of Organic Semiconductors 876
3.2 Structural Defects in Small-Molecule Crystals 877
3.3 Defects in Polymers 886
4 Summary 890
References 891

Abstract

Amorphous and organic semiconductors have strong topological irregularities with respect to specific ideal structures, which depend on the particular class of such semiconductors. Most of these defects are rather gradual displacements from an ideal surrounding. The disorder leads to defects levels with a broad energy distribution which extends as band tails into the bandgap. Instead of a sharp band edge known from crystalline solids a mobility edge

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_20

exists separating between extended states in the bands and localized states in the band tails.

Amorphous semiconductors, also referred to as semiconducting glasses, comprise the classes of amorphous chalcogenides and tetrahedrally bonded amorphous semiconductors. Amorphous chalcogenides are structurally floppy solids with low average coordination numbers and pronounced pinning of the Fermi level near midgap energy. The more rigid tetrahedrally bonded amorphous semiconductors have larger coordination numbers. They may be well doped *p*-type and *n*-type much like crystalline semiconductors.

Organic semiconductors comprise small-molecule crystals and polymers. Both have weak intermolecular bonds favoring deviations from ideal alignment. In small-molecule semiconductors the structure of thin films grown on substrates usually deviates from the structure of bulk crystals, with a substantially different molecule ordering at the interface and a strong dependence on the dielectric properties of the substrate. Polymers consist of long chain-like molecules packed largely uniformly in crystalline domains separated by amorphous regions with tangled polymer chains. Besides chemical structure of the chains crystallinity depends on the molecular length.

Keywords

Amorphous chalcogenides · Anderson localization · Anderson-Mott localization · Band tails · Coordination number · Dangling bonds · Defects · Doping · Localization · Grain boundary · Mobility edge · Organic semiconductors · Paracrystallinity · Polymers · Semiconducting glasses · Small-molecule crystals · Point defects · Tailing of states · Trap states · Tetrahedrally bonded amorphous semiconductors · Thin-film phase

1 Tailing States in Disordered Semiconductors

Amorphous inorganic semiconductors and organic semiconductors show similarities in their electronic properties despite differences in bonding and structure. Both types of solids have substantial disorder in their structure induced by topological irregularities and potentially also by alloying. In Sect. 3 of ▶ Chap. 17, "Optical Properties of Defects" we have shown that heavily disordered crystalline semiconductors have band tails extending into the bandgap. Band tails, induced by disorder, are also referred to as *Lifshitz tails* (Lifshitz 1964).

Doping of disordered semiconductors for achieving controlled *n* or *p*-type conductivity is often hampered by a pinning of the Fermi level deep in the bandgap. It was originally believed that the transfer of electrons from donor- into acceptor-like tailing states would cause such a pinning, but the small density of respective defects and transport properties (see ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects") do not support this explanation. Anderson (1975) explained this finding in terms of a model in which the coupling of electrons to the lattice creates an attractive interaction between two electrons of opposite spin, which occupy the same

one-electron defect state; this leads to negative-U centers (Sect. 2.8 in ▶ Chap. 19, "Deep-Level Centers"). The simple model was improved by Street and Mott (1975) and Kastner et al. (1976), and is outlined in the next section.

There is a continuous transition from band states that are *not localized* to *localized states* of major defects within the bandgap. Depending on the degree of disturbance of the lattice potential, the eigenstate of an arbitrary host atom may lie in the band, near the band edges, or further separated within the bandgap. An important question relates to the distribution of these levels in a typical amorphous semiconductor. More specifically, we are interested when such levels can be considered part of the band, and when they become defect levels in the bandgap. This distinction is easily visualized in a crystalline semiconductor, in which band states can be occupied by electrons that are described by nonlocalized Bloch wavefunctions, as opposed to gap states described by localized electron eigenfunctions. In amorphous semiconductors and in many cases also in organic semiconductors, this is not possible since $\mathbf{k}$ is no longer a good quantum number due to the large topological disorder.

1.1 The Anderson Model

An instructive approach to model the level distribution in a disordered semiconductor is to change from a strictly periodic potential, representing a crystal and yielding bands separated by gaps in the classical sense, to a perturbed potential, which can be made less and less periodic, and analyze the resulting eigenfunctions (Anderson 1958). Anderson starts from a *three-dimensional* Kronig-Penney potential, which yields simple bands interspaced with bandgaps, much like the one-dimensional case discussed in Sect. 1.2 of ▶ Chap. 6, "The Origin of Band Structure". The band width ΔE_{B} can be expressed as

$$\Delta E_{\mathrm{B}} = 2\overline{m}I, \tag{1}$$

where $\overline{m}$ is the mean coordination number[1] and I is the transfer integral

$$I = \int \psi^*(\mathbf{r} - \mathbf{R}_n) H \psi(\mathbf{r} - \mathbf{R}_{n+1})\, d\mathbf{r} \tag{2}$$

between states localized at $\mathbf{R}_n$ and $\mathbf{R}_{n+1}$, and where H is an appropriate one-particle Hamiltonian. The transfer integral can be approximated in the form

$$I = I_0 \exp(-r/r_0); \tag{3}$$

[1] In amorphous structures the average number of next neighbors may substantially differ from that of the corresponding crystalline counterpart, see Sect. 3.1 of ▶ Chap. 3, "The Structure of Semiconductors".

it decreases with increasing well depth $V = (1/e)\, I$. Here r is the distance from the well center and r_0 is the fall-off radius of the transfer integral. For a hydrogen-like potential well, I_0 is given by

$$I_0 = \left[\frac{3}{2}\left(1 + \frac{r}{r_0}\right) + \frac{1}{6}\left(\frac{r}{r_0}\right)^2\right] \frac{\sqrt{2me^2E_0}}{\hbar}, \tag{4}$$

with E_0 as the ground state in a single well and

$$r_0^2 = \hbar^2/(2mE_0). \tag{5}$$

With periodic wells, the ground state E_0 broadens to a band with band width ΔE_B – see Eq. 1. It results in a rather narrow band when the wells are sufficiently deep and spaced widely enough (Fig. 1a).

This model has a wide range of application. For instance, we can use the results to explain that donors, when spaced close enough to each other, produce a narrow impurity band rather than a sharp ground-state level. The application of this model to amorphous semiconductors will become evident in the next section.

Let us first explore what happens when the model is applied to a periodic lattice potential, however, with changes from the strict periodicity, introduced by changing the interatomic distance or the potential. Anderson superimposes a random potential V with a spread of $\pm\Delta V/2$ onto the average well depth V_0 (Fig. 1b) and consequently obtains a broader level distribution with tails beyond the original band edges. When ΔV is very small compared to the well depth, only small deviations from the periodic Bloch-type solutions occur. These result in some scattering of essentially free Bloch electrons within this band, with a mean free path given by (Mott and Massey 1965):

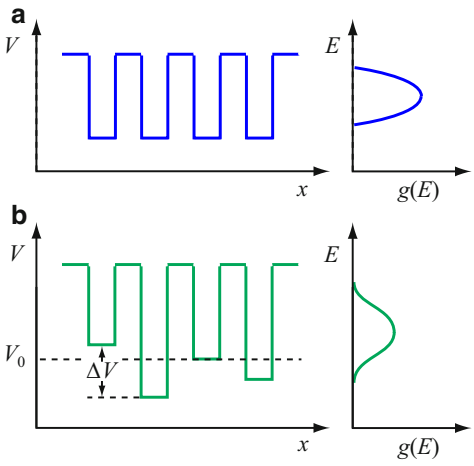


Fig. 1 Anderson model: (**a**) periodic potential and resulting level distribution. (**b**) Anderson potential with random potential V added to potential-well depth and resulting level distribution

$$\lambda = \frac{\hbar}{\pi}\left(\frac{2}{\Delta V}\right)^2 \frac{v_e}{a^3\, g(E)}, \quad \text{where } v_e = \frac{\hbar}{m}\, k. \tag{6}$$

When $\lambda \gg a$, all levels are extended band levels.[2] The fluctuating potential then results only in a slight perturbation of the band edges.

1.2 Anderson and Anderson-Mott Localizations

From Eq. 6 we see that the mean free path λ decreases with increasing spread of the fluctuating potential ΔV. When the mean free path is reduced to the distance between the wells, all states within the band become localized. An estimate for the relative mean free path can be obtained from Eq. 6, using the classical formula for the density of states $g(E)$ as an approximation near the band edge:

$$g(E)\ dE = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} \sqrt{E}\ dE, \tag{7}$$

yielding

$$\frac{\lambda}{a} = 32\pi\left(\frac{I}{\Delta V}\right)^2 = \frac{8\,\pi}{\overline{m}^2}\ \left(\frac{\Delta E_{\mathrm{B}}}{\Delta V}\right)^2. \tag{8}$$

Localization occurs when $\lambda/a = 1$, i.e., when ΔV increases to $0.7 \times \Delta E_{\mathrm{B}}$ for an average coordination number $\overline{m} = 6$. The electron is no longer free to move within the band all over the solid but is localized within the radius of any one atom.

A serious problem of Anderson localization is quite subtle. Anderson determines, as a criterion for localization, the absence of quantum diffusion in the disordered system as a function of energy and strength of disorder. This corresponds to a more stringent decrease of the wavefunction with increasing distance from a center as expressed by Eq. 8, so that the remaining overlap is insufficient for diffusion from neighbor to neighbor of an electron in such a center. This yields for an *Anderson localization* a potential fluctuation

$$\Delta V_{\mathrm{A}} \gtrsim \left(\sqrt{8\pi}/(\kappa_{\mathrm{A}}\,\overline{m})\right)\ \Delta E_{\mathrm{B}}, \tag{9}$$

[2]It should be noted that the definition of a band state is related to the coherence length of an electron wave, which is essentially the same as the mean free path λ.

with a numerical factor $\kappa_A \cong 6$. Others have obtained values of κ_A between 1.3 and 5. A review of these estimations is given by Thouless (1974). The area of Anderson localization and the transition from localized to delocalized states has been active for many years; for a review, see Lee and Ramakrishna (1985).

Mott has applied the Anderson idea to randomly distributed defects as they may exist in heavily doped crystals or for defect states in amorphous or organic semiconductors, which results in an *impurity band* of width ΔE_{imp} - see Sect. 1 of ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects". Instead of a random distribution of the potential well depth, a random distribution of centers *in space* is now assumed. If the density of the randomly placed centers is sufficiently large, we can think of these as forming an amorphous semiconductor with lateral disorder.

With a density of N_{imp} of such centers, the average distance of two neighboring centers is given by

$$r_{imp} = \sqrt[3]{2/N_{imp}}. \tag{10}$$

When close enough, they interact with each other and influence their eigenstates as given by the *transfer integral*

$$I_{imp} \propto V_{imp} = V_0 \exp\left(-r_{imp}/r_0\right). \tag{11}$$

Equating this V_{imp} (which fluctuates with r_{imp}) with the fluctuating potential in the Anderson model and assuming according to Eq. 9 that localization occurs when

$$V_{imp} \cong \Delta V_A \cong 2\Delta E_{imp}, \tag{12}$$

we have with Eq. 1

$$V_0 \exp\left(-r_{imp}/r_0\right) \cong 4\overline{m} V_0 \exp(-r_A/r_0). \tag{13}$$

For an average coordination number $\overline{m} \cong 5$, we obtain from Eq. 13

$$\left(r_A - r_{imp}\right)/r_0 \cong \ln(4\overline{m}) \cong 3. \tag{14}$$

A somewhat lower coordination number for covalent crystals or amorphous semiconductors has only minor influence on the numerical value of condition (17) given below. With an Anderson-Mott density, below which localization occurs,

$$\left(N_{imp}\right)_{loc} = \left(\frac{4\pi}{3} r_A^3\right)^{-1}, \tag{15}$$

we now obtain from Eq. 14 with Eqs. 10 and 15:

$$\left(N_{\text{imp}}\right)_{\text{loc}} = N_{\text{A-M}} \cong 8 \times 10^{-3} r_0^{-3}. \tag{16}$$

Assuming hydrogen-like centers, this yields with $r_0 = a_{\text{qH}}$ the condition for *Anderson-Mott localization* - see Mott and Davis (1979):

$$N_{\text{A-M}}^{1/3} \times a_{\text{qH}} \cong 0.2. \tag{17}$$

This is essentially the same condition that was used for an insulator-metal transition when the density N described donors in a crystalline semiconductor – see the Mott transition discussed in Sect. 1.2 of ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects".

1.3 Band Tails, Localization, and Mobility Edge

We will now apply the concept of Anderson-Mott localization to a heavily distorted semiconductor. Most of the states well within the conduction or valence band are similar to the states within a crystal. This can be justified by the measured $g(E)$ distribution of an amorphous semiconductor shown in Fig. 43 of ▶ Chap. 8, "Bands and Bandgaps in Solids". We may consider the tailing states extending into the bandgap (caused by defects or deviations from ideal topology) as disorder, resulting in a continuous distribution of states. Deeper states occur less frequently, because the centers which produce such states are less probable. When deep enough, each type of defect center will produce a localized level. Shallower centers are, however, close enough: Their levels will broaden into bands, and the resulting states are no longer localized. Overlapping levels and narrow bands will all melt into the tailing states. Thus one expects a transition from localized to delocalized states as one goes from deep tail states to the band states at a critical energy, referred to as the *mobility edge*.

There is a smooth transition in the density of states between localized and delocalized states. It is hence difficult to define a band edge in the classical sense. However, with the help of the localization criterion, a pseudo-band-edge or mobility edge can be defined as the energy of the defect level at which localization occurs. Fig. 2 compares the density of states in a crystalline semiconductor showing a well-defined band edge with that of a highly disordered semiconductor exhibiting strong tailing of states into the bandgap; a mobility edge is indicated for n-type conductivity. Carriers which are thermally activated above the mobility edge contribute to charge transport, while carriers at lower energy are localized in defect states. Consequences of band tailing are treated in Sects. 2 and 4 of ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects", when carrier transport is discussed - see also Götze (1981), and Lee and Ramakrishna (1985). For optical transitions involving states below and above the mobility edge of valence and conduction bands see Orapunt and O'Leary (2008).

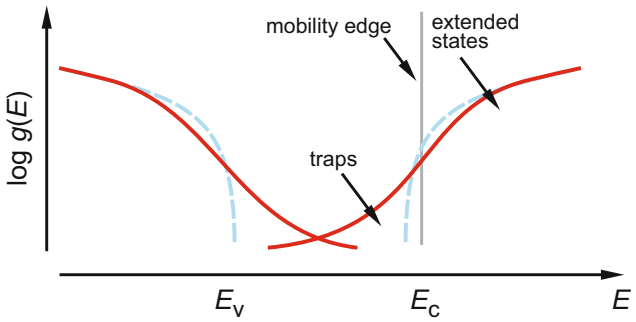


Fig. 2 Tailing of states into the bandgap of a heavily disordered semiconductor (*red solid lines*) with a mobility edge separating extended and localized states. *Dashed lines* indicate the density of states in a crystalline semiconductor

1.4 Simulation of Defect States in Amorphous Semiconductors

Computer simulations have reached a level of realism required for explaining experimental observations and provided microscopic pictures of processes in amorphous semiconductors, such as the subtle effects of photo-induced structural changes affecting electron-lattice coupling (e.g., for α-Se by Drabold et al. 2001; for α-Si:H by Pfanner et al. 2013). First-principles simulations on highly distinct amorphous solids (α-Si, α-Se, α-GaN) indicate that there is generally a strong correlation between the localization of electron states and the magnitude of thermally induced fluctuations in energy eigenvalues; light-induced changes are intimately connected to changes in cell charge states (Prasai et al. 2016).

The first step in any simulation study of electronic states in a disordered semiconductor is an appropriate atomistic model. Several levels of model potentials are being applied to arrive at realistic modeling of the disordered system (Wooten and Weaire 1989; Varshishta et al. 1996; Djordjevic et al. 1995; Mousseau and Barkema 2000). In smaller systems also accurate ab initio calculations were performed (e.g., for α-Si:H by Jarolimek et al. 2009 and Meidanshahi et al. 2019).

An interesting result from such structural modeling demonstrates that an amorphous structure is not the same as that of an equilibrated liquid, particularly for systems which are not glass formers. Thus, for α-Si, a direct quench from the melt yields a dip in the electronic density of states instead of a state-free optical bandgap. This is because the liquid, which is predominantly like a sixfold coordinated metal, is topologically quite distinct from amorphous (or crystalline) Si, which is tetrahedrally bonded. Thus quenching a liquid in simulation studies to obtain an amorphous structure gives rise to many overcoordinated defects instead of dangling bonds.

Once an atomistic model is established, one can perform electronic structure calculations applying either model Hamiltonians and tight-binding methods (Lewis and Mousseau 1998) or first-principle methods like density-functional theory (Sanchez et al. 1997). Some significant findings from such studies are the following.

In α-Si, the midgap-defect states originate from dangling bonds. The floating-bond conjecture discussed in Sect. 2.2.3 assuming that fivefold coordinated atoms are important has been disproved, though such states are shifted well toward the edge of the conduction band and are not as localized as the dangling bonds.

Cluster simulations with systems containing 4096 or more atoms showed that states approaching the middle of the bandgap are exponentially localized, whereas states deep in the valence-band (and conduction-band) tail are well-extended and essentially uniform throughout the amorphous semiconductor.

Simulations yield the following qualitative picture of the Anderson transition: Since severe distortions are rare, respective clusters are isolated from each other and lead to localized "cluster states". For less severe distortions the abundance increases, and cluster states of similar energy not far away from each other get likely. Two, three, or more cluster states of nearly equal energy and appreciable overlap become possible and the electronic states become delocalized. This is the Anderson transition, which is basically a quantum-percolative transition.

2 Defects in Amorphous Semiconductors

In some respects, ideal amorphous semiconductors are topologically equivalent to their crystalline counterparts except for small variations in bond length and bond angles (see Sects. 3.1.1 and 3.2.2 in ▶ Chap. 3, "The Structure of Semiconductors" and Tanaka 1998). Therefore, it is reasonable to explain some defect properties in amorphous semiconductors also with the concept of tailing states into the bandgap (Mott 1969). Before exploring the specific nature of defects in amorphous semiconductors, we emphasize the distinction of two classes of amorphous semiconductors.

2.1 Classes of Amorphous Semiconductors

There are at least two different classes of covalent amorphous semiconductors, distinguished by their electronic response to changes in the defect structure.[3]

The *first class* are *amorphous chalcogenides*, which are alloys containing elements of group VI and others, such as elements of groups IV and V, as the main glass-forming components, such as α-As_2 components, for instance, α-$Te_{40}As_{35}Si_{15}Ge_7P_3$, a well-known material because of its switching[4] capability (Cohen et al. 1969) reviewed by Hudgens (2012). The class also includes less complex compounds, such as α-As_2Se_3 and monatomic α-Se. Some of these systems have both covalent and ionic components to their bonds. Typically, the

[3]A third class, somewhat in between these two, contains α-P and α-As.

[4]Bipolar devices made from such materials are capable of switching at high speed from a low to a high conducting state at a critical bias (Ovshinsky 1968).

chalcogenides have a low average coordination number, below 2.4. For comparison, the average coordination numbers of some crystals are, e.g., 4 for Si, Ge, GaAs, etc., 3 for As or GeTe, 2.7 for $GeTe_2$, 2.4 for As_2Se_3, and 2 for Se (Adler 1985). It should be noted that structures with average coordination number $\overline{m}$ exceeding 2.4 are structurally rigid, whereas those with smaller values are structurally floppy. For a review on the average-coordination number concept in chalcogenide glasses see Varshneya et al. (1993).

In amorphous chalcogenides, it is nearly impossible to move the Fermi level from its near-midgap position by doping. Some chemical modification, i.e., changing its conductivity to become extrinsic while maintaining the bandgap, however, was achieved in a few chalcogenide glasses by adding modifying elements. These should not be confused with dopants, since the necessary concentration of the modifier is relatively large to become effective. As such modifiers, the transition metals Ni, Fe, and Co can be used and in some instances W, B, or C (Ovshinsky 1977, 1980). For a review, see Adler (1985) and Adler and Fritzsche (1985).

Amorphous chalcogenides such as GeTe are prominent compounds for phase-change materials used for rewritable optical data storage, and are also considered promising for nonvolatile electrical data storage (Wuttig and Yamada 2007, Jia et al. 2020). GeTe has a low resistivity in its crystalline rhombohedral phase (*c*-GeTe), changing reversibly and fast to high resistivity for amorphous α-GeTe. In *c*-GeTe both Ge and Te atoms are threefold coordinated, with a lone pair being localized on each of them. The Ge vacancy is a prominent defect with low formation energy due to a strong reconstruction; this defect pins the Fermi energy to the top of valence band (Huang and Robertson 2012). In α-GeTe most of the defects states in the bandgap result in high resistivity; they originate from groups of Ge atoms close to each other with at least one over- or under-coordinated Ge atom with respect to those in *c*-GeTe (Zipoli et al. 2016). The resistance of the amorphous phase slowly increases over time by an evolution of the bond network towards a first-neighbor topology more similar to that of the crystalline phase. By a recently suggested mechanism of valence-alternation pairs, defects comprising Ge–Ge chains form as the result of Ge–Te bond breaking, with low energy cost and preserving the total number of bonds (Li and Robertson 2020). This model accounts for most of previous experimental and theoretical findings.

The *second class* is composed of *tetrahedrally bonded amorphous semiconductors*, such as α-Si:H, α-GaAs, and α-$CdGeAs_2$, with an average coordination number $\overline{m} > 2.4$; for details on α-Si:H, see Street (1991). These amorphous semiconductors behave substantially different: When properly prepared (all Si atoms fourfold coordinated in α-Si:H), the solids respond easily to doping with a shift in their Fermi level much like a crystalline semiconductor. These semiconductors can be made *n-type* or *p-type* by doping with donors or acceptors, respectively. Nevertheless, α-Si:H also has a strong tailing of defect states into the bandgap, although these tails do not overlap significantly near the center of the bandgap. For an example of such a band tailing in α-Si:H, see Fig. 17 in ► Chap. 13, "Band-to-Band Transitions".

Aside from a pinning of the Fermi level due to overlapping tail states for the first class, and sensitivity to doping for the second class, there are other experimental distinctions for these two major groups. These relate to the strength of *electron-spin resonance* (ESR) signal, which indicates the existence of unpaired spin electrons (charged defects), the steepness of the *optical absorption edge*, which indicates the degree of band tailing, and other properties relating to carrier transport as discussed in ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects".

2.2 Defect Types in Amorphous Semiconductors

Defects are easily identified in *crystal lattices*, where vacancies, interstitials, and even small deviations from the periodic structure can be identified. This is much more difficult in amorphous semiconductors, where deviations from the *coordination number* $\overline{m}$, *bond angle* θ, and *average bond length* a are the principal defect features. In general, many defects in glasses are of a gradual rather than distinct nature and may be classified into:

- Local strain-related defects, i.e., variations of a, θ, and $\overline{m}$
- Deviation from an optimal bonding configuration
- Incorporation of small concentrations of impurities
- Dangling or floating bonds
- Microcrystallite boundaries, observed in some glasses
- Variably sized small voids

Most of these defects cause changes in the electron-energy spectrum by deforming the band edge and extending states into the bandgap (see, e.g., Tanaka and Nakayama 1999). Depending on the type of amorphous semiconductor, the ensuing defect spectrum may extend nearly exponentially from the band edge into the bandgap or into the Urbach tail, or may produce well-defined (although broad) peaks of the distribution function within the bandgap (Fig. 3).

2.2.1 Strain-Related Defects

The *local strain-related defects* may be seen as similar to acoustic phonon-induced deformations (see Sect. 2.2.1 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields") of the band edge. Here, however, they are caused by a stationary, frozen-in strain, often also with a larger amplitude than for thermal phonons. Each of these stretched bonds or deformed bond angles can produce a level in the band tail when the deformation from the ideal values is sufficiently large.

In tetrahedrally bonded semiconductors, deviations of the bonding angle and coordination number can also result in different types of bonding between neighboring atoms. For instance, in amorphous Si, an sp^3 hybrid, an sp^2 hybrid, or a p^3-configuration produces respectively a neutral, positively, or negatively charged dangling bond, while an s^2p^2 hybrid produces a twofold coordinated Si atom. These bonds may be formed to relieve some of the stress. The corresponding bond angles are 109.5° (sp^3), 120° (sp^2), and 95° (p^3 and s^2p^2) (Adler 1985).

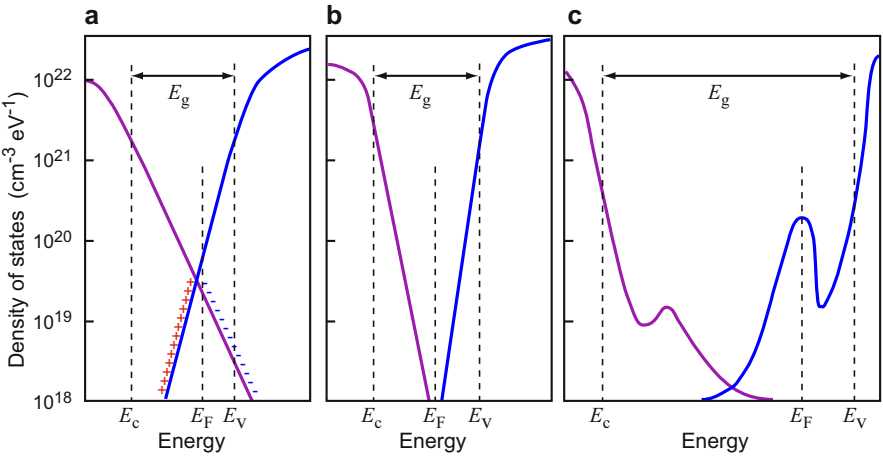


Fig. 3 Exponential distribution of states into the bandgap: (**a**) in chalcogenide glasses, (**b**) in tetravalent glasses, and (**c**) in tetravalent glasses with distinct peaks

Often, these defect centers are identified as $A_{\overline{m}}^{q}$, with A identifying the chemical species: T for tathogen, P for pnictogen, and C for chalcogen, i.e., elements of the IV, V, or VI groups, respectively; $\overline{m}$ is the average coordination number and q is the charge character relative to the lattice (2+, +, 0, −, or 2–). For instance, T_4^0 represents a neutral, fourfold coordinated Si atom, while T_3^0 describes a neutral dangling bond, both for the sp^3 ground state. As possible defects in α-Si, all of T_2^{2+}, T_2^{+}, T_2^{0}, T_2^{-}, and T_2^{2-}, as well as T_3^{+}, T_3^{0}, and T_3^{-}, are being considered as centers with lower-than-normal coordination and as possible alternatives for local stress-relief.

2.2.2 Under- and Overcoordinated Defects

There is evidence that two types of "intrinsic defects" are prevalent in α-Si:H – the threefold coordinated Si-atom, which is equivalent to a vacancy in crystalline Si, and the fivefold *overcoordinated* Si atom, equivalent to a self-interstitial. These two types can be created in pairs, as indicated in Fig. 4, somewhat similar to a Frenkel-pair creation. This reaction may be initiated by light and provides a possible mechanism for the Staebler-Wronski effect (Sect. 3.2.2 of ▶ Chap. 3, "The Structure of Semiconductors"; Staebler and Wronski 1977). Another mechanism was suggested from molecular dynamics simulations of α-Si:H, indicating a pathway to a preferential formation of metastable SiH_2 conformations in the photo-excited state (Abtew and Drabold 2006). Other photochemical processes in chalcogenide glasses can be observed as darkening under light, see, e.g., Li et al. (1989).

2.2.3 Dangling and Floating Bonds

Dangling bonds do not seem to play a major role in most semiconducting glasses of technical interest. The elimination of these dangling bonds can be obtained, for

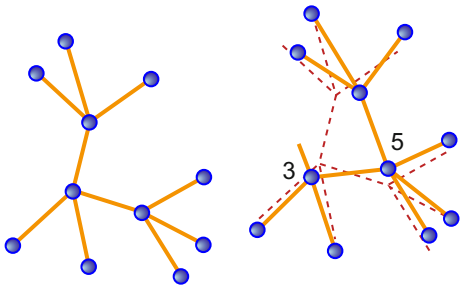


Fig. 4 Pair creation (and recombination) of dangling (3) and floating (5) bonds (After Pantelides 1989)

example, by H or F in α-Si:H or α-Si:F. Such a removal results in the major differences between the amorphous Si and the amorphous Si:H or Si:F alloys. The description as an alloy is used here since a large atomic fraction (>10%) of H or F is incorporated. As a consequence of the dangling-bond removal, α-Si:H or α-Si:F can be doped and turn *n*- or *p*-type, similar to crystalline Si, as shown in Fig. 5a.

The dominant intrinsic defect in α-Si:H, the *D center*, characterized by paramagnetic resonance ($g = 2.0055$), was initially assigned to the dangling bond, i.e., a threefold coordinated Si atom. An alternative explanation was given in terms of a *floating bond*, i.e., the fivefold coordinated Si atom (Pantelides 1986); more recent studies favor the dangling bond, indicating a substantial influence of the local geometric and electronic structure of the immediate surrounding (Pfanner et al. 2013). The close relation of these two centers is shown in Fig. 6. In contrast to the dangling bond, the floating bond is highly mobile and is of interest to interstitial-mediated diffusion - see Sect. 3.2 in ▶ Chap. 15, "Crystal Defects".

2.2.4 Deviation from Optimal Bonding Configuration

The *deviations from an optimal bonding configuration* occur predominately in chalcogenide glasses and may be understood by comparing the relative bonding strengths of various bonds. For example, in amorphous Ge_xTe_{1-x}, a configuration wherein the stronger Ge-Ge bond appears most frequently while the number of the weaker Te-Te bond is minimized in the entire material, may be termed an *ideal amorphous* Ge_xTe_{1-x} *structure.* Any deviation from it may be identified as a defect of the structure and has a lower overall binding energy.

Another deviation from optimal bonding relates to a *valence-alternation pair,* in which the valency of nearby atoms in amorphous chalcogenide semiconductors is changed (see Zipoli et al. 2016). Charged dangling bonds may also be formed, in which the lowest-energy neutral defect is a threefold-coordinated site that is unstable with respect to the creation of singly and threefold coordinated charged atoms according to the sum reaction $2C_3^0 \rightarrow C_3^+ + C_1^-$, with a chalcogen *C* (Kastner et al. 1976; see also Sect. 2.2.1). However, these bonds exactly compensate each other and form "pairs", so that no electron-spin resonance signal results – see Sect. 2.2.7. As charged, but compensated defects, they effectively pin the Fermi level.

In chalcogenide glasses, defect states with negative correlation energy, i.e., *negative-U centers*, can occur (Anderson 1975, Ovshinsky 1976, Huang and

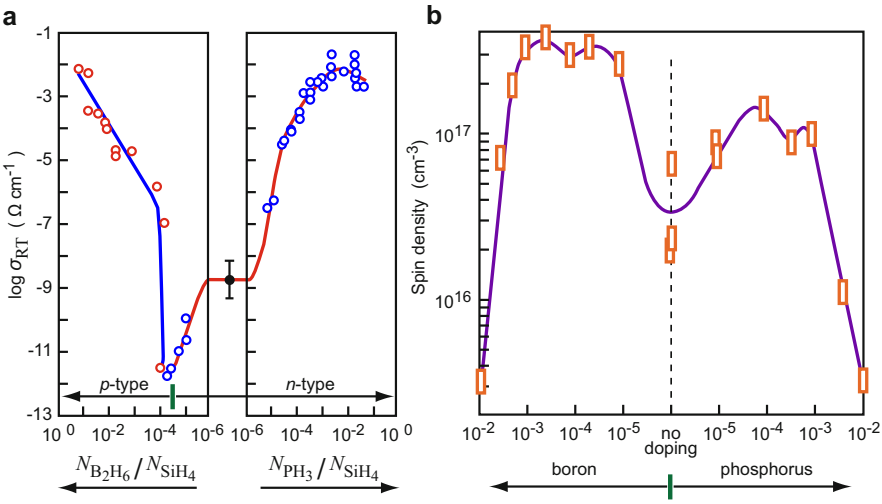


Fig. 5 (**a**) Room temperature electrical conductivity of α-Si:H as a function of the phosphine or diborane concentration during deposition (After LeComber and Spear 1976). (**b**) Spin density optically induced in α-Si:H as a function of doping gas pressure (After Knights et al. 1977)

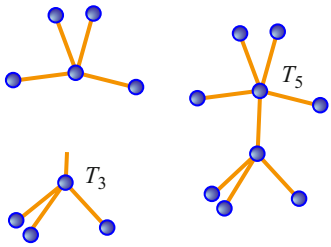


Fig. 6 Relationship of threefold (T_3) and fivefold (T_5) coordinated Si atoms with possible interconversion of these dangling and floating bonds (After Pantelides 1986)

Robertson 2012). These states can bind two electrons, the second with a larger binding energy than the first – see Sect. 2.8 in ▶ Chap. 19, "Deep-Level Centers". This can be explained when a strong electron-lattice interaction exists, and the energy released by the lattice deformation near the polarized defect (i.e., the defect state occupied by two electrons) is larger than the Coulomb repulsion between the two electrons. As a result, we expect in a system with N defect states and n electrons that $n/2$ states are doubly occupied in the lowest energy state, causing a pinning of the Fermi level. This also explains why these materials are diamagnetic.

2.2.5 Doping in Semiconducting Glasses

When the density of *impurity atoms* is small enough, and the atoms are not incorporated as part of the glass-forming matrix in tetragonal glasses, the ensuing defect may result in a distinct level similar to that in a crystalline semiconductor. Depending on the actual surrounding, however, the resulting energy level of a deep-level defect is different. The same defect may have even a donor- or an acceptor-like

character or, with external excitation, act as a recombination center in a different microscopic environment of the host.

In the tetrahedrally bonded amorphous semiconductors (e.g., in α-Si:H), the chemical nature of shallow defects determines their electronic defect behavior with less ambiguity. These act as *dopants,* with a similar effect as in crystals, though with a more complicated configuration. Incorporation of a P atom is likely to occur in an sp^3 configuration rather than a p^3 bonding. It is observed to shift the Fermi level to within 0.1 eV of the conduction-band edge (or better, the electron-mobility edge – see Fig. 2 and ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects") and thereby act as an effective donor. The incorporation of a B atom can be accomplished in an sp^3 or an sp^2 configuration, with bond angles of 109.5° or 120° and a coordination number of 4 or 3, respectively, or as a complex with bridging H atoms (Adler 1985). It acts as an acceptor. Other local bonding configurations can occur with the incorporation of N (pnictogen), 0, or S (chalcogen), resulting in different sets of donor levels within the bandgap.

Amorphous oxide semiconductors such as $InGaZnO_x$ have strong ionic bonding; in addition, the states of their conduction-band edge consist of metal *s* orbitals, which are much less affected by disorder than the *p*-like band-edge states of α-Si:H. This is favorable to form a sharp and shallow conduction-band tail. They also have a larger Bohr radius than α-Si:H, yielding a low density of states at the mobility edge and hence a high carrier mobility (Hosono 2006). Since they do not suffer from a Staebler–Wronski-like instability, these oxides are reasonably stable. Most amorphous oxide semiconductors based on ZnO or In_2O_3 have a conduction band deep below the vacuum level and are therefore only *n*-type; they are convenient for large-area unipolar devices like TFTs. A few oxides with fittingly located conduction and valence band are also bipolar dopable; for a review, see Robertson (2012).

2.2.6 Microcrystalline Boundaries and Voids

Microcrystallite boundaries seem to be absent in true glasses. However, there is some evidence that in a few semiconducting glasses precrystallization takes place. This is possibly the result of imperfect growth techniques and caused by the formation of small crystallite nuclei. In this case, the internal strain of the nuclei is relaxed by the creation of a boundary (microsurfaces) that resembles crystal boundaries with localized strain. Another mechanism of strain relaxation was pointed out by Meidanshahi et al. (2019), where H atoms preferentially bond to Si atoms with strained Si–Si bonds rather than dangling bonds in the most energetically stable configurations; thereby the midgap states in the electronic structure of α-Si are efficiently reduced. An example of recrystallization was observed for amorphous silicon on *c*-Si; amorphous Si layers on crystalline Si recrystallize with an activation energy of ~2.5 eV (Lietoila et al. 1982). Such a recrystallization may occur through diffusion of dangling bonds (Mosley and Paesler 1984), mediated by floating bonds (Pantelides 1989), converting all rings into six-member rings. On the other hand, a study of the effect of hydrogen on the recrystallization of *nanocrystalline* silicon indicates that H tends to migrate from the crystalline to the amorphous phase and

increases the interface thickness by inducing local disorder; immobile Si_nH_m hydrides are formed and impede recrystallization (Fugallo and Mattoni 2014).

Small voids of variable size can be identified in the center of large-number rings (see Fig. 1 of ▶ Chap. 1, "Properties and Growth of Semiconductors") in certain glasses. These defects show some similarities to vacancies in a crystal lattice, although they vary in size and bond reconfiguration.

2.2.7 Spin Density of Defects

Electron-spin resonance yields additional information about a defect (see Sect. 3.4.2 of ▶ Chap. 18, "Shallow-Level Centers"). When the defect possesses an electron unpaired with another one of opposite spin (short *unpaired spin*), the resulting magnetic momentum can be picked up by a spin-flip electromagnetic resonance experiment. The frequency and line shape of the resonance is influenced by the surrounding of the defect and yields more detailed information in crystalline solids. This hyperfine structure is washed out in amorphous semiconductors. However, from the density of unpaired spins, we still obtain valuable information. This density is small ($\sim 10^{16}$ cm^{-3}) in α-Si:H and verifies a low density of dangling bonds, T_3^0 (sp^3), which represent unpaired spins (see Adler 1985).

Coexistent with T_3^0 centers are pairs of T_3^+ and T_3^- centers, which have compensated spins. Their total energy depends on the relative distance (Kastner et al. 1976) and has its minimum value when they are nearest neighbors (*intimate charge-transfer defect* – Adler and Yoffa 1977).

The spin density can be used as a measure of uncompensated donors or acceptors and is shown for doped α-Si:H in Fig. 5b. However, the observed decrease of unpaired spin density for high-doping densities causes some problems in explanation. In contrast, *non*hydrogenated α-Si shows spin densities which are substantially larger and increase with damaging ion bombardment up to 10^{19} cm^{-3}. With subsequent annealing, the spin density is reduced as expected (Stuke 1976).

3 Defects in Organic Semiconductors

3.1 Classes of Organic Semiconductors

Organic semiconductors comprise two classes: small-molecule crystals and polymers. A characteristic common feature is the bonding of carbon atoms by a system of conjugated π electrons; see Sect. 3.3 in ▶ Chap. 2, "Crystal Bonding".

In *small-molecule crystals* molecules represent the building blocks of the solid. In bulk crystals these molecules are often arranged in a herringbone packing; see Fig. 17 and Sect. 1.5 in ▶ Chap. 3, "The Structure of Semiconductors". The weak intermolecular bonding between rigid molecules and the orientational degree of freedom favors the formation of local disorder defects and polymorphic structures. In devices instead of bulk crystals usually thin films of organic semiconductors are employed. The structure of thin-film phases deviates from the bulk

structure and depends on properties of the substrate and on film thickness. Carrier mobilities are particularly high in some organic bulk single-crystals such as Rubrene and Pentacene and orders of magnitude lower in respective thin films due to a very high density of defects. The π electrons are delocalized over the individual molecules, but conductivity of the crystal largely depends on the transfer integral between neighboring molecules that is strongly affected by the defects.

Polymers are interesting for applications due to their favorable large-scale processing ability and their robustness. The structure is generally more distorted, and consequently the mobility of carriers is usually lower than in small-molecule crystals. The wavefunction of the π electrons is not delocalized over the entire polymer backbone but localized in the double (or triple) bonds, yielding the alternating bond lengths of the conjugated π electron system. Defects in the bond alternation sequence yield unpaired electrons as discussed in Sect. 3.3; furthermore, strong disorder in the alignment of the polymer chains generally lead to low carrier mobility. In Sects. 3.2.1 and 3.2.2 the two classes of organic semiconductors are discussed separately.

The orientation of molecules as building blocks of organic crystals and variations in their internal structure provide additional degrees of freedom in the assembly of the solid compared to inorganic semiconductors. This and the weak intermolecular bonding forces give rise to orientational disorder in organic crystals. In addition, impurities introduce localized levels in organic semiconductors. All these imperfections are considered in the following.

3.2 Structural Defects in Small-Molecule Crystals

3.2.1 Molecule Crystals Grown on Substrates

Applications of organic crystals usually imply the deposition of *thin films* on either inorganic or organic substrates; we thus put some focus on thin films grown on substrates. Charge transport in organic materials is generally limited by numerous defects, which introduce a high density of trap states. Due to the weak intermolecular bonding organic crystals grown on a substrate do not form layers with significantly accumulated strain as observed for inorganic heterostructures (Sect. 1 in ▶ Chap. 16, "Crystal Interfaces"). Since epitaxy of organic crystals is performed at low depositions temperatures where many materials are stable, many layer/substrate combinations also may be realized. The large variety of alignment of molecules on a substrate and variations in conformations within individual molecules provide many degrees of freedom, giving rise to a large number of distortions and polymorphs in organic thin films.

Formation of an ordered organic layer depends on the interaction between substrate atoms and molecules of the layer. The structure of the interface is basically determined by the dielectric properties of the substrate material. At least the *initial* structure of the organic layer growing on a substrate is different for growth on metals and on insulators. For reviews on nucleation of organic semiconductors see Hooks et al. (2001), Virkar

et al. (2010), Evans and Spalenka (2015), Simbrunner and Sitter (2015), and Chen et al. (2021).

To illustrate the general features we consider the widely studied *pentacene*. This organic semiconductor is attractive due to the very high mobility of carriers [up to ~10 $cm^2/(Vs)$] found in bulk crystals and represents a model and benchmark system for organic small-molecule semiconductors. Pentacene ($C_{22}H_{14}$) is a polycyclic aromatic hydrocarbon composed of five fused benzene rings in a linear planar arrangement (Fig. 12a). Bulk crystals grow in the herringbone structure discussed in Sect. 1.5 of ▶ Chap. 3, "The Structure of Semiconductors"; the thin-film phases deviate from the bulk structure, representing largely modifications of the herringbone packing. Thin pentacene films were studied on many substrate materials.

On metals the interaction between substrate and the delocalized electrons of the layer molecules is strong. The contact of individual molecules with the surface is hence large, leading mostly to a *lying-down configuration* of the first molecule layer(s): The molecular long axis is then parallel to the surface. A lateral ordering aligned to the periodically arranged metal atoms may occur at suitable growth conditions; the respective molecular spacing in the organic layer may substantially differ from that in an organic bulk crystal. For sufficiently strong intermolecular interactions often a continuous variation toward the bulk value with increasing distance to the interface is observed, including various structural phases.

The structure of a monolayer thick pentacene layer deposited at room temperature on a Ni(111) surface is shown in Fig. 7. The molecules are strongly adsorbed on the metal surface and aligned along the $\langle 1\bar{1}0 \rangle$ directions of the Ni substrate, with an approximately equal fraction for the three possible orientations (Dinca et al. 2015). The arrangement is referred to as "random tiling" phase, since the alignment of a given molecule cannot be concluded from the directional alignment of its neighboring molecules. Pentacen layers subsequently deposited on the flat-lying molecules align in a tilted arrangement yielding a pronounced islanding (Käfer et al. 2007).

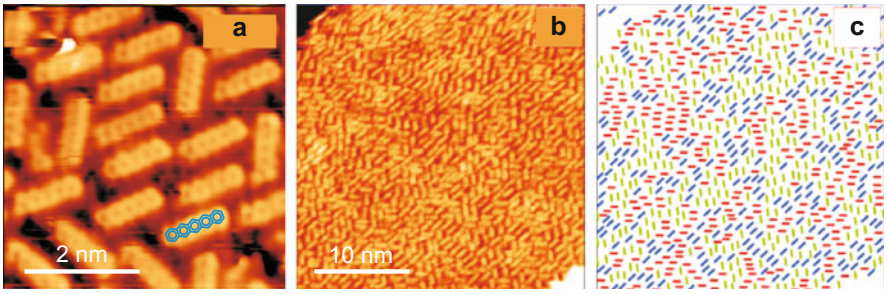


Fig. 7 Scanning-tunneling micrographs of pentacene molecules adsorbed on Ni(111). (**a**) Individual molecules with a superimposed structural model. (**b**) Overview image showing orientational disorder in the molecule alignment. (**c**) Same as (**b**) with color-coded orientations (After Dinca et al. 2015)

Organic molecules may be inserted as an interlayer between the metal substrate and the organic semiconductor to reduce the strong interaction. A monolayer of benzenethiol on an Au(111) surface was, e.g., applied for growing a pentacene film, yielding a crystalline bulk-like *standing-up phase* with the *ab* plane parallel to the interface as illustrated in Fig. 8a (Kang et al. 2005). Using such layers it was shown that defect states may occur even if the two-dimensional crystalline packing with the herringbone motiv is maintained. Scanning-tunneling microscopy images of two monolayer thick pentacene layers showed a broad Gaussian distribution of slight thickness variations in the range of 1–2 Å, well below the thickness of one monolayer (~15 Å). A corresponding displacement of pentacene molecules along the long molecular axis illustrated in Fig. 8 breaks the translation symmetry in a crystal; calculations show that such a disorder is sufficient to create defect states in the bandgap as shown in Fig. 8b.

On insulating substrates the interaction between the molecules of an organic layer and the atoms of a substrate is much weaker. Still the effect of the substrate atoms can lead to crystal structures distinct from the bulk structure. Usually the lateral alignment of layer molecules with respect to the substrate orientation is poor, due to the weak interaction and a much larger, normally incommensurate unit cell. Independent locations of nucleation with different orientations lead to the creation of extended defects during coalescence. Typically organic thin films exhibit a texture structure with a preferred crystallographic plane parallel to the interface.

Pentacene on silicon dioxide is a particularly interesting system for organic thin-film transistors; since the charge transport in such devices predominantly occurs in the first few monolayers above the dielectric (Lang et al. 2004, Park et al. 2007), carrier mobility largely depends on the molecular packing in the first monolayer which determines the overlap of orbitals between neighboring molecules. The two

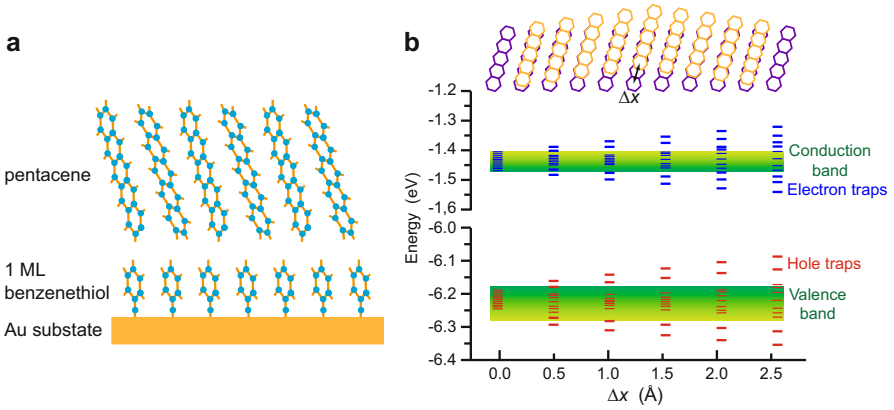


Fig. 8 (**a**) Orientation of pentacene molecules deposited at 295 K on a Au(111) surface passivated with a monolayer of benzenethiol. (**b**) Creation of trap states in the bandgap for holes and electrons due to the displacement of pentacene molecules by an amount Δx along the long molecular axis in a pentacene crystal (After Kang et al. 2005)

molecules in the thin-film unit cell of pentacene are vertically oriented on SiO_2 (also on other substrates), yielding a more rectangular cell ($\gamma = 89.8°$) than the bulk unit cell ($\gamma = 86°$, see Table 3 in ▶ Chap. 3, "The Structure of Semiconductors"), and molecules of the first monolayer adsorb exactly vertically on SiO_2 ($\gamma = 90°$, Mannsfeld et al. 2009). The *a*-*b* layers exhibit a herringbone packing (viewed along the interface normal) like the bulk structure. The upright alignment on the substrate indicates a somewhat stronger molecular-substrate interaction than the intermolecular interaction in the layer and yields in the first monolayer a slightly compressed in-plane unit cell with a shorter *a*-axis.

The insertion of a thin organic interlayer between the substrate and the organic semiconductor similar to the passivation layer on a metal shown in Fig. 8a proved also beneficial for pentacene layers on SiO_2 substrates. Measurements of the density of states in 10 nm thick pentacene films grown on bare SiO_2 substrates and on SiO_2 treated with hexamethyldisilaxane (HMDS) yield a density of gap states reduced by 1 order of magnitude for the latter (Yogev et al. 2013). Even less gap states were achieved on a treated AlO_x substrate. Fig. 9 demonstrates that the Fermi level moves deep into the bandgap at a high density of gap states. The high density of deep states is detrimental for carrier transport and leads to a pinning of the Fermi level, which controls band alignment at interfaces and doping efficiency.

On an organic substrate due to the weak bonds generally no coherently strained interface is formed. General rules reviewed by Hooks et al. (2001) distinguish substrate-layer combinations which show epitaxy-like behavior and those which are linearly or axially commensurate; for the latter the lateral vectors a_S and b_S on the substrate side are related to the respective vectors a_L and b_L of the layer by a coefficient matrix given by

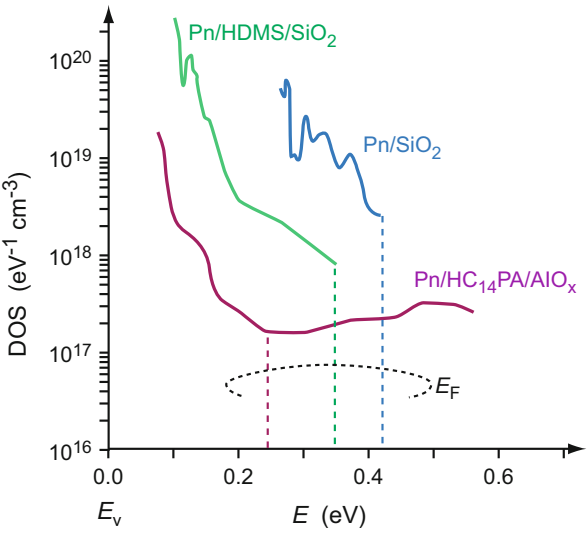


Fig. 9 Density of states (DOS) in pentacene (Pn) thin films grown on various substrates, measured using Kelvin-probe force microscopy. *Dashed lines* signify the Fermi energy in the pentacene layers under equilibrium conditions (After Yogev et al. 2013)

$$\begin{pmatrix} a_L \\ b_L \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} a_S \\ b_S \end{pmatrix}. \tag{18}$$

The organic substrate and layer crystals are commensurate if all coefficients m_{ij} are integers. If the coefficients are rational numbers, the layer may adopt a coincidence structure. For irrational coefficients the layer is incommensurate to the substrate. It may still have a preferred normal orientation but looses the epitaxial relationship to the substrate and will show a textured structure. A study of a large number of interfaces showed that the favored alignment of the layer can be estimated from geometric considerations [Hillier and Ward (1996), Forrest (1997)].

Organic molecules may also be inserted *as interlayer* between an inorganic substrate and the organic semiconductor to control the initially formed structure, as shown above for growth on metals and insulators; this also applies for substrates of inorganic semiconductors. Pentacene molecules form strong multiple covalent bonds on Si(001) (Meyer zu Heringdorf et al. 2001) or Si(111) surfaces (Al-Mahboob et al. 2008). The pentacene growth hence commences with a disordered layer of strongly bonded molecules which form coalescing crystalline islands in a textured organic solid. Passivating the Si surface prior to pentacen growth by an exposure to organic molecules such as styrene, cyclopentane, or hexane allows to yield monolayer-thick pentacene films with low defect densities (Hamers 2008; Seo and Evans 2009).

Graphite is another interesting substate which avoids strong molecule bonding. Its lattice on the basal plane is furthermore nearly identical to the carbon frame of polycyclic aromatic hydrocarbons like pentacene, favoring the formation of highly ordered monolayers. Subsequent molecular layers are gradually vertically tilted with increasing coverage. In the first pentacene monolayer the molecules are adsorbed in closely spaced rows with their plane parallel to the surface; the long axis is oriented along one of the azimuth $\langle \bar{1}2\bar{1}0 \rangle$ directions, and neighboring molecules are slightly shifted along the long axis yielding an oblique surface unit-cell and a coincidence $\begin{pmatrix} 7 & 0 \\ -1 & 3 \end{pmatrix}$ superstructure (Götzen et al. 2010). A grain boundary between single-crystalline domains in a monolayer-thick pentacene layer is shown in Fig. 10. Pentacen molecules are quite mobile at room temperature due to the weak adsorption to the substrate: the ovals in panels (a) and (b) mark changes occurring by detachment from and attachment to existing islands on a time scale of minutes. Still the boundary persists and cannot be closed even in thicker layers and after thermal treatment. Fig. 10c shows an atomic force micrograph of a thicker layer. The organic film consists of flat islands separated by narrow, deep crevices.

Grain boundaries like those visible in Fig. 10 strongly deteriorate the conductivity of the organic layers. The effect of such boundaries on electronic properties was studied in a self-consistent polarization-field approach (Verlaak and Heremans 2007); the optimum arrangement of the molecules at the boundary was calculated using molecular-dynamic forcefields. Results shown in Fig. 11 display the energy of a free hole at a considered boundary-near molecule with respect to its energy in a perfect crystal; at increased energy the molecule acts as scattering center, at decreased energy as a trap. The calculation gives evidence that both traps and

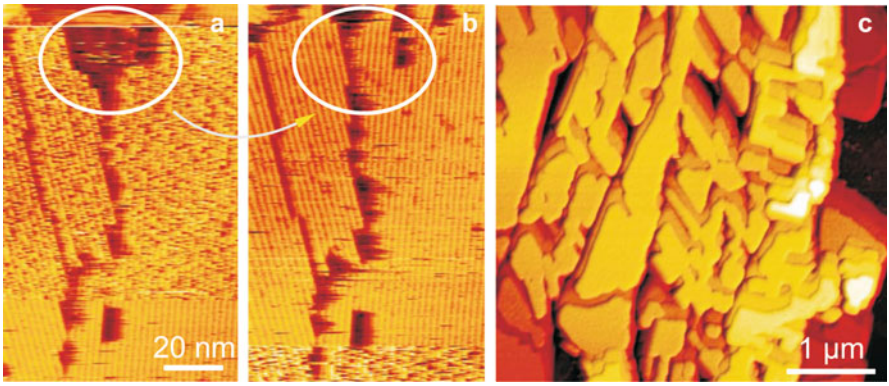


Fig. 10 (**a**) Scanning-tunneling micrographs of a monolayer-thick pentacene layer adsorbed on (0001) graphite; (**b**) same as (**a**) recorded some minutes later. Changes to a *grainy* part of the image indicate sudden tip switches due to molecule pick-up and release at the scanning tip. (**c**) Morphology of a layer with 35 nm nominal thickness (After Götzen et al. 2010)

scattering centers for free holes appear at the boundary. The hole energy is determined by the particular environment of a molecule near the boundary; the boundary has no periodicity on a short range, leading to a substantial variation in hole energies along the interface. At larger angles the energies shift gradually from trap states at the left grain and scattering centers at the right grain to the inverse case with trap states at the right grain (not shown in the figure); this is traced back to the interaction between the strongly polarizing hole and the quadrupole moments of neighboring molecules[5] as pointed out by Verlaak and Heremans (2007). The study indicates that the potential landscape at a grain boundary in organic crystals is complicated and has a granular structure on the molecular level; charge transport across such extended defects is not described by a simple model.

3.2.2 Point Defects in Molecule Crystals

The defect types classified for inorganic crystals in ▶ Chap. 15, "Crystal Defects", also occur in organic crystals: planar defects like stacking faults and the grain boundaries pointed out above, line defects like screw and edge dislocations, and point defects [Seo and Evans (2009), Maeda et al. (2001)]. The point defects comprise intrinsic defects like vacancies and impurities. Most organic crystals are not grown under ultraclean conditions like those applied for inorganic semiconductors. Oxygen, hydrogen, and water are prominent contaminants known to affect also organic crystals. We again focus on pentacene as a model system for the effect of such impurities.

[5]Polycyclic aromatic hydrocarbons like pentacene have a positively charged planar backbone of atom cores and negatively charged π electrons in front and at the back of this plane, yielding a permanent quadrupole moment.

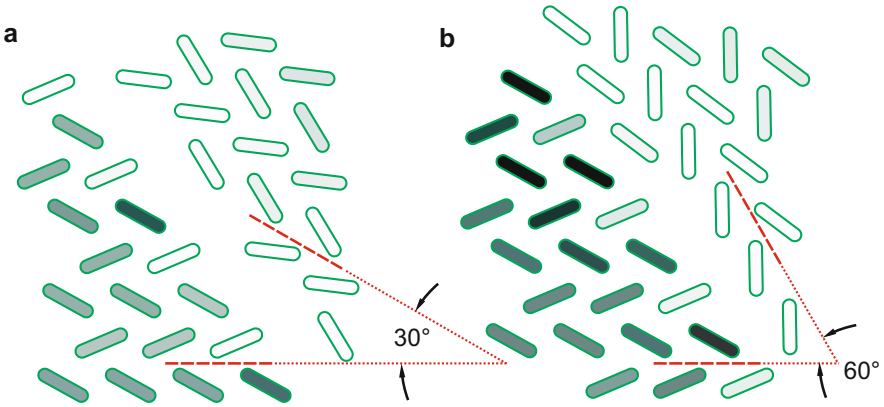


Fig. 11 Effect of boundaries between grains in a pentacene crystal misaligned by (**a**) 30° and (**b**) 60°. The molecules are viewed along their long axis; their plane is perpendicular to the image plane. The *shading* marks the energy of a hole at a molecule: states with energies below −0.25 eV (*dark shading*) are traps, states with energies above 0.25 eV (*bright shading*) act as scattering centers (After Verlaak and Heremans 2007)

Oxygen is easily introduced into a pentacene crystal due to a calculated energy drop of 0.13eV per O_2 molecule (Tsetseris and Pantelides 2007). Studies on pentacene thin-film transistors clearly prove that oxygen only creates defects states if the Fermi energy is high in the bandgap (Knipp and Northrup 2009). Total-energy calculations within the local DFT scheme yield two prominent oxygen-related defects forming trap states in the gap (Fig. 12). Breakup of the O_2 molecule leads to the energetic favorable complex Pn-2O illustrated in Fig. 12b, where the two dangling O atoms are bond to C atoms at positions 6 and 13 (depicted in panel *a*) of the same molecule (pentacenequinone). Oxygen removes a p_z orbital of the respective C atom from the planar π system of the molecule by forming a double bond and creates a localized gap state (Northrup and Chabinyc 2003); filled with two electrons, the energy of the Pn-2O state is computed to lie 0.36 eV above the valence-band edge [transition level (−/2–)]. The creation of the twofold charged defect yields a formation energy of $2(E_g - 0.36\ \text{eV})$ with a Pn bandgap energy of 2.0 eV, greatly favoring defect formation at high Fermi level. A shallow (0/–) acceptor energy of 0.08 eV was determined for this complex as a possible source of *p*-type doping. The dangling oxygen at position C6 may build an intermolecular bridge to a C13 atom of a neighboring Pn molecule (Fig. 12c), thereby forming the Pn-$2O_{II}$ complex, gaining 0.61 eV formation energy, and creating a deeper acceptor level at $E(0/–) = 0.29$ eV above the valence-band edge. Such findings agree with the broad distribution (0.16 eV FWHM) of trap states peaking 0.28 eV above valence-band maximum (mobility edge) measured with pentacene thin-film transistors which were exposed to photo-oxidation (Kalb et al. 2008). Since the π electron system of pentacenequinone is smaller than that of Pn, it has a larger bandgap energy; the molecule is hence also believed to act as a scattering center.

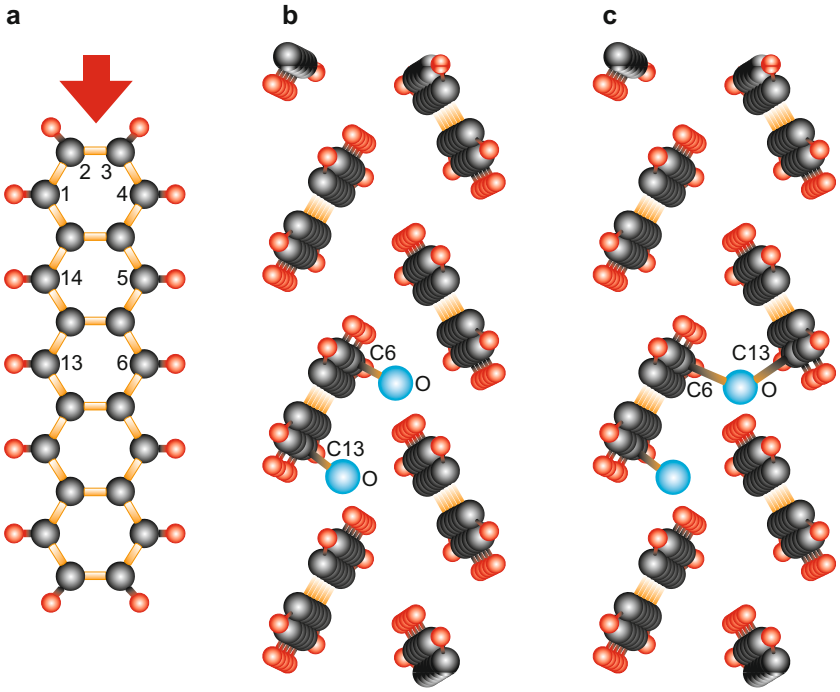


Fig. 12 (**a**) Numbering of the 14 outer C atoms in a pentacene (Pn) molecule; the *arrow* on top indicates the viewing directions in (**b**) and (**c**). (**b**) The Pn-2O complex with two dangling O atoms bond to a single Pn molecule. (**c**) The Pn-$2O_{II}$ complex with an intermolecular O bridge and a dangling O atom (After Knipp and Northrup 2009)

Hydrogen added to a pentacene molecule acts quite similar to an added oxygen atom (Northrup and Chabinyc 2003). Adding H_2 to a pentacene molecule is exothermic; again the center ring of pentacene is most reactive, and bonding of two additional hydrogen atoms at the 6 and 13 positions (Fig. 12a) yielding dihydropentacene is favorable (Mattheus et al. 2002). The calculated (+/0) level for *one* additional H atom at C6 position lies ~0.34 eV above the valence-band maximum and the (0/–) level occurs at ~0.80 eV, yielding a 0.5-eV-wide range of the Fermi energy for the stable neutral defect; adding *two* H atoms at the 6 and 13 positions yields a decreased valence-band edge and an increased conduction-band edge at the dihydropentacene molecule, so that no gap states are expected in this case (Northrup and Chabinyc 2003). Such a molecule will therefore also act as a scattering center.

Water is also easily introduced into a pentacene crystal; an energy gain of 0.55 eV was calculated for the incorporation of a H_2O molecule at a favorable position

between pentacene layers (Tsetseris and Pantelides 2007); the study indicates that the molecule does not dissociate. Furthermore, neither the structure of the crystal nor the electronic density of states shows significant changes for intercalated water. Scattering of carriers and hence reduction of carrier mobility are the main effects expected from the incorporation of water into pentacene.

To summarize, impurities give rise to traps in the bandgap with a broad distribution of states, similar to the effect of structural disorder. The density of trap states was experimentally derived from various measurements; results from the evaluation of the space-charge limited current in transistor structures are shown in Fig. 13. We observe a strong tailing from the valence-band edge (mobility edge at $E = 0$) into the bandgap. Particularly low trap densities have pentacene and rubrene single crystals (*sc bulk*) and field-effect transitors made with such crystals (*sc FET*); the increased trap density in FET structures is attributed to defects at the interface to the gate dielectric. The low trap density corresponds to the large carrier mobility found in such semiconductors. Polycrystalline thin films have significantly higher trap densities due to the high density of defects discussed in Sect. 3.3. We note that the defect density in organic small-molecule transistors is comparable to that in hydrogenated amorphous silicon.

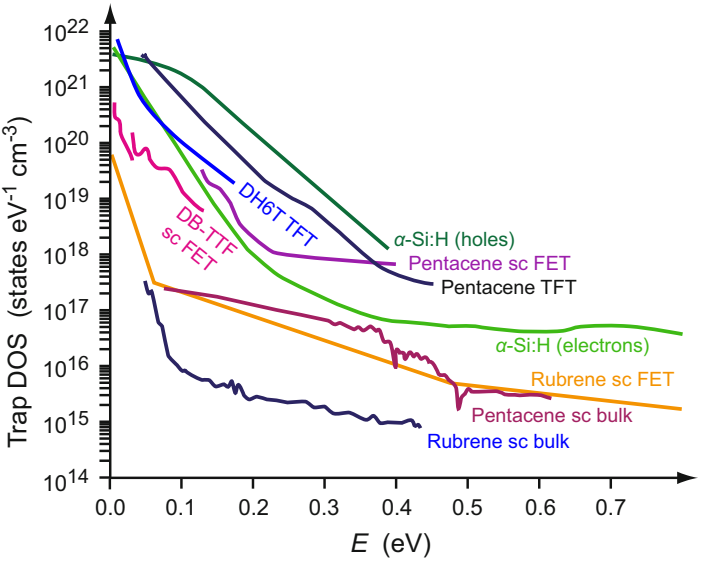


Fig. 13 Typical trap densities in small-molecule semiconductors (*black*, *red*, and *blue curves*) and hydrogenated amorphous silicon (*green curves*). *Blue curves* refer to bulk single-crystals, *red curves* to field-effect transistors made with single crystals; *black curves* give results of polycrystalline thin-film transistors. Energy values are relative to the valence-band edge E_v (for electrons in α-Si:H relative to the conduction-band edge E_c) (After Kalb et al. 2010)

3.3 Defects in Polymers

Polymers are organic compounds consisting of long, chain-like molecules with typically 10^2 to 10^4-fold repeated molecular units. In semiconducting polymers with π bonding in a conjugated chain the bonds have alternating lengths along the backbone. In the double (or triple) bonds π electrons of the p_z orbitals are localized and form an occupied valence band (or HOMO, highest occupied molecular orbital), separated by a bandgap from the conduction band (LUMO, lowest unoccupied molecular orbital).

Due to alternating bond lengths an irregularity may occur termed *bond-alternation defect* as illustrated for polyacetylene in Fig. 14. The terminal repeat units must have single bonds; thus if a finite polymer chain does not have the correct length, a mismatch in bond-alternation occurs, yielding the backbone structure shown in Fig. 14c. The C atom at this bond-alternation defect has two single C–C bonds and one unpaired π electron. The polymer chain remains electrically neutral, but the electron at the defect has an unpaired spin, giving rise to a finite paramagnetism of the otherwise diamagnetic chain. Due to similarities in the theoretical description the unpaired electron, which may move along the chain, is also referred to as *soliton*. The soliton represents a boundary which separates domains in the phase of the π bonds in the polymer backbone.

In a solid, the long polymer molecules are generally packed together non-uniformly, building both crystalline and highly disordered amorphous domains. The amorphous regions are composed of coiled and tangled chains, whereas in crystalline (albeit still distorted) regions linear polymer chains are oriented in a three-dimensional matrix.

Polythiophene and its derivatives are interesting polymers due to their very high carrier mobility (up to ~10 cm^2/Vs at the mobility edge). Fig. 15 shows a schematic of the polymer structure for polythiophene; in crystalline regions the molecules are

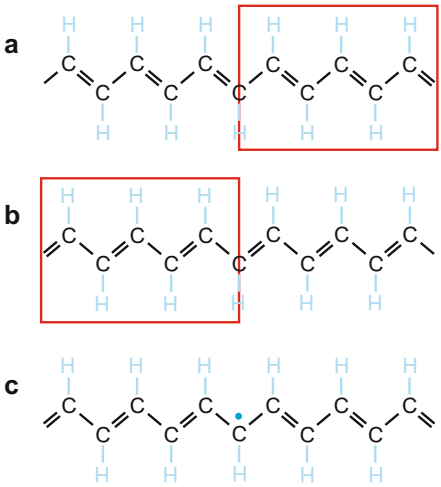


Fig. 14 Different bond-alternation order in sequences (**a**) and (**b**) of a polyacetylene chain; (**c**) bond-alternation defect occurring when sequences of (**a**) and (**b**) marked by *red boxes* meet. The *dot* on top of the central C atom represents an unpaired electron created at the defect

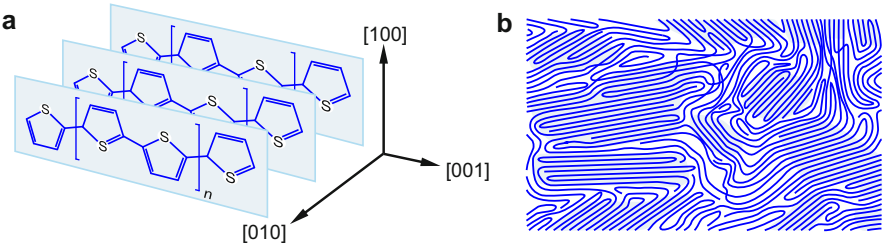


Fig. 15 (**a**) Packing motiv of a polymer semiconductor in crystalline regions. The index at the bracket indicates the *n*-fold repetition of the unit along the polymer chain. (**b**) Tangled polymer chains in the bulk with crystalline (*shaded*) and amorphous regions

arranged in long sheets, which are oriented along the conjugated backbone direction labeled [001] and packed along the $\pi - \pi$ stacking direction [010]. The sheets are additionally stacked along the [100] lamella-stacking direction. Each crystallite represents a grain in the polymer material. Grain boundaries are anisotropic regions of strong disorder with bend chains; low-angle grain boundaries are somewhat less distorted than high-angle grain boundaries.

Trap states appear in polymers predominantly in the most disordered regions; in most cases the structural nature of such states is unknown. Still the distortions in the crystalline regions are important, since these crystallites determine the carrier transport in a high-carrier-mobility polymer, where percolation effects allow for a dominating conduction in the ordered regions. Polymer segments extending from one ordered domain into another can enhance the mobility significantly if the $\pi-\pi$ coupling is sufficiently large (Northrup 2015). If no perculation by crystalline regions occurs, conduction is primarily through amorphous regions with very low mobility (Street et al. 2005).

Significant distortions also occur in the crystalline regions of polymers; these regions are hence referred to as *paracrystalline*, denoting an intermediate state between crystalline and amorphous (Hindeleh and Hosemann 1991). Such states can be characterized by a crystallographic *paracrystalline distortion parameter g*, defined by the average separation $\langle d_{hkl} \rangle$ of adjacent repetition units along a specific [*hkl*] direction (considered ideal) and the relative statistical deviations of actual separations:

$$g^2 = \left(\langle d_{hkl}^2 \rangle - \langle d_{hkl} \rangle^2 \right) / \langle d_{hkl} \rangle^2 = \langle d_{hkl}^2 \rangle / \langle d_{hkl} \rangle^2 - 1. \tag{19}$$

If several orders of a reflection peak are measured, the paracrystallinity parameter *g* can be calculated from the slope *s* of the plot $\delta b(l^2)$, where δb is the widths of the diffraction peaks calculated from $\delta b = 1/L_{\mathrm{coh}}$ (L_{coh} is the coherence length or paracrystal size) and *l* is the order of the diffraction (Hindeleh and Hosemann 1988). This analysis assumes peak broadening due to paracrystalline disorder proportional to the square of order and hence Lorentzian-shaped peaks, while broadening due to

lattice-parameter fluctuations leads to Gaussian-shaped peaks with linear l dependence. The parameter g is then given by

$$g^2 = s\, d_{hkl}/\pi^2. \tag{20}$$

Often only a *single* diffraction peak is used to estimate the parameter g. If the shape of the diffraction peak is described by a Voigt profile with a mixing parameter between 0 (Gauss) and 1 (Lorentz), peaks with Lorentz character express an only small contribution from lattice-parameter fluctuations. Such peaks yield a good estimate with

$$g \cong \frac{1}{2\pi}\sqrt{\delta b\, d_{hkl}}. \tag{21}$$

The assumption for the validity of Eq. 21 is fulfilled for many polymers; for a discussion, see Peng et al. (2022).[6]

The paracrystallinity parameter g allows to classify materials with different degrees of distortions as listed in Table 1 (Hosemann and Hindeleh 1995); values below 1% are indicative for crystalline properties, the range between 1% and 10% characterizes paracrystalline materials, and values of 10–15% represent a glass or a melt.

Distortions in the (para-) crystalline regions of a polymer consist, e.g., of random variations in the spacings between the backbone lamella illustrated in Fig. 15a. Such variations create band tailing into the bandgap comparable to the distortions shown in Fig. 8 for molecule shifts in a pentacene crystal. The magnitude of such variations was measured for aligned ribbons of fused-ring polythiophene[7] by analyzing data of

Table 1 Values of the paracrystalline distortion parameter g in various microparacrystals and other substances

Substance	g (%)
Single crystal	0
Crystallite in polymer	2
Bulk polymer	3
Graphite, coal tar	6
SiO_2 glass	12
Molten metals	15
Boltzmann gas	100

[6]More than a single g parameter, referring to different *hkl* directions and related fluctuations, may be required to characterize a sample. Also a description in terms of a g parameter accounting for the lattice disorder within paracrystals and another parameter for the degree of paracrystallinity, which measures the mass/volume fraction of ordered material as discussed by Marina et al. (2022), may be convenient.

[7]The repetition unit of fused-ring polythiophene comprises four rings like those shown in Fig. 15a with the inner two rings fused by commonly sharing a two carbon atoms with a C = C double bond (see inset Fig. 16).

grazing incidence x-ray diffraction; the resulting effect on the density of states was calculated using density-functional theory and is shown in Fig. 16 (Rivnay et al. 2011). In the calculations the distance between the chains (ideally a) was varied according to a Gaussion distribution with a standard deviation σ, yielding a paracrystallinity $g = \sigma/a$. The model predicts an exponential dependence of the distribution of trap-tailing states within the bandgap, similar to results of disordered inorganic solids discussed with the Kane model in Sect. 3 and Eq. 83 of ▶ Chap. 17, "Optical Properties of Defects". The characteristic energy E_0 in the DOS function $g(E) = g_0 \exp(-E/E_0)$ has values near 10 meV for well-ordered and 100 meV for strongly disordered polymers.

Properties of polymers related to the macromolecular order (such as crystallinity) and consequential characteristics depend also on the length of the molecular chains, in addition to the chemical structure and the applied processing. With increasing length of the backbone chain (and hence molecular weight) most polymers transform from a highly ordered chain alignment to a structure composed of crystalline and amorphous regions. All structures have lamellae like those shown in Fig. 15a as building blocks, but their size, perfection, and molecular interconnectivity differ. A qualitative picture for the development of various material properties with increasing molecular length is given in Fig. 17 (Virkar et al. 2010). The material properties addressed in the blue curve refer to crystallinity, density, melting temperature, or Young's modulus, which is related to the stiffness of the solid. Values of these quantities increase as the molecular chains gets longer and get maximum at a critical molecular weight (i.e., length) when chains start to entangle. At this length a two-phase morphology with crystalline and amorphous regions appears, and often the considered values eventually decrease. For a respective study on charge transport see Koch et al. (2013). Other mechanical quantities such as tensile strength and

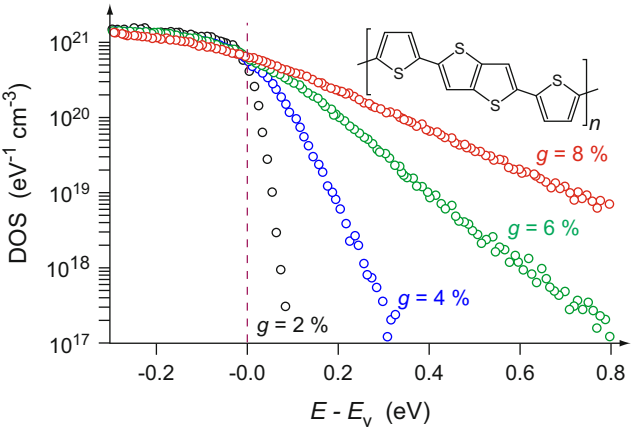


Fig. 16 Calculated density of states (DOS) of fused-ring polythiophene for different degrees of paracrystalline disorder expressed by the paracrystallinity parameter g. The valence-band maximum E_v refers to perfectly crystalline region (After Rivnay et al. 2011)

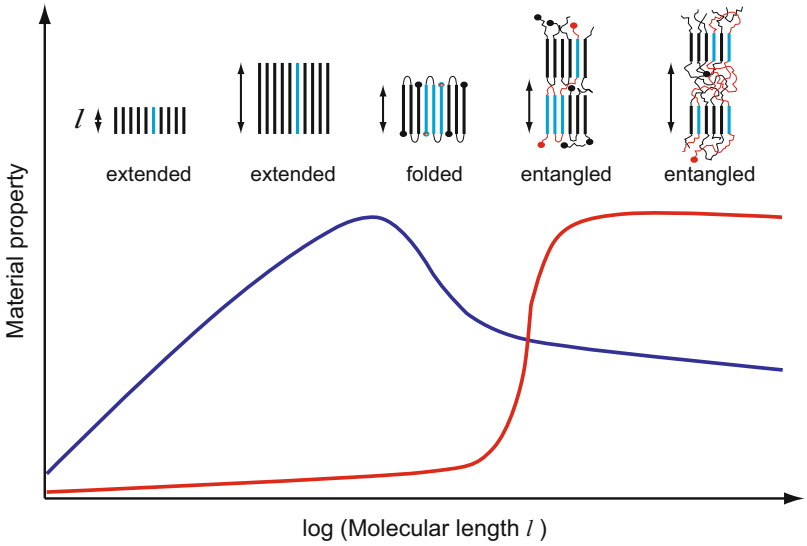


Fig. 17 Qualitative dependence of properties like crystallinity or melting temperature (*blue curve*) and mechanical strength (*red curve*) from the length of the molecular chains in a polymer. The *inset* on top illustrates the respective macromolecular structure (After Virkar et al. 2010)

elongation at break increase when the chains entangle (red curve). For a review on the connection between polymer conformation and materials properties see Virkar et al. (2010).

4 Summary

The most prevalent feature of all amorphous and organic semiconductors is the broad distribution of defect levels and consequently a pronounced tail of states extending into the bandgap. There is no sharp band edge as in crystalline solids; instead, another edge – the mobility edge – can be defined between extended and localized states, which, for typical semiconducting glasses with $a_{qH} = 30$ Å, lies at a level density of $\sim 10^{18}$ cm^{-3}.

For *amorphous semiconductors* (semiconducting glasses) two main classes can be distinguished: the amorphous chalcogenides, which are often alloys of several elements, and the tetrahedrally bonded amorphous semiconductors, of which α-Si:H is a typical representative. Intrinsic defects in both classes are more gradual than in crystals, where vacancies and interstitials can be identified unambiguously. The *intrinsic* defects in semiconducting glasses may be associated with deviations from an ideal amorphous structure and can be compared with frozen-in local stresses, causing more or less local deformation. *Extrinsic* defects (impurities) in *chalcogenide glasses* are often absorbed and become part of the glass matrix without producing

well-defined defect levels in the bandgap. Even at larger densities they are not able to shift the Fermi-level from its pinned position close to the center of the bandgap. In *tetrahedrally bonded amorphous semiconductors*, however, many impurities act as donors or acceptors, as in crystalline semiconductors, and can easily render these materials *n*- or *p*-type. Although gradual in nature, the defect-induced tails in the bandgap have a dominating influence on electrical and optical properties in most semiconducting glasses. Saturating dangling bonds, with H or F in α-Si:H or α -Si:F, clean out most of the bandgap and permit well-defined doping, resulting in devices of high technical interest.

The main classes of *organic semiconductors* are small-molecule crystals and polymers. For *small-molecule semiconductors* thin films grown on substrates are technologically most interesting. The structure of their polycrystalline grains represent usually modifications of the bulk structure, with a substantially different molecule ordering at the interface. The strong interaction to metal substrates mostly results in a laying-down configuration of molecules, while weaker interactions to insulating or molecular substrates favor a bulk-like standing-up phase. Grain boundaries are a major source of traps and scattering centers and strongly degrade electronic properties. The long backbone chains *of semiconducting polymers* are ordered in paracrystalline grains, which are separated by amorphous regions with tangled polymer chains. Band tailing increases as the distortion parameter *g*, which has values near 3 in bulk polymers, increases. Starting with the macromolecular order for short molecular chains, crystallinity increases for longer chains up to an optimum length, beyond which amorphous regions with entangled chain segments degrades crystallinity and consequential characteristics. Advances in processing techniques enabled specific polymers to catch up in carrier-transport properties with good data of small-molecule crystals, paving the way for interesting low-cost applications.

References

Abtew TA, Drabold DA (2006) Atomistic simulation of light-induced changes in hydrogenated amorphous silicon. J Phys Condens Matter 18:L1

Adler D (1985) Chemistry and physics of covalent amorphous semiconductors. In: Adler D, Schwartz BB, Steele MC (eds) Physical properties of amorphous materials. Plenum Press, New York, pp 1–103

Adler D, Fritzsche H (eds) (1985) Tetrahedrally bonded amorphous semiconductors. Springer, New York

Adler D, Yoffa EJ (1977) Localized electronic states in amorphous semiconductors. Can J Chem 55:1920

Al-Mahboob A, Sadowski JT, Fujikawa Y, Nakajima K, Sakurai T (2008) Kinetics-driven anisotropic growth of pentacene thin films. Phys Rev B 77:035426

Anderson PW (1958) Absence of diffusion in certain random lattices. Phys Rev 109:1492

Anderson PW (1975) Model for the electronic structure of amorphous semiconductors. Phys Rev Lett 34:953

Chen S, Li Z, Qiao Y, Song Y (2021) Solution-processed organic semiconductor crystals for field-effect transistors: from crystallization mechanism towards morphology control. J Mater Chem C 9:1126

Cohen MH, Fritzsche H, Ovshinsky SR (1969) Simple band model for amorphous semiconducting alloys. Phys Rev Lett 22:1065
Dinca LE, De Marchi F, MacLeod JM, Lipton-Duffin J, Gatti R, Ma D, Perepichkab DF, Rosei F (2015) Pentacene on Ni(111): room-temperature molecular packing and temperature-activated conversion to graphene. Nanoscale 7:3263
Djordjevic BR, Thorpe MF, Wooten F (1995) Computer model of tetrahedral amorphous diamond. Phys Rev B 52:5685
Drabold DA, Nakhmanson S, Zhang X (2001) Electronic structure of amorphous insulators and phoro-structural effects in chalcogenide glasses. In: Thorpe MF, Tichý L (eds) Properties and applications of amorphous materials. Kluwer, Dordrecht, pp 221–250
Evans PG, Spalenka JW (2015) Epitaxy of small organic molecules. In: Kuech TF (ed) Handbook of crystal growth of thin films and epitaxy: basic techniques, vol 3 part A, 2nd edn. Elsevier, Amsterdam, pp 509–554
Forrest SR (1997) Ultrathin organic films grown by organic molecular beam deposition and related techniques. Chem Rev 97:1793
Fugallo G, Mattoni A (2014) Thermally induced recrystallization of textured hydrogenated nanocrystalline silicon. Phys Rev B 89:045301
Götze W (1981) The conductor-nonconductor transition in strongly disordered three-dimensional systems. In: Devreese JT (ed) Recent development in condensed matter physics. Plenum Press, New York, pp 133–154
Götzen J, Käfer D, Wöll C, Witte G (2010) Growth and structure of pentacene films on graphite: weak adhesion as a key for epitaxial film growth. Phys Rev B 81:085440
Hamers RJ (2008) Formation and characterization of organic monolayers on semiconductor surfaces. Annu Rev Anal Chem 1:707
Hillier AC, Ward MD (1996) Epitaxial interactions between molecular overlayers and ordered substrates. Phys Rev B 54:14037
Hindeleh AM, Hosemann R (1988) Paracrystals representing the physical state of matter. J Phys C 21:4155
Hindeleh AM, Hosemann R (1991) Microparacrystals: the intermediate stage between crystalline and amorphous. J Mater Sci 26:5127
Hooks DE, Fritz T, Ward MD (2001) Epitaxy and molecular organization on solid substrates. Adv Mater 13:227
Hosemann R, Hindeleh AM (1995) Structure of crystalline and paracrystalline condensed matter. J Macromol Sci Phys B 34:327
Hosono H (2006) Ionic amorphous oxide semiconductors: material design, carrier transport, and device application. J Non-Cryst Sol 352:851
Huang B, Robertson J (2012) Nature of defects and gap states in GeTe model phase change materials. Phys Rev B 85:125305
Hudgens S (2012) Progress in understanding the Ovshinsky Effect: Threshold switching in chalcogenide amorphous semiconductors. Phys Stat Sol B 249:1951
Jarolimek K, de Groot RA, de Wijs GA, Zeman M (2009) First-principles study of hydrogenated amorphous silicon. Phys Rev B 79:155206
Jia S, Li H, Gotoh T, Longeaud C, Zhang B, Lyu J, Lv S, Zhu M, Song Z, Liu Q, Robertson J, Liu M (2020) Ultrahigh drive current and large selectivity in GeS selector. Nat Commun 11:4636
Käfer D, Ruppel L, Witte G (2007) Growth of pentacene on clean and modified gold surfaces. Phys Rev B 75:085309
Kalb WL, Mattenberger K, Batlogg B (2008) Oxygen-related traps in pentacene thin films: energetic position and implications for transistor performance. Phys Rev B 78:035334
Kalb WL, Haas S, Krellner C, Mathis T, Batlogg B (2010) Trap density of states in small-molecule organic semiconductors: a quantitative comparison of thin-film transistors with single crystals. Phys Rev B 81:155315
Kang JH, da Silva Filho D, Bredas J-L, Zhu X-Y (2005) Shallow trap states in pentacene thin films from molecular sliding. Appl Phys Lett 86:152115

Kastner M, Adler D, Fritzsche H (1976) Valence-alternation model for localized gap states in lone-pair semiconductors. Phys Rev Lett 37:1504
Knights JC, Biegelsen DK, Solomon I (1977) Optically induced electron spin resonance in doped amorphous silicon. Solid State Commun 22:133
Knipp D, Northrup JE (2009) Electric-field-induced gap states in pentacene. Adv Mater 21:2511
Koch FPV, Rivnay J, Foster S, Müller C, Downing JM, Buchaca-Domingo E, Westacott P, Yu L, Yuan M, Baklar M, Fei Z, Luscombe C, McLachlan MA, Heeney M, Rumbles G, Silva C, Salleo A, Nelson J, Smith P, Stingelin N (2013) The impact of molecular weight on microstructure and charge transport in semicrystalline polymer semiconductors–poly(3-hexylthiophene), a model study. Prog Polym Sci 38:1978
Lang DV, Chi X, Siegrist T, Sergent AM, Ramirez AP (2004) Amorphouslike density of gap states in single-crystal pentacene. Phys Rev Lett 93:086802
LeComber PG, Spear WE (1976) Electronic properties of doped amorphous Si and Ge. Am Inst Phys Conf Proc 31:284
Lee PA, Ramakrishna TV (1985) Disordered electronic systems. Rev Mod Phys 57:287
Lewis LJ, Mousseau N (1998) Tight-binding molecular-dynamics studies of defects and disorder in covalently bonded materials. Comput Mater Sci 12:210
Li JM, Pfeiffer G, Paesler MA, Sayers DE, Fontaine A (1989) Photon intensity-dependent darkening kinetics in optical and structural anisotropy in a-As_2S_3: a study of X-ray absorption spectroscopy. J Non Cryst Solids 114:52
Lietoila A, Wakita A, Sigmon TW, Gibbons JF (1982) Epitaxial regrowth of intrinsic, *p*-doped and compensated (P + B-doped) amorphous Si. J Appl Phys 53:4399
Lifshitz IM (1964) The energy spectrum of disordered systems. Adv Phys 13:483
Maeda T, Kobayashi T, Nemoto T, Isoda S (2001) Lattice defects in organic crystals revealed by direct molecular imaging. Philos Mag B 81:1659
Mannsfeld SCB, Virkar A, Reese C, Toney MF, Bao Z (2009) Precise structure of pentacene monolayers on amorphous silicon oxide and relation to charge transport. Adv Mater 21:2294
Marina S, Gutierrez-Fernandez E, Gutierrez J, Gobbi M, Ramos N, Solano E, Rech J, You W, Hueso L, Tercjak A, Ade H, Martin J (2022) Semi-paracrystallinity in semi-conducting polymers. Mater Horiz 9:1196
Mattheus CC, Baas J, Meetsma A, de Boer JL, Kloc C, Siegrist T, Palstra TTM (2002) A 2:1 cocrystal of 6,13-dihydropentacene and pentacene. Acta Crystallogr E 58:o1229
Meidanshahi RV, Bowden S, Goodnick SM (2019) Electronic structure and localized states in amorphous Si and hydrogenated amorphous Si. Phys Chem Chem Phys 21:13248
Meyer zu Heringdorf F-J, Reuter MC, Tromp RM (2001) Growth dynamics of pentacene thin films. Nature 412:517
Mosley LE, Paesler MA (1984) Electronic effect on crystallization growth velocities produced by charged dangling bonds in a-Si. Appl Phys Lett 45:86
Mott NF (1969) Charge transport in non-crystalline semiconductors. In: Madelung O (ed) Festkörperprobleme/Advances in solid state physics, vol 9. Vieweg, Braunschweig, pp 22–45
Mott NF, Davis EA (1979) Electronic processes in non-crystalline materials, 2nd edn. Oxford University Press, Oxford, UK
Mott NF, Massey HSW (1965) The theory of atomic collisions. Claredon Press, Oxford, UK
Mousseau N, Barkema GT (2000) Activated mechanisms in amorphous silicon: an activation-relaxation-technique study. Phys Rev B 61:1898
Northrup JE (2015) Mobility enhancement in polymer organic semiconductors arising from increased interconnectivity at the level of polymer segments. Appl Phys Lett 106:023303
Northrup JE, Chabinyc ML (2003) Gap states in organic semiconductors: hydrogen- and oxygen-induced states in pentacene. Phys Rev B 68:041202
Orapunt F, O'Leary SK (2008) Optical transitions and the mobility edge in amorphous semiconductors: A joint density of states analysis. J Appl Phys 104:073513

Ovshinsky SR (1968) Reversible electrical switching phenomena in disordered structures. Phys Rev Lett 21:1450
Ovshinsky SR (1976) Localized states in the gap of amorphous semiconductors. Phys Rev Lett 36:1469
Ovshinsky SR (1977) Chemical modification of amorphous chalcogenides. In: Proceedings of the 7th international conference on amorphous and liquid semiconductors, Edinburgh, pp 519–523
Ovshinsky SR (1980) The chemistry of glassy materials and their relevance to energy conversion. J Non Cryst Solids 42:335
Pantelides ST (1986) Defects in amorphous silicon: a new perspective. Phys Rev Lett 57:2979
Pantelides ST (1989) The nature of defects and defect dynamics in amorphous silicon. In: Fritzsche H (ed) Amorphous silicon and related materials. World Scientific Publishing, New York, pp 541–556
Park B-N, Seo S, Evans P (2007) Channel formation in single-monolayer pentacene thin film transistors. J Phys D 40:3506
Peng Z, Ye L, Ade H (2022) Understanding, quantifying, and controlling the molecular ordering of semiconducting polymers: from novices to experts and amorphous to perfect crystals. Mater Horiz 9:577
Pfanner G, Freysoldt C, Neugebauer J, Inam F, Drabold D, Jarolimek K, Zeman M (2013) Dangling-bond defect in a-Si:H: characterization of network and strain effects by first-principles calculation of the EPR parameters. Phys Rev B 87:125308
Prasai K, Biswas P, Drabold DA (2016) Electrons and phonons in amorphous semiconductors. Semicond Sci Technol 31:073002
Rivnay J, Noriega R, Northrup JE, Kline RJ, Toney MF, Salleo A (2011) Structural origin of gap states in semicrystalline polymers and the implications for charge transport. Phys Rev B 83:121306
Robertson J (2012) Properties and doping limits of amorphous oxide semiconductors. J Non-Cryst Sol 358:2437
Sanchez ML, Aguilar MA, de Valle FJO (1997) Study of solvent effects by means of averaged solvent electrostatic potentials obtained from molecular dynamics data. J Comput Chem 18:313
Seo S, Evans PG (2009) Molecular structure of extended defects in monolayer-scale pentacene thin films. J Appl Phys 106:103521
Simbrunner C, Sitter H (2015) Organic van der Waals epitaxy versus templated growth by organic-organic heteroepitaxy. In: Kuech TF (ed) Handbook of crystal growth of thin films and epitaxy: basic techniques, vol 3 part A, 2nd edn. Elsevier, Amsterdam, pp 483–508
Staebler DL, Wronski CR (1977) Reversible conductivity changes in discharge-produced amorphous Si. Appl Phys Lett 31:292
Street RA (1991) Hydrogenated amorphous silicon. Cambridge University Press, Cambridge, UK
Street RA, Mott NF (1975) States in the gap in glassy semiconductors. Phys Rev Lett 35:1293
Street RA, Northrup JE, Salleo A (2005) Transport in polycrystalline polymer thin-film transistors. Phys Rev B 71:165202
Stuke J (1976) In: Kolomiets BT (ed) Electronic phenomena in non-crystalline solids. USSR Academy of Sciences, Leningrad, pp 193–202
Tanaka K (1998) Medium-range structure in chalcogenide glasses. Jpn J Appl Phys 37:1747
Tanaka K, Nakayama S-i (1999) Band-tail characteristics in amorphous semiconductors studied by the constant-photocurrent method. Jpn J Appl Phys 38:3986
Thouless DJ (1974) Electrons in disordered systems and the theory of localization. Phys Rep 13:93
Tsetseris L, Pantelides ST (2007) Intercalation of oxygen and water molecules in pentacene crystals: first-principles calculations. Phys Rev B 75:153202
Varshishta P, Kalia RK, Nakano A, Li W, Ebbsjö I (1996) Molecular dynamics methods and large-scale simulations of amorphous materials. In: Thorpe MF, Mitkova MI (eds) Amorphous insulators and semiconductors. NATO ASI ser 3 high technology, vol 23. Kluwer Academic Publishers, Dordrecht, p 151

Varshneya AK, Seeram AN, Swiler DR (1993) A review of the average coordination number concept in multicomponent chalcogenide glass systems. Phys Chem Glasses 34:179
Verlaak S, Heremans P (2007) Molecular microelectrostatic view on electronic states near pentacene grain boundaries. Phys Rev B 75:115127
Virkar AA, Mannsfeld S, Bao Z, Stingelin N (2010) Organic semiconductor growth and morphology considerations for organic thin-film transistor. Adv Mater 22:3857
Wooten F, Weaire D (1989) Modelling tetrahedrally bonded random networks by computer. In: Ehrenreich H, Turnbull D (eds) Solid state physics, vol 40. Academic Press, New York, pp 1–42
Yogev S, Matsubara R, Nakamura M, Zschieschang U, Klauk H, Rosenwaks Y (2013) Fermi level pinning by gap states in organic semiconductors. Phys Rev Lett 110:036803
Wuttig M, Yamada N (2007) Phase-change materials for rewriteable data storage. Nat Mater 6:824
Zipoli F, Krebs D, Curioni A (2016) Structural origin of resistance drift in amorphous GeTe. Phys Rev B 93:115201

Part VI

Transport

Equilibrium Statistics of Carriers

21

Karl W. Böer and Udo W. Pohl

Contents

1 The Intrinsic Semiconductor 900
1.1 Level Distribution Near the Band Edge 901
1.2 Statistical Distribution Functions 903
1.3 Intrinsic Carrier Densities in Equilibrium 904
1.4 Density-of-State Effective Mass and Fermi Energy 906
1.5 Intrinsic Carrier Generation 910
2 The Extrinsic Semiconductor 910
2.1 The Position of the Fermi Level 912
2.2 Temperature Dependence of the Fermi Level 915
2.3 Carrier Density in Extrinsic Semiconductors 917
2.4 Intrinsic and Minority Carrier Densities 918
2.5 Self-Activated Carrier Generation 919
2.6 Frozen-In Carrier Densities 920
3 Phase Transitions at High Carrier Densities 922
3.1 The Mott Transition 922
3.2 Electron–Hole Condensation 923
4 Summary 928
References 929

Abstract

Electrons and holes are the carriers of currents in semiconductors. The density of these carriers in equilibrium is obtained from the Fermi–Dirac statistics. The Fermi energy E_F as a key parameter can be obtained from quasineutrality; it lies near the middle of the bandgap for intrinsic and near the donor or acceptor level

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_21

for doped semiconductors. The difference between the respective band edge and the Fermi level represents the activation energy of a Boltzmann factor, whose product with the joint density of energy states yields the carrier density. The density of minority carriers may be frozen-in in semiconductors with a large bandgap and represented by a quasi-Fermi energy.

At high quasi-particle densities and low temperature, phase transitions take place with substantial changes in the optical and electronic behavior. An insulator-metal transition occurs above a critical Mott density of dopants. A similar process is initiated by sufficient optical generation of electrons and holes, leading to an electron–hole plasma and – at suitable conditions – to a condensation into an electron–hole liquid.

Keywords

Boltzmann distribution · Bose–Einstein distribution · Carrier density · Degenerate semiconductor · Density of states · Density-of-state mass · Effective density of states · Electron–hole liquid · Electron–hole plasma · Exciton Mott transition · Extrinsic semiconductor · Fermi–Dirac distribution · Fermi energy · Intrinsic carrier density · Intrinsic semiconductor · Mott density · Mott transition

1 The Intrinsic Semiconductor

Electrons and holes are the carriers of a current in semiconductors. Their distribution as a function of temperature and doping is analyzed in this chapter by an equilibrium statistical approach. The influence of the host material on the electronic properties will be discussed first. A pure semiconductor with no dopants is called an *intrinsic semiconductor*; the influence from lattice defects on its carrier density is negligible. The electronic properties of such a semiconductor are determined by the *mutual* generation or recombination of electrons and holes.

Most of the carriers usually originate from *lattice defects*, which also determine the transport properties of these carriers (▶ Chap. 22, "Carrier-Transport Equations"). The dominating influence of lattice defects on the density of electrons and holes will be evaluated in Sect. 2.

Carrier Densities in Equilibrium If semiconductors are exposed only to thermal excitation and are kept long enough at a constant temperature, *thermodynamic* (short "thermal") *equilibrium* is established. For the following discussion, it is assumed that such an equilibrium is always established, no matter how long it takes to reach it – see Sect. 2.6. When equilibrium is established, the same temperature T, which determines the distribution function, characterizes all subsystems, e.g., electrons, holes, and phonons (i.e., lattice atoms). The density distribution of electrons in the conduction band in a small energy range dE can be obtained from statistical arguments as the product of the density of energy states $g(E)$ and the statistical distribution function describing the occupation of these states $f(E)$:

$$n(E)\,dE = g(E)f(E)\,dE. \tag{1}$$

The density distribution for holes, i.e., of missing electrons, in the valence band is similarly obtained from

$$p(E)\,dE = g(E)\,[1 - f(E)]\,dE. \tag{2}$$

Since the probability of finding an electron at energy E is proportional to $f(E) \leq 1$, with $f(E) = 1$ as the certainty, the probability of finding an electron *missing* is simply $1-f(E)$. The total density of electrons n or holes p is then obtained from integrating over all energies of the respective band. We hence need both the density of states, i.e., the distribution of energy levels within the band, and the statistical distribution function to yield the temperature-dependent carrier density in this band.

1.1 Level Distribution Near the Band Edge

In Sect. 4.2.2 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals", the density of states within a band was derived in general terms. For the description of carrier transport, an explicit expression of this distribution is required. Such an expression can be given easily near the *edge* of the conduction or valence band. In an isotropic parabolic band, the dispersion equation of Bloch electrons is given by

$$E(k) = E_c + \frac{\hbar^2}{2m_n}k^2 + \ldots \tag{3}$$

These electrons are free to move within the semiconductor but are confined to the outer bounds. This can be described as confinement within a "box", for definiteness a cube of dimensions l, and requires standing-wave boundary conditions with nodes at the box surface.[1] For the three components of an electron wavevector, we have

$$k_i = \nu_i \frac{\pi}{l_i} \text{ with } \quad \nu_i = 1, 2, 3, \quad \ldots \text{ and } i = (x, y, z). \tag{4}$$

The wavevector and therefore the energy are represented by a set of discrete values given by the triplet of integers ν_x, ν_y, and ν_z. In addition, m_s is used to identify the spin quantum state. The energy increases monotonically with these integers.[2]

At low temperatures, the energy states are filled sequentially; each state is quantified by ν_x, ν_y, and ν_z and can be occupied by a maximum of two electrons with opposite spin (*Pauli principle*). These states are sequentially filled up to a radius

[1] Or cyclic boundary conditions; here energy and particle number are conserved by demanding, that with the passage of a particle out of a surface, an identical one enters from the opposite surface (Born-von Karman boundary condition). The two conditions are mathematically equivalent.

[2] The values of ν_i are positive integers. Negative values do not yield linear independent waves and $\nu_i = 0$ yields waves which cannot be normalized.

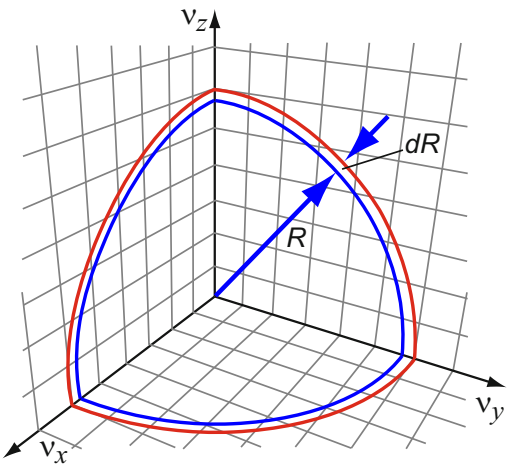


Fig. 1 Octant of spheres with constant energy in quantum-number space spanned by ν_x, ν_y, ν_z

$R = R(\nu_x, \nu_y, \nu_z)$, which is determined by the number of available electrons. The number of states G within an energy range from E to $E + dE$ is obtained from the volume of the spherical shell of the octant depicted in Fig. 1, permitting only positive values of ν_i:

$$G_{\nu}^{(dR)} = \frac{1}{8}\frac{4\pi}{3}\left(R_2^3 - R_1^3\right) = \frac{\pi}{6}d^3R = \frac{\pi}{2}R^2dR. \tag{5}$$

Since $R^2 = \nu_i^2$ in quantum-number space and using Eq. 4, we can replace R with k and obtain in the momentum space the number of states in the momentum interval dk:

$$G_k^{(dk)} = \frac{\pi}{2}\frac{l^3}{\pi^3}k^2dk \tag{6}$$

within the volume of the semiconductor $l^3 = V$. Finally, replacing k with the energy from Eq. 3, we obtain

$$G_E^{(dE)} = \frac{V}{2\pi^2}k^2dk = \frac{V}{2\pi^2}\left(\frac{2m_n}{\hbar^2}\right)^{3/2}\frac{1}{2}\sqrt{E - E_c}dE \tag{7}$$

as the number of states between E and $E + dE$. Dividing by the crystal volume and permitting double occupancy due to spin up and down for each state, we obtain the density of states for electrons in the conduction band with $E \geq E_c$:

$$g_n(E)\,dE = \frac{1}{2\pi^2}\left(\frac{2m_n}{\hbar^2}\right)^{3/2}\sqrt{E - E_c}\;\;dE. \tag{8}$$

This density is zero at the lower edge of the conduction band and increases proportionally to the square root of the energy (measured from E_c) near the band edge (Fig. 2). A similar square-root dependence results for the density of states near the upper edge of the valence band (energy E_v), where a quasi-free hole picture

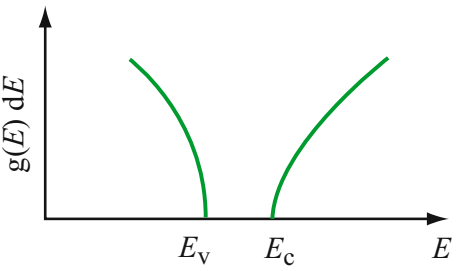


Fig. 2 Square-root dependence of the density of states near the band edge, where the quasi-free electron or hole model can be applied

can be applied.[3] Replacing m_n with m_p in Eq. 8, and shifting the energy axis by $E_g = E_c - E_v$, yields for holes with $E \leq E_v$:

$$g_p(E)\,dE = \frac{1}{2\pi^2}\left(\frac{2m_p}{\hbar^2}\right)^{3/2}\sqrt{E_v - E}\;\;dE. \tag{9}$$

The result of this simple model is carried through most of the common discussions of carrier transport in solids. It is valid near both edges of the band and is an acceptable approximation as long as the bands are parabolic in $E(k)$. At higher energies the nonparabolicity affects the effective mass; for respective modification, see Sect. 1.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids".

1.2 Statistical Distribution Functions

The distribution function which determines the occupation of the energy levels depends on the spin of the considered subsystem. Electrons are *fermions* with spin ± ½ and follow the *Fermi–Dirac distribution* function

$$f_{FD}(E) = \frac{1}{\exp\left(\dfrac{E - E_F}{kT}\right) + 1} \tag{10}$$

which is shown in Fig. 3 with the temperature as family parameter. This distribution function has a boxlike shape for $T = 0$, yielding complete occupancy of states ($f_{FD} = 1$) for all levels $E < E_F$ and completely unoccupied states ($f_{FD} = 0$) for $E > E_F$. Here E_F is the *Fermi energy*. For $T > 0$, the degree of filling decreases exponentially with increasing E and reaches 50% at the Fermi energy: the corners of the box are "rounded off". E_F is defined by $f_{FD}(E_F) = 0.5$. For the evaluation of the Fermi–Dirac distribution at energy (several kT) above the Fermi energy, the distribution can be approximated by a *shifted Boltzmann distribution*

[3] In the hole picture, an increasing energy corresponds to a *decreasing electron energy*. The energy axis in Fig. 2 refers to an electron energy.

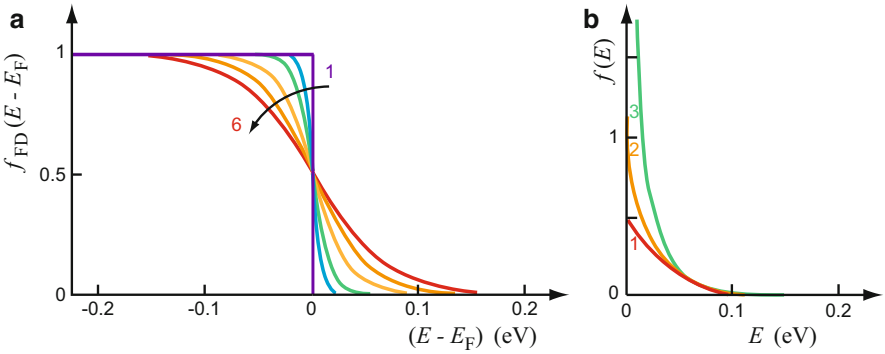


Fig. 3 (**a**) Fermi–Dirac distribution function with the temperature as family parameter: 0, 50, 100, 200, 300, and 400 K for curves 1–6, respectively. (**b**) Comparison of Fermi–Dirac, Boltzmann, and Bose–Einstein distributions for curves 1–3, respectively, at $T = 300$ K. f_{FD} is drawn at a scale shifted by E_F

$$f_B(E) = \exp\left(-\frac{E - E_F}{kT}\right). \tag{11}$$

The degree of deviation from the Fermi–Dirac distribution is indicated in Fig. 3b.

It is instructive to compare these distributions with the *Bose–Einstein distribution*

$$f_{BE}(E) = \frac{1}{\exp\left(\frac{E}{kT}\right) - 1}, \tag{12}$$

which describes the distribution of *bosons*, which have integer spin, such as phonons, photons, or excitons. The distribution has a singularity at $E = 0$; this means it increases dramatically ($f_{BE} \to \infty$) as $E \to 0$, while for a vanishing argument, i.e., for $E - E_F \to 0$, we have $f_B \to 1$ and $f_{FD} \to 0.5$. This is shown in Fig. 3b. The tail of all three distributions for $E/kT > 3$ is practically identical.

1.3 Intrinsic Carrier Densities in Equilibrium

The electron distribution $n(E)$ within a band is obtained as the product of the density of energy states and the Fermi distribution:

$$n(E)\,dE = g_n(E) f_{FD}(E)\,dE \tag{13}$$

and is shown schematically in Fig. 4 for the lower part of the conduction band. The electron distribution is zero within the bandgap since the level distribution vanishes in the gap for an ideal semiconductor.

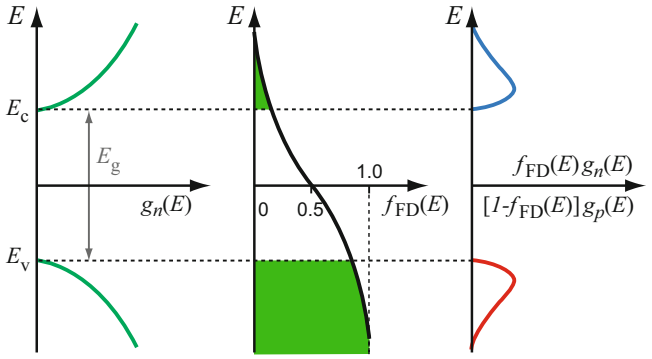


Fig. 4 Level density (*left, green curves*), Fermi–Dirac function (*middle*), and electron density distribution (*right, blue curve*) within the lower part of the conduction band and corresponding hole density distribution within the upper part of the valence band (*right, red curve*). *Green shading* in the middle diagram indicates the occupation with electrons

Because of the steep decrease of the Fermi–Dirac distribution function, most of the conduction electrons are located very close to the bottom of the conduction band: 95% of the electrons are within $3kT$ from E_c, except for high electric fields when substantial electron heating occurs. Since a typical band width is on the order of 10 eV, i.e., ~$400kT$ at room temperature, we can assign to the vast majority of the conduction electrons in a semiconductor the same effective mass.[4] Therefore, it is justified to replace the electron distribution function $n(E)$ with a simple *electron density* in the conduction band

$$n = \int_{E_c}^{\infty} g_n(E) f_{FD}(E)\, dE. \tag{14}$$

Because of the steep decrease of $f(E)$ with E, the integration to the upper edge of the conduction band is replaced with ∞, with negligible error in the result. When evaluating Eq. 14, we can rewrite this equation by defining an *effective level density* N_c at the lower edge of the conduction band, multiplied by a shifted Boltzmann distribution (Eq. 11) used instead of the Fermi–Dirac distribution, yielding

$$n = N_c f_B(s), \quad \text{where } s = -(E_c - E_F)/(kT). \tag{15}$$

The pre-exponential factor N_c for a simple parabolical band with spherical cross section near $k = 0$ and for $E_c - E_F > 3kT$ is then

[4]This is justified since m_n does not change much near the bottom of the conduction band (see Fig. 15 of ▶ Chap. 6, "The Origin of Band Structure"); moreover, m_n increases with increasing E, which usually renders higher energy electrons less important for a number of low-field transport properties.

$$N_c = 2\left(\frac{m_n kT}{2\pi\hbar^2}\right)^{3/2} \tag{16}$$

or

$$N_c = 2.5 \cdot 10^{19}\left(\frac{m_n}{m_0}\right)^{3/2}\left(\frac{T(\mathrm{K})}{300}\right)^{3/2} \quad (\mathrm{cm}^{-3}). \tag{17}$$

For a more exact solution (if $E_c - E_F < 3kT$) see Eq. 29 and Fig. 6 below.

The hole density can be obtained using the same arguments as for electrons, with

$$p = \int_{-\infty}^{E_v} g_p(E)\,(1 - f_{FD}(E))\,dE = N_v f_B(\tilde{s}), \tag{18}$$

where $\tilde{s} = -(E_F - E_v)/(kT)$, and $1-f_{FD}$ is the probability of finding a hole there:

$$\begin{aligned} f_{FD}^{(p)}(E) &= 1 - f_{FD}^{(n)}(E) \\ &= 1 - \frac{1}{\exp\left(\dfrac{E - E_F}{kT}\right) + 1} = \frac{1}{\exp\left(\dfrac{E_F - E}{kT}\right) + 1}. \end{aligned} \tag{19}$$

Here, N_v is the *effective level density* for holes at the upper edge of a single parabolic valence band of spherical symmetry:

$$N_v = 2\left(\frac{m_p kT}{2\pi\hbar^2}\right)^{3/2}. \tag{20}$$

The total electron density in the conduction band and the total hole density in the valence band are consequently given by

$$n = N_c \exp\left(-\frac{E_c - E_F}{kT}\right) \text{ and } p = N_v \exp\left(-\frac{E_F - E_v}{kT}\right). \tag{21}$$

1.4 Density-of-State Effective Mass and Fermi Energy

For real semiconductors, band dispersion must be taken into account, particularly for valence bands and for anisotropic semiconductors such as wurtzites. In the conduction band of anisotropic semiconductors, the mass has tensor properties. In general, the equi-energy surface can be described by an ellipsoid with three different axes, hence with three different curvatures and therefore three different effective masses in the direction of the main axes, resulting in a density-of-state mass

$$m_{nds} = \sqrt[3]{m_1 m_2 m_3}. \tag{22}$$

As discussed in Sect. 1.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids", the *satellite valleys* of cubic semiconductors are described by a rotational ellipsoid with $m_3 = m_{nl}$ (the longitudinal effective mass along the ⟨100⟩ direction in Si) and

$m_1 = m_2 = m_{nt}$ (the transverse effective mass, perpendicular to the main axis) with spherical symmetry – see Figs. 10 and 16b of ▶ Chap. 8, "Bands and Bandgaps in Solids". With a coordinate transformation in momentum space, we can reduce these ellipsoids to spheres and apply the commonly used calculation for the density of states. If there is more than one ellipsoid with identical energy (six in the example of conduction bands for Si), a respective *degeneracy factor* ν_d must be employed. This results in a *density-of-state mass for electrons* of

$$m_{n\mathrm{ds}} = \nu_\mathrm{d}^{2/3} \left(m_{n\mathrm{l}} m_{n\mathrm{t}}^2\right)^{1/3}. \tag{23}$$

Sometimes an *anisotropy factor* $K_\mathrm{a} = \mathrm{m}_{nl}/m_{nt}$ is used, rendering the density-of-state effective mass for electrons $m_{n\mathrm{ds}} = m_{n\mathrm{l}} K_\mathrm{a}^{-2/3} \nu_\mathrm{d}^{2/3}$. For warped (nonparabolic) conduction bands, we have

$$m_{n\mathrm{ds}\pm} = \frac{m_0}{A \pm B'} \left(1 + 0.0333\ \ \Gamma + 0.0106\ \ \Gamma^2 + \ldots\right); \tag{24}$$

for the parameters A, B', and Γ – see Sect. 4.5 of ▶ Chap. 22, "Carrier-Transport Equations" and Sect. 1.2.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids".

The distribution of holes between the *lh* (light hole) and *hh* (heavy hole) bands is proportional to their respective density of states:

$$N_{\mathrm{v}\,lh} = 2\left(\frac{m_{p\,lh}\,kT}{2\pi\hbar^2}\right)^{3/2} \text{and } N_{\mathrm{v}\,hh} = 2\left(\frac{m_{p\,hh}\,kT}{2\pi\hbar^2}\right)^{3/2}. \tag{25}$$

There are more states per energy interval in a heavy-hole band than in a light-hole band near the band edge because of the lower curvature of $E(k)$. Hence, when filling the band to a certain energy, more holes are in the heavy-hole band. Since both densities are additive, these equations may be used to introduce an *effective combined density-of-state mass for holes*

$$m_{p\,\mathrm{ds}}^{3/2} \cong m_{p\,lh}^{3/2} + m_{p\,hh}^{3/2}, \tag{26}$$

neglecting the contribution from the spin-orbit split-off band and any deviations from spherical $E(k)$ behavior in the valence band. If the spin-orbit split-off band is taken into account, we obtain for cubic semiconductors instead of Eq. 26

$$m_{p\,\mathrm{ds}} = \left[m_{p\,lh}^{3/2} + m_{p\,hh}^{3/2} + e^{-\Delta_0/(kT)} m_{p\,so}^{3/2}\right]^{2/3}, \tag{27}$$

where Δ_0 is the valence-band splitting energy listed in Table 5 of ▶ Chap. 8, "Bands and Bandgaps in Solids". The leading exponential in the bracket of Eq. 27 accounts for the (usually not pronounced) occupation of the split-off band with holes at temperature T. The effective combined hole mass of Eq. 27 corresponds to the density-of-states mass of an effective *single* equivalent parabolic valence band. In the density-of state

masses of silicon shown in Fig. 5b, the slight slopes for the heavy hole and split-off hole reflect the temperature-dependent change of the curvature at the respective band edges, while the shape for the light hole is mainly determined by nonparabolicity. For the influence of warping, see Sect. 4.5 of ▶ Chap. 22, "Carrier-Transport Equations".

With increasing temperature, higher energy states are filled with a decreased curvature of $E(\mathbf{k})$ and therefore with an increased effective mass of the carriers; the corresponding density-of-state masses are shown in Fig. 5 for electrons and holes. The density of states at the edge of the valence or conduction band (Eqs. 20 and 16) is then obtained by replacing m_p or m_n with the density-of-state masses $m_{p\mathrm{ds}}$ or $m_{n\mathrm{ds}}$, which contain potential degeneracy factors according to Eq. 23:

$$N_\mathrm{v} = 2\left(\frac{m_{p\,\mathrm{ds}}\,kT}{2\pi\hbar^2}\right)^{3/2} \text{ and } N_\mathrm{c} = 2\left(\frac{m_{n\,\mathrm{ds}}\,kT}{2\pi\hbar^2}\right)^{3/2}. \tag{28}$$

Use of the Fermi Integral When the Boltzmann approximation cannot be applied in evaluating Eq. 14 because of the close proximity of E_F to the band edge, one must use the *Fermi integral*

$$F_j(s) = \int_0^\infty \frac{x^j\ dx}{\exp(x-s)+1}, \tag{29}$$

with j as the order of the Fermi integral. One now obtains

$$n = \frac{1}{2\pi^2}\left(\frac{2m_n}{\hbar^2}\right)^{3/2} F_j\left(\frac{E_\mathrm{c}-E_\mathrm{F}}{kT}\right). \tag{30}$$

The Fermi integral is tabulated by McDougall and Stoner (1938) or Blakemore (1962). For $j = 1/2$ it can be approximated (Ehrenberg 1950) by:

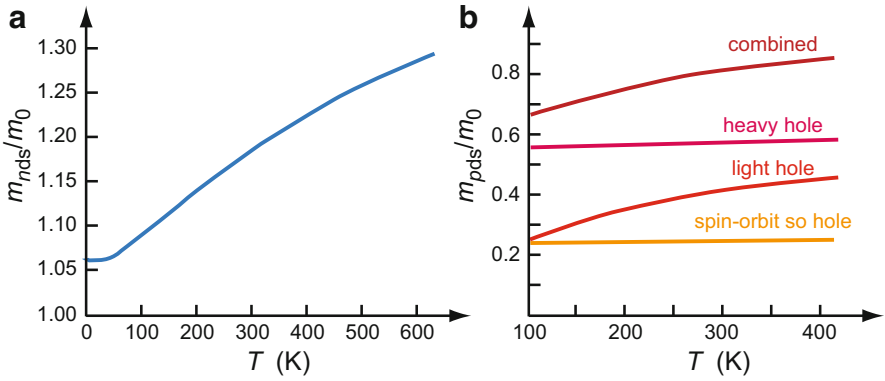


Fig. 5 Density-of-state mass for Si as a function of the temperature, (**a**) for electrons and $N_D = 5 \times 10^{17}$ cm^{-3} (After Barber 1967), (**b**) for holes and $N_A = 10^{14}$ cm^{-3} (After Li 1979)

$$F_{1/2}(s) = \frac{2\sqrt{\pi}\exp s}{4 + \exp s}. \tag{31}$$

A comparison between the classical approximation (Eq. 15), the approximation (Eq. 31), and the exact form (Eq. 29) is shown in Fig. 6. They agree for $s < -2$, but deviate substantially from each other when the Fermi level moves closer than $2kT$ toward or beyond the band edge. For a review of approximations, see Blakemore (1982).

The Fermi Energy In order to obtain the position of the Fermi level, additional input is needed. In an undoped ideal homogeneous semiconductor, electrons and holes are created in pairs. Therefore, the density of electrons in the conduction band must be equal to the density of holes in the valence band:

$$n = p. \tag{32}$$

Using this *neutrality condition*, we obtain with Eqs. 21 and 28 the relation

$$E_F = \frac{E_c + E_v}{2} + kT \ln\left(\frac{m_{p\,ds}}{m_{n\,ds}}\right)^{3/4}, \tag{33}$$

which puts E_F essentially in the middle of the bandgap for a sufficiently wide bandgap, except for a shift due to the ratio of m_{nds}/m_{pds}.

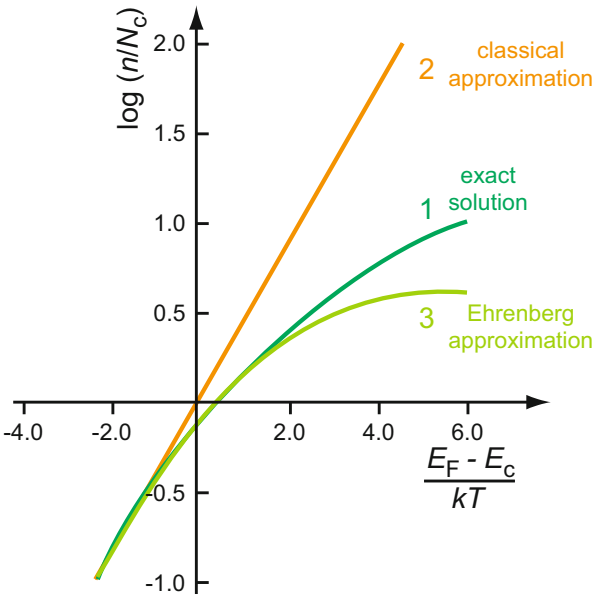


Fig. 6 Relative carrier density as a function of the normalized energy difference $s = -(E_c - E_F)/(kT)$ for the exact solution, the classical approximation, and the Ehrenberg approximation (curves 1–3, respectively)

1.5 Intrinsic Carrier Generation

In intrinsic semiconductors, both electrons and holes contribute to the electronic properties. The generation of these carriers by thermal excitation (▶ Chap. 29, "Carrier Generation") across the bandgap is termed *intrinsic carrier generation*. The carrier density is obtained by introducing E_F from Eq. 33 into Eq. 15. For sufficiently wide bandgap material, with E_g in excess of ~$6kT$ at room temperature, where the "1" in the denominator of the Fermi–Dirac distribution can be neglected, one obtains

$$n = p = n_i = N_c \left(\frac{m_p}{m_n}\right)^{3/4} \exp\left(-\frac{E_c - E_v}{2kT}\right) \tag{34}$$

or, in a more symmetrical form, for the *intrinsic carrier density*

$$n_i = \sqrt{N_c N_v}\, \exp\left(-\frac{E_c - E_v}{2kT}\right). \tag{35}$$

The intrinsic carrier density increases exponentially with a slope of one-half of the bandgap energy. This density is shown in Fig. 7 as a function of temperature with the bandgap energy as family parameter. For a review, see Blakemore (1962) or Shklovskii and Efros (1984).

2 The Extrinsic Semiconductor

Extrinsic semiconductors contain lattice defects that determine the density of electrons or holes in the conduction or valence band, since the intrinsic densities n_i at room temperature are very small: $n_i = 2.4 \times 10^{13}$, 1.45×10^{10}, and 1.79×10^{6} cm^{-3} for

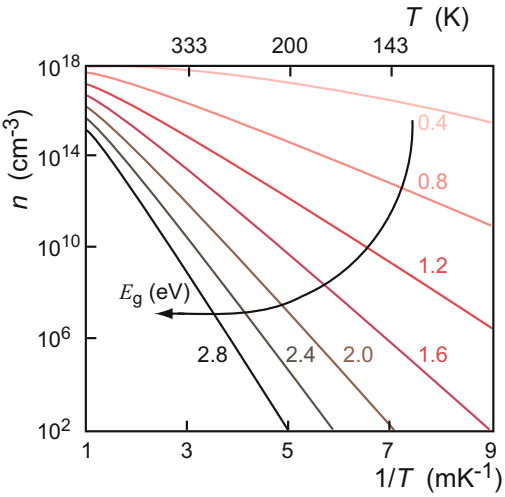


Fig. 7 Intrinsic electron density in a semiconductor as a function of the reciprocal temperature. The bandgap energy is the family parameter; the electron mass m_n is assumed equal to the mass of the free electron m_0

Ge, Si, and GaAs, respectively. The usually relevant densities of electrons or holes of about 10^{15}–10^{19} cm^{-3} are generated almost exclusively by donors or acceptors. The introduction of such donors or acceptors into the host crystal is mostly done by adding a small density of the appropriate impurities to the semiconductor. This process is referred to as *doping*, see Sect. 1.2 of ▶ Chap. 15, "Crystal Defects".

***n*- or *p*-Type Semiconductors** Doped semiconductors have one predominant type of carriers. With *donor doping*, $n \gg p$, the semiconductor is termed *n*-type. In homogeneous materials, essentially all transport properties are determined by these *majority carriers*, here electrons in the conduction band. The influence of *minority carriers*, here holes in the valence band, is negligible with homogeneous doping. This is no longer true in inhomogeneous materials, where the influence of minority carriers can play an important role. For example, the current in most solar cells is almost entirely provided by minority carriers. Semiconductors with *acceptor doping* ($p \gg n$) are called *p-type.* In these, the influence of electrons is negligible.

A rather simple way to determine experimentally whether a semiconductor is *n*- or *p*-type is based on minority-carrier *injection* from a point contact. A high field causes such an injection in this geometry, whereas, at a large area contact, the field is insufficient for injection (Henisch 1957). Therefore, for a bias that injects minority carriers (*forward bias* for these carriers), a large current will flow, while with opposite polarity (*reverse bias*), only the small number of already-present minority carriers can be collected, resulting in a much smaller current (Fig. 8). Other means to identify *n*- or *p*-type conductivity employ the Hall effect or thermo-emf[5] measurements (Sect. 1.2.1 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients").

Carrier Densities in Doped Semiconductors The density of electrons and holes in a semiconductor, with given densities of donors and acceptors of a known energy and charge character, can easily be determined by again using statistical arguments. As shown in Sect. 1.3, the carrier density is given by

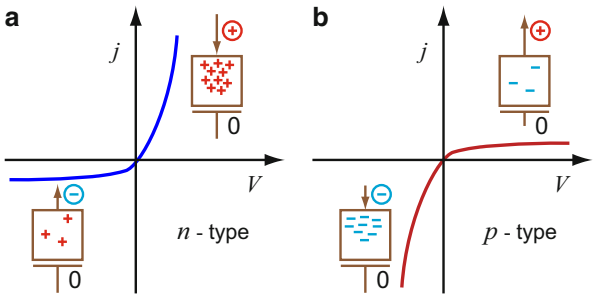


Fig. 8 Schematics of semiconductors with point (top) and large-area (bottom) contacts, showing rectifying characteristics of (**a**) *n*-type and (**b**) *p*-type semiconductors. The current is carried by minority carriers which are depicted in the figure

[5]When a hot wire touches the semiconductor, the wire becomes charged oppositely to the carrier type. For example, when contacted to an *n*-type semiconductor, the wire becomes positively charged with respect to the semiconductor since electrons are emitted into the semiconductor.

$$n = N_c \exp\left(-\frac{E_c - E_F}{kT}\right) \text{ and } p = N_v \exp\left(-\frac{E_F - E_v}{kT}\right). \tag{36}$$

In contrast to the intrinsic behavior discussed in Sect. 1.3, both of these densities are no longer equal to each other in an extrinsic semiconductor. Due to an excess of either donors or acceptors, the Fermi level moves substantially away from the middle of the bandgap toward the edge of the conduction or valence band, respectively.

2.1 The Position of the Fermi Level

The position of the Fermi level can be determined from the neutrality condition

$$n + N_A^- = p + N_D^+, \tag{37}$$

where N_A^- is the density of negatively charged acceptors, i.e., acceptors which have accepted an electron, and N_D^+ is the density of positively charged donors, i.e., donors that have donated their electron – see Fig. 9. Since the total density of a dopant equals the sum of its charged and neutral parts, the charged acceptors and donors are given by

$$N_A^- = N_A - N_A^0 \text{ and } N_D^+ = N_D - N_D^0 \tag{38}$$

with N_A and N_D as the total density of acceptors or donors, independent of their occupation.

The degree of filling of donors (N_D^0/N_D) and acceptors (N_A^0/N_A) is given by the Fermi–Dirac distribution functions

$$N_D^0 = \frac{N_D}{1 + \frac{1}{\nu_{dD}} \exp\left(\frac{E_D - E_F}{kT}\right)} \text{ and } N_A^0 = \frac{N_A}{1 + \frac{1}{\nu_{dA}} \exp\left(\frac{E_F - E_A}{kT}\right)}, \tag{39}$$

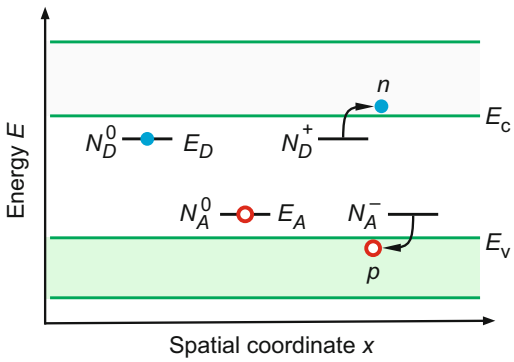


Fig. 9 Band model with neutral and ionized donors and acceptors

with ν_{dD} and ν_{dA} as the degeneracy factors of the donor and acceptor. The values of the degeneracy factors depend on the electronic structure of the defect – see Landsberg (1982). In the simplest case, ν_{dD} or $\nu_{dA} = 2$ for defects which have an initially empty or paired state and a final state with an unpaired electron. Alternatively, ν_{dD} or $\nu_{dA} = ½$ if the initial state was occupied with the unpaired electron, while the final state is empty or is occupied with paired electrons – see also van Vechten (1980).

Equations 36 through 39 present a set of algebraic equations for determining E_F. It is helpful to use a graphical method for obtaining E_F as a function of doping and temperature (Shockley 1950).[6] In Fig. 10, n, N_A^-, p, and N_D^+ are plotted in a semilogarithmic representation as a function of E_F for given temperature and doping densities – here, predominant acceptor doping is assumed. The intersection of $n + N_A^-$ and $p + N_D^+$ (heavy curves) yields the solutions $E_F^{(sol)}$ and $p^{(sol)}$.

From the example shown in Fig. 10, the Fermi level is determined by p and N_A^- only and lies between the acceptor level E_A and the valence-band energy E_v. The material in this example is p-type. The determining equation here can be simplified to $p = N_A^-$, and neglecting the "1" in the denominators of Eq. 39, we obtain

$$N_v \exp\left(\frac{E_v - E_F}{kT}\right) = \nu_{dA} N_A \exp\left(-\frac{E_A - E_F}{kT}\right), \tag{40}$$

yielding for the p-type semiconductor

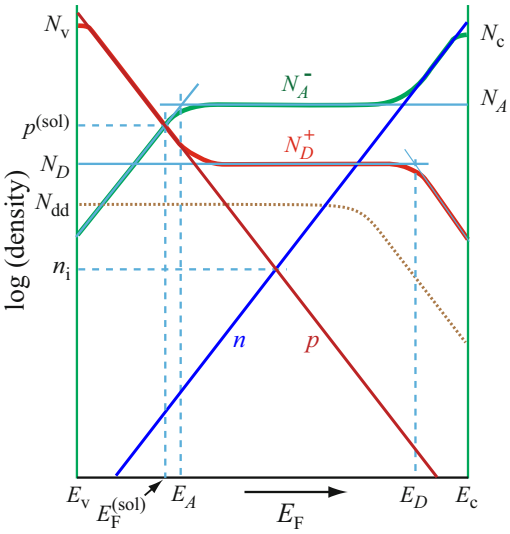


Fig. 10 Carrier densities in bands (n and p) and in donors or acceptors (N_D^+ and N_A^-) as a function of the Fermi energy. A graphical solution (sol) of Eq. 37 is indicated by the arrow ($E_F^{(sol)}$) with the corresponding hole density in the valence band $p^{(sol)}$. N_{dd} is the density of a deep donor not affecting the solution

[6]The graphical presentation is similar to the Brouwer diagram introduced in Sect. 2.6 of ▶ Chap. 15, "Crystal Defects". In the Brouwer diagram, ln(n) is plotted versus $1/kT$, while here we plot ln(n) versus E. In both presentations, small contributions are neglected due to the logarithmic density scale.

$$E_F = \frac{E_v + E_A}{2} - \frac{kT}{2}\ln\left(\frac{N_v}{\nu_{dA} N_A}\right). \tag{41}$$

When the density of donors is increased for partial compensation, keeping all other parameters constant, essentially no changes in E_F is observed until N_D approaches N_A. Then E_F moves away from the valence band until, when N_D reaches N_A, the Fermi level jumps to the middle of the gap; here compensation is complete. When N_D even slightly exceeds N_A, E_F further moves to a position near E_D. The semiconductor has turned from *p*-type to *n*-type, and the Fermi level for this *n*-doped semiconductor is determined by

$$E_F = \frac{E_D + E_c}{2} + \frac{kT}{2}\ln\left(\frac{N_c}{\nu_{dD} N_D}\right). \tag{42}$$

Figure 10 also shows that, in general, additional doping, indicated for a deeper donor level N_{dd} by the dotted curve, has little influence on E_F as long as the density is substantially below that of the most prevalent level of the same type. Observe the addition of concentrations on the logarithmic density scale in Fig. 10.

Defect Compensation In the presence of both donors and acceptors, surplus electrons from the donors recombine with holes from the acceptors, with the overall effect of reducing the surplus density of electrons and holes, i.e., bringing the concentration closer to the intrinsic, undoped case. Such a semiconductor is called a *partially compensated semiconductor* (see Fig. 11); the effective density of defects is given by the difference of the concentration of donors and acceptors:

$$N_{eff} = N_D - N_A. \tag{43}$$

This effective defect density will be donor-like if $N_D > N_A$ or acceptor-like if $N_A > N_D$. The carrier density for a partially compensated semiconductor is then given by an implicit equation:

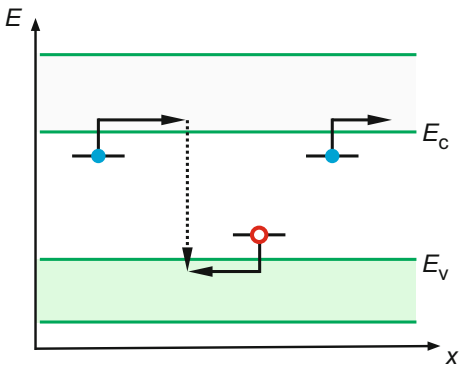


Fig. 11 Recombination of surplus electrons and holes from donors and acceptors in a partially compensated semiconductor

$$\frac{n(n+N_A)}{N_D-N_A-n}=\nu_{dD}N_c\exp\left(-\frac{E_c-E_D}{kT}\right); \tag{44}$$

see, e.g., Blakemore (1962) and Landsberg (1982). For an additional inclusion of excited donor states, see Landsberg (1956).

2.2 Temperature Dependence of the Fermi Level

The temperature dependence can be obtained from the graphical representation, shown in Fig. 12 for three temperatures. At low temperature, curve set 1, the carrier densities between acceptor and donor interact. The governing quasi-neutrality equation can be simplified to $N_A^- = N_D^+$ or, as seen from Fig. 12,

$$N_D=\frac{\nu_{dA}}{\nu_{dD}}N_A\exp\left(-\frac{E_A-E_F}{kT}\right), \tag{45}$$

resulting in a Fermi level that in this example lies close to the acceptor level; this is a partially compensated case:

$$E_F=E_A-kT\ln\left(\frac{\nu_{dA}}{\nu_{dD}}\frac{N_A}{N_D}\right). \tag{46}$$

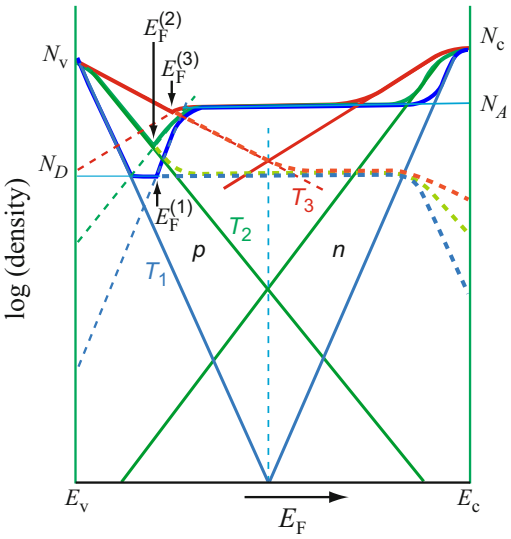


Fig. 12 Carrier densities as in Fig. 10, but for three different temperatures, $T_1 < T_2 < T_3$, for curve sets 1, 2, and 3, respectively, and corresponding Fermi energies

With increasing temperature, curve set 2, the carrier densities in the acceptors and in the valence band interact, yielding the set of equations discussed in the previous section. The effect of compensation is reduced. The Fermi level lies about halfway between the defect level and the corresponding band.

At still higher temperature, curve set 3, the acceptor becomes depleted, and the quasi-neutrality equation becomes

$$p = N_{\mathrm{v}} \exp\left(\frac{E_{\mathrm{v}} - E_{\mathrm{F}}}{kT}\right) = \nu_{\mathrm{dA}} N_A, \tag{47}$$

with the Fermi level shifting closer to the center of the bandgap:

$$E_{\mathrm{F}} = E_{\mathrm{v}} - kT \ln\left(\frac{N_{\mathrm{v}}}{\nu_{\mathrm{dA}} N_A}\right). \tag{48}$$

Finally, at still higher temperatures (not shown in Fig. 12), electron and hole densities from the two bands are the dominating partners; this is the intrinsic case, described in Sect. 1.3 and given by Eq. 33. The Fermi level has now approached the center of the gap.

A typical temperature dependence of the Fermi level in a doped semiconductor is summarized in Fig. 13 for *n*- and *p*-type Si. This figure also indicates the dependence of the bandgap energy on the temperature – see Sect. 2.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids". For donor doping, one of the upper curves is selected; for acceptor doping, one of the lower curves is selected, according to the donor or acceptor density.

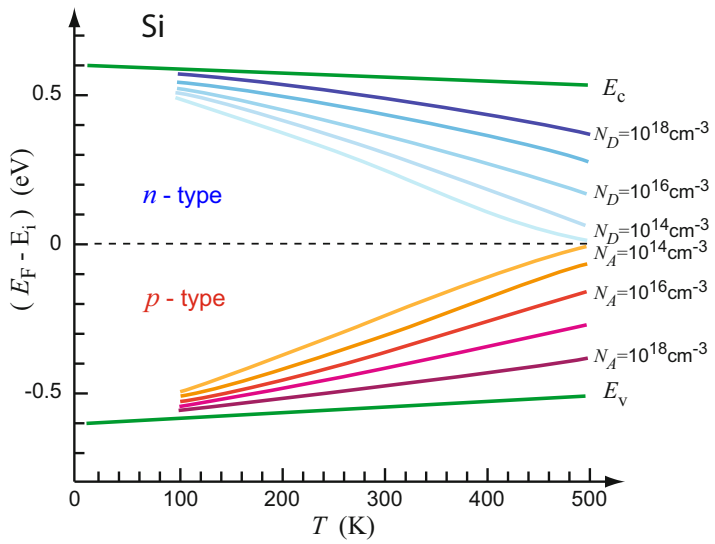


Fig. 13 Fermi energy in Si as a function of the temperature for five different doping levels, also showing the decrease of the bandgap energy at higher temperatures (After Grove 1967)

2.3 Carrier Density in Extrinsic Semiconductors

Introducing the Fermi level, as obtained in Sect. 2.2, into the equations for the carrier densities (Eq. 36) yields the extrinsic branches in the $n(T)$ or $p(T)$ curves.

We distinguish different temperature ranges for partially compensated or for essentially uncompensated semiconductors.

1. Low temperatures: from Eqs. 36 and 48, we obtain for partially compensated semiconductors

$$n = \frac{\nu_{dD}}{\nu_{dA}} \frac{N_c(N_D - N_A)}{N_A} \exp\left(-\frac{E_c - E_D}{kT}\right) \tag{49}$$

and for uncompensated semiconductors

$$n = \sqrt{\nu_{dD} N_c N_D} \exp\left(-\frac{E_c - E_D}{2kT}\right). \tag{50}$$

The factor 1/2 in the exponent is characteristic of uncompensated semiconductors.

2. Medium temperatures: from Eqs. 36 and 42, we obtain for compensated or uncompensated semiconductors

$$n = \nu_{dD} N_D - \nu_{dA} N_A, \tag{51}$$

i.e., carrier depletion from donors.

3. High temperatures: one reaches the intrinsic branch with

$$n = \sqrt{N_v N_c} \exp\left(-\frac{E_c - E_v}{2kT}\right). \tag{52}$$

This behavior is depicted in Fig. 14 for partially compensated and uncompensated semiconductors.

A similar set of equations is obtained for p-type semiconductors by replacing N_c with N_v and ($\nu_{dD} N_D$, E_D) with ($\nu_{dA} N_A$, E_A) in Eqs. 49, 50, 51, and 52.

Defects in Gapless Semiconductors Donors or acceptors also influence the conductivity of gapless semiconductors such as $Hg_{1-x}Cd_xTe$, although the defect states lie inside one of the bands and therefore become resonant states – see Sect. 1 of ▶ Chap. 18, "Shallow-Level Centers". These defects cause a shift of the Fermi level away from the touching point in $E(k)$ (see, e.g., center diagrams of Figs. 4 and 5a in ▶ Chap. 8, "Bands and Bandgaps in Solids") making the material n- or p-type. The position of the Fermi level is determined by the quasineutrality condition Eq. 37. See Tsidilkovski et al. (1985) for a review.

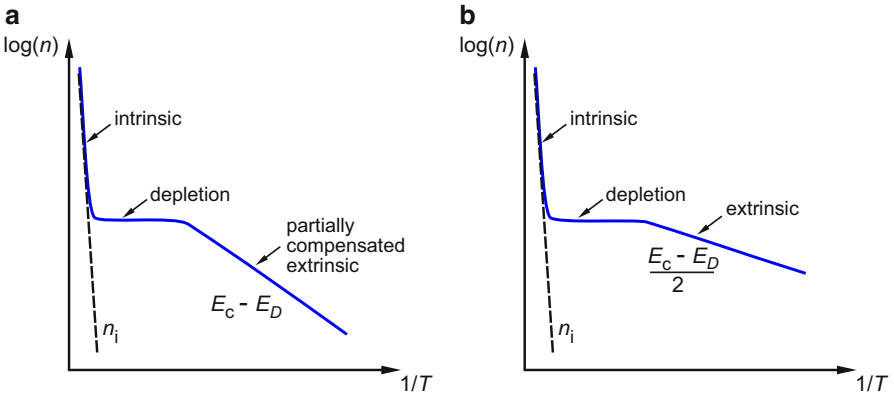


Fig. 14 Carrier density as a function of the temperature in (**a**) partially compensated and (**b**) uncompensated semiconductors

2.4 Intrinsic and Minority Carrier Densities

The density of *minority carriers*, i.e., electrons in a *p*-type or holes in an *n*-type semiconductor, is obtained from the equilibrium equation once the position of the Fermi level is known. For instance, in the example given in the previous section for an *n*-type semiconductor, with E_F given by Eq. 42 and n given by Eq. 50, the minority carrier density p is given by

$$p = N_v \exp\left(-\frac{E_F - E_v}{kT}\right). \tag{53}$$

The product of minority and majority carrier densities in *thermal equilibrium* yields the *intrinsic carrier density*

$$n_i^2 = np = N_c N_v \exp\left(-\frac{E_c - E_v}{kT}\right). \tag{54}$$

In graded junctions or highly doped inhomogeneous materials (see Sects. 2.3 of ▶ Chap. 8, "Bands and Bandgaps in Solids", and 3.1 of ▶ Chap. 22, "Carrier-Transport Equations"), a similar relationship holds except that $E_g = E_g(x) = E_{g0} + \Delta E_g(x)$, where x is the composition parameter of the alloy; hence $n_i^2 = n_i^2(x)$ with

$$n_i^2(x) = N_c N_v \exp\left(-\frac{\Delta E_g(x)}{kT}\right) \exp\left(-\frac{E_{g0}}{kT}\right), \tag{55}$$

and thus is independent of doping. With the knowledge of n_i, we can easily obtain the corresponding equilibrium density of minority carriers for any given majority carrier-density. For example

$$p = n_i^2/n, \tag{56}$$

with n as given in Eq. 49. This relation, however, no longer holds when another excitation such as light is employed or when the minority carrier-density is frozen-in – see Sect. 2.6; then $n\ p > n_i^2$ – see ▶ Chap. 31, "Photoconductivity". For deviations in degenerate semiconductors, see Blakemore (1962).

2.5 Self-Activated Carrier Generation

Doping with acceptors and donors to provide extrinsic conductivity, as discussed in the previous section, is usually done at a constant level, independent of temperature. However, vacancies and interstitials can also act as donors or acceptors (Sect. 2.4 of ▶ Chap. 15, "Crystal Defects") and thereby contribute to the carrier generation. Since the generation of these native defects from Schottky or Frenkel disorder is an intrinsic process, they may be considered as influencing an intrinsic carrier generation (i.e., a generation which is not doping dependent).[7]

In a monatomic semiconductor with vacancy generation in thermodynamic equilibrium, the influence on the carrier generation is unambiguous. When vacancies act as acceptors, the density of the acceptors increases with temperature according to

$$N_A = N_\mathrm{L} \exp\left(-\frac{E_\mathrm{Schottky}}{kT}\right) \tag{57}$$

where N_L is the density of lattice sites and E_S is the Schottky energy.

Introducing this relationship into the carrier–density relation Eq. 53, we obtain

$$p = \sqrt{N_\mathrm{v} N_\mathrm{L}} \exp\left(-\frac{E_A - E_\mathrm{v} + E_\mathrm{Schottky}}{2kT}\right). \tag{58}$$

The pre-exponential factor of Eq. 58 is larger than for an extrinsic carrier generation (Eq. 53). The behavior resembles electronic intrinsic generation (Eq. 52), since $E_A - E_\mathrm{v} + E_\mathrm{Schottky}$ can be on the order of the bandgap energy and is not determined by doping except for compensation, which is neglected here.

This self-activated conductivity can be distinguished from a purely electronic, intrinsic conductivity by a larger carrier density at the intercept for $T \to \infty$, since N_L is typically on the order of 10^{22} cm^{-3}, while N_v and N_c are on the order of 10^{19} cm^{-3} (Fig. 15). In addition, one has to consider the temperature dependence of the bandgap energy, which causes a contribution in the pre-exponential factor.

An experimental determination of self-activated conductivity by comparing the slope with the optically measured bandgap energy E_g is handicapped by ambiguities.

[7] The dependence of their formation energy in the *charged* state on the Fermi level must, however, be taken into account; see Sects. 2.2 and 2.3 of ▶ Chap. 19, "Deep-Level Centers".

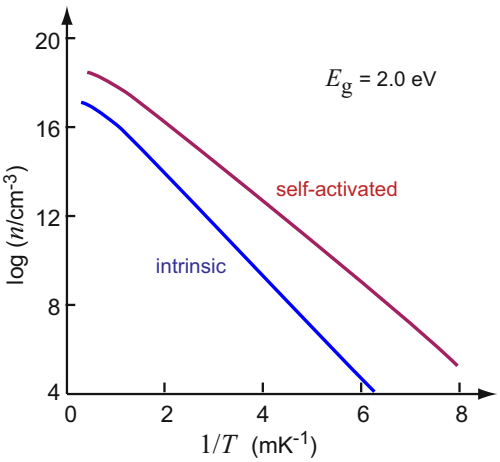


Fig. 15 Self-activated carrier density as a function of the inverse temperature, computed for a bandgap energy of 2 eV, $E_{\text{Schottky}} = 1.5$ eV, and $E_A - E_v = 0.1$ eV; $N_L = 10^{22}$ cm^{-3} and $N_v = 10^{19}$ cm^{-3}

These ambiguities are due to levels in the Urbach tail (see Sect. 3 of ▶ Chap. 17, "Optical Properties of Defects", and Sect. 1 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors") and by the unknown magnitude of the Franck–Condon shift (see Sect. 2.3.1 of ▶ Chap. 17, "Optical Properties of Defects"), which permits thermal ionization with a lower energy than the optical band–edge transition. In addition, pairs of intrinsic defects in binary or higher compounds are created at densities which do not yield exact compensation – see Sect. 2.3 of ▶ Chap. 15, "Crystal Defects". This does not eliminate the influence of intrinsic defects, as an exact compensation would, but makes an analysis more complex.

2.6 Frozen-In Carrier Densities

The minority-carrier density for sufficiently doped, wider bandgap materials becomes unreasonably small. For instance, in *n*-type GaAs with a bandgap energy of 1.424 eV at 300 K, the intrinsic carrier density is 1.8×10^6 cm^{-3}. With a density of shallow donors of 10^{17} cm^{-3} yielding $n \cong 10^{17}$ cm^{-3} at room temperature, the density of minority carriers in thermal equilibrium would be $p \cong 3 \times 10^{-5}$ cm^{-3} (Eq. 54). This density is substantially below a value probably maintained by background cosmic radiation, and it would take an extremely long time to approach equilibrium without such radiation. Therefore, a different approach is necessary to describe the actual behavior for minority carriers in wider-gap semiconductors.

The minority-carrier response time can be estimated from

$$\tau_p = \frac{1}{\nu_p} \exp\left(\frac{E_F - E_v}{kT}\right), \tag{59}$$

with ν_p as the attempt-to-escape frequency of a *hole* from a shallow *acceptor*. With ν_p typically on the order of 10^{10}–10^{13} s^{-1} (see Sect. 1.2.2 of ▶ Chap. 29, "Carrier Generation"), one estimates that it will take more than 10^5 s to achieve equilibrium when the Fermi level is farther away than ~1 eV from the band edge. For most electronic experiments, the present state is considered *frozen-in* if a delay of more than one day (~10^5 s) is necessary to achieve a *steady state*, that is, whenever the Fermi level is more than 1 eV away from the respective band edge. This means that minority carrier densities at room temperatures are not expected to drop below a frozen-in (f-i) density of

$$p_{\mathrm{f-i}} \cong N_{\mathrm{v}} \exp\left(-\frac{1\ \mathrm{eV}}{kT}\right) \cong 10^2\ \ \mathrm{cm}^{-3}. \tag{60}$$

When, for consistency, one wants to maintain the Fermi-level concept, then it requires the introduction of a *quasi-Fermi level* for minority carriers, here $E_{\mathrm{F}p}$ for holes, which is formally introduced by

$$p_{\mathrm{f-i}} = N_{\mathrm{v}} \exp\left(-\frac{E_{\mathrm{F}p} - E_{\mathrm{v}}}{kT}\right). \tag{61}$$

This concept can be justified as reasonable under certain conditions[8] – see Sect. 3.3 of ▶ Chap. 22, "Carrier-Transport Equations". It is a helpful approximation for a variety of discussions dealing with an additional excitation (e.g., optical) to thermal excitation.

When defining a quasi-Fermi level, it is customary to also convert the notation of the Fermi level relating to the majority carrier, although essentially unchanged from E_{F},[9] to another quasi-Fermi level: here $E_{\mathrm{F}n} \cong E_{\mathrm{F}}$. Deviations from the thermal equilibrium then result simply in a split of E_{F} into $E_{\mathrm{F}n}$ and $E_{\mathrm{F}p}$. Extensive use of this concept will be made in Sect. 3.3 of ▶ Chap. 22, "Carrier-Transport Equations".

It will suffice here to summarize that quasi-Fermi levels in normal observation of semiconductors cannot be farther separated from the respective bands than by at most ~40kT, i.e., about 1 eV for room temperature due to freezing-in:

$$(E_{\mathrm{c}} - E_{\mathrm{F}n}) < 40\,kT \text{ and } \left(E_{\mathrm{F}p} - E_{\mathrm{v}}\right) < 40\,kT. \tag{62}$$

In wider-bandgap materials ($E_{\mathrm{g}} > 1$ eV) or at lower temperatures, freezing-in becomes an important consideration. Freezing-in also applies to thermal excitation from deep defect levels in the bandgap. However, since the frequency factor depends on the type of defect center (see ▶ Chap. 29, "Carrier Generation"), the freezing-in

[8] Justification in steady state, e.g., with optical excitation, is reasonable. The use, however, of quasi-Fermi levels close to a frozen-in situation becomes questionable when different types of defects are involved.

[9] Through mutual generation and recombination of electrons and holes, a change in minority carrier density also causes a change in majority carrier density; but, because of the much larger density of majority carriers, the change is truly negligible.

depth varies between $20kT$ and $40kT$ for Coulomb-repulsive to Coulomb-attractive centers, respectively, for a freezing-in time of 10^5 s. In devices, usually much shorter response times are required. For freezing-in, this causes a much smaller distance of the quasi-Fermi level to the corresponding bands and requires a more detailed discussion regarding which of the trap levels can follow with their population and which cannot.

3 Phase Transitions at High Carrier Densities

3.1 The Mott Transition

At very low temperature ($T \rightarrow 0$), hydrogen-like donors and acceptors are neutral, states in the conduction and valence bands are respectively unoccupied and occupied, and the semiconductor behaves like an insulator (see Fig. 14). At very high doping concentrations, however, energy states in the bandgap lose their localization if their mutual energy is close enough (see band tailing in Sect. 3 of ▶ Chap. 17, "Optical Properties of Defects", and Sect. 1 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors"). These states then form a narrow band and, when partially filled, contribute to electronic conduction even for $T \rightarrow 0$, i.e., the semiconductor behaves like a metal (Mott 1984).

The critical value of doping concentration for such an insulator–metal transition is reached when the average distance between hydrogen-like impurities (Sect. 1 of ▶ Chap. 18, "Shallow-Level Centers") approaches their diameter. Their Coulomb potentials start to overlap, and as a consequence carriers can be transferred to adjacent impurities by either thermionic emission over the separating barrier or by tunneling. These processes increase for decreasing interimpurity separation. In addition, the ionization energy of the shallow impurities is lowered by the screening of their potential due to the high free-carrier concentration. This produces more free carriers, which contribute to screening, consequently causing (even at low temperature) ionization of yet unionized hydrogen-like impurities, in turn producing more free carriers. As a consequence, a quite abrupt transition occurs for increased doping, the *Mott transition* (Mott 1990), which causes an increase in conductivity and makes the semiconductor metal-like. Such semiconductor–metal transitions, following Eq. 63, are observed for a variety of semiconductors (Edwards and Sienko 1978). This transition can be achieved in thermal equilibrium: a semiconductor turns into a metal at high doping densities – it becomes a *degenerate semiconductor*. Beyond such a transition, the electrical conductivity remains high and nearly constant down to $T \rightarrow 0$, as shown in Fig. 16b for Si with a donor density above ~3.2×10^{18} cm^{-3}. The semiconductor then behaves like a metal. Similar results were found for the scarcely soluble chalcogens Se (Ertekin et al. 2012) or Te (Wang et al. 2020), introduced into Si by nonequilibrium hyperdoping, i.e., ion implantation and nanosecond annealing. DFT modelling demonstrates an insulator–metal transition by merging of defect and conduction bands and confirms the experimentally obtained critical dopant concentration.

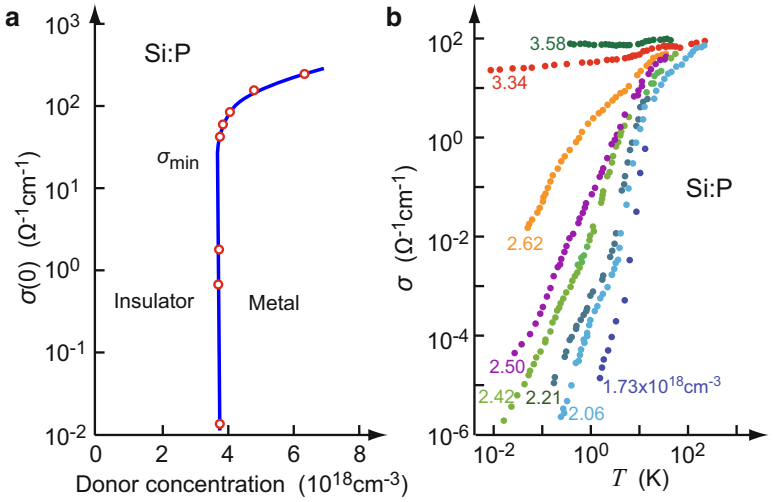


Fig. 16 (**a**) Electrical conductivity of *n*-type Si:P extrapolated to $T = 0$ as a function of the P donor concentration (After Rosenbaum et al. 1980). (**b**) Resistivity of Si:P as a function of the temperature with the donor density as family parameter (After Sasaki 1980). A semiconductor–metal transition is accomplished when a finite conductivity is maintained for $T \to 0$

The critical density for the Mott transition is closely related to the Anderson–Mott localization density (Eq. 17 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors") and is attained when the average distance $\overline{r} \cong n_{\text{Mott}}^{-1/3}$ between shallow donors or acceptors is approximately four times their quasi-hydrogen radius a_{qH}:

$$a_{\text{qH}}\, n_{\text{Mott}}^{1/3} \cong 0.24 \tag{63}$$

(Mott and Davis 1979; Mott 1987, 1990); see also Sect. 1.2.2 of ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects". Figure 17 indicates good agreement with the experiment for a variety of semiconductors. These critical densities are within the accuracy of the estimate on the same order of magnitude as the densities for electron–hole droplet formation discussed in the following section.

The dependence of Eq. 63 assumed a common Thomas-Fermi screening length in metal and insulator and neglected exchange interaction; an enhanced effective-mass approach yields a temperature dependence of n_{Mott} in agreement with conductivity data of *n*-GaAs and *n*-GaN (Romanets and Sachenko 2016).

3.2 Electron–Hole Condensation

There are also other ways to initiate metal-like behavior in semiconductors. Rather than generating free carriers by the overlap of hydrogen-like *impurities* creating an impurity band, a similar critical process can be initiated by a sufficient *optical*

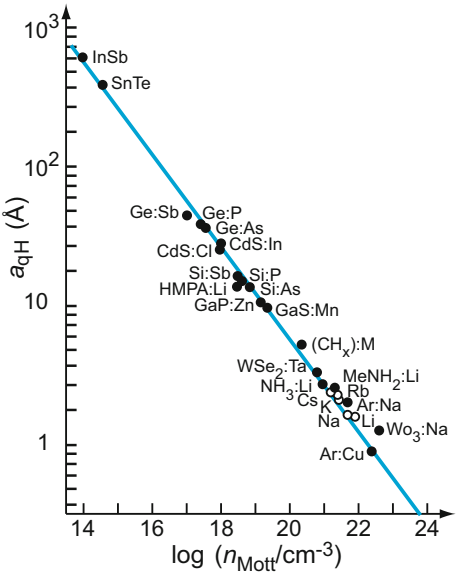


Fig. 17 Mott density of donors for the insulator–metal transition as a function of the quasi-hydrogen donor radius. The straight line is given by Eq. 63 (After Edwards and Sienko 1981)

generation of free electrons and holes. When their density is large enough, the interaction between electrons and holes becomes strong, and they can no longer be regarded as a gas of independent particles; at low enough temperatures, their Coulomb energy becomes larger than their thermal energy:

$$\frac{e^2}{4\pi\varepsilon_{\text{stat}}\varepsilon_0} n^{1/3} > kT; \tag{64}$$

then condensation of the carrier ensemble into a liquid phase takes place, which causes substantial changes in the behavior of the semiconductor and renders it metal-like. Such a condensation within an electron–hole system is similar to the phase transition of normal gasses into liquids; at low temperatures and sufficient densities, electrons and holes can occupy a state as interpenetrating fluids with a lower energy than that of excitons or biexcitons.

Optically created electrons and holes first thermalize with the lattice, typically within $\tau_{\text{th}} \cong 1$ ns, before the carriers recombine, typically within $\tau_{\text{rec}} \cong 10$ μs. Therefore, these carriers can be regarded as having acquired lattice temperature. At low enough temperatures, they tend to form excitons or, at high excitation density, also biexcitons (exciton molecules). At even higher density, i.e., when $n_{\text{exc}} > n_{\text{crit}} \cong \left(2a_{\text{qH}}\right)^{-3}$, excitons overlap and no individual electron–hole binding persists. The exciton gas then transforms to an *electron–hole plasma* with an average energy per electron–hole pair below that of a single exciton E_{exc}.

At this transition, referred to as *exciton Mott transition*, excitons are rapidly ionized; the binding energies and oscillator strengths of the lowest excitonic bound states are reduced by orders of magnitude due to increased screening. However, theoretical (Schleife et al. 2012) and experimental studies (Suzuki and Shimano

2012, Sekiguchi and Shimano 2015) indicate that bound excitonic states persist in the metallic phase above the Mott density.

The electron–hole plasma is stabilized in indirect semiconductors like Si and Ge by the multivalley degeneracy of the conduction bands, which lowers the kinetic energy. At sufficiently high density and low temperature, the condensation into the even more stable state of an *electron–hole liquid* occurs (Landau and Lifshitz 1976, Reinecke and Ying 1979, Sibeldin 2017). Such a condensation is similar to that of forming a liquid metal from its vapor phase; during the condensation, the orbits of valence electrons disappear, changing their behavior into that of free electrons. However, the heavy ion cores are replaced here by holes which have a mass on the same order of magnitude as the electrons, causing a breakdown of the adiabatic approximation. As a result, large zero-point vibrations with amplitudes of the exciton diameter prevent solidification of the electron–hole liquid at $T = 0$ K and yield a much lower binding energy of the liquid than for the core atoms in a corresponding metal.

The resulting electron–hole liquid has metallic properties where electrons and holes are free to move as interpenetrating Fermi liquids. The condensation into an electron–hole liquid was originally proposed by Keldysh (1968). It can be described with a diagram with liquid and gas phases, similar to the one for first-order phase transitions for vapor to liquid condensation.

Phase Diagram The properties of a system consisting of electrons, holes, excitons, biexcitons, and the electron–hole liquid at any given temperature can be described by a phase diagram. It is instructive to plot such a phase diagram in a T versus $\ln(n/n_{\mathrm{crit}})$ representation (Fig. 18), which is rather universal when scaled to the individual n_{crit}, T_{crit} values. In any isothermal process below T_{crit}, one observes at low densities a gas which, in equilibrium, contains excitons and excitonic molecules (biexcitons). At higher densities, it will appear as a mixture of electron–hole plasma (vapor) and liquid. Beyond the second phase boundary, it will show only the liquid phase. An experimental phase diagram is given in Fig. 19.

The critical parameters of an electron–hole liquid depend on the binding energy per carrier pair E_{pair} given by $E_{\mathrm{pair}} = E_{\mathrm{kin}}^{(n+p)} + E_{\mathrm{xc}}^{(n+p)}$. The correlation energy $E_{\mathrm{xc}}^{(n+p)}$

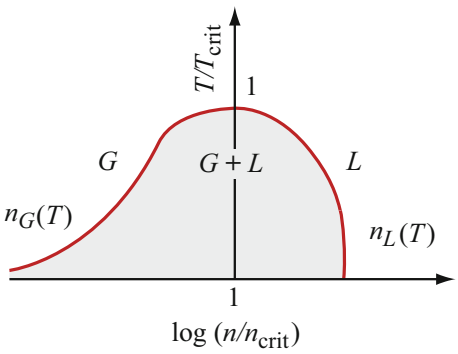


Fig. 18 Phase diagram for an exciton gas G and an electron–hole liquid L (After Keldysh 1986)

stems from the Pauli exclusion principle. It causes an increase of the distance between carriers of the same spin and consequently reduces the attractive and dominating Coulomb energy, but prevails when the interparticle distance decreases much below a_{qH}. The relevant critical parameters are listed in Table 1 for a number of semiconductors. The critical temperatures are on the order of 6 . . . 60 K for Ge and CdS, respectively. The critical density is related to the quasi-hydrogen radius a_{qH} (see Eq. 63) and the critical temperature to the binding energy E_{pair} as

$$n_{crit}^{1/3} \cong 0.2\, a_{qH}^{-1} \text{ and } kT_{crit} \cong 0.1\, E_{pair}. \tag{65}$$

Typical values for n_{crit} range from 10^{15} to 10^{20} cm^{-3}, and typical binding energies range from below 1 meV for GaAs to $\geq$ 50 meV (Thomas and Timofeev 1980). Binding energies, hence T_{crit}, are increased in materials with highly anisotropic

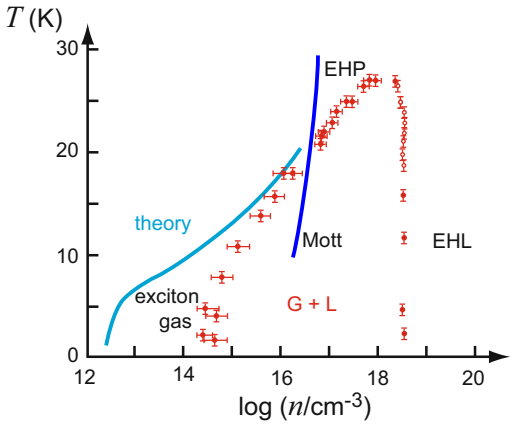


Fig. 19 Experimental phase diagram of Si for an electron–hole plasma (*EHP*), electron–hole liquid (*EHL*), and the intermediate range of coexisting gas and liquid ($G + L$) (After Dite et al. 1977). The diagram also contains the theoretical curve (*theory*) and the curve for Mott condensation (*Mott*)

Table 1 Critical parameters for electron–hole droplets. E_{pair} is the binding energy per carrier pair, n_{crit} is the critical carrier density in the electron–hole plasma, and $n_{L,0}$ is the density of the electron–hole liquid in equilibrium with the gas phase (After Tikhodeev 1985)

Solid	E_{pair}	T_{crit} (K)	n_{crit} (cm^{-3})	$n_{L,0}$ (cm^{-3})
Si	23	28	$1.2 \cdot 10^{18}$	$3.5 \cdot 10^{18}$
Ge	6	6.7	$6 \cdot 10^{16}$	$2.3 \cdot 10^{17}$
GaP	6			$7 \cdot 10^{18}$
GaS	9			$4.5 \cdot 10^{20}$
AlAs	14.5			$12 \cdot 10^{19}$
GaAs	1	6.5	$4 \cdot 10^{15}$	$1 \cdot 10^{16}$
CdS	14	64	$7.8 \cdot 10^{17}$	$5.5 \cdot 10^{18}$
CdSe	5	30	$1.2 \cdot 10^{17}$	$8.3 \cdot 10^{17}$
CdTe	0.9	18	$4.4 \cdot 10^{16}$	$2.9 \cdot 10^{17}$
ZnO	22	70	$4 \cdot 10^{17}$	$2.8 \cdot 10^{19}$
ZnS	12	79	$1.4 \cdot 10^{17}$	$8.3 \cdot 10^{17}$
ZnSe	5			$3.2 \cdot 10^{18}$
ZnTe	3			$6.6 \cdot 10^{17}$
AgBr	30			$1 \cdot 10^{19}$

effective mass and in multi-valley semiconductors. At low temperatures, electrons rarely exchange between different valleys; therefore, the electron–hole liquid can be regarded as a multicomponent Fermi liquid.

The condensation is characterized by a well-defined critical density, binding energy per electron–hole pair, and surface tension of the liquid; the latter causes *droplets* to form during condensation, with typically a few μm in diameter and a sharp boundary between the gas and the liquid phase. The vapor condenses to liquid droplets to fill only part of the total volume, much like a liquid condensing from a gas in a container of fixed volume. Such drops can be detected directly because of their different refractive index, which gives rise to increased light scattering for photons with an energy below the bandgap energy (Hensel et al. 1978), or by their specific luminescence, shown in Fig. 20. Strain confinement may give rise to larger drops up to 1 mm diameter (Wolfe et al. 1975). The strain produces a band deformation which acts as a well for the electron–hole drops (Gourley and Wolfe 1978).

Carriers are confined within such drops as electrons are within a metal. They have to overcome a work function given by the difference of E_{pair} and the exciton Rydberg energy E_{exc} to exit from the drops. With increasing temperature, electron–hole pairs can evaporate from the drops by forming excitons. When the optical excitation is turned off, this proceeds until all droplets have evaporated, provided the recombination lifetime is longer than the time for evaporation (as often occurring in indirect-gap semiconductors).

Bose–Einstein Condensation The formation of excitons implies the possibility of another phase transition referred to as *Bose–Einstein condensation*. While electrons and holes are fermions, in their mutually bound state they form composite bosons.[10] Bosons do not obey the Pauli exclusion principle, i.e., they do not tend to avoid each other. At sufficiently low temperature and high density, the entire ensemble of bosons can condensate into a single quantum-mechanical entity,

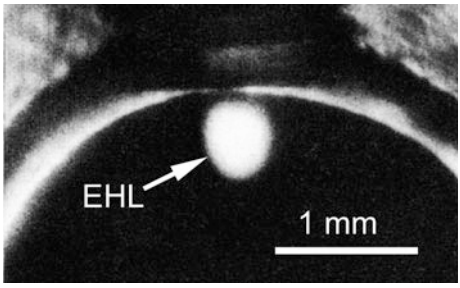


Fig. 20 Electron–hole liquid (*EHL*) with 300 μm radius in a strained Ge disk at 2 K, viewed through a (001) face. The bright region originates from *e*–*h* recombination luminescence at 1.75 μm in the liquid (After Markiewicz et al. 1977)

[10]With strong photon coupling, however, one has polaritons rather than excitons. They cannot accumulate near $K = 0$ because of their photon nature. Therefore, Bose–Einstein condensation of these quasi-particles in bulk material is impossible. However, for exciton–polaritons coupled to the mode of a microcavity mode, such a condensation appears meanwhile established as discussed by Deng et al. (2010).

described by a unique wavefunction on a near-macroscopic scale. Prominent phenomena related to such condensation are the superconductivity with zero electrical resistance (▶ Chap. 26, "Superconductivity") and the superfluidity of helium with zero friction of the flow in the liquid state. While the ^{4}He isotope is a genuine boson, the ^{3}He isotope and the electrons need to form pairs to become composite bosons.

In contrast to the electron–hole liquid, which is described by interpenetrating Fermi fluids, the boson gas of the excitons is expected to coalesce into a Bose–Einstein condensate when the de Broglie wavelength λ_{DB} of the exciton at thermal velocity becomes equal to the interparticle distance (Blatt et al. 1962; Keldysh and Kopaev 1965)

$$\lambda_{DB}(T_{BEC}) = \frac{h}{\sqrt{2m_{exc}\,kT_{BEC}}} \leq \frac{1}{\sqrt[3]{n_{exc}}}, \tag{66}$$

where m_{exc} is the effective mass of the exciton and T_{BEC} is the critical temperature for Bose–Einstein condensation. This temperature is expected to be on the order of 10 K at $n_{exc} \cong 10^{17}$ cm^{-3} for a typical exciton mass. At low enough temperatures, excitons can associate to biexciton molecules which are also bosons and should also undergo a Bose–Einstein condensation (Keldysh 1968).

Despite half a century of efforts to demonstrate excitonic Bose–Einstein condensation, a compelling experimental evidence is missing. The experiments particularly focused on small-gap alloy semiconductors with a bandgap energy tuned below the exciton binding energy. The exciton ground state is then expected to become unstable against formation of excitonic correlations (Zimmermann 2007). The fermionic constituents lead, however, always to a phase transition into an electron–hole plasma or an electron–hole liquid. Whether or not an excitonic Bose–Einstein condensation exists is still controversial. Recent studies focus on the condensation of *dark* exciton states, which are discussed in Sects. 2.3.2 and 3.2.1 of ▶ Chap. 14, "Excitons", and on spatially indirect excitons; for a review see Combescot et al. (2017). For further contributions, see publications in Solid State Communications vol. 134 (2005), Combescot and Snoke (2008), Deveaud (2015), Combescot et al. (2015), and Phan et al. (2016).

4 Summary

The density of electrons and holes as carriers of currents in semiconductors can be obtained from equilibrium statistics – the Fermi–Dirac statistics. The key parameter of this statistics is the Fermi level E_F, which can be obtained from the quasineutrality condition. E_F lies near the middle of the bandgap for intrinsic semiconductors and near the donor or acceptor level for doped (extrinsic) semiconductors. The carrier density is determined by the product of the joint density of states, i.e., the product of defect level and effective band-level densities, and the Boltzmann factor, with an activation energy given by the difference between band edge and Fermi energy. In doped

semiconductors, both majority and minority carrier-densities are of interest for semiconducting devices. The minority carrier-density of wider-gap semiconductors is often frozen-in at densities in excess of 10^2 cm^{-3}, represented by a quasi-Fermi level of typically not more than 1 eV above the minority carrier band-edge at room temperature. In materials with dominating intrinsic defects, i.e., Schottky or Frenkel disorder, self-activated conductivity must be considered at elevated temperatures. Here, the density of these defects, as well as the carrier densities, is temperature dependent.

At high doping concentrations above the Mott density, the Coulomb potentials of hydrogen-like impurities start to overlap, and free carriers are produced. This insulator–metal transition yields a degenerate semiconductor with a metal-like conductivity even at lowest temperature.

Further phase transitions take place at high quasi-particle densities and low temperatures. One of these transitions involves an optically generated exciton gas which transforms to an electron–hole plasma at high carrier density. At even higher density, the carriers condense to an electron–hole liquid. The condensation from the exciton gas may also directly form electron–hole droplets, which can be optically observed.

References

Barber HD (1967) Effective mass and intrinsic concentration in silicon. Solid State Electron 10:1039

Blakemore JS (1962) Semiconductor statistics. Pergamon Press, Oxford

Blakemore JS (1982) Approximations for Fermi-Dirac integrals, especially the function F12(η) used to describe electron density in a semiconductor. Solid State Electron 25:1067

Blatt JM, Böer KW, Brandt W (1962) Bose-Einstein condensation of excitons. Phys Rev 126:1691

Combescot M, Snoke DW (2008) Stability of a Bose-Einstein condensate revisited for composite bosons. Phys Rev B 78:144303

Combescot M, Combescot R, Alloing M, Dubin F (2015) Effects of fermion exchange on the polarization of exciton condensates. Phys Rev Lett 114:090401

Combescot M, Combescot R, Dubin F (2017) Bose–Einstein condensation and indirect excitons: a review. Rep Prog Phys 80: 066501

Deng H, Haug H, Yamamoto Y (2010) Exciton-polariton Bose-Einstein condensation. Rev Mod Phys 82:1489

Deveaud B (2015) Exciton-polariton Bose-Einstein condensates. Annu Rev Condens Matter Phys 6:155

Dite AF, Kulakovskii VD, Timofeev VB (1977) Gas–liquid phase diagram in a nonequilibrium electron–hole system in silicon. Sov Phys JETP 45:604

Edwards PP, Sienko MJ (1978) Universality aspects of the metal-nonmetal transition in condensed media. Phys Rev B 17:2575

Edwards PP, Sienko MJ (1981) Phase separation in metal solutions and expanded fluid metals. J Am Chem Soc 103:2967

Ehrenberg W (1950) The electric conductivity of simple semiconductors. Proc Phys Soc A 63:75

Ertekin E, Winkler MT, Recht D, Said AJ, Aziz MJ, Buonassisi T, Grossman JC (2012) Insulator-to-metal transition in selenium-hyperdoped silicon: observation and origin. Phys Rev Lett 108: 026401

Gourley PL, Wolfe JP (1978) Spatial condensation of strain-confined excitons and excitonic molecules into an electron–hole liquid in silicon. Phys Rev Lett 40:526

Grove AS (1967) Physics and technology of semiconductor devices. Wiley, New York

Henisch HK (1957) Rectifying semiconductor contacts. Clarendon, Oxford

Hensel JC, Phillips TG, Thomas GA (1978) The electron–hole liquid in semiconductors: experimental aspects. In: Seitz F, Turnbull D (eds) Solid state physics, vol 32. Academic Press, New York, pp 87–314

Keldysh LV (1968) Closing remarks. In: Ryvkin SM, Shmartsev YV (eds) Proceedings of the 9th international conference on the physics of semiconductors, Moscow. Nauka Publ., Lenigrad, pp 1303–1311

Keldysh LV (1986) The electron–hole liquid in semiconductors. Contemp Phys 27:395

Keldysh LV, Kopaev YV (1965) Possible instability of the semimetal state toward Coulomb interaction. Sov Phys Solid State 6:2219

Landau LD, Lifshitz EM (1976) Statisticheskaya fizika. Nauka Publ., Moscow (Statistical physics, in Russian)

Landsberg PT (1956) Defects with several trapping levels in semiconductors. Proc Phys Soc B 69:1056

Landsberg PT (1982) Semiconductor statistics. In: Moss TS, Paul W (eds) Handbook of semiconductors, vol 1, Band theory and transport. North Holland Publishing Company, Amsterdam, pp 359–449

Li SS (1979) The theoretical and experimental study of the temperature and dopant density dependence of hole mobihty, effective mass, and resistivity in boron-doped silicon, Nat Bur Stand (US) special publication 400-47. US Government Printing Office, Washington, DC

Markiewicz RS, Wolfe JP, Jeffries CD (1977) Strain-confined electron–hole liquid in germanium. Phys Rev B 15:1988

McDougall J, Stoner EC (1938) The computation of Fermi-Dirac functions. Philos Trans A 237:67

Mott NF (1984) Metals, nonmetals and metal-nonmetal transitions: some recollections. Rep Prog Phys 47:909

Mott NF (1987) Conduction in non-crystalline materials. Claredon Press, Oxford

Mott NF (1990) Metal-insulator transitions. Taylor & Francis, London

Mott NF, Davis EA (1979) Electronic processes in non-crystalline materials. Claredon Press, Oxford

Phan V-N, Becker KW, Fehske H (2016) Ground-state and spectral signatures of cavity exciton-polariton condensates. Phys Rev B 93:075138

Reinecke TL, Ying SC (1979) Scaling relations for electron–hole-droplet condensation in semiconductors. Phys Rev Lett 43:1054

Romanets PN, Sachenko AV (2016) Electron transport near the Mott transition in *n*-GaAs and *n*-GaN. Phase Transitions 89: 52

Rosenbaum TF, Andres K, Thomas GA, Bhatt RN (1980) Sharp metal-insulator transition in a random solid. Phys Rev Lett 45:1023

Sasaki W (1980) Metal non-metal transition in phosphorus-doped silicon. Philos Mag B 42:725

Schleife A, Rödl C, Fuchs F, Hannewald K, Bechstedt F (2012) Optical absorption in degenerately doped semiconductors: Mott transition or Mahan excitons? Phys Rev Lett 107: 236405

Sekiguchi F, Shimano R (2015) Excitonic correlation in the Mott crossover regime in Ge. Phys Rev B 91: 155202

Shklovskii BI, Efros AL (1984) Electronic properties of doped semiconductors. Springer, Berlin

Shockley W (1950) Electrons and holes in semiconductors. D van Norstrand Co, New York

Sibeldin NN (2017) Electron-hole liquid in semiconductors and low-dimensional structures. Phys-Usp 60: 1147

Suzuki T, Shimano R (2012) Exciton Mott transition in Si revealed by terahertz spectroscopy. Phys Rev Lett 109: 046402

Thomas GA, Timofeev VB (1980) A review of $N = 1$ to ∞ particle complexes in semiconductors. In: Moss TS, Balkanski M (eds) Handbook of semiconductors, vol 2, Optical properties of solids. North Holland Publishing Company, Amsterdam, pp 45–63

Tikhodeev SG (1985) The electron–hole liquid in a semiconductor. Sov Phys Usp 28:1

Tsidilkovski IM, Harus GI, Shelushinina NG (1985) Impurity states and electron transport in gapless semiconductors. Adv Phys 34:43

van Vechten JA (1980) A simple man's view on the thermochemistry of semiconductors. In: Moss TS, Keller SP (eds) Handbook of semiconductors, vol 3, Materials properties and preparation. North Holland Publishing Company, Amsterdam, pp 1–111

Wang M, Debernardi A, Zhang W, Xu C, Yuan Y, Xie Y, Berencén Y, Prucnal S, Helm M, Zhou S (2020) Critical behavior of the insulator-to-metal transition in Te-hyperdoped Si. Phys Rev B 102: 085204

Wolfe JP, Hansen WL, Haller EE, Markiewicz RS, Kittel C, Jeffries CD (1975) Photograph of an electron–hole drop in germanium. Phys Rev Lett 34:1292

Zimmermann R (2007) Bose-Einstein condensation of excitons: promise and disappointment. In: Ivanov AL, Tikhodeev SG (eds) Problems of condensed matter physics. Oxford University Press, Oxford, pp 281–296

Carrier-Transport Equations

22

Karl W. Böer and Udo W. Pohl

Contents

1 Carriers in Semiconductors 934
1.1 Bloch Electrons and Holes 935
1.2 The Polaron 936
2 Conductivity and Mobility of Carriers 945
2.1 Electronic Conductivity 947
2.2 Electron Mobility 949
3 Currents and Electric Fields 951
3.1 Drift Current in an Electric Field 951
3.2 Diffusion Currents and Total Currents 954
3.3 Electrochemical Fields and Quasi-Fermi Levels 958
3.4 Carrier Distributions in External and Built-In Fields 961
4 The Boltzmann Equation 965
4.1 The Boltzmann Equation for Electrons 965
4.2 The Boltzmann Equation for Phonons 968
4.3 The Relaxation-Time Approximation 968
4.4 Carrier Scattering and Energy Relaxation 970
4.5 The Mobility Effective Mass 974
4.6 Momentum and Energy Relaxation 975
4.7 Phonon and Electron Drag 979
5 Summary 979
References 981

Abstract

The current in semiconductors is carried by electrons and holes. Their lattice polarization modifies the effective mass, expressed as a change to polarons. While

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_22

for large polarons the effect is small, semiconductors with narrow bands and large lattice polarization show a significant effect described by small polarons. The total current is composed of a drift and a diffusion current of electrons and holes. The drift current is determined by the electric field, and the energy obtained by carrier acceleration is given to the lattice by inelastic scattering, which opposes the energy gain, causing a constant carrier-drift velocity and Joule's heating. The diffusion current is proportional to the gradient of carrier density up to a limit given by the thermal velocity. Proportionality factor of both drift and diffusion currents is the carrier mobility, which is proportional to a relaxation time and inverse to the mobility effective mass. Currents are proportional to negative potential gradients, with the conductivity as the proportionality factor. In spatially inhomogeneous semiconductors, both an external field, impressed by an applied bias, and a built-in field, due to space-charge regions, exist. Only the external field causes carrier heating by shifting and deforming the carrier distribution from a Boltzmann distribution to a distorted distribution with more carriers at higher energies.

The Boltzmann equation permits a detailed analysis of the carrier transport and the carrier distribution, providing well-defined values for transport parameters such as relaxation times. The Boltzmann equation can be integrated in closed form only for a few special cases, but approximations for small applied fields provide the basis for investigating scattering processes; these can be divided into essentially elastic processes with mainly momentum exchange and, for carriers with sufficient accumulated energy, into inelastic scattering with energy relaxation.

Keywords

Bloch oscillation · Boltzmann transport equation · Built-in electric field · Carrier heating · Collision integral · Conductivity · Diffusion current · Drag · Drift current · Drift velocity · Effective mass · Einstein relation · Electron drag · Energy relaxation · Fröhlich coupling · Inelastic and elastic scattering · Joule's heating · Mean free path · Mobility · Mobility effective mass · Momentum relaxation · Phonon drag · Polaron · Polaron mass · Polaron self-energy · Quasi-Fermi level · Relaxation time · Relaxation-time approximation

1 Carriers in Semiconductors

Electrons and holes in the conduction and valence bands are quasi-free to move in space and energy and to accept energy from an external field. Carrier transport in semiconductors with nonideal periodicity[1] is subjected to scattering with a mean free path λ between scattering events. Such a scattering reduces the effective volume to a value of λ^3, in which there is coherence of the electron wave. Only within such a limited volume does any one electron experience lattice periodicity and is

[1] A crystal with nonideal lattice periodicity is a solid which contains lattice defects (e.g., impurities) and oscillatory motions (i.e., phonons).

nonlocalized. Scattering introduces a loss, i.e., a damping mechanism counteracting the energy gain from an electric field. All scattering events change the carrier momentum, i.e., its direction of motion. However, only some of them, the *inelastic scattering* events, *significantly* change the energy of the carrier. Usually, several *elastic* scattering events are followed by one *inelastic* event after the carrier has gained sufficient energy from the field to permit inelastic scattering, usually by generating optical phonons. A large variety of scattering events can be distinguished; they are listed in Sect. 4.4 and discussed in ▶ Chap. 23, "Carrier Scattering at Low Electric Fields". The sum of all of these determines the carrier motion, its *mobility,* which will be defined later.

First a quasiclassical picture will be used in this chapter to describe the basic elements of the carrier motion in a semiconductor. The quantum-mechanical part is incorporated by using an effective mass rather than the rest mass, i.e., by dealing with Bloch electrons or Bloch holes as quasiparticles. Before we discuss carrier transport, we need to specify in a more refined model what we mean by carriers in semiconductors.

The excitation of electrons from the valence band (e.g., by absorption of photons) creates a certain concentration n of electrons in the conduction bands. These electrons interact with lattice imperfections, such as phonons, impurities, or other deviations from an ideal periodicity. This interaction is termed a *scattering event.* The scattering tends to bring the electrons into *thermal equilibrium* with the lattice and, in doing so, to the lowest valley of the lowest conduction band.

1.1 Bloch Electrons and Holes

Near the bottom of the conduction band, the electron is described as a *Bloch electron*,[2] i.e., as an electron with an effective mass given by the curvature of $E(\mathbf{k})$, as defined in Sect. 2.2 of ▶ Chap. 6, "The Origin of Band Structure", Eq. 20.

In an analogous description, the hole is described as a Bloch-type quasiparticle, residing near the top of the uppermost valence bands, with an effective mass given by their curvatures – see Sect. 1.1.1 of ▶ Chap. 8, "Bands and Bandgaps in Solids".

The band picture, however, which is the basis for this discussion, results from a series of approximations listed in ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals". The most severe one is the adiabatic approximation, which limits the electron-phonon interaction. Within the band model, the Bloch electron interacts with a *static potential* of the nuclei, while only the electrons surrounding each nucleus are polarized dynamically. With sufficient coupling between the moving electron and the lattice (i.e., a sufficiently large coupling constant a_c – see Sect. 1.2.1, this is no longer justified. For these cases, therefore, the Bloch electron picture needs to be augmented. The corresponding quasiparticle derived from a higher approximation is the *polaron.*

[2]We are adopting here the picture of a localized electron. Such a localization can be justified in each scattering event. In this model, we use a gas-kinetic analogy with scattering cross-sections, e.g., for electron-phonon interaction. This is equivalent to a description of the interaction of delocalized electrons and phonons when calculating scattering rates.

1.2 The Polaron

The previous discussion of free carriers with an effective mass was based on an ideal periodic lattice. There are many reasons why a real crystal lattice shows deviations from this periodicity. The perturbations, which cause major changes in the carrier trajectories, can be described as local scattering centers discussed in ▶ Chap. 23, "Carrier Scattering at Low Electric Fields".

Another interaction that steadily *accompanies* the carrier throughout its entire motion,[3] however, is better incorporated in the effective-mass picture. It also has an influence on the scattering effectiveness and thus on the mobility, which will be discussed in this section. This interaction involves lattice polarization and causes deviations from the *Bloch electron* picture discussed above. Deviations are significant in lattices with a large coupling of carriers to the lattice, such as in ionic crystals (typically alkali halides), crystals with a strong ionic character and large bandgap energy (e.g., transition-metal oxides), or organic crystals. Also many typical semiconductors contain at least a fraction of ionic bonding, especially compound semiconductors with higher iconicity (e.g., nitrides or II-VI compounds). Moreover, a strong polarization of the lattice occurs also in the neighborhood of certain impurities or in highly disordered semiconductors. In all these cases the interaction of electrons with lattice polarization becomes important. Such an interaction can be static (for an electron at rest) or dynamic (accompanying a moving electron).

The interaction involves the part of the Hamiltonian not considered in the band theory when describing the Bloch electron. This adiabatic approximation neglects the interaction of electrons with the induced motion of lattice atoms. In order to include the motion due to the lattice polarization by the electron, several approaches can be taken, all of which relate to the coupling of the electron with the lattice. Such a coupling can be expressed as (Hayes and Stoneham 1984):

- Fröhlich coupling, i.e., via interaction with longitudinal optical phonons
- Deformation-potential coupling, i.e., via the electric field produced by the strain field in acoustic oscillations
- Piezoelectric coupling, i.e., via the electric field produced by acoustic phonons in piezoelectric semiconductors

The first coupling effect is dominant. It can be described as absorption or emission of virtual LO phonons by the electron, which thereby lowers its eigenstate – see, e.g., Comas and Mora-Ramos (1989).

Another approach to describe the interaction is a static description in which the electron induces a shift of the surrounding ions due to its own Coulomb field (Fig. 1a). The eigenstates of the electron in this Coulomb funnel are hydrogen-

[3]This interaction is also commonly described as an electron-phonon interaction, however, of a different kind than that responsible for scattering. As a lattice deformation relates to phonons, the interaction of electrons with the lattice causing a specific deformation can formally be described by continuously absorbing and emitting phonons (see the following sections).

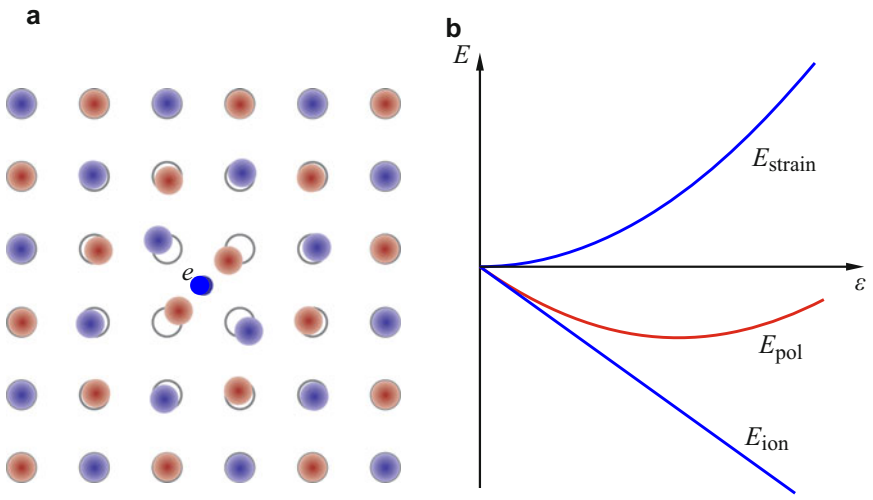


Fig. 1 (**a**) Schematic of the polarization cloud in the vicinity of an electron (e) in a ionic lattice. *Gray circles* indicate regular lattice sites. (**b**) Polaron energy E_{pol} given by the sum of elastic strain energy E_{strain} and electrical polarization energy E_{ion}

like, similar to that of a hydrogen-like donor (Sect. 1 of ▶ Chap. 18, "Shallow-Level Centers"); the ground state describes its *self-energy*. This lowers the energy of a Bloch electron accordingly.

The energy lowering can be estimated from a simple consideration. The strain of the lattice environment illustrated in Fig. 1a requires an elastic work

$$E_{strain} = (C_1/2)\,\varepsilon^2 \Delta V, \tag{1}$$

where ε describes the relative displacements of the ions within the deformation volume ΔV and the constant C_1 is the energy density for a strain $\varepsilon = 1$. This energy cost is overbalanced by an energy gain E_{ion} due to the polarization:

$$E_{ion} = -C_2\,\varepsilon\ \ \Delta V. \tag{2}$$

The constant C_2 is an energy density, which describes the change of the charge balance at a strain $\varepsilon = 1$. The sum of Eqs. 1 and 2 yields the net energy gain E_{pol} by the formation of the polaron state as shown in Fig. 1b:

$$E_{pol}/\Delta V = (C_1/2)\,\varepsilon^2 - C_2\,\varepsilon, \tag{3}$$

with a minimum $-C_2^2/2C_1$ at the strain $\varepsilon = C_2/C_1$.

For a sufficiently large coupling to the lattice, the self-energy is large enough so that the electron is actually trapped in its own potential well: it becomes a *self-trapped electron*. This was suggested by Landau (1933) and Frenkel (1936) for bulk crystals, and reviewed for alkali halides by Lushchik and Lushchik (2018). Self-trapped carriers were also reported for semiconductor confined systems such as quantum wires (e.g.,

Muljarov and Tikhodeev 1997). Pekar (1954) pointed out, however, that when the electron moves to a neighboring lattice position, the same trap level appears. Thus, one can describe this as a *continuous* virtual level or a band below the conduction band in which the electron can move. This picture applies for lattices with smaller coupling constant.

The state of the electron with its surrounding polarization cloud can then be described as a quasiparticle, the *polaron.* The distinguishing parameter between the self-trapped and the mobile polaron is the strength of the electron-lattice interaction, which also can be related to the polaron size. The *small polaron* is tightly bound and self-trapped: it moves via hopping between neighboring ions. The *large polaron* moves much like an electron described by the Boltzmann equation with scattering events, but with a larger effective mass caused by the polarization cloud carried along by the polaron (for a review, see Christov 1982). One can also describe the polaron as an electron surrounded by a cloud of virtual phonons, which represents its surrounding polarization. One refers to these polarons as *phonon-dressed electrons*, since attached to them is the "fabric" of the surrounding lattice. They can be described by a Fröhlich Hamiltonian, which explicitly accounts for the electron-phonon (LO phonons at $k = 0$) interaction (Fröhlich et al. 1950, advanced by Lee et al. 1953). Alternatively, it can be described with the Feynman Hamiltonian (*Feynman path integral,* Feynman 1955), which simulates the virtual phonons by a fictitious particle that interacts with the electron via a harmonic potential (Peeters and Devreese 1984). Large and small polarons were more recently also studied by a computational approach from first principles, employing electronic band structures, phonon dispersion relations, and electron-phonon matrix elements calculated in the crystal unit cell (Sio et al. 2019a, b).

The size of the polaron is measured by the extent of the lattice distortion caused by the electron. Large polarons extend substantially beyond nearest-neighbor distances, small polarons do not. For reviews, see Alexandrov and Devreese (2010), and Franchini et al. (2021).

1.2.1 Large Polarons and Fröhlich Coupling

In the conventional band model, the polarization of Bloch electrons is included by using the *optical* dielectric constant, which describes the interaction with the electrons of each lattice atom. In the discussion of polarons, their polarization also considers the shift in the position of the nuclei: the *static* dielectric constant is then used to account for the more intensive shielding of the Coulomb interaction:

$$\left(\frac{e^2}{8\pi\,\varepsilon_{\text{opt}}\varepsilon_0\,r}\right)_{\text{el}} \rightarrow \left(\frac{e^2}{8\pi\,\varepsilon_{\text{stat}}\varepsilon_0\,r}\right)_{\text{pol}} = \frac{e^2}{8\pi\,\varepsilon_0\,r}\left[\frac{1}{\varepsilon_{\text{opt}}} - \left(\frac{1}{\varepsilon_{\text{opt}}} - \frac{1}{\varepsilon_{\text{stat}}}\right)\right]. \tag{4}$$

The net difference unaccounted for in the Bloch electron picture describes the Coulomb energy of the polaron:

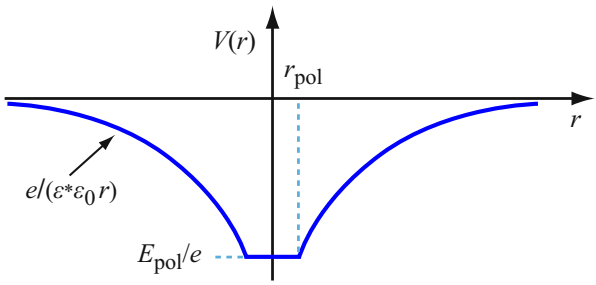


Fig. 2 Net potential distribution assigned to a polaron, indicating the polaron radius r_{pol}, below which the uncertainty relation precludes further extrapolation of the quasi-Coulomb potential

$$E_{\mathrm{pol,\,Coul}} = \frac{e^2}{8\pi\,\varepsilon^*\varepsilon_0\,r_{\mathrm{pol}}} \text{ with } \frac{1}{\varepsilon^*} = \frac{1}{\varepsilon_{\mathrm{opt}}} - \frac{1}{\varepsilon_{\mathrm{stat}}}. \tag{5}$$

The resulting net Coulomb potential of the polaron is shown in Fig. 2. Within this potential funnel, the *polaron self-energy* can be expressed as the 1*s* quasi-hydrogen ground-state energy

$$E_{\mathrm{pol}} = E_{\mathrm{qH}}\left(\frac{\varepsilon_{\mathrm{stat}}}{\varepsilon^*}\right)^2 \frac{m_{\mathrm{pol}}}{m_n}. \tag{6}$$

It is distinguished from the quasi-hydrogen ground-state energy by replacing $\varepsilon_{\mathrm{stat}}$ with ε^*, and the electron effective mass with that of the polaron.

The radius of the polaron can be defined as the corresponding 1*s* quasi-hydrogen radius

$$r_{\mathrm{pol}} = a_{\mathrm{qH}}\frac{\varepsilon^*}{\varepsilon_{\mathrm{stat}}}\frac{m_n}{m_{\mathrm{pol}}} = \frac{4\pi\,\varepsilon^*\varepsilon_0\,\hbar^2}{m_{\mathrm{pol}}\,e^2}. \tag{7}$$

When the polaron eigenfunctions are overlapping, the polaron self-energy (Eq. 6) broadens from a sharp level into a polaron band within which polarons can move through the crystal. The width of this band can be estimated from the uncertainty by absorbing or emitting virtual LO phonons as $\pm\hbar\omega_{\mathrm{LO}}$. The corresponding uncertainty radius of such a polaron is given by the uncertainty distance of finding a particle that interacted with an LO phonon: it has an energy $\hbar^2k^2/(2m_{\mathrm{pol}})$ with uncertainty $\pm\hbar\omega_{\mathrm{LO}}$. Therefore, it has an uncertainty in wavenumber of $\pm\sqrt{2m_{\mathrm{pol}}\omega_{\mathrm{LO}}}/\hbar$, the reciprocal of which is its corresponding uncertainty in position

$$r^*_{\mathrm{pol}} = \frac{\hbar}{\sqrt{2m_{\mathrm{pol}}\hbar\omega_{\mathrm{LO}}}}, \tag{8}$$

which is also used as the radius of a large polaron. For typical compound semiconductors, the polaron radius is on the order of several lattice constants.

The Fröhlich Coupling Constant For interaction with LO phonons, it is convenient to express the Fröhlich coupling in terms of the *coupling constant* α_{c}. It is given as the ratio of the Coulomb energy of a polaron, which describes the electron-phonon

interaction, to the energy of the LO phonon, i.e., the predominantly interacting phonon:

$$\alpha_c = \sqrt{\frac{\frac{1}{2}\left(\frac{e^2}{4\pi\,\varepsilon^*\varepsilon_0\,r_{pol}}\right)}{\hbar\omega_{LO}}}. \tag{9}$$

Entering the expression for r_{pol} Eq. 7 into Eq. 9 and replacing ε^* from Eq. 5, we obtain for the Fröhlich coupling constant α_c:

$$\alpha_c = \frac{e^2}{8\pi\,\varepsilon_0\,\hbar}\sqrt{\frac{2\,m_{pol}}{\hbar\omega_{LO}}}\left(\frac{1}{\varepsilon_{opt}} - \frac{1}{\varepsilon_{stat}}\right), \tag{10}$$

which is below unity for good semiconductors. Here α_c can be interpreted as twice the number of virtual phonons surrounding – that is, interacting with – a slowly moving carrier in the respective band.

Semiconductors without ionic bonding (Si, Ge, α-Sn) have $\varepsilon_{opt} = \varepsilon_{stat}$ and consequently $\alpha_c = 0$. Fröhlich coupling constants are therefore much larger for ionic than for covalent semiconductors and increase with increasing effective charge and decreasing strength of the lattice binding forces.[4] Some values of α_c for typical semiconductors are given in Table 1.

Polaron Energy and Effective Mass Using α_c, we can express the polaron energy as a fraction of the LO phonon energy[5]:

$$E_{pol} = E_c - \frac{p^2}{2m_{pol}} \cong -\left(\alpha_c + 0.01592\;\alpha_c^2 + \;\ldots\right)\;\hbar\omega_{LO}. \tag{11}$$

Assuming a small perturbation of the parabolic band, Lee et al. (1953) obtain for the energy dispersion within the band

$$E(k) = \frac{\hbar^2k^2}{2\,m_n} - \alpha_c\left(\hbar\omega_{LO} + \frac{\hbar^2k^2}{12\,m_n} + \;\ldots\right) \cong -\alpha_c\;\hbar\omega_{LO} + \frac{\hbar^2k^2}{2\,m_{pol}}, \tag{12}$$

which yields for the polaron mass

[4] A decrease of lattice-bonding forces corresponds also to a decreased Debye temperature; see also Sect. 1.1.2 of ▶ Chap. 5, "Phonon-Induced Thermal Properties".

[5] This result is obtained from Fröhlich et al. (1950) and with a variational method from Lee et al. (1953). It can be used up to $\alpha_c \gtrsim 1$. Earlier results from Pekar, using an adiabatic approximation, yielded $E_{pol} = -\left(\alpha_c^2/3\pi\right)\;\hbar\omega_{LO}$, which gives a lower self-energy than Eq. 11 in the range of validity $\alpha_c < 1$.

Table 1 Fröhlich coupling constants α_c of various solids

Solid	α_c	Solid	α_c	Solid	α_c	Solid	α_c
InSb	0.02	GaP	0.201	CdTe	0.35	ZnS	0.63
InAs	0.05	GaN	0.48	CdSe	0.46	ZnO	1.19
InP	0.15	AlSb	0.023	CdS	0.51		
InN	0.24	AlAs	0.126	CdO	0.74	KI	2.50
GaSb	0.025	AlP	0.49	ZnTe	0.33	KBr	3.05
GaAs	0.068	AlN	0.65	ZnSe	0.43	KCl	3.44

$$m_{\mathrm{pol,\,large}} \cong \frac{m_n}{1-\frac{\alpha_c}{6}}\,. \tag{13}$$

Equation 13 may be used for small α_c, i.e., a weak electron-phonon interaction. For larger values $\alpha_c \gg 1$, the following approximation is used for the polaron mass (Feynman 1955):

$$m_{\mathrm{pol,\,large}} \cong \frac{16\ \alpha_c^4\ m_n}{81\pi^4}. \tag{14}$$

Materials with large polarons that show a significant increase in the effective mass are silver halides. These have an intermediate coupling constant (1.60 and 1.91 for AgBr and AgCl, respectively).

The variational method of Lee et al. (1953) is used to compute intermediate coupling. Many II-VI and some III-V compounds have α_c values that make these large polarons sufficiently distinct from electrons (see Table 1 and Evrard 1984). Their effective mass can be determined by cyclotron resonance (Peeters and Devreese 1984). The Landau levels appearing in a magnetic field are shifted by $\Delta E \cong \alpha_c \hbar\omega_{\mathrm{LO}}$ for $\alpha_c \ll 1$; when the cyclotron resonance frequency approaches ω_{LO}, this level splits into two peaks indicating the strength of the electron-lattice interaction. For a more exact approximation using a path integral formulation, see Feynman (1955). For reviews, see Kartheuser et al. (1979), Bogoliubov and Bogoliubov Jr. (1986), and Devreese (1984).

In Table 2, the properties of large polarons for a number of crystals are listed. We note the inverse relation of r_{pol} and E_{pol}; for III-V semiconductors, the polaron mass is close to the electron rest-mass, while it increases for II-VI semiconductors and gets large for halides. Correspondingly m_n/m_{pol} reflects basically the effective electron mass in typical semiconductors.

In the presence of a magnetic field, the cyclotron behavior of polarons within the corresponding Landau levels is observed instead of the bare Bloch band mass of carriers. For a transition from Landau level n to $n+1$, Bajaj (1968) obtained for polarons

$$\omega_{c,\,\mathrm{pol}} = \omega_c\left(1-\frac{\alpha_c}{6}\right) - \frac{3}{20}\alpha_c\frac{\omega_c^2}{\omega_{\mathrm{LO}}}\left(\frac{\hbar k^2}{2\,m_n\,\omega_c}+n+1\right), \tag{15}$$

where ω_c is the cyclotron-resonance frequency of electrons (Eqs. 24 and 25 of ▶ Chap. 8, "Bands and Bandgaps in Solids").

Table 2 Polaron parameters of various solids

Solid	r_{pol} (Å)	E_{pol} (meV)	m_{pol}/m_0	m_n/m_{pol}
InSb	105	0.5	1.00	0.014
InAs	73.9	1.5	1.01	0.023
InP	33.8	5.2	1.02	0.078
GaSb	52.2	0.9	1.01	0.047
GaAs	38.5	2.6	1.01	0.068
GaP	21.1	6.5	1.02	0.175
CdTe	42.8	8.1	1.07	0.107
CdSe	33.1	12.2	1.08	0.14
CdS	25.3	20.2	1.10	0.126
ZnTe	30.5	8.4	1.06	0.169
ZnSe	27.0	13.1	1.08	0.183
ZnS	17.7	30.8	1.14	0.31
ZnO	14.8	64.8	1.18	0.276
KI	25.6	44.8	1.71	0.56
KBr	22.1	64.3	2.03	0.75
KCl	18.2	92.2	2.26	0.97

1.2.2 Small Polarons and Criteria for Different Polarons

Small polarons were introduced by Tjablikov (1952) and further analyzed by Holstein (1959). They are observed when the coupling constant is larger than 5. This strong interaction causes self-trapping: the wavefunction corresponds to a localized electron; the tight-binding approximation is appropriate for a mathematical description (Tjablikov 1952). Small polarons are distinguished from electrons by a rather large polaron self-energy E_{pol}. The *optical energy* necessary to bring a small polaron into the band, i.e., to free the self-trapped electron, was estimated by Pekar to be

$$E_{pol,opt,small} \cong 0.14 \ \alpha_c^2 \ \hbar\omega_{LO}. \tag{16}$$

Their *thermal ionization energy* $E_{pol,th,small}$ is considerably smaller, typically ~(1/3) × $E_{pol,opt,small}$, because of the strong lattice coupling; see Sects. 1.2.1 and 1.2.2 of ▶ Chap. 17, "Optical Properties of Defects". For further discussion of E_{pol}, see the review of Devreese (1984).

The effective mass of small polarons is given by (see Appel 1968):

$$m_{pol,small} \cong \frac{m_n \alpha_c^2}{48}. \tag{17}$$

Examples for materials with small polarons are narrow-band semiconductors with large α_c values (~10), e.g., transition-metal oxides such as NiO or the molecular crystals described in Sect. 3 of ▶ Chap. 6, "The Origin of Band Structure" with their particularly small band dispersion (see Sect. 4.3 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals"). Due to this narrow bandwidth, the band conductivity is usually disturbed by phonon scattering in organic crystals at room temperature, and hopping mechanisms may prevail even if the material has crystalline structure; see Sects. 4.3 and 5.1.2 of ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects".

Polarons are particularly well studied in TiO_2 and halide perovskites, which are interesting materials for photovoltaics and optoelectronics. In anatase TiO_2, shallow states below the conduction band are assigned to large electron polarons, while rutile TiO_2 is prone to the formation of small electron polarons. Both small and large polarons can also form in metal-halide perovskites, which are soft, polar materials with a strong driving force for polaron formation. The ABX_3 structure of halides is composed of a large mono- or divalent cation $A^{+/2+}$, a multivalent cation B^{n+} usually smaller than $A^{+/2+}$, and an anion X^{m-}. The ideal cubic structure has a 6-fold coordinated B atom forming BX_6 octahedra, surrounding 12-fold coordinated A cations; additionally, there are orthorhombic, tetragonal, and other structures (Ghosh et al. 2020). First-principles calculations indicate that short metal-halide bonds favor the formation of small polarons, as concluded from ionic substitution in compounds with A = CH_3NH_3 (= methylammonium MA) or Cs or HC $(NH_2)_2$ (= formamidinium FA), B = Pb or Sn, and X = I or Br (Neukirch et al. 2018; Mahata et al. 2019; Zhou et al. 2019). The binding energy of electron-polarons in iodine-based compounds gets larger as the A-site cation is changed from FA to Cs to MA. The local softness of the lattice in iodide compounds leads to smaller neutral reorganization energy than in bromide compounds. In the studied halides, electron polarons and hole polarons induce different structural deformations. Electron-polaron formation is related to Jahn-Teller-like distortions in the central octahedron, with apical PbI bonds expanding about 15-30% and equatorial PbI bonds expanding about 2%. In contrast, hole polarons cause the central octahedron to uniformly contract by about 5%. Modeling shows that local lattice symmetry, electronic structure, and electron-phonon coupling are interrelated in polaron formation. For a review, see Franchini et al. (2021).

A review for small polarons is given by Emin (1973); for a summary of polaron mobility due to various scattering mechanisms, see Appel (1968) and Evrard (1984); formulas to compute large-polaron mobilities are summarized by Frost (2017) and mobility calculations by Irvine et al. (2021). For a first-principles approach, see also Kang and Han (2018). In the strong-coupling limit, the scattering rates may exceed the thermal energy, $h\tau_{sc}^{-1} > (1/2)kT$; in this regime, the quasiparticle model of a Fröhlich polaron may be too simplified for describing the carrier mobility; see Mishchenko et al. (2019).

When the coupling to the lattice is strong enough so that the polarizing electron produces a significant Coulomb funnel, a second electron with opposite spin may be trapped within the same funnel (Chakraverty and Schlenker 1976). This *bipolaron* again can move through the lattice at an energy below that of a free electron and with an effective mass somewhat larger than that of two free electrons (Böttger and Bryksin 1985, Hai et al. 2018). For bipolarons in organic polymers, see Brazovskii et al. (1998). A significant difference, compared with two free electrons or two independent polarons, is the fact that the new quasiparticle *bipolaron* has zero spin and consequently acts as a boson. It is also referred to as a *Cooper pair*, and its formation is used to explain superconductivity; see Sect. 1.2 of ▶ Chap. 26, "Superconductivity".

Existence Criteria for Different Polarons There are three different energies, the relative magnitudes of which determine the preferred existence of large polarons, small polarons, or bipolarons. These are[6]

- The relaxation energy $E_{\text{relax}} = \alpha_c \ \hbar\omega_{\text{LO}}$
- The transfer energy $J = \hbar^2/(n_{\text{coord}}\, m^* a^2)$
- The Hubbard correlation energy $U = U_0 - \lambda^2/\beta$ (Sect. 2.8 of ▶ Chap. 19, "Deep-Level Centers")

Here n_{coord} is the coordination number, m^* is the effective mass, a is the nearest-neighbor distance, λ is an electron-lattice coupling constant, and β is an elastic restoring term.

We distinguish:

1. The competition between band formation and lattice relaxation
   (a) $E_{\text{relax}} \ll J$ large polaron
   (b) $E_{\text{relax}} \gg J$ small polaron
2. The competition between lattice relaxation and carrier correlation
   (a) $E_{\text{relax}} \ll U$ large polarons remain at separate sites.
   (b) $E_{\text{relax}} \gg U$ electrons prefer to share sites.
3. Sign change in carrier-correlation energy
   $U < 0$ bipolaron formation

For further discussion, see Toyozawa (1981), Franchini et al. (2021).

Electrons or Polarons in Semiconductors In a rigorous presentation, all electrons or holes near the band edges in equilibrium with the lattice should be replaced by polarons. Since for most semiconductors the difference between electrons and polarons is very small (Table 2), it is justified to proceed with the conventional description, using instead Bloch electrons and holes. Even in such semiconductors, this is no longer sufficient in the neighborhood of lattice defects with strong electron coupling. Here, the Huang-Rhys factor S, rather than the Fröhlich coupling constant α_c, is used, and the polaron picture is applied in order to obtain reasonable results for the defect and lattice relaxation in agreement with the experiment – see Sect. 1.2.2 of ▶ Chap. 17, "Optical Properties of Defects". Polaron effects are also significant in some amorphous semiconductors (Cohen et al. 1983, Zhang and Drabold 2010).

[6]More precisely, these three energies are: E_{relax}, the energy given to phonons during lattice relaxation; J, the bandwidth of a band created by free, uncoupled carriers of the given density; and U, the energy necessary to put two carriers with opposite spin on the same lattice site.

2 Conductivity and Mobility of Carriers

Carrier transport proceeds under external forces,[7] resulting in drift, and under internal quasiforces, resulting in diffusion. This may involve different charged particles which contribute additively to the current or have an indirect effect when it involves neutral particles, e.g., excitons (see Sect. 1.2 of ▶ Chap. 14, "Excitons"). Exciton diffusion plays a major role in devices fabricated from organic semiconductors; for a review see Mikhnenko et al. (2015).

Carrier transport occurs in bands near the band edges, i.e., near E_c for electrons and near E_v for holes. For materials with a sufficiently large defect density, carrier transport may proceed also via tunneling between trapping states. It may also involve carriers hopping from traps into the band or hopping of self-trapped small polarons. Trapped carriers travel a short distance in the band and later are recaptured, then reemitted, and so on; an analogous process occurs with small polarons – see ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects".

All of these processes add up to produce the total current and usually have vastly different magnitudes. Ordinarily, only one transport process predominates in homogeneous semiconductors. In nonhomogeneous materials, however, at least two and frequently four contributions are important in different regions of the devices. These are drift and diffusion currents of electrons and holes. First, a rather simple picture of the carrier transport is presented, which serves as guidance for a more sophisticated approach in later chapters.

At finite temperatures, carriers are found above the edge of the respective band according to their statistical distribution function. In semiconductors, they usually follow the Boltzmann distribution function within the bands when they are not degenerate, i.e., when the carrier densities are below $0.1N_c$ or $0.1N_v$ (Eqs. 16 and 20 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"). Their thermal velocity, the root mean square velocity,[8] is obtained from the equipartition principle, i.e., kinetic energy = ½kT per degree of freedom:

hence

$$\frac{m_n}{2}\langle v^2\rangle = \frac{3}{2}kT, \tag{18}$$

$$\sqrt{\langle v^2\rangle} = v_{\mathrm{rms}} = \sqrt{\frac{3kT}{m_n}}, \tag{19}$$

or

[7] Strictly speaking, steady-state carrier transport is due to external forces *only*. The diffusion current originates from a deformed carrier-density profile due to external forces and is a portion of the conventionally considered diffusion component. The major part of the diffusion is used to compensate the built-in field and has no part in the actual carrier transport: both drift and diffusion cancel each other and are caused by an artificial model consideration – see Sect. 3.4.

[8] The *rms* (*root mean square*) *velocity* v_{rms}, which is commonly used, should be distinguished from the slightly different *average velocity* $v_{\mathrm{av}} = \langle|v|\rangle$ and from the *most probable velocity* v_{mp}. Their ratios are $v_{\mathrm{rms}} : v_{\mathrm{av}} : v_{\mathrm{mp}} = \sqrt{3/2} : \sqrt{4/\pi} : 1 = 1.2247 : 1.1284 : 1$, as long as the carriers follow Boltzmann statistics. For a distinction between these different velocities, see Fig. 3.

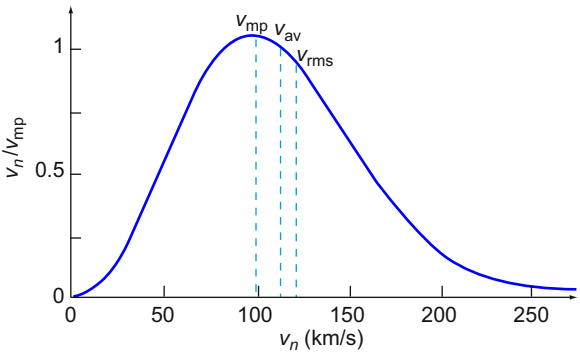


Fig. 3 Classical velocity distribution, with root mean square, average, and most probable velocities identified

$$v_{\text{rms}} = 1.18 \times 10^7 \sqrt{\frac{m_0}{m_n} \frac{T\ (\text{K})}{300}} (\text{cm/s}). \tag{20}$$

The velocity distribution is illustrated in Fig. 3. In thermal equilibrium and in an isotropic lattice, the motion of the carriers is random.

The quantum-mechanical model of a periodic potential teaches that, in contrast to a classical model, an ideal lattice is transparent for electrons or holes within their respective bands. That is, the carriers belong to the entire semiconductor and as waves are not localized: their position cannot be identified, except stating that n carriers (per cm^3) are within the given crystal. *There is no scattering of carriers within such an ideal crystal.* This behavior of carriers is unexpected in a classical model, which visualizes the filling of space with atomic spheres and expects only very limited possibilities for an electron traversing between these spheres without being scattered.

The introduction of phonons and crystal defects provides centers for scattering. In this way, a carrier motion results, which can be described as a *Brownian motion*[9] with a *mean free path* commensurate with the average distance between scattering centers. This distance is several hundred angstroms in typical crystalline semiconductors, i.e., the mean free path extends to distances much longer than the interatomic spacing.

When a carrier responds to an external field, it is accelerated in the direction of the electric field. Many important features of the carrier motion can be explained by assuming only inelastic scattering. Since it takes place at defects, which themselves are in thermal equilibrium with the lattice, the carrier tends to lose the excess energy gained between the scattering events. Figure 4 illustrates the typical motion of a carrier with and without an external electric field, assuming for both cases identical scattering events. The changes due to the field are exaggerated; under normal external fields, the changes from the random walk without field are very small perturbations barely being visible in the scale of Fig. 4.

[9]This motion resembles a random walk (Chandrasekhar 1943).

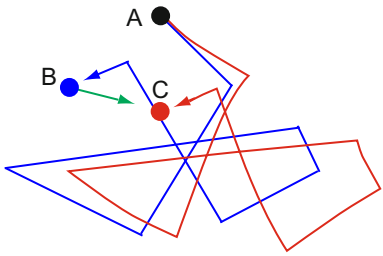


Fig. 4 Random walk of a carrier with (*red curves*) and without (*blue lines*) an external electric field. The *green arrow* from *B* to *C* indicates the relative displacement in field direction after the indicated nine scattering events

Sign Conventions In previous chapters, the elementary charge is used as $e = |e|$. When the *transport* of electrons and holes is discussed, it is instructive to discuss the proper signs: $-e$ for electrons and $+e$ for holes. This has an influence on derived parameters, e.g., the mobility, as will be discussed in Sect. 2.2.

The electric field[10] F is defined as the negative gradient of the vacuum level. The bias V is conventionally labeled + for the anode and – for the cathode, while the electrostatic potential ψ has the opposite signs: electrons have a larger potential energy at the cathode than at the anode. The relation to the field F is therefore

$$F = \frac{dV}{dx} = -\frac{d\psi}{dx}. \tag{21}$$

With $E_c = |e|\psi + \text{const}$, it yields a positive field when the band slopes downward from the cathode toward the anode, giving the visual impression that electrons "roll downhill" and holes "bubble up". We will use the proper signs in the following sections, however, reverting back to the commonly used $e = |e|$ later in order to avoid confusion in comparison to familiar descriptions.

When expressing forces, we need to distinguish the sign of the carrier; therefore, for an accelerating force we have

$$\text{Force} = (-eF)_n = (+eF)_p, \tag{22}$$

with subscripts n and p for electrons and holes.

2.1 Electronic Conductivity

In following the arguments introduced by Drude (1900; later refined by Lorentz 1909, and Sommerfeld 1928), electrons are accelerated in an electric field **F** by the force $-e\mathbf{F}$:

$$m_n \frac{d\mathbf{v}}{dt} = -e\,\mathbf{F}. \tag{23}$$

[10] In chapters dealing with carrier transport, F is chosen for the field, since E is used for the energy.

During a free path, the electron gains an incremental velocity, for an arbitrarily chosen field in x direction $\mathbf{F} = (F_x, 0, 0)$:

$$\Delta v_x = -\frac{e}{m_n} F_x \tau_{\mathrm{sc}}. \tag{24}$$

After averaging the *incremental velocity* between collisions and replacing the scattering time τ_{sc} with the average time $\overline{\tau}$ between scattering events, we obtain the *drift velocity* v_{d}:

$$v_{\mathrm{d}} = -\frac{e}{m_n} \overline{\tau} F_x. \tag{25}$$

With an electron density n and a charge $-e$, we obtain for the *current density* for electrons

$$j_n = e\, n\, v_{\mathrm{d}}, \tag{26}$$

or, introducing the *electron conductivity* σ_n,

$$j_n = \sigma_n F, \text{ with } \sigma_n = \frac{e^2}{m_n} \overline{\tau}_n n\,. \tag{27}$$

In a homogeneous semiconductor, the external field is given by the bias *V,* divided by the electrode distance *d,* yielding *Ohm's law*

$$j_n = \sigma_n F = \sigma_n \frac{V}{d} = \frac{V}{AR}, \tag{28}$$

with A as the area of the semiconductor normal to the current and the resistance

$$R = \frac{\rho_n d}{A} = \frac{d}{A \sigma_n}, \tag{29}$$

where ρ_n is the *specific resistivity* $\rho_n = 1/\sigma_n$.

Joule's Heating The additional energy gained by the accelerated electrons from the external electric field is delivered to the lattice during inelastic collisions, generating phonons. This *Joule's heating* is given by the power density p which n electrons transfer to the lattice while travelling at the average *constant* drift velocity v_{d} despite the accelerating force – eF; applying the sign convention used above we obtain

$$p = n\, e\, v_{\mathrm{d}} F = j_n F = \sigma_n F^2 = \frac{j_n^2}{\sigma_n}. \tag{30}$$

Bloch Oscillations When a Bloch electron is accelerated in an external field, it moves up relative to the conduction-band edge. When the band is narrow and the field is high enough, the electron can cross the center of the conduction band; as it approaches the upper band edge before scattering occurs, its effective mass becomes negative (Sect. 2.2 of ▶ Chap. 6, "The Origin of Band Structure"). It is consequently decelerated and finally reflected at the upper conduction-band edge. It hence moves down in the conduction band until the process repeats, and the Bloch electron undergoes an oscillatory motion, the *Bloch oscillation*. This was proposed early (Esaki and Tsu 1970), but can hardly be observed in bulk semiconductors, because conduction bands are usually too wide for an undisturbed motion high into the band before scattering occurs; *partial* Bloch oscillations in GaAs were demonstrated by optically induced high fields (Kuehn et al. 2010). In superlattices, however, minibands are narrow enough to permit the observation of Bloch oscillations; see, e.g., Shah (1994), Leo (1996), Waschke et al. (1994), Unuma et al. (2010), and Ivanov et al. (2015).

2.2 Electron Mobility

The quantity

$$\frac{e}{m_n}\overline{\tau}_n = \mu_n \tag{31}$$

is the *electron mobility*, since carriers are more *mobile* when they experience less scattering, i.e., the time between collisions is larger, and when their effective mass is smaller, i.e., they can be accelerated more easily. With $-e$ for electrons and $+e$ for holes, the mobility is negative for electrons and positive for holes, while the conductivity ($\propto e^2$, see Eq. 27) is always positive. Conventionally, however, $\mu_n = |\mu_n|$ is used, and we will follow this convention here.

The *electron conductivity* and *hole conductivity* are given by

$$\sigma_n = e\,\mu_n\,n \text{ and } \sigma_p = e\,\mu_p\,p. \tag{32}$$

Gas-Kinetic Model for Electron Scattering Different types of lattice defects are effective to a differing degree in carrier scattering. In a simple gas-kinetic model, *scattering centers* have a well-defined *scattering cross-section* s_n; a *scattering event*, i.e., a marked deflection[11] from an otherwise straight carrier path with exchange of momentum and/or energy, takes place when the carrier approaches the scattering center within its cross section. A *mean free path* λ_n can then be derived by constructing a cylinder of cross section s_n around an arbitrary straight carrier path and computing the average distance from the last scattering center to which this cylinder will extend until it incorporates the centerpoint of the next scattering center.

[11]Often a minimum scattering angle of 90° is used to distinguish scattering events with loss of memory from forward scattering events – see Sect. 4.6.1.

At this length the cylinder volume $\lambda_n\, s_n$ equals the average volume $1/N_{sc}$ (cm^3) that one of these centers occupies, where N_{sc} is the density of scattering centers; hence

$$\lambda_n = \frac{1}{s_n N_{sc}}. \tag{33}$$

Consequently, the time between scattering events is given by

$$\tau_{sc} = \frac{1}{v_{rms}\, s_n N_{sc}}. \tag{34}$$

This time is used to obtain an estimate for the carrier mobility in the Drude approximation:

$$\mu_n = \frac{e}{m_n}\tau_{sc} = \frac{e}{m_n}\frac{\lambda}{v_{rms}}, \tag{35}$$

or

$$\mu_n = 1.8 \times 10^{15}\frac{m_0}{m_n}\tau_{sc\ (s)}\left(cm^2/Vs\right), \tag{36}$$

or, using the expression of Eq. 19 for v_{rms}:

$$\mu_n = 1.5 \times \lambda(\text{Å})\ \left(\frac{m_0}{m_n}\right)^{3/2}\left(\frac{300}{T\ (K)}\right)^{1/2}\left(cm^2/Vs\right). \tag{37}$$

The dependence of the carrier mobility on the mean scattering time and mean free path is shown in Fig. 5. The application of this simple gas-kinetic model, however,

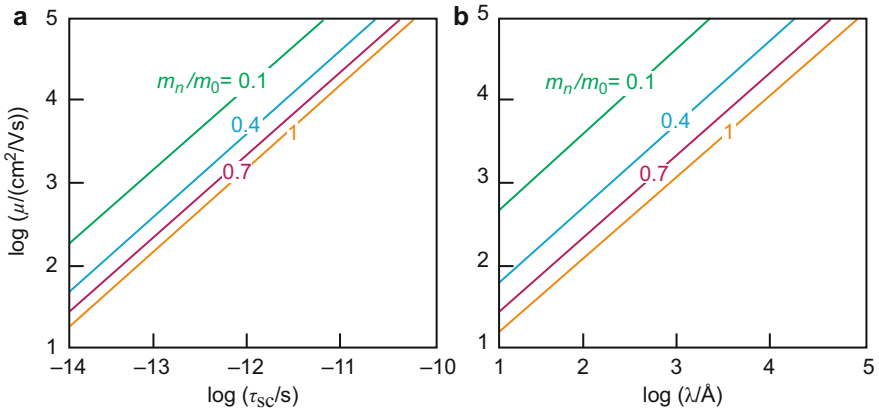


Fig. 5 Carrier mobility as a function of (**a**) the time between scattering events τ_{sc} and of (**b**) the mean free path λ, with the effective electron mass m_n as family parameter

has to be taken with caution because of its simplified assumptions. Generally, it yields too large densities of tolerable scattering centers.

3 Currents and Electric Fields

3.1 Drift Current in an Electric Field

The drift current is the product of the elementary charge, the carrier mobility, the single carrier density (derived in ▶ Chap. 21, "Equilibrium Statistics of Carriers"), and the electric field. For electrons or holes, it is

$$j_{n,\,\mathrm{drift}} = e n \mu_n F \text{ or } j_{p,\,\mathrm{drift}} = e p \mu_p F. \tag{38}$$

External Electric Field In *homogeneous semiconductors*, disregarding space-charge effects near interfaces, and for steady-state conditions – assumed with few exceptions throughout this book – the electric field is given by the applied voltage (*bias*) divided by the distance between the electrodes in a one-dimensional geometry; see Fig. 6:

$$F = \frac{V}{d}. \tag{39}$$

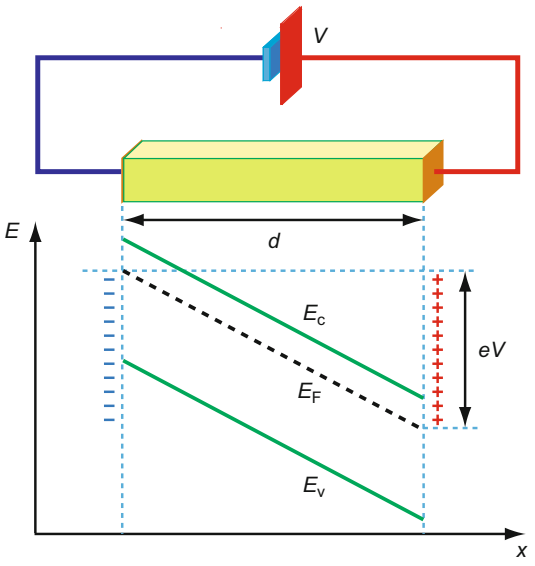


Fig. 6 Preferred quasi-one-dimensional geometry with band diagram subject to an external bias *V*, resulting in band tilting

For the field concept to apply, the distance d between electrodes must also be large compared to the interatomic spacing. The field can then be expressed by the macroscopic sloping of the bands[12]:

$$F = \frac{1}{e}\frac{dE_c}{dx} = \frac{1}{e}\frac{dE_v}{dx}. \tag{40}$$

It is also given, and more importantly so, by the slope of the Fermi potential (see Sect. 3.3) which, within the homogeneous material, is the same as the slope of the bands:

$$F = \frac{1}{e}\frac{dE_F}{dx}. \tag{41}$$

The bias is expressed as the difference of the Fermi levels between both electrodes; compare Fig. 6.

When using the electrostatic potential ψ_n with

$$-\frac{d\psi_n}{dx} = F, \tag{42}$$

the drift current can be expressed as a product of the electrical conductivity and the negative gradient of this potential:

$$j_{n,\text{drift}} = -\sigma_n \frac{d\psi_n}{dx} \quad \text{or} j_{p,\text{drift}} = -\sigma_p \frac{d\psi_p}{dx}. \tag{43}$$

For reasons to become apparent below, *two* electrostatic potentials are introduced: ψ_n and ψ_p for conduction and valence bands, respectively, with $e(\psi_n - \psi_p) = E_g$.

In a homogeneous semiconductor in steady state and with vanishing space charge, these drift currents are the total currents, and the slopes of both potentials are the same. There are special cases, however, in which the band edges of the valence and conduction bands are no longer parallel to each other. One of these will be mentioned briefly in the following section.

The Built-In Electric Field A semiconductor with a graded composition produces a position-dependent, varying bandgap energy. If this composition varies smoothly without steps, one or both bands are sloped without an applied bias, representing built-in fields – see Sect. 3.4. As an example, in $ZnSe_uS_{1-u}$, there is complete miscibility in the entire range ($0 \leq u \leq 1$), with the S-Se sublattice being a statistical alloy.[13] The bandgap

[12] As a reminder: here and in all following sections, $|e|$ is used when not explicitly stated differently.

[13] The atoms in the alloyed lattice (or here in the anion sublattice) are statistically arranged. Strictly speaking, this causes random fluctuation of the composition in a microscopic volume element of the crystal and results in a local fluctuation of the bandgap energy due to a locally varying parameter x; as a result, extended or localized states with energies $E_{c,v}(x)$ close to the band edge $E_{c,v}(\bar{x})$ of the mean composition $\bar{x}$ are formed, leading to some tailing of the band edges (Sect. 3.1 of ▶ Chap. 17, "Optical Properties of Defects"). The composition parameter u used in the text actually refers to the mean composition, and the effect of band tailing is neglected here.

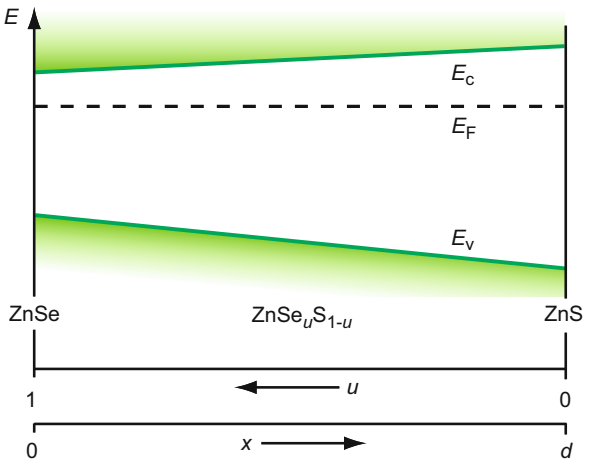


Fig. 7 Band diagram for a $ZnSe_uS_{1-u}$ mixed crystal with linearly varying composition u along the x axis. At $x = 0$, the material is ZnSe; at $x = d$, the material is ZnS

energy changes linearly[14] from 2.45 eV for ZnSe at the left side of the crystal shown in Fig. 7 to 3.6 eV for ZnS at its right side – see also Fig. 24 of ▶ Chap. 8, "Bands and Bandgaps in Solids". ZnSe and ZnS are both *n*-type materials. Depending on doping, the Fermi level in ZnSe can be shifted easily between 0.8 and 0.2 e V below E_c and in ZnS between 1.0 and 0.4 eV. Depending on the doping profile in the mixed-composition region, a wide variety of relative slopes, including nonmonotonic slopes, of valence and conduction bands can be designed for a vanishing bias, i. e., for a horizontal Fermi energy E_F. In Fig. 7, an example with opposite and linear sloping of $E_c(x)$ and $E_v(x)$ is shown, resulting effectively in a built-in field (see Sect. 3.4) of opposite sign for electrons and holes. In thermodynamic equilibrium, however, there is no net current in spite of the sloping bands. This is accomplished by exact compensation of finite drift currents with opposing diffusion currents, which self-consistently determine the slopes of the bands. An example for the application of such a graded $Al_uGa_{1-u}As$ composition in the base of heterojunction bipolar transistors is given in Horio et al. (1999). Other examples are GRINSCH lasers with a graded-index separate confinement heterostructure (Kwak et al. 2019), and solar cells with graded-bandgap absorber layer (Yang et al. 2016, Panahi et al. 2021).

The change in the bandgap energy can be expressed as

$$E_c(x) = E_v(x) + E_{g0} + \Delta E_g(x), \tag{44}$$

[14]Major deviations from linearity of E_g with composition are observed when the conduction-band minimum lies at a different point in the Brillouin zone for the two end members. One example is the alloy of Ge and Si. Other deviations (bowing – see Sect. 2.1 of ▶ Chap. 8, "Bands and Bandgaps in Solids") are observed when the alloying atoms are of substantially different size and electronegativity.

where x is a spatial coordinate. Using a conventional asymmetry factor A_E, which measures the fraction of the bandgap change $\Delta E_g(x)$ occurring in the conduction band relative to the horizontal Fermi level, Eq. 44 is also written

$$E_c(x) = E_c(x=0) + A_E \Delta E_g(x) = e\psi_n(x) \tag{45}$$

and

$$E_v(x) = E_v(x=0) - (1 - A_E)\,\Delta E_g(x) = e\psi_p(x). \tag{46}$$

The corresponding built-in fields for electrons and holes are given by

$$F_n = -A_E \frac{\partial \Delta E_g(x)}{e\partial x} = -\frac{\partial \psi_n}{\partial x} \tag{47}$$

and

$$F_p = -(A_E - 1)\frac{\partial \Delta E_g(x)}{e\partial x} = -\frac{\partial \psi_p}{\partial x}, \tag{48}$$

justifying the introduction of separate electrostatic potentials for electrons and holes: with $\Delta E_g(x) \neq 0$ and $A_E \neq 1/2$, we have $F_n \neq F_p$.

3.2 Diffusion Currents and Total Currents

Carrier diffusion by itself can be observed when the external field vanishes and a concentration gradient exists. An example in which these conditions are approximately fulfilled is the diffusion of minority carriers created by an inhomogeneous optical excitation (see, e.g., Najafi et al. 2015). The *diffusion current* is proportional to the diffusion coefficient D and to the carrier-density gradient; for electrons or holes, it is

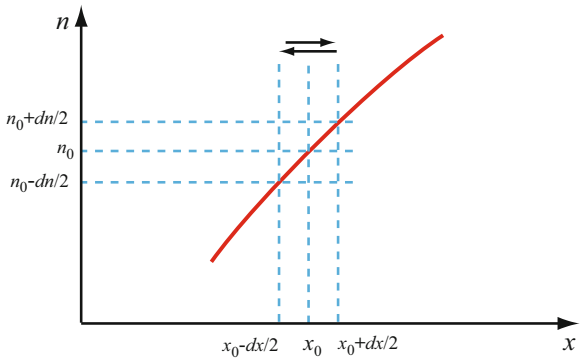


Fig. 8 Illustration of the derivation of the diffusion current in a medium with inhomogeneous carrier density $n(x)$. *Arrows* indicate diffusion currents

$$j_{n,\text{diff}} = eD_n \frac{dn}{dx} \quad \text{or} \quad j_{p,\text{diff}} = -eD_p \frac{dp}{dx}. \tag{49}$$

The negative sign of the hole current is due to the fact that in both equations $\pm e = |e|$ is used. The diffusion current can be derived as the difference between two currents caused by a completely random motion of carriers originating in adjacent slabs with slightly different carrier densities (Fig. 8). The current originating at $x_0 + dx/2$ and crossing the interface at x_0 from right to left, is caused by the Brownian motion of electrons of a density $n_0 + dn/2$. It is given by

$$\overleftarrow{j}_{n,\text{diff}} = e\left(n_0 + \frac{dn}{2}\right)\frac{v_{\text{rms}}^2}{3}\frac{\tau_n}{dx}. \tag{50}$$

The current crossing the boundary from left to right is accordingly given by

$$\overrightarrow{j}_{n,\text{diff}} = e\left(n_0 - \frac{dn}{2}\right)\frac{v_{\text{rms}}^2}{3}\frac{\tau_n}{dx}. \tag{51}$$

The current is proportional to the carrier velocity v_{rms} and the carrier mean free path λ_n. In turn, λ_n is given by $v_{\text{rms}}\tau_n$. The factor $\frac{1}{3}$ arises from gas-kinetic arguments when the root mean square velocity is obtained from an isotropic velocity distribution: $\overline{v^2} = v_{\text{rms}}^2 = v_x^2 + v_y^2 + v_z^2$; with $v_x^2 = v_y^2 = v_z^2$, we obtain the relation $v_x^2 = \frac{1}{3}v_{\text{rms}}^2$ for the x-component used in Eq. 50.

The difference of both currents (Eqs. 50 and 51) is the net diffusion current

$$j_{n,\text{diff}} = \overrightarrow{j}_{n,\text{diff}} - \overleftarrow{j}_{n,\text{diff}} = e\,\frac{v_{\text{rms}}^2\tau_n}{3}\frac{dn}{dx}, \tag{52}$$

with the diffusion coefficient given by

$$D_n = \frac{v_{\text{rms}}^2\tau_n}{3}. \tag{53}$$

By using $v_{\text{rms}}^2 = 3kT/m_n$ (Eq. 19), we obtain the more commonly used equation for the diffusion current

$$j_{n,\text{diff}} = \mu_n kT\frac{dn}{dx} \text{ and } j_{p,\text{diff}} = -\mu_p kT\frac{dp}{dx}. \tag{54}$$

Both diffusion currents for electrons and holes have the same negative sign for a positive gradient of $n(x)$ or $p(x)$ when recognizing that μ_n is negative and μ_p is positive. However, since the conventional notation with $\mu_n = |\mu_n|$ is used, the difference in signs appears.

Maximum Diffusion Currents As the gradient of the carrier density increases, the diffusion current increases proportionally to it (Eq. 49). However, this proportionality is limited, when the density gradient becomes so steep that the reverse current (Eq. 51) becomes negligible compared to the forward current (Eq. 50).

When increasing the distance dx to the mean free path λ_n, we obtain from Eq. 52 with $\vec{j}_{n,\mathrm{diff}} \ll \overleftarrow{j}_{n,\mathrm{diff}}$ for the maximum possible diffusion current through a planar surface

$$j_{n,\mathrm{diff,max}} = e\, n_0\, \frac{v_{\mathrm{rms}}^2}{3}\, \frac{\tau_n}{\lambda_n}, \tag{55}$$

or, for carriers following Boltzmann statistics and within a device with planar geometry,

$$j_{n,\mathrm{diff,max}} = \frac{e\, n}{\sqrt{6\pi}}\, v_{\mathrm{rms}}. \tag{56}$$

This current is known as the *Richardson-Dushman current* (Dushman 1930). It is equal to the *thermionic emission current* into the vacuum if the semiconductor is cut open at x_0 (Fig. 8) and if a vanishing work function is assumed – that is, if all electrons in the conduction band at x_0, with a velocity component toward the surface, could exit into the vacuum (see also Simoen et al. 1998).

The Einstein Relation Comparing the diffusion equations, Eq. 49 with Eq. 54, we obtain a relation between the diffusion constant and the carrier mobility:

$$D_{n,p} = \frac{\mu_{n,p}\, kT}{e}, \tag{57}$$

which is known as the *Einstein relation* and holds for systems that follow Boltzmann statistics. This can be seen from the following arguments. In thermal equilibrium, the total current, as well as each carrier current, vanishes: $j \equiv j_n \equiv j_p \equiv 0$. The electron current is composed of drift and diffusion currents – see Eq. 61; hence,

$$\mu_n n F + D_n \frac{dn}{dx} = 0, \tag{58}$$

which can be integrated to yield

$$n(x_0 + \Delta x) = n(x_0)\exp(-\mu_n F \Delta x / D_n). \tag{59}$$

On the other hand, electrons obey the Boltzmann distribution in equilibrium in the conduction band of a semiconductor. Their surplus energy, obtained in an electric field at a distance Δx, is $\Delta E = e\Delta V = e\, F\Delta x$ (Fig. 9), yielding a density

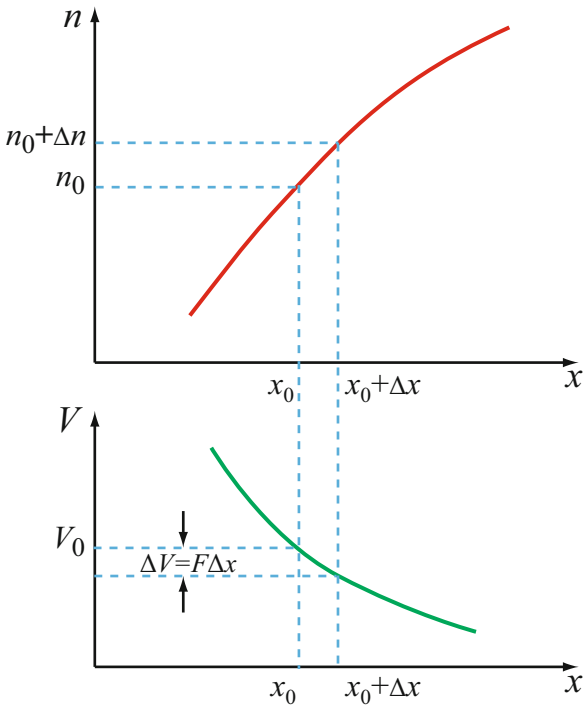


Fig. 9 Electron density and electrostatic potential distribution in the Boltzmann region in thermal equilibrium (schematic)

$$n(x_0 + \Delta x) = n(x_0)\exp\left(-\frac{eF\Delta x}{kT}\right). \tag{60}$$

A comparison of the exponents in Eqs. 59 and 60 yields the Einstein relation (Eq. 57). In case the Boltzmann distribution is not fulfilled (degeneracy), a generalization of Eq. 57 can be derived involving Fermi integrals (Landsberg 1952).

The assumptions used beyond the Boltzmann distribution are that of a one-carrier model near equilibrium and that the total current is small compared to drift and diffusion currents; hence, Eq. 58 holds. At high fields, one or more of these conditions are no longer fulfilled. Consequently, the Einstein relation needs to be modified – see Sect. 1 of ▶ Chap. 24, "Carrier Scattering at High Electric Fields", and Kan et al. (1991). For nonparabolic bands, see Landsberg and Cheng (1985). For hot-carrier diffusion at low temperature in quantum wells, see Chattopadhyay et al. (1989). The diffusion of hot carriers is an important process in hot-carrier solar cells, where multiple quantum wells can reduce the cooling rate and increase the hot-carrier lifetime (Conibeer et al. 2017); this allows for the extraction of carriers with excess energy (Kahmann and Loi 2019).

Total Currents The *total current* is given as the sum of drift and diffusion currents. For electrons, we have

$$j_n = j_{n,\text{drift}} + j_{n,\text{diff}} = e\mu_n nF + D_n \frac{dn}{dx}, \tag{61}$$

and for holes

$$j_p = j_{p,\text{drift}} + j_{p,\text{diff}} = e\mu_p pF - D_p \frac{dp}{dx}. \tag{62}$$

The *total carrier current* is the sum of both,

$$j = j_n + j_p. \tag{63}$$

In homogeneous semiconductors, only one of the four components is usually predominant, while in a *pn* junction with sufficient bias, each one becomes predominant within a different region.

3.3 Electrochemical Fields and Quasi-Fermi Levels

For evaluating the total current, we need to consider the gradient of $\psi(x)$ for the drift component and the gradient of $n(x)$ for the diffusion component. It is instructive to deduce the electron and hole current for thermal equilibrium. In thermal equilibrium, n is given by the Fermi distribution (Eq. 10 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"). When the Fermi energy is separated by several kT from the band edge, we can disregard the 1 in the denominator of Eq. 10 in ▶ Chap. 21, "Equilibrium Statistics of Carriers" and approximate this equation with the Boltzmann distribution, yielding

$$E_c(x) - E_F(x) = kT \ln \frac{N_c}{n(x)}. \tag{64}$$

Replacing $E_c(x)$ with $-e\psi(x) + c$ and differentiating both sides of Eq. 64 with respect to x, we obtain after division by e

$$-\frac{d\psi_n}{dx} - \frac{1}{e}\frac{dE_F}{dx} = -\frac{kT}{e}\frac{1}{n}\frac{dn}{dx}. \tag{65}$$

After multiplying both sides with $\sigma_n = e\,\mu_n n$ and rearranging, we obtain

$$\sigma_n \frac{1}{e}\frac{dE_F}{dx} = -\sigma_n \frac{d\psi_n}{dx} + \mu_n kT \frac{dn}{dx}. \tag{66}$$

The right-hand side is the total electron current; thus, the left-hand side must also be equal to j_n:

$$j_n = -\sigma_n \frac{1}{e} \frac{dE_F}{dx}. \tag{67}$$

Since in thermal equilibrium, i.e., with vanishing external field, the Fermi level must be constant (horizontal in an $E(x)$ presentation), we conclude that the electron *and* hole current must vanish separately in equilibrium:

$$\frac{dE_F}{dx} \equiv 0 \quad \Rightarrow \quad j_n \equiv j_p \equiv 0. \tag{68}$$

Quasi-Fermi Levels In *steady state*, e.g., with a constant external excitation (▶ Chap. 29, "Carrier Generation"), the electron and hole densities deviate from thermodynamic equilibrium values. Nevertheless, we may use the Fermi distribution to describe their density in the bands, using the *quasi-Fermi levels* E_{Fn} and E_{Fp} according to the definition equations

$$n \equiv N_c \frac{1}{\exp\left(\frac{E_c - E_{Fn}}{kT}\right) + 1}, \tag{69}$$

and

$$p \equiv N_v \frac{1}{\exp\left(\frac{E_{Fp} - E_v}{kT}\right) + 1}, \tag{70}$$

with $E_{Fn} \neq E_{Fp}$. This rather useful approximation introduces errors which may or may not be acceptable depending on the cause for deviation from the thermal equilibrium. In general, the error is quite small for optical excitation and for low external fields. For high external fields, the distribution function is substantially deformed (see ▶ Chap. 24, "Carrier Scattering at High Electric Fields"), and a more sophisticated approximation is required.

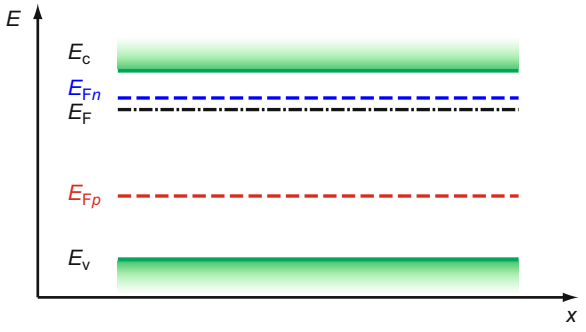


Fig. 10 Band model with external excitation resulting in a split of the Fermi level E_F into two quasi-Fermi levels E_{Fn} and E_{Fp}. The scheme represents the zero-field case

With optical or field-induced carrier generation discussed in ▶ Chap. 29, "Carrier Generation", n and p are increased above their thermodynamic equilibrium value; hence, $E_{Fp} < E_F < E_{Fn}$, resulting in a decreased distance of both quasi-Fermi levels from their corresponding bands; see Fig. 10. In certain cases, the recombination may be increased above the equilibrium value, as, for instance, in a *pn* junction in reverse bias; here, E_{Fn} can drop below E_{Fp}.

In using the same algebraic procedure as described in the previous section, we have for the total electron current in steady state:

$$j_n = -\sigma_n \frac{1}{e} \frac{dE_{Fn}}{dx}; \tag{71}$$

this means, the *total electron current* is proportional to the negative slope of the quasi-Fermi potential, as the drift current is proportional to the negative slope of the electrostatic potential (Eq. 43). For both currents, the conductivity is the proportionality constant.

In order to emphasize this similarity, we define the *electrochemical potentials* for electrons and holes:

$$\varphi_n = \frac{1}{e} E_{Fn} \text{ and } \varphi_p = \frac{1}{e} E_{Fp}. \tag{72}$$

The total currents can now be expressed as

$$j_n = -\sigma_n \frac{\partial \varphi_n}{\partial x} \text{ and } j_p = -\sigma_p \frac{\partial \varphi_p}{\partial x}. \tag{73}$$

For homogeneous semiconductors with homogeneous generation of carriers, these currents become the drift currents, and Eq. 73 becomes equal to Eq. 43.

In steady state, the total current is divergence-free, i.e., $j_n + j_p = \text{const}$ (for included carrier generation see ▶ Chap. 31, "Photoconductivity"). Therefore,

$$\sigma_n \frac{\partial \varphi_n}{\partial x} + \sigma_p \frac{\partial \varphi_p}{\partial x} \equiv \text{const} \tag{74}$$

or

$$\mu_n n(x) \frac{\partial \varphi_n}{\partial x} + \mu_p p(x) \frac{\partial \varphi_p}{\partial x} \equiv \text{const}. \tag{75}$$

Since a semiconductor is predominantly either n- or p-type, except for the inner part of a junction, we usually can neglect one part of the sum. For example, for the n-type region,

$$n(x) \frac{\partial \varphi_n}{\partial x} \equiv \text{const}; \tag{76}$$

i.e., if there is a gradient in the carrier density, then the highest slope in $\varphi_n(x)$ is expected where the carrier density is lowest for an inhomogeneous $n(x)$ distribution.

From Eq. 75, we also conclude that for vanishing currents in steady state, the slopes of the quasi-Fermi potentials must be opposite to each other. The lower the corresponding carrier densities, the higher the slopes:

$$\mu_n n(x) \frac{\partial \varphi_n}{\partial x} = -\mu_p p(x) \frac{\partial \varphi_p}{\partial x}. \tag{77}$$

Summary: Potential Gradients and Currents The various currents in a semiconductor can be expressed in a similar fashion. They are proportional to the negative gradient of electrostatic or electrochemical potentials with the conductivity as a proportionality factor:

$$\begin{aligned} j_n &= -\sigma_n \frac{\partial \varphi_n}{\partial x}, j_p = -\sigma_p \frac{\partial \varphi_p}{\partial x}, \\ j_{n,\,\mathrm{drift}} &= -\sigma_n \frac{\partial \psi_n}{\partial x}, j_{p,\,\mathrm{drift}} = -\sigma_p \frac{\partial \psi_p}{\partial x}, \\ j_{n,\,\mathrm{diff}} &= -\sigma_n \frac{\partial (\varphi_n - \psi_n)}{\partial x}, j_{p,\,\mathrm{diff}} = -\sigma_p \frac{\partial (\varphi_p - \psi_p)}{\partial x}. \end{aligned} \tag{78}$$

3.4 Carrier Distributions in External and Built-In Fields

An external bias resulting in a surface charge on the two electrodes leads to an *external field* with no space charge within the semiconductor (Fig. 6). However, if there are inhomogeneities in the distribution of charged donors or acceptors[15] or spatially inhomogeneous compositions in an alloy semiconductor (Fig. 7), regions with a *space-charge density* ρ exist within the semiconductor. This charge density causes the development of an *internal field* F_{int} according to the Poisson equation

$$\frac{dF_{\mathrm{int}}}{dx} = \frac{\rho}{\varepsilon \varepsilon_0}. \tag{79}$$

The acting resulting field is the sum of both internal and external fields,

[15] *pn* junctions are the best studied intentional space-charge regions. Inhomogeneous doping distributions – especially near surfaces, contacts, or other crystal inhomogeneities – are often unintentional and hard to eliminate.

$$F = F_{\text{int}} + F_{\text{ext}}. \tag{80}$$

An external and internal field of equal magnitude results in the same slope of the bands. Therefore, this distinction between internal and external fields is usually not made, and the subscripts at the fields are omitted. There are, however, drawbacks in such a general description of fields, which can best be seen from carrier heating in an electric field. *Carrier heating* is used to describe the field dependence of the mobility (see Sects. 2 and 3 of ▶ Chap. 24, "Carrier Scattering at High Electric Fields") in a microscopic model: accelerated carriers are shifted up to higher energies within a band; consequently, their effective mass and the scattering probability change. Usually the effective mass increases and the creation of phonons becomes easier. The mobility hence becomes field dependent and usually decreases with increasing field.

Such carrier heating *is absent in thermal equilibrium:* the carrier gas and the lattice with its phonon spectrum are in equilibrium within each volume element; thus, carrier and lattice temperatures remain the same (Stratton 1969). No energy can be extracted from an internal field, i.e., from a sloped band, due to a space charge *in equilibrium.*[16] This situation may be illustrated with an example replacing electrical with gravitational forces: a sloping band due to a space-charge region looks much like a mountain introduced on top of a sea-level plane, the Fermi level being equivalent to the sea level. As the introduction of the mountain does little to the distribution of molecules in air, the introduction of a sloping band does little to the

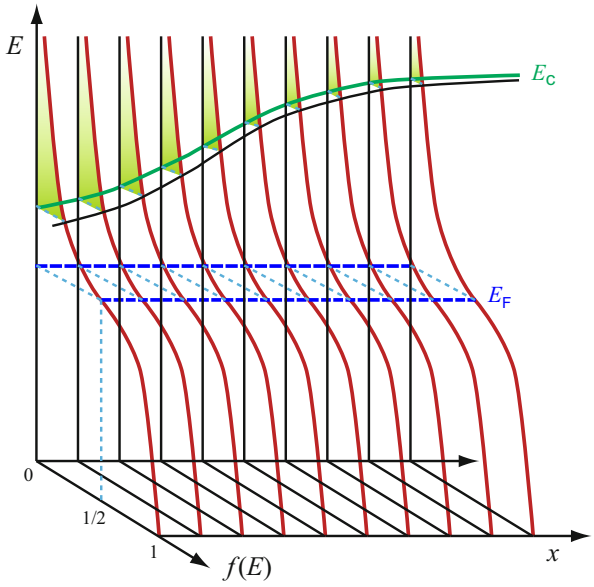


Fig. 11 Fermi distribution for different positions in a semiconductor at zero-applied bias with a built-in field region due to a junction or a gradient in composition (After Böer 1985a)

[16]This argument no longer holds with a bias, which will modify the space charge; partial heating occurs, proportional to the fraction of external field. This heating can be related to the tilting of the quasi-Fermi levels.

distribution of electrons in the conduction band. Since there are fewer molecules above the mountain, the air pressure is reduced, just as there are fewer electrons in a band where it has a larger distance from the Fermi level as illustrated in Fig. 11.

However, when one wants to conveniently integrate over all altitudes (energies) in order to arrive at a single number – the air pressure (or electron density) – one must consider additional model consequences to prevent winds from blowing from the valleys with high pressure to the mountain top with low pressure by following only the pressure gradient. Neither should one expect a current of electrons from the regions of a semiconductor with the conduction band close to the Fermi level, which results in a high electron density, to a region with low electron density in the absence of an *external field.* To prevent such currents in the electron-density model, one uses the *internal fields*, i.e., the built-in fields, and balances the diffusion current with an exactly compensating drift current. The advantage of this approach is the use of a simple carrier density and a simple transport equation. The penalty is the need for some careful definitions of transport parameters, e.g., the mobility, when comparing external with built-in fields and evaluating the ensuing drift and diffusion currents when the external fields are strong enough to cause carrier heating.

Carrier Concentrations in Built-In or External Fields The carrier distribution and mobility are different in built-in or external fields. The carrier distribution is determined relative to the Fermi level. For vanishing bias, the distribution does not depend on the position; the Fermi level is constant (horizontal). The distribution remains unchanged when a junction with its built-in field is introduced.[17] The sloping bands cut out varying amounts from the lower part of the distribution, much like a mountain displaces its volume of air molecules at lower altitudes; see Fig. 11. The carrier concentration n becomes space dependent through the space dependence of the lower integration boundary, while the energy distribution of the carriers $n(E)$ remains independent in space:

$$n(x) = \int_{E_c(x)}^{\infty} n(E)\ dE. \tag{81}$$

This is similar to the velocity distribution of air molecules, which is the same at any given altitude, whether over a mountain or an adjacent plane; whereas the integrated number, i.e., the air pressure near the surface of the sloping terrain, is not. This does not cause any macroscopic air motion, since at any stratum of constant altitude, the molecular distribution is the same; hence, the molecular motion remains totally random.

[17]With bias, the Fermi level in a junction is split into two quasi-Fermi levels which are tilted, however, with space-dependent slope. Regions of high slope within the junction region will become preferentially heated. The formation of such regions depends on the change of the carrier distribution with bias and its contribution to the electrochemical potential (the quasi-Fermi level). Integration of transport, Poisson, and continuity equations yields a quantitative description of this behavior (Böer 1985b).

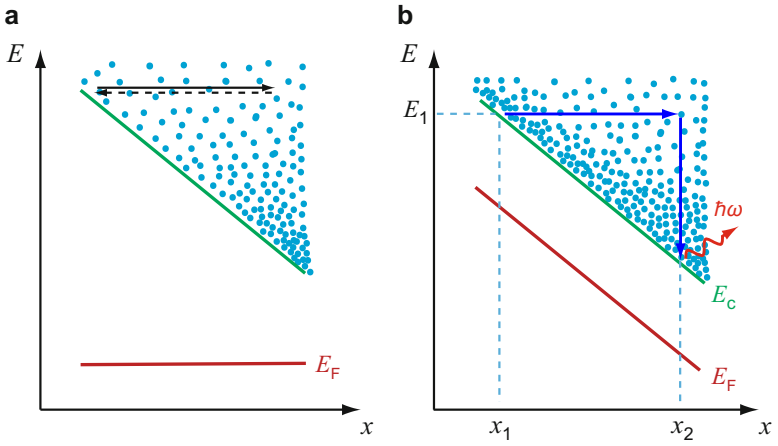


Fig. 12 Sloping band due to (**a**) an internal (built-in) field with horizontal Fermi level E_F and (**b**) due to an external field with parallel sloping of both, bands and Fermi level. The electron distribution is indicated by a *dot* distribution, and the action of field and scattering by *arrows* (After Böer 1985a)

In a similar fashion, electrons at the same distance above the Fermi level are surrounded by strata of constant electron density; within such strata their motion must remain random. During scattering in thermal equilibrium, the same amounts of phonons are generated as are absorbed by electrons, except for statistical fluctuations: on the average, all events are randomized. Electron and hole currents both vanish in equilibrium *for every volume element.* Figure 12a illustrates such a behavior.

In an *external field*, however, bands *and Fermi level* are tilted parallel to each other; this means, with applied bias, the carrier distribution becomes a function of the spatial coordinate as illustrated in Fig. 12b. When electrons are accelerated in the field, they move from a region of higher density $n(E_1 - E_F)_{x_1}$ to a region of lower density $n(E_1 - E_F)_{x_2}$. These electrons can dissipate their net additional energy to the lattice by emitting phonons and causing lattice (Joule's) heating. In addition, while in net motion, electrons fill higher states of the energy distribution, thereby causing the carrier temperature to increase. The carrier motion in an *external field* is therefore no longer random; it has a finite component in field direction: the *drift velocity* $v_d = \mu F_{ext}$ and the collisions with lattice defects are at least partially inelastic. A net current and lattice heating result.

Field Dependence of Mobilities At higher fields the *carrier mobility* becomes field dependent. The difference between the built-in and the external fields relates to the influence of carrier heating on the mobility, since the *averaging process* for determining the mobility uses the corresponding distribution functions. For instance, with an electric field in the x direction, one obtains for the drift velocity of electrons

$$v_d = \mu_n F_x = \overline{v}_x = \frac{\int v_x f(v)\, g(v)\, d^3v}{\int f(v)\, g(v)\, d^3v}, \tag{82}$$

where $g(v)$ is the density of states in the conduction band per unit volume of velocity space and d^3v is the appropriate volume element in velocity space. If F_x is the built-in field F_{int}, then the distribution function is the Boltzmann function $f_B(v)$. If F_x is the external field F_{ext}, the distribution function is modified due to carrier heating according to the field strength $f_{F_{\text{ext}}}(v)$ – see Sects. 4.4 and 4.6. The averaging process involves the distribution function, which is modified by both scattering and effective-mass contributions. For a review, see Nag (1980); see also Seeger (1973) and Conwell (1967). In contrast, when only a built-in field is present, the averaging must be done with the undeformed Boltzmann distribution, since lattice and electron temperatures remain the same at each point of the semiconductor.

A more detailed discussion of differences in carrier transport for external, built-in, and mixed fields is postponed to ▶ Chap. 24, "Carrier Scattering at High Electric Fields" after an explicit introduction of carrier heating.

4 The Boltzmann Equation

There are several simplifying assumptions in the Drude-Sommerfeld approach introduced in Sect. 2.1 that are not generally valid.

First, most of the scattering events are not inelastic. A carrier usually accumulates energy during several mean free paths. Each path is interrupted by a mostly elastic collision until it dissipates the increased energy in one inelastic collision; then it continues to accumulate energy, and so on. Different kinds of collisions must be distinguished. This requires replacing the average time between collisions with a relaxation time, which is typical for the decay of a perturbation introduced by an applied external force, e.g., an external electric field.

Second, the interaction with a scattering center often depends on the energy of the electron, e.g., the scattering cross-section of ions is energy dependent, and only more energetic electrons can be scattered inelastically. Therefore, the assumption of a constant, energy-independent time between scattering events needs to be refined. In this chapter, we will introduce the basic Boltzmann equation which permits an analysis of a more advanced description of carrier scattering.

4.1 The Boltzmann Equation for Electrons

A formalism that permits a refinement of the carrier-transport analysis must account for the change in the population of carriers in space and energy or momentum when exposed to external forces. This population is described by a distribution function: in equilibrium, these are the Boltzmann or Fermi functions. Under the influence of a

field, this distribution is modified. It is the purpose of an advanced theory to determine the modified distribution function. From it, other transport parameters can be derived.

A formalism first proposed by Liouville (1838, see Ferziger and Kaper 1972) is too cumbersome for the evaluation of carrier transport. A more useful approach can be derived from the Liouville equation as a zeroth-order approximation[18]; this was suggested by Boltzmann, based on empirical arguments and will be described below.

Conventionally, one uses an accounting procedure for carriers in *phase space*, i.e., in a six-dimensional space-and-momentum representation (x, y, z, k_x, k_y, k_z). The population of electrons in phase space is given by the distribution function $f(\mathbf{r}, \mathbf{k}, t)$; it changes with time. A group of electrons within a volume element of phase space will move and reside in different volume elements as time progresses. Such a motion is described by df/dt. To express the total differential by the local differential, one must consider the deformation of the distribution function due to the time dependence of $\mathbf{r}$ and $\mathbf{k}$ and obtains, using only the first term of a Taylor expansion:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \dot{\mathbf{k}} \cdot \frac{\partial f}{\partial k} + \dot{\mathbf{r}} \cdot \frac{\partial f}{\partial r} = \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}, \tag{83}$$

where the index *coll* indicates the collisions. The first term accounts for the local change of the distribution in time, the second term for the change in momentum space, and the third term for the change of the distribution in real space. The sum of these changes must be equal to the changes of the distribution caused by collisions. This simplified Liouville equation is called the *Boltzmann equation.*

In *steady state* ($\partial f/\partial t \equiv 0$), the Boltzmann equation reads

$$\left(\frac{\partial f(\mathbf{r}, \mathbf{k}, t)}{\partial t}\right)_{\text{coll}} = \dot{\mathbf{k}} \cdot \frac{\partial f}{\partial k} + \dot{\mathbf{r}} \cdot \frac{\partial f}{\partial r} \tag{84}$$

with the first term determined by the forces acting on free electrons, and $\dot{\mathbf{k}}$ given by

$$\dot{\mathbf{k}} = -\frac{e}{\hbar}\mathbf{F}. \tag{85}$$

Here $\mathbf{F}$ is the electric field. The second term is proportional to the spatial gradient of the carrier distribution and to the group velocity

$$\dot{\mathbf{r}} = \frac{1}{\hbar}\frac{\partial E(\mathbf{k})}{\partial \mathbf{k}} = \mathbf{v}. \tag{86}$$

[18]The most severe approximation is the linear relation in time, which eliminates memory effects in the Boltzmann equation (Nag 1980).

This basic Boltzmann equation contains all the dependences necessary for analyzing carrier transport.[19] Some of these dependences, such as the temperature dependence, are contained implicitly. The important part for the carrier transport is the innocent-looking left-hand side of the Boltzmann Eq. 84, which contains the contribution of the more or less inelastic collisions that provide the "friction" for the carrier transport.

The collision term, also referred to as the *collision integral*, describes the transition of an electron from a state E_k, $\mathbf{k}$ to a state $E_{k'}, \mathbf{k}'$. This can be expressed as the difference between electrons scattered from the state $\mathbf{k}$, occupied according to the Fermi-Dirac distribution function $f_{\mathrm{FD}}(\mathbf{k})$, into the state $\mathbf{k}'$, unoccupied according to $1 - f_{\mathrm{FD}}(\mathbf{k}')$, minus the reverse process, and integrated over all possible states $\mathbf{k}'$, into and from which such scattering is possible:

$$\left(\frac{\partial f(\mathbf{k})}{\partial t}\right)_{\mathrm{coll}} = \frac{V}{(2\pi)^3}\int \{f_{\mathrm{FD}}(\mathbf{k})[1 - f_{\mathrm{FD}}(\mathbf{k}')]S(\mathbf{k},\mathbf{k}') - f_{\mathrm{FD}}(\mathbf{k}')[1 - f_{\mathrm{FD}}(\mathbf{k})]S(\mathbf{k}',\mathbf{k})\}\, d\mathbf{k}', \tag{87}$$

where V is the crystal volume and S is the scattering probability

$$S(\mathbf{k},\mathbf{k}') = \frac{2\pi}{\hbar}\ |M(\mathbf{k},\mathbf{k}')|^2\ \delta(E_k - E_{k'} \pm \Delta E). \tag{88}$$

$M(\mathbf{k}, \mathbf{k}')$ is the matrix element for the scattering event and ΔE is the fractional change in electron energy during the partially inelastic scattering. The matrix elements can be expressed as

$$M(\mathbf{k},\mathbf{k}') = \int_v \psi^*_{q',k'}\, \Delta V \psi_{q,k}\ dv \tag{89}$$

where ΔV is the perturbation potential inducing the scattering event, dv is the volume element, and $\psi_{q,k}$, $\psi_{q',k'}$ are the wavefunctions before and after scattering. The perturbation potential depends on the type of scattering event and could be the deformation potential for scattering on acoustic or optical phonons; see ▶ Chap. 23, "Carrier Scattering at Low Electric Fields":

$$\Delta V = \begin{cases} \Xi_{\mathrm{c}}(\partial/\partial\mathbf{r}) \cdot \mathbf{u} & \text{acoustic phonons} \\ D_{\mathrm{o}}\ u & \text{optical phonons} \end{cases} \tag{90}$$

with $\mathbf{u}$ as the displacement of the lattice atoms and Ξ_{c} or D_{o} as the appropriate deformation potentials. Other examples will be given in ▶ Chap. 23, "Carrier Scattering at Low Electric Fields", where also the scattering potential and matrix elements for some of the most important scattering centers are tabulated.

[19]Here discussed for electrons, although with a change of the appropriate parameters, it is directly applicable to holes, polarons, etc; applications to excitons, polaritons, and *e-h* plasmas are reviewed by Snoke (2011). The influence of other fields, such as thermal or magnetic fields, is neglected here; for such influences, see ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients".

4.2 The Boltzmann Equation for Phonons

A Boltzmann-type equation similar to Eqs. 83 and 84 can be set up for the phonon system, which interacts with the electron system. Since the only driving forces for the phonon system are those of diffusion due to thermal gradients (neglecting drag effects discussed in Sect. 4.7), we obtain for steady-state conditions

$$\left(\frac{\partial f(\mathbf{q})}{\partial t}\right)_{\text{coll}} = \dot{\mathbf{r}} \cdot \frac{\partial f(\mathbf{q})}{\partial \mathbf{r}} \quad \text{with} \quad \dot{\mathbf{r}} = \frac{\partial \omega(\mathbf{q})}{\partial \mathbf{r}}; \tag{91}$$

here $\dot{\mathbf{r}}$ is the group velocity of phonons, i.e., the sound velocity in the low $\mathbf{q}$ acoustic branch. The gradient of the phonon distribution function $f(\mathbf{r}, \mathbf{q}, T(\mathbf{r}), t)$ contains the thermal gradient. If undisturbed, the phonon distribution is described by the Bose-Einstein function $f_{\text{BE}}(\mathbf{q})$ – see Eq. 12 of ▶ Chap. 21, "Equilibrium Statistics of Carriers". The collision term contains all phonon-phonon and phonon-electron interactions. We will regard the former as less important to the present discussion. Interaction of optical with acoustic phonons, however, can become quite important, e.g., for cooling of a heated electron ensemble – see ▶ Chap. 32, "Dynamic Processes". In a fashion similar to that given for the electron collision term (Eq. 87), we obtain the transition rate by taking the product of the densities of the occupied and the empty states and the matrix element for each transition, integrated over all possible transitions for absorption and a similar term for emission of phonons,

$$\begin{aligned}\left(\frac{\partial f(\mathbf{k},\mathbf{q})}{\partial t}\right)_{\text{coll}} = \frac{V}{(2\pi)^3}\int\int \{&S(\mathbf{k}+\mathbf{q},\mathbf{k})[1+f_{\text{BE}}(\mathbf{k})]\; f_{\text{BE}}(\mathbf{k}+\mathbf{q}) \\ &- S(\mathbf{k},\mathbf{k}+\mathbf{q})\,[1+f_{\text{BE}}(\mathbf{k}+\mathbf{q})]\; f_{\text{BE}}(\mathbf{k})\}\, d\mathbf{k}\, d\mathbf{q}\,.\end{aligned} \tag{92}$$

In equilibrium, the right side vanishes as transitions from $\mathbf{k}$ to $\mathbf{k} + \mathbf{q}$ equal those from $\mathbf{k} + \mathbf{q}$ to $\mathbf{k}$. Only when a perturbation is introduced, either from the electron ensemble interacting with phonons or from a temperature gradient, will the right side remain finite. The collision term dealing with the interaction of phonons and electrons can be evaluated after linearization.

In order to obtain numerical values, however, one needs to introduce specific assumptions about the microscopic collision process between phonons and electrons. The analysis of such collisions will fill the major part of ▶ Chap. 24, "Carrier Scattering at High Electric Fields".

4.3 The Relaxation-Time Approximation

In order to further discuss carrier transport, we have to solve the Boltzmann equation; that is, we have to obtain an expression for $f(\mathbf{r}, \mathbf{k}, t)$. Since the Boltzmann equation is a nonlinear integrodifferential equation, it cannot be integrated

analytically and requires the use of approximations or of numerical methods.[20] Both will be mentioned later (Sect. 4.4). However, in order to see some of the important relations, a simplified approach is introduced first: the *relaxation-time approximation*.

The balance between gain due to all forces and loss due to collisions of a perturbation, induced by external forces, produces a steady state with a deformed electron distribution. When such forces are suddenly removed, the distribution rapidly returns to its unperturbed state according to

$$\frac{\partial f}{\partial t} = \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}. \tag{93}$$

Assuming that this collision term is linear in the deviation from the unperturbed distribution f_0, we obtain

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{f - f_0}{\tau_{\text{m}}}. \tag{94}$$

Equation 93 can then be integrated. It yields an exponential return from the steady-state, perturbed function $f = f_0 + \delta f$ (with a time-independent δf) to the undisturbed distribution in equilibrium f_0 with the *momentum relaxation time* τ_{m} as the characteristic time constant (see Sect. 4.6.1):

$$f(t) - f_0 = \delta f \exp\left(-\frac{t}{\tau_{\text{m}}}\right). \tag{95}$$

In this linearized form, the deformed distribution function will be used first. In the following sections, an example with zero magnetic field and vanishing gradients in n and T is discussed. Here the second term of Eq. 84 vanishes.

In a homogeneous semiconductor with a force produced by a constant electric field $\mathbf{F}$, we obtain from Eqs. 94 and 84 for a small perturbation of the distribution function

$$\frac{e\mathbf{F}}{\hbar} \frac{\partial f}{\partial \mathbf{k}} = \frac{\delta f}{\tau_{\text{m}}} \tag{96}$$

or, with $\hbar\mathbf{k} = m_n\mathbf{v}$, hence $\partial/\partial\mathbf{k} = (\hbar/m_n)\ \partial/\partial\mathbf{v}$, we obtain

$$\frac{e}{m_n}\tau_{\text{m}}\mathbf{F} \frac{\partial f}{\partial \mathbf{v}} = \delta f. \tag{97}$$

This shows that the change in the distribution function is proportional to the drift velocity $\mathbf{v}_{\text{d}} = (e/m_n)\ \tau_{\text{m}}\mathbf{F}$. With f_0 given by the Boltzmann distribution

[20] An alternative approach considers a hydrodynamic transport as, e.g., discussed for electrons by Bringuier (2019) and for phonons by Sendra et al. (2021).

$$f_0 \propto \exp\left(-\frac{\frac{m_n v^2}{2}}{kT}\right), \tag{98}$$

we obtain for a small perturbation

$$\frac{\partial f}{\partial \mathbf{v}} \cong \frac{\partial f_0}{\partial \mathbf{v}} = -\frac{m_n \mathbf{v}}{kT} f_0; \tag{99}$$

and with Eq. 97, we have the following equation as the final result for the *deformed Boltzmann distribution* due to an external field, in the *relaxation time approximation*:

$$f = f_0\left(1 - \frac{e}{kT}\tau_{\mathrm{m}}\mathbf{F}\cdot\mathbf{v}\right). \tag{100}$$

An illustration of the relaxation process within this approximation is given in Fig. 13. A constant and small homogeneous electric Field **F** shifts the Fermi sphere in **k** space a small amount δ**k** along the field direction. After switching off the field, the return to equilibrium proceeds by inelastic scattering processes from occupied states of the displaced sphere to unoccupied states *of the undisturbed distribution in equilibrium*. Therefore, only electrons in states with $|\mathbf{k}| > k_{\mathrm{F}}$ scattered to states with $|\mathbf{k}| < k_{\mathrm{F}}$ can reestablish equilibrium, with k_{F} being the radius of the Fermi sphere illustrated in Fig. 13. These are all electrons with energy close to the Fermi energy E_{F}, which have states near the surface of the Fermi sphere and hence a speed near $v_{\mathrm{F}} = \hbar k_{\mathrm{F}}/m_n$.

4.4 Carrier Scattering and Energy Relaxation

For a homogeneous semiconductor, $\partial f/\partial \mathbf{r} \equiv 0$ holds, and we obtain from Eqs. 83 and 87

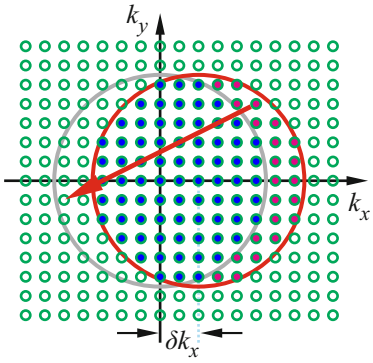


Fig. 13 At an applied electric field, the Fermi sphere of occupied states in **k** space (*green-filled circles*) is shifted by an amount δk_x, yielding the sphere indicated by the *red circle*; *magenta filling* indicates states occupied due to the applied field. Relaxation to equilibrium occurs by inelastic scattering into unoccupied states (*red arrow*) of the equilibrium sphere (*gray circle*)

$$\frac{\partial f(\mathbf{k})}{\partial t} = -\frac{e\mathbf{F}}{\hbar}\cdot\frac{\partial f(\mathbf{k})}{\partial \mathbf{k}} - \frac{V}{8\pi^3}\int \{f(\mathbf{k})[1-f(\mathbf{k}')]\,S(\mathbf{k},\mathbf{k}') - f(\mathbf{k}')[1-f(\mathbf{k})]\,S(\mathbf{k}',\mathbf{k})\}\; d\mathbf{k}', \quad (101)$$

with the electric field **F** producing a deformation from the equilibrium distribution and the collision integral, i.e., the second term in Eq. 101, counteracting this deformation. For an analysis, see Haug (1972).

In equilibrium, the solution of the Boltzmann equation is the Boltzmann or Fermi function $f_0(\mathbf{k})$. With an applied electric field **F**, the distribution is shifted by the drift velocity (see Fig. 13) and is slightly deformed. If the effective mass m_n is isotropic, the scattering probability S (Eq. 88) generally depends only on the magnitude $|\mathbf{k}-\mathbf{k}'|$ and hence on the angle included by the vectors, but not on the individual orientations of **k** and **k**′. The distribution function is then conveniently expressed by a series development using Legendre polynomials (P_n):

$$f(\mathbf{k}) = f_0(\mathbf{k}) + \sum_{n=1}^{\infty} f_n(\mathbf{k})\,P_n(\cos\theta), \quad (102)$$

where θ is the angle between **F** and **k.** After introducing $f(\mathbf{k})$ into Eq. 101, we obtain for the steady-state ($\partial f/\partial t \equiv 0$) a set of n equations to determine $f(\mathbf{k})$.

For small fields, only the first two terms of the development of Eq. 102 are taken and yield

$$f(\mathbf{k}) = f_0(\mathbf{k}) + \cos\theta f_1(\mathbf{k}). \quad (103)$$

The perturbation term $f_1(\mathbf{k})$ of the distribution function is often expressed in terms of a function $\phi(E)$

$$f_1(\mathbf{k}) = \frac{e\hbar}{m_n}\mathbf{F}\cdot\mathbf{k}\;\frac{\partial f_0(E)}{\partial E}\;\phi(E), \quad (104)$$

with $E = E(\mathbf{k})$. This permits a simplified expression for the collision integral (Nag 1980)

$$\left(\frac{\partial f(\mathbf{k})}{\partial t}\right)_{\text{coll}} = -\frac{e\hbar}{m_n kT}\mathbf{F}\cdot\mathbf{k}\;\frac{V}{8\pi^3}\int f_0(E)\,[1-f_0(E')] \times\left[\phi(E) - \frac{k'\cos\theta_k}{k}\phi(E')\right] S(\mathbf{k},\mathbf{k}')\,d\mathbf{k}', \quad (105)$$

where θ_k is the angle between **k** and **k**′ and E' is the energy corresponding to the wavevector **k**′.

Any further simplification of the collision integral requires assumptions of the specific scattering event, which will be listed in Sect. 1 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields" and dealt with sequentially in the following

sections. However, some general remarks here will assist in categorizing the different scattering types.

Elastic Scattering Elastic scattering keeps the electron energy during the scattering event unchanged: $E' = E$. This simplifies Eq. 105 to

$$\left(\frac{\partial f(\mathbf{k})}{\partial t}\right)_{\text{coll}} = -\frac{e\hbar}{m_n kT}\mathbf{F}\cdot\mathbf{k} \times f_0(E)\left[1-f_0(E)\right]\frac{V}{8\pi^3}\int (1-\cos\theta_k)\,S(\mathbf{k},\mathbf{k}')\,d\mathbf{k}'. \tag{106}$$

Elastic scattering events are:

1. All acoustic phonon scattering events, such as deformation-potential scattering and piezoelectric scattering
2. All defect scattering events, such as scattering at neutral impurities, ionized impurities, and larger-defect scattering
3. Alloy scattering

Inelastic Scattering Inelastic scattering does not permit further simplification of Eq. 105. Here $E' \neq E$, and in each case the collision integral must be evaluated separately. Such inelastic scattering events are

1. Optical phonon scattering, such as nonpolar and polar optical scattering
2. Intervalley scattering

The total scattering term is given as the sum over the different scattering types:

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \sum_i \left(\frac{\partial f}{\partial t}\right)_{\text{coll},i}. \tag{107}$$

The Carrier Current The current of carriers can be obtained from the deformed Boltzmann distribution Eq. 100 by summation over all n_{el} carriers and velocities $\mathbf{v}$,

$$\mathbf{j} = e\sum_{i=1}^{n_{\text{el}}}\sum_{\mathbf{v}} e\,\mathbf{v}\,\delta f = e^2\sum_{i=1}^{n_{\text{el}}}\sum_{\mathbf{v}}\frac{\mathbf{v}\mathbf{F}\cdot\mathbf{v}f_0\tau_{\text{m}}}{kT}. \tag{108}$$

Assuming a spherical equi-energy surface for $E(\mathbf{k})$, the summation over $\mathbf{v}\mathbf{F}\cdot\mathbf{v}$ can be carried out,[21] using for $\mathbf{v}\,\mathbf{v}$ the averages $\langle v_x v_y\rangle = \langle v_y v_z\rangle = \langle v_z v_x\rangle = 0$ and $\langle v_x^2\rangle = \langle v_y^2\rangle = \langle v_z^2\rangle = v^2/3$; this yields

$$\mathbf{j} = \frac{e^2}{3kT}\sum_{\mathbf{v}} v^2 \tau_{\mathrm{m}} f_0 \mathbf{F}. \tag{109}$$

Considering that

$$n = \sum_{i=1}^{n_{\mathrm{el}}}\sum_{\mathbf{v}} f_0, \tag{110}$$

we obtain

$$\mathbf{j} = e n \frac{e\mathbf{F}}{3kT}\frac{\sum_{\mathbf{v}} v^2 \tau_{\mathrm{m}} f_0}{\sum_{\mathbf{v}} f_0} = e n \frac{e}{3kT}\langle v^2 \tau_{\mathrm{m}}\rangle \mathbf{F}, \tag{111}$$

which gives the electron mobility as

$$\mu_n = \frac{e}{3kT}\langle v^2 \tau_{\mathrm{m}}\rangle. \tag{112}$$

With $\langle E\rangle = (3/2)\ kT = m_n\langle v^2\rangle/2$, we can replace 3 kT by $m_n\langle v^2\rangle$, yielding

$$\mu_n = \frac{e}{m_n}\frac{\langle v^2 \tau_{\mathrm{m}}\rangle}{\langle v^2\rangle} = \frac{e}{m_n}\frac{\langle E\,\tau_{\mathrm{m}}\rangle}{\langle E\rangle}. \tag{113}$$

This result replaces the average time between scattering events obtained from the Drude theory with the energy-weighted average of the relaxation time τ_{m}. Dropping the requirement of spherical equi-energy surfaces, the end result (Eq. 113) remains the same, except that m_n is replaced by the anisotropic mobility effective mass; for more detail, see Conwell (1982).

[21] The quantities τ_{m} and f_0 in Eq. 108 are functions of E only and hence do not change over a constant-energy surface.

4.5 The Mobility Effective Mass

Following external forces, the carriers are accelerated proportionally to their effective masses (Eq. 23). Their anisotropy is taken into consideration by introducing a *mobility effective mass.* For a three-axes ellipsoid, this effective mass is given by the inverse average of the effective masses along the main axes:

$$\frac{1}{m_{n,\mu}} = \frac{1}{3}\left(\frac{1}{m_1} + \frac{1}{m_2} + \frac{1}{m_3}\right). \tag{114}$$

In general, a mobility *tensor* is introduced for each of the ν_d satellite valley $E(\mathbf{k})$ ellipsoids, identified by the index i:

$$\underline{\boldsymbol{\mu}}^{(i)} = \frac{e\overline{\tau}_m}{\hbar^2}\begin{pmatrix} \frac{\partial^2 E^{(i)}}{\partial k_x^2} & \frac{\partial^2 E^{(i)}}{\partial k_x \partial k_y} & \frac{\partial^2 E^{(i)}}{\partial k_x \partial k_z} \\ \frac{\partial^2 E^{(i)}}{\partial k_y \partial k_x} & \frac{\partial^2 E^{(i)}}{\partial k_y^2} & \frac{\partial^2 E^{(i)}}{\partial k_y \partial k_z} \\ \frac{\partial^2 E^{(i)}}{\partial k_z \partial k_x} & \frac{\partial^2 E^{(i)}}{\partial k_z \partial k_y} & \frac{\partial^2 E^{(i)}}{\partial k_z^2} \end{pmatrix}. \tag{115}$$

In Si, there are three pairs of ellipsoids with different ⟨1 0 0⟩ orientations; see Fig. 10a of ▶ Chap. 8, "Bands and Bandgaps in Solids". All of these ellipsoids have their $E(\mathbf{k})$ minima at equal energies and are therefore equally populated at vanishing external forces with $N_0^{(i)} = \frac{1}{6}N_0$. Their effect on the total mobility is obtained by adding its components, which results in an isotropic mobility tensor

$$\underline{\boldsymbol{\mu}}_\mathrm{tot} = \frac{e\overline{\tau}_m}{\hbar^2}\begin{pmatrix} \frac{1}{3}\left(\frac{2}{m_{n,\mathrm{t}}} + \frac{1}{m_{n,\mathrm{l}}}\right) & 0 & 0 \\ 0 & \frac{1}{3}\left(\frac{2}{m_{n,\mathrm{t}}} + \frac{1}{m_{n,\mathrm{l}}}\right) & 0 \\ 0 & 0 & \frac{1}{3}\left(\frac{2}{m_{n,\mathrm{t}}} + \frac{1}{m_{n,\mathrm{l}}}\right) \end{pmatrix}, \tag{116}$$

with indices l and t indicating transversal and longitudinal components with respect to the ⟨1 0 0⟩ axes. Consequently, the *electron mobility effective mass* for Si is given by

$$\frac{1}{m_{n,\mu}} = \frac{1}{3}\left(\frac{2}{m_{n,\mathrm{t}}} + \frac{1}{m_{n,\mathrm{l}}}\right). \tag{117}$$

The *hole mobility effective mass* can be derived in a similar fashion. Assuming spherical $E(\mathbf{k})$ surfaces around $k = 0$ (for warped bands, see below) and disregarding the deeper spin-orbit band, one obtains

$$\frac{1}{m_{p,\mu}} = \frac{1}{2}\left(\frac{1}{m_{p,lh}} + \frac{1}{m_{p,hh}}\right); \tag{118}$$

here, $m_{p,lh}$ and $m_{p,hh}$ represent the light and heavy hole masses in the corresponding bands.

Hole-Mobility Mass in Warped Bands The valence bands are significantly warped; see Sects. 1.2.2 and 1.2.3 of ▶ Chap. 8, "Bands and Bandgaps in Solids". In cubic semiconductors, the $E(\mathbf{k})$ surfaces can be represented by an empirical expression (Eq. 15 of ▶ Chap. 8, "Bands and Bandgaps in Solids"); using the abbreviations $B' = \sqrt{B^2 + \frac{1}{6}C^2}$ and $\Gamma_\pm = \mp C^2/\{2B' \times (A \pm B')\}$ in this equation, we obtain, after taking the second derivative of $E(\mathbf{k})$ for the effective masses, a useful approximation

$$m_{p,\pm} = \frac{m_0}{A \pm B'}\left(1 + 0.333\,\Gamma + 0.0106\,\Gamma^2 + \quad \ldots\right) \tag{119}$$

for the mobility effective mass. The variable $m_\pm$ represents the *light* or *heavy hole* mass if the upper or lower sign, respectively, is used in Eq. 119. The values of the constants A, B, and C can be obtained from the Luttinger parameters (Eq. 16 of ▶ Chap. 8, "Bands and Bandgaps in Solids"), which are given in Table 5 of ▶ Chap. 8, "Bands and Bandgaps in Solids" for a number of typical semiconductors. More recent calculations of the carrier mobility from first principles without parameters are reviewed by Poncé et al. (2020).

4.6 Momentum and Energy Relaxation

With each scattering event, momentum is exchanged; the carrier changes the direction of its path. In addition, more or less energy is exchanged, with the carrier losing or gaining energy from the scattering center. There are different rates for momentum and energy relaxation, and the relaxation times depend on the specific scattering mechanism.

4.6.1 The Average Momentum Relaxation Time

The average momentum relaxation time, which was already introduced in Eq. 95, is defined as

$$\langle \tau_m \rangle = \frac{\int \tau_m(v)\ \ v_x\,(\partial f/\partial v_x)\,d^3v}{\int f_0\,d^3v} = \frac{m_n}{kT}\frac{\int \tau_m v_x^2 f_0\,d^3v}{\int f_0\,d^3v}. \tag{120}$$

It can be obtained from the net increment of the electron momentum, which is proportional to the average drift velocity,

$$\overline{v}_x = \frac{\int v_x f(v)\, g(v)\, d^3v}{\int f(v)\, g(v)\;\; d^3v}, \tag{121}$$

where $g(v)$ is the density of states. Assuming only small changes from the thermal distribution, $g(v)$ can be expressed as the effective density of states at the edge of the band (Eq. 16 of ▶ Chap. 21, "Equilibrium Statistics of Carriers") and cancels out in Eq. 121. This yields

$$\overline{v}_x = -\frac{eF_x}{3kT}\frac{\int_0^\infty v^4 \tau_\mathrm{m} f_0(v)\, dv}{\int_0^\infty v^2 f_0(v)\, dv} = -\frac{eF_x}{3kT}\left\langle v^2 \tau_\mathrm{m}(v)\right\rangle. \tag{122}$$

Using the equipartition law for a Boltzmann gas of electrons $\frac{m}{2}\left\langle v^2\right\rangle = \frac{3}{2}kT$, we obtain for the average drift velocity

$$\overline{v}_x = -\frac{eF_x}{3kT}\frac{\left\langle v^2 \tau_\mathrm{m}(v)\right\rangle}{\left\langle v^2\right\rangle} = -\frac{e}{m_n}\langle \tau_\mathrm{m}\rangle F_x. \tag{123}$$

This result is closely related to the Drude equation, however, having replaced $\overline{\tau}$ in Eq. 25 with the *average momentum relaxation time*

$$\langle \tau_\mathrm{m}\rangle = \frac{\left\langle v^2 \tau_\mathrm{m}(v)\right\rangle}{\left\langle v^2\right\rangle}. \tag{124}$$

For an evaluation of Eq. 124, we need the distribution function f_0 and the actual scattering mechanism to determine $\tau_\mathrm{m}(v)$ – see ▶ Chap. 23, "Carrier Scattering at Low Electric Fields" and Seeger (1973).[22]

After a collision, the electron path changes by an angle θ, and the fractional change of angle per collision is on the average $\langle 1 - \cos\theta\rangle$. The momentum relaxation time is the time after which the electron path is totally randomized, i.e., its "memory" is lost; hence,

$$\tau_\mathrm{m} = \frac{\tau_\mathrm{sc}}{\langle 1 - \cos\theta\rangle}, \tag{125}$$

[22] A recent study indicates that the widely applied relaxation-time approximation to the relativistic Boltzmann equation is not compatible with conservation laws if the relaxation time depends on energy (Rocha et al. 2021); see also Bringuier (2019).

where τ_{sc} is the average time between two collisions. Scattering with $\theta \geq 90°$ is memory erasing (Sect. 3.1 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields"). Only the collisions in which all angles θ are equally probable result in $\langle \cos\,\theta \rangle = 0$ and, therefore, yield $\tau_m = \tau_{sc}$. For small-angle scattering events, one needs several scatterings before the momentum is relaxed, leading to $\tau_m > \tau_{sc}$.

The Mean Free Path of Carriers Between collisions, the carrier traverses one free path. The mean free path is obtained by averaging

$$\overline{\lambda} = \frac{\langle v^2 \lambda(v) \rangle}{\langle v^2 \rangle}. \tag{126}$$

$\overline{\lambda}$ is related to the momentum relaxation time:

$$\overline{\lambda} = \langle \tau_m \rangle \frac{\langle v^2 \rangle}{\langle v \rangle} = \sqrt{\frac{3\pi}{8}} \langle \tau_m \rangle v_{rms}. \tag{127}$$

4.6.2 The Average Energy Relaxation Time

The energy loss or gain due to scattering of electrons with phonons is given by

$$\frac{dE}{dt} = \frac{V}{8\pi^3} \int \left\{ \left[\hbar\omega_q S(\mathbf{k}, \mathbf{k}') \right]_{\text{emi},\,q} - \left[\hbar\omega_q S(\mathbf{k}', \mathbf{k}) \right]_{\text{abs},\,q} \right\} d\mathbf{k}'. \tag{128}$$

The subscript *emi,q* stands for emission and *abs,q* for absorption of a phonon. When multiplying the Boltzmann equation with E and integrating over $\mathbf{k}$, we obtain (Seeger 1973)

$$\frac{d\langle E \rangle}{dt} = e\,\mathbf{F}\ \langle \mathbf{v} \rangle - \frac{\langle E \rangle - E_L}{\tau_e}, \tag{129}$$

with E_L the equilibrium energy at lattice temperature,

$$\langle E \rangle = \frac{\int E f\, d\mathbf{k}}{\int f\, d\mathbf{k}}, \text{ and } \langle \mathbf{v} \rangle = \frac{\int \mathbf{v} f\, d\mathbf{k}}{\int f\, d\mathbf{k}}. \tag{130}$$

The *energy relaxation time* is then obtained from Eq. 129 after switching off the field, yielding

$$\tau_e = \frac{\langle E \rangle - E_L}{\dfrac{d\langle E \rangle}{dt}}; \tag{131}$$

this means, τ_e is given by the ratio of the *average surplus energy* to the rate of *energy loss* due to scattering and is a function of *E;* the energy-loss rate is not a simple exponential function. It shows a maximum when the electron energy equals the optical phonon energy.

The rate of momentum or energy loss depends on the actual scattering mechanism. From gas-kinetic arguments, one obtains for collisions between an electron and a *lattice defect* of mass *M* an energy-exchange rate of

$$\left(\frac{\tau_e}{\tau_{sc}}\right)_{ion} = \frac{m_n}{M}. \tag{132}$$

The energy loss is negligible in one scattering event if the scattering center is a defect atom, since $M \gg m_n$.

The fraction of energy lost by an electron in a collision with *acoustic phonons* can also be obtained from an effective-mass ratio. Using the *equivalent phonon mass*

$$m_{phonon} = \frac{kT}{v_s^2}, \tag{133}$$

where v_s is the sound velocity, we obtain

$$\left(\frac{\tau_e}{\tau_{sc}}\right)_{ac\ phonon} = \frac{m_n}{m_{phonon}} = \frac{m_n v_s^2}{kT} = \frac{3v_s^2}{v_{rms}^2}, \tag{134}$$

which is on the order of 10^{-3}. In other words, only 0.1% of the electron energy can be lost to an acoustic phonon during any one-scattering event.

In contrast, the ratio of energy relaxation time to scattering time for *optical phonons* is

$$\left(\frac{\tau_e}{\tau_{sc}}\right)_{opt\ phonon} = \frac{\hbar\omega_0}{kT}, \tag{135}$$

which is on the order of 1 at room temperature. This means that many scattering events usually pass before the accumulated energy obtained from the field can be dissipated to the lattice by emitting one optical phonon, while the momentum is relaxed after one or only a few collisions. This modifies the rather crude model given in Sect. 2.1 by introducing the momentum relaxation time for evaluating the mobility and the energy relaxation time for Joule's heating.[23] A more detailed discussion is given in ▶ Chaps. 23, "Carrier Scattering at Low Electric Fields", and ▶ 24, "Carrier Scattering at High Electric Fields", when an analysis of the different scattering mechanisms and a better estimate of the magnitude of energy obtained from the field is given.

[23]In two-dimensional structures, significant momentum relaxation is found due to near-surface acoustic phonon scattering; see Pipa et al. (1999).

The effect of either considering the momentum-relaxation time or the energy-relaxation time when evaluating the carrier mobility in Si or GaAs is discussed by Ma et al. (2018).

4.7 Phonon and Electron Drag

Interacting electrons and phonons exchange energy and momentum. A drift motion superimposed on the random motion of one ensemble transfers part of the net momentum to the other ensemble during scattering. This means that electrons drifting in an external field tend to push phonons in the same direction, which causes a slight temperature gradient in the field direction, superimposed on the homogeneous Joule's heating. This process is called *electron drag* (Hubner and Shockley 1960).

Similarly, a temperature gradient tends to push electrons from the warm to the cold end of a semiconductor. This is known as *phonon drag*. The drag effect can be quite large, e.g., up to a factor of 6 compared to simple thermopower in *p*-type Ge at 20 K, as shown by Herring (1954); for the phonon-drag contribution in Si see Mahan et al. (2014).

Ab initio calculations of coupled Boltzmann equations for electrons and phonons show that a *mutual* drag occurs, which enhances the thermopower and indicates a dominance by phonons, rather than by electrons as traditionally assumed; the drag also strongly enhances the intrinsic electron mobility and thermal conductivity (Protik and Kozinsky 2020; Xu et al. 2021). A review of the impact of electron-phonon interaction on phonon transport is given by Quan et al. (2021).

The drag effect also creates a force which acts on mobile crystal defects. A general scheme for the effect on nanoscale defects is discussed by Swinburne and Dudarev (2015), a review on mechanisms acting in the motion of dislocations is given by Gurrutxaga-Lerma et al. (2020).

When phonons propagate as acoustic waves, ac-electric fields can be induced; or, vice versa, when sufficiently high electric fields are applied, coherent phonon waves can be generated when the drift velocity of electrons surpasses the (sound) velocity of the phonon waves (McFee 1966). These *acousto-electric effects* have technical application for creating current oscillators (Bray 1969).

5 Summary

Electrons and holes are carriers of the current in semiconductors. Their effective mass is modified by polarizing the lattice, but this influence, expressed as a change to polarons, is negligible for most semiconductors. It is contained in the effective mass obtained by cyclotron resonance and is commonly listed as the effective mass of the carriers, i.e., of electrons or holes. Only for small polarons, which occur in semiconductors with narrow bands and large lattice polarization, the effective mass is significantly affected and self-trapping occurs.

The current through a semiconductor is composed of a drift and a diffusion current of electrons and holes. In homogeneous semiconductors, only one of these four components is dominant. The *drift current* is determined by the electric field, which acts as a slight perturbation of an essentially random walk of carriers, except for very high fields. The additional energy obtained by carrier acceleration from the field is given to the lattice by inelastic scattering, causing Joule's heating. The *diffusion current* is proportional to the carrier gradient up to a maximum diffusion current, which is limited by the thermal velocity of carriers. Proportionality factor of both drift and diffusion currents is the carrier mobility, which is proportional to a relaxation time and inversely proportional to an effective mass tensor, the mobility effective mass.

Drift and total currents are proportional to negative potential gradients: the first one being the electrostatic potential and the second the electrochemical potential. The proportionality factor of both is the conductivity. In spatially inhomogeneous semiconductors, both an external field, impressed by an applied bias, and a built-in field, due to space-charge regions, exist within the semiconductor. The external field causes carrier heating by shifting and deforming the carrier distribution from a Boltzmann distribution to a distorted distribution with more carriers at higher energies within the band. In contrast, the built-in field leaves the Boltzmann distribution of carriers unchanged; the carrier gas remains unheated at exactly the same temperature as the lattice at every volume element of the crystal, except for statistical fluctuations. A consequence of the difference between external and built-in fields is the difference in determining the field dependence of the mobility, which requires an averaging over carriers with different energies within the band. For a built-in field, the averaging follows a Boltzmann distribution; in an external field, there are more electrons at higher energies, and the distribution is distorted accordingly. This can have significant impact for the evaluation of device performances when high fields are considered.

The Boltzmann equation permits a sophisticated analysis of the carrier transport, including changes in the carrier distribution. Such a change in the distribution substantially influences the averaging, which is necessary to arrive at well-defined values for a number of transport parameters – most importantly the relaxation times. Significant differences can be defined between the time between two scattering events τ_{sc}, the momentum relaxation time τ_m, and the energy relaxation time τ_e. Although the Boltzmann equation cannot be integrated in closed form except for a few special cases, the deformed distribution function can be approximated for small applied fields. It provides the basis for the investigation of various scattering processes, which can be divided into essentially elastic processes with mainly momentum exchange and, for carriers with sufficient accumulated energy, into inelastic scattering with energy relaxation. The latter becomes more prevalent at elevated temperatures and higher electric fields.

References

Alexandrov AS, Devreese JT (2010) Advances in polaron physics. Springer, Berlin/Heidelberg
Appel J (1968) Polarons. In: Seitz F, Turnbull D, Ehrenreich H (eds) Solid state physics, vol 21. Academic Press, New York, pp 193–391
Bajaj KK (1968) Polaron in a magnetic field. Phys Rev 170:694
Böer KW (1985a) High-field carrier transport in inhomogeneous semiconductors. Ann Phys 497:371
Böer KW (1985b) Current-voltage characteristics of diodes with and without light. Phys Status Solidi A 87:719
Bogoliubov NN, Bogoliubov NN Jr (1986) Aspects of polaron theory. World Scientific Publishing, Singapore
Böttger H, Bryksin VV (1985) Hopping conduction in solids. Phys Status Solidi B 96:219
Bray R (1969) A Perspective on acoustoelectric instabilities. IBM J Res Dev 13:487
Brazovskii S, Kirova N, Yu ZG, Bishop AR, Saxena A (1998) Stability of bipolarons in conjugated polymers. Opt Mater 9:502
Bringuier E (2019) The Boltzmann equation and relaxation-time approximation for electron transport in solids. Eur J Phys 40:025103
Chakraverty BK, Schlenker C (1976) On the existence of bipolarons in Ti_4O_7. J Physique (Paris) Colloq 37:C4–C353
Chandrasekhar S (1943) Stochastic problems in physics and astronomy. Rev Mod Phys 15:1
Chattopadhyay D, Rakshit PC, Kabasi A (1989) Diffusion of one- and two-dimensional hot electrons in semiconductor quantum-well structures at low temperatures. Superlattice Microstruct 6:399
Christov SG (1982) Adiabatic polaron theory of electron hopping in crystals: a reaction-rate approach. Phys Rev B 26:6918
Cohen MH, Economou EN, Soukoulis CM (1983) Electron-phonon interactions near the mobility edge in disordered semiconductors. J Non-Cryst Solids 59/60:15
Comas F, Mora-Ramos ME (1989) Polaron effect in single semiconductor heterostructures. Physica B 159:413
Conibeer G, Zhang Y, Bremner SP, Shrestha S (2017) Towards an understanding of hot carrier cooling mechanisms in multiple quantum wells. Jpn J Appl Phys 56:091201
Conwell EM (1967) High-field transport in semiconductors. Academic Press, New York
Conwell EM (1982) The Boltzmann equation. In: Paul W, Moss TS (eds) Handbook of semiconductors vol 1: band theory and transport properties. North Holland Publishing, Amsterdam, pp 513–561
Devreese JT (1984) Some recent developments on the theory of polarons. In: Devreese JT, Peeters FM (eds) Polarons and excitons in polar semiconductors and ionic crystals. Plenum Press, New York, pp 165–183
Drude P (1900) Zur Elektronentheorie der Metalle. Ann Phys 1:566. (On the electron theory of metals, in German)
Dushman S (1930) Thermionic emission. Rev Mod Phys 2:381
Emin D (1973) On the existence of free and self-trapped carriers in insulators: an abrupt temperature-dependent conductivity transition. Adv Phys 22:57
Esaki L, Tsu R (1970) Superlattice and negative differential conductivity in semiconductors. IBM J Res Dev 14:61
Evrard R (1984) Polarons. In: Di Bartolo B (ed) Collective excitations in solids. Plenum Press, New York, pp 501–522
Ferziger JD, Kaper HG (1972) Mathematical theory of transport processes in gasses. North Holland Publishing, Amsterdam
Feynman RP (1955) Slow electrons in a polar crystal. Phys Rev 97:660
Franchini C, Reticcioli M, Setvin M, Diebold U (2021) Polarons in materials. Nat Rev Mater 6:560

Frenkel JI (1936) On the absorption of light and the trapping of electrons and positive holes in crystalline dielectrics. Sov Phys J 9:158
Fröhlich H, Pelzer H, Zienau S (1950) Properties of slow electrons in polar materials. Philos Mag (Ser 7) 41:221
Frost JM (2017) Calculating polaron mobility in halide perovskites. Phys Rev B 96:195202
Ghosh D, Welch E, Neukirch AJ, Zakhidov A, Tretiak S (2020) Polarons in halide perovskites: a perspective. J Phys Chem Lett 11:3271
Gurrutxaga-Lerma B, Verschueren J, Sutton AP, Dini D (2020) The mechanics and physics of high-speed dislocations: a critical review. Int Mater Rev 66(4):215
Hai GQ, Cândido L, Brito BGA, Peeters FM (2018) Electron pairing: from metastable electron pair to bipolaron. J Phys Commun 2:035017
Haug A (1972) Theoretical solid state physics. Pergamon Press, Oxford
Hayes W, Stoneham AM (1984) Defects and defect processes in nonmetallic solids. Wiley, New York
Herring C (1954) Theory of the thermoelectric power of semiconductors. Phys Rev 96:1163
Holstein T (1959) Studies of polaron motion Part 1: the molecular-crystal model. Ann Phys 8:325
Horio K, Okada T, Nakatani A (1999) Energy transport simulation for graded HBT's: importance of setting adequate values for transport parameters. IEEE Trans Electron Devices 46:641
Hubner K, Shockley W (1960) Transmitted phonon drag measurements in silicon. Phys Rev Lett 4:504
Irvine LAD, Walker AB, Wolf MJ (2021) Quantifying polaronic effects on the scattering and mobility of charge carriers in lead halide perovskites. Phys Rev B 103:220305
Ivanov KA, Petrov AG, Kaliteevski MA, Gallant AJ (2015) Anharmonic Bloch oscillation of electrons in biased superlattices. JETP Lett 102:796
Kahmann S, Loi MA (2019) Hot carrier solar cells and the potential of perovskites for breaking the Shockley-Queisser limit. J Mater Chem C 7:2471
Kan EC, Ravaioli U, Chen D (1991) Multidimensional augmented current equation including velocity overshoot. IEEE Electron Device Lett 12:419
Kang Y, Han S (2018) Intrinsic carrier mobility of cesium lead halide perovskites. Phys Rev Applied 10:044013
Kartheuser E, Devreese JT, Evrard R (1979) Polaron mobility at low temperature: a self-consistent equation-of-motion approach. Phys Rev B 19:546
Kuehn W, Gaal P, Reimann K, Woerner M, Elsaesser T, Hey R (2010) Coherent ballistic motion of electrons in a periodic potential. Phys Rev Lett 104:146602
Kwak J, Park J, Park J, Baek K, Choi A, Kim T (2019) 940-nm 350-mW transverse single-mode laser diode with AlGaAs/InGaAs GRIN-SCH and asymmetric structure. Curr Opt Photonics 3:583
Landau LD (1933) Electron motion in crystal lattices. Sov Phys J 3:664
Landsberg PT (1952) On the diffusion theory of rectification. Proc R Soc Lond A Math Phys Sci 213:226
Landsberg PT, Cheng HC (1985) Activity coefficient and Einstein relation for different densities of states in semiconductors. Phys Rev B 32:8021
Lee TD, Low F, Pines D (1953) The motion of slow electrons in a polar crystal. Phys Rev 90:297
Leo K (1996) Optical investigations of Bloch oscillations in semiconductor superlattices. Phys Scripta T68:78
Liouville J (1838) Note sur la théorie de la variation des constantes arbitraires. J Math Pures Appl 3:349. (Note on the theory of the variation of arbitrary constants, in French)
Lorentz HA (1909) The theory of electrons. Teubner Verlag, Leipzig
Lushchik C, Lushchik A (2018) Evolution of anion and cation excitons in alkali halide crystals. Phys Sol State 60:1487
Ma J, Nissimagoudar AS, Li W (2018) First-principles study of electron and hole mobilities of Si and GaAs. Phys Rev B 97:045201

Mahan GD, Lindsay L, Broido DA (2014) The Seebeck coefficient and phonon drag in silicon. J Appl Phys 116:245102
Mahata A, Meggiolaro D, De Angelis F (2019) From large to small polarons in lead, tin, and mixed lead-tin halide perovskites. J Phys Chem Lett 10:1790
McFee JH (1966) Transmission and amplification of acoustic waves. In: Mason W (ed) Physical acoustics, vol 4A. Academic Press, New York, pp 1–44
Mikhnenko OV, Blom PWM, Nguyen T-Q (2015) Exciton diffusion in organic semiconductors. Energy Environ Sci 8:1867
Mishchenko AS, Pollet L, Prokof'ev NV, Kumar A, Maslov DL, Nagaosa N (2019) Polaron mobility in the "beyond quasiparticles" regime. Phys Rev Lett 123:076601
Muljarov EA, Tikhodeev SG (1997) Self-trapped excitons in semiconductor quantum wires inside a polar dielectric matrix. Phys Status Solidi A 164:393
Nag BR (1980) Electron transport in compound semiconductors. Springer, Berlin
Najafi E, Scarborough TD, Tang J, Zewail A (2015) Four-dimensional imaging of carrier interface dynamics in *p-n* junctions. Science 347:164
Neukirch AJ, Abate II, Zhou L, Nie W, Tsai H, Pedesseau L, Even J, Crochet JJ, Mohite AD, Katan C, Tretiak S (2018) Geometry distortion and small polaron binding energy changes with ionic substitution in halide perovskites. J Phys Chem Lett 9:7130
Panahi SRFS, Abbasi A, Ghods V, Amirahmadi M (2021) Improvement of CIGS solar cell efficiency with graded bandgap absorber layer. J Mater Sci: Mater Electron 32:2041
Peeters FM, Devreese JT (1984) Theory of polaron mobility. In: Ehrenreich H, Turnbull D (eds) Solid state physics, vol 38. Academic Press, Orlando, pp 81–133
Pekar SI (1954) Untersuchungen über die Elektronentheorie der Kristalle. Akademie Verlag, Berlin. (Investigations on the electron theory of crystals, in German)
Pipa VI, Vagidov NZ, Mitin VV, Stroscio M (1999) Momentum relaxation of 2D electron gas due to near-surface acoustic phonon scattering. Physica B 270:280
Poncé S, Li W, Reichardt S, Giustino F (2020) First-principles calculations of charge carrier mobility and conductivity in bulk semiconductors and two-dimensional materials. Rep Prog Phys 83:036501
Protik NH, Kozinsky B (2020) Electron-phonon drag enhancement of transport properties from a fully coupled *ab initio* Boltzmann formalism. Phys Rev B 102:245202
Quan Y, Yue S, Liao B (2021) Impact of electron-phonon interaction on thermal transport: a review. Nanosc Microsc Thermophys Engin 25:73
Rocha GS, Denicol GS, Noronha J (2021) Novel relaxation time approximation to the relativistic Boltzmann equation. Phys Rev Lett 127:042301
Seeger K (1973) Semiconductor physics. Springer, Wien/New York
Sendra L, Beardo A, Torres P, Bafaluy J, Alvarez FX, Camacho J (2021) Derivation of a hydrodynamic heat equation from the phonon Boltzmann equation for general semiconductors. Phys Rev B 103:140301
Shah J (1994) Bloch oscillations in semiconductor superlattices. Proc SPIE 2145:144
Simoen E, Claeys C, Czerwinski A, Katcki J (1998) Accurate extraction of the diffusion current in silicon *p-n* junction diodes. Appl Phys Lett 72:1054
Sio WH, Verdi C, Poncé S, Giustino F (2019a) Polarons from first principles, without supercells. Phys Rev Lett 122:246403
Sio WH, Verdi C, Poncé S, Giustino F (2019b) *Ab initio* theory of polarons: formalism and applications. Phys Rev B 99:235139
Snoke DW (2011) The quantum Boltzmann equation in semiconductor physics. Ann Phys (Berlin) 523:87
Sommerfeld A (1928) Zur Elektronentheorie der Metalle auf Grund der Fermischen Statistik. Z Phys 47:1 (On the electron theory due to Fermi's statistics, in German)
Stratton R (1969) Carrier heating or cooling in a strong built-in electric field. J Appl Phys 40:4582
Swinburne TD, Dudarev SL (2015) Phonon drag force acting on a mobile crystal defect: full treatment of discreteness and nonlinearity. Phys Rev B 92:134302

Tjablikov SV (1952) On the energy spectrum of electrons in a polar crystal. Zh Eksp Teor Fiz 23:381

Toyozawa Y (1981) Charge transfer instability with structural change. I. Two-sites two-electrons system. J Phys Soc Jpn 50:1861

Unuma T, Ino Y, Kuwata-Gonokami M, Bastard G, Hirakawa K (2010) Transient Bloch oscillation with the symmetry-governed phase in semiconductor superlattices. Phys Rev B 81:125329

Waschke C, Leisching P, Haring Bolivar P, Schwedler R, Brüggemann F, Roskos HG, Leo K, Kurz H, Köhler K (1994) Detection of Bloch oscillations in a semiconductor superlattice by time-resolved terahertz spectroscopy and degenerate four-wave mixing. Solid State Electron 37:1321

Xu Q, Zhou J, Liu T-H, Chen G (2021) First-Principles study of all thermoelectric properties of Si-Ge alloys showing large phonon drag from 150 to 1100 K. Phys Rev Appl 16:064052

Yang KJ, Son DH, Sung SJ, Sim JH, Kim YI, Park S-N, Jeon D-H, Kim JS, Hwang D-K, Jeon C-W, Nam D, Cheong H, Kang J-K, Kim D-H (2016) A band-gap-graded CZTSSe solar cell with 12.3% efficiency. J Mater Chem A 4:10151

Zhang ML, Drabold DA (2010) Phonon driven transport in amorphous semiconductors: transition probabilities. Eur Phys J B 77:7

Zhou L, Katan C, Nie W, Tsai H, Pedesseau L, Crochet JJ, Even J, Mohite AD, Tretiak S, Neukirch AJ (2019) Cation alloying delocalizes polarons in lead halide perovskites. J Phys Chem Lett 10:3516

Carrier Scattering at Low Electric Fields

23

Karl W. Böer and Udo W. Pohl

Contents

1 Types of Scattering Centers 986
2 Electron Scattering with Phonons 989
2.1 General Properties of Scattering with Phonons 989
2.2 Scattering with Acoustic Phonons 991
2.3 Scattering with Optical Phonons 995
3 Scattering at Coulomb Potentials 999
3.1 Ionized-Impurity Scattering 999
3.2 Coulomb Scattering in Anisotropic Semiconductors 1004
3.3 Scattering at Dislocations 1005
3.4 Carrier-Carrier and Carrier-Plasmon Scattering 1006
4 Scattering by Neutral Defects 1007
4.1 Scattering by Neutral Lattice Defects, Intrinsic Point Defects, and Defect Clusters 1007
4.2 Influence of External Surfaces 1009
4.3 Influence of Microcrystallite Boundaries 1010
4.4 Alloy Scattering 1011
5 Multivalley Carrier Transport 1013
5.1 Processes in Intervalley Scattering 1013
5.2 Mobility for Intervalley Scattering 1014
6 Summary 1017
References 1017

Abstract

Carrier scattering, originating from deviations from ideal lattice periodicity, acts as a damping process for carrier motion. Both elastic and inelastic scattering

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_23

involve a large variety of scattering centers. Carriers are scattered by acoustic and optical phonons, at neutral or charged impurities, at interfaces, and at other scattering centers. Most scattering events are elastic, changing only the momentum of a carrier but not its energy. Inelastic scattering involves optical phonons and intervalley scattering; in these processes carriers lose much of their energy to the lattice.

At low electric fields, many elastic scattering events precede an inelastic event. The dominating type of scattering changes with lattice temperature. Usually, ionized-impurity scattering prevails at low temperatures and scattering at phonons at high temperatures. The type of carrier scattering determines the relaxation time and with it the carrier mobility.

Keywords

Acoustic-phonon scattering · Alloy scattering · Carrier-carrier scattering · Carrier mobility · Deformation potential · Drag · Elastic scattering · Inelastic scattering · Intervalley scattering · Mobility · Momentum-relaxation time · Optical-phonon scattering · Piezoelectric scattering · Scattering centers · Surface-induced scattering

1 Types of Scattering Centers

At low electric fields[1] the carrier-distribution function is only little deformed at carrier scattering with respect to the equilibrium state, and the linearization discussed in Sect. 4.3 of ▶ Chap. 22, "Carrier-Transport Equations", is appropriate. The goal of this chapter is an estimation of the relaxation time for the return to equilibrium, which is a measure of the carrier mobility. When carriers are accelerated in an external electric field, their increased momenta and energies relax according to a multitude of scattering events. It is impossible to account for these events in a global fashion, and various approximations, which are different for different types of scattering centers, are required.

The carrier mobility introduced in Sect. 2.2 of ▶ Chap. 22, "Carrier-Transport Equations" is, by definition, the ratio of the electron group-velocity to the applied electric field. It may also be expressed by the ratio of the electrical conductivity (here discussed for electrons) to the carrier density,

$$\mu_n = \sigma_n/(e\,n), \tag{1}$$

where e is the electron charge. In semiconductors, the mobility is a more convenient quantity to characterize scattering events than the electrical conductivity, since the

[1]These are electric fields $F \ll v_{\mathrm{rms}}/\mu$, where carriers gain only a small fraction of kT between scattering events; this means $F \ll 5\times10^4$ V/cm for a typical semiconductor with $m_n = 0.05\ m_0$ and $\mu = 1000\ \mathrm{cm}^2/(\mathrm{Vs})$.

carrier density depends strongly on temperature in addition to sample purity. The mobility is closely related to the effective mass by

$$\mu_n = \frac{e}{m_n}\tau, \tag{2}$$

where τ is the time between two scattering events in the simple Drude approximation or a weighted relaxation time in the more refined analysis as discussed in Sects. 2.2 and 4.4 of ▶ Chap. 22, "Carrier-Transport Equations", respectively. Eq. 2 shows that the mobility depends (via m_n) on the band structure of the semiconductor, as well as on the scattering mechanisms affecting the carriers. Due to the dependence on band structure we must distinguish between semiconductors with a Fermi surface shaped as a single body and those with carriers in multiple valleys yielding multiple equi-energy surfaces at the band edge. Examples for the former are III-V compounds and for the latter group IV semiconductors.

Often the temperature dependence of the mobility is used to determine the dominant scattering mechanism. Figure 1 illustrates that this dependence may span several orders of magnitude and that it is a complex function due to the effect of various scattering types. A positive temperature coefficient of the mobility indicates scattering by charged impurities, while a negative coefficient is usually indicative of carrier scattering by phonons.

In this chapter we will introduce step by step some estimates of the various relaxation mechanisms, providing corresponding relations for the temperature-dependent carrier mobilities. The following sections are organized following

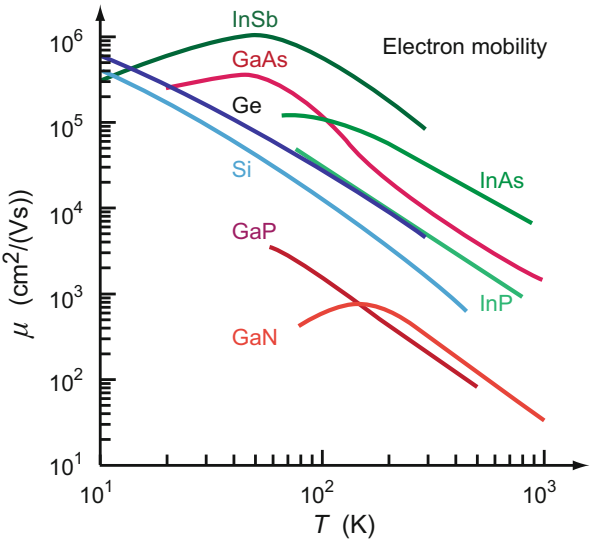


Fig. 1 Electron mobility of III–V and group IV semiconductors at low electron density

the nature of the scattering mechanisms. The considered types of scattering centers are

1. Phonons
   (a) Acoustic phonons, interacting with electrons via deformation potentials
   (b) Acoustic phonons with piezoelectric interaction
   (c) Optical phonons with deformation-potential interaction
   (d) Optical phonons, interacting with electrons via the polarization of the lattice
2. Coulomb potentials
   (a) Ionized-impurities
   (b) Dislocations
   (c) Carrier-carrier scattering
   (d) Carrier-plasmon scattering
3. Neutral defects
   (a) Neutral lattice defects
   (b) External surfaces
   (c) Microcrystallite boundaries
   (d) Intrinsic point defects
   (e) Local fluctuations in alloys
4. Transport in multivalley semiconductors and intervalley scattering

Each of these scattering centers will shorten the relaxation time τ. When estimating the effect of several types of scattering centers, each related to a specific τ_i, the total time between independent scattering events can be estimated from *Mathiessen's rule*[2]

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i}, \tag{3}$$

since the collision term in the Boltzmann equation (Eq. 83 of ▶ Chap. 22, "Carrier-Transport Equations") is additive. Consequently, the inverse mobilities calculated for single, independent types of scattering are also added to result in the inverse total carrier mobility – see also Debye and Conwell (1954)

$$\frac{1}{\mu} = \sum_i \frac{1}{\mu_i}. \tag{4}$$

[2]An error up to 20% can occur when applying Eq. 3 because of nonlinearities and interaction of different scattering events as shown by Rode and Knight (1971).

From Eq. 3, it is clear, however, that only the centers which influence the carriers *the most* need to be considered in an actual crystal under a given condition. Depending on the type of crystal bonding, crystal preparation (growth and treatments), doping, temperature, and other external influences – such as light, strain, and electric and magnetic fields – the predominant scatterer may vary from sample to sample. A review by Nag (1984) summarizes the different scattering mechanisms and gives tables for the expressions of the relaxation times; see also Seeger (2004). The calculation of scattering in first-principles approaches by 2017 is reviewed by Giustino (2017); progress in this rapidly developing field is impressive, more recent results are discussed by Poncé et al. (2020, 2021), along with an assessment of common approximations.

The scattering events described in the next sections leave the carrier within its valley (for intervalley scattering see Sect. 5). We will first assume that such scattering events take place with electrons near the Γ point.

2 Electron Scattering with Phonons

Most carriers have an energy near (3/2) kT. At low electric fields ($F \ll v_{\mathrm{rms}}/\mu$), they gain only a small fraction of additional energy compared to their thermal energy between scattering events. Carriers can interact with various types of phonons by absorbing or emitting a phonon. Before discussing this scattering in detail, a few general remarks will provide some overall guidance.

2.1 General Properties of Scattering with Phonons

The phonon-dispersion relation of most semiconductors shows that optical phonons have an energy larger than kT at room temperature, see Table 6 in ▶ Chap. 4, "Elasticity and Phonons"; therefore, only above 300 K an appreciable density of optical phonons is thermally created. Also the creation of optical phonons by scattering is unlikely since most of the carriers do not have sufficient energy. Therefore, carriers scatter predominantly with the lower-energy acoustic phonons, which are plentiful at room temperature. During such a scattering event, the electron energy is changed by only a small fraction (Eq. 132 of ▶ Chap. 22, "Carrier-Transport Equations"); this means, the scattering is an *essentially elastic* event.

In *semiconductors with a direct bandgap*, only phonons near the center of the Brillouin zone have a high probability of scattering. During such events, substantial changes in the direction of motion can occur: $\mathbf{k}$ can easily change its sign.

In *semiconductors with indirect bandgap*, electrons are in a valley of a relatively large $\mathbf{k}$ value. Elastic scattering with phonons of low energy and momentum tend to leave the electrons within their valley, with only small changes of their momentum, which changes their direction insignificantly. Intervalley scattering, which will be discussed in Sect. 5, requires a higher phonon momentum, i.e., higher-energy acoustic or optical phonons. Electrons near the surface of the Brillouin zone can, however, undergo Umklapp processes; in such a process the electron wavevector is

changed from an initial state $\mathbf{k}$ to a final state $\mathbf{k}'$ in the *neighboring* Brillouin zone, requiring only small values of $\mathbf{q}$ to reverse the direction of the electron motion during scattering. This occurs most easily in Ge, where the conduction-band minimum lies at this surface, rather than in Si, where the minima of $E(\mathbf{k})$ for conduction electrons lie at $0.8\pi/a$ – for more detail, see Sect. 5.1.

Elastic and Inelastic Scattering When electrons scatter with phonons, energy and momentum conservation laws must be fulfilled. At room temperature, electrons have an average energy of kT ($\cong$ 25 meV). Near the Γ point, optical phonons are of the same order of magnitude, while acoustic phonons start from $\hbar\omega = 0$ at $q = 0$ and have energies $\hbar\omega \ll kT$ in its vicinity. Therefore, almost no energy is exchanged when scattering with such acoustic phonons; hence, such a scattering is an *elastic scattering* event as noted above.

In contrast, only the faster electrons have enough energy to create an optical phonon, and thereby lose almost all of their excess energy. This type of scattering is an *inelastic scattering* event. One needs, however, to be careful with an assignment to phonons when scattering with phonons of higher momentum is considered. Since at higher $\mathbf{q}$ the energy of acoustic phonons approaches, in order of magnitude, that of optical phonons, *both* types of phonons cause *inelastic scattering.*

Phonon Generation and Annihilation When interacting with electrons, phonons can be generated, thereby cooling the electron ensemble, or annihilated, thereby heating it.

We now consider scattering that brings electrons with an initial momentum $\mathbf{k}$ to a final momentum $\mathbf{k}' = \mathbf{k} + \mathbf{q}$. The probability W for such transitions is proportional to the corresponding matrix elements (see Eq. 89 of ▶ Chap. 22, "Carrier-Transport Equations", with ΔV given by the deviation from the periodic potential), and the population of initial and final states. For the absorption or emission of a phonon (see Sect. 4.2 of ▶ Chap. 22, "Carrier-Transport Equations") we have

$$W(\mathbf{k},\mathbf{k}+\mathbf{q}) = \frac{2\pi}{\hbar}\left|M_{\mathbf{k},\mathbf{k}+\mathbf{q}}\right|^2 f(\mathbf{k})[1-f(\mathbf{k}+\mathbf{q})]\,\{f(\mathbf{q})\;\delta\left[E(\mathbf{k}+\mathbf{q})-E(\mathbf{k})-\hbar\omega_{\mathbf{q}}\right] + [1-f(-\mathbf{q})]\;\delta\;\left[E(\mathbf{k}+\mathbf{q})-E(\mathbf{k})+\hbar\omega_{\mathbf{q}}\right]\}, \tag{5}$$

where $M_{\mathbf{k},\mathbf{k}+\mathbf{q}}$ is the matrix element for the scattering event, f is the distribution function, and δ is the Dirac delta function. After integration over all possible transitions, we obtain the collision term of the Boltzmann equation (see Eq. 87 of ▶ Chap. 22, "Carrier-Transport Equations")

$$\left(\frac{\partial f(\mathbf{k})}{\partial t}\right)_{\text{coll}} = \frac{V}{8\pi^3}\int W(\mathbf{k},\mathbf{k}')\;dk \tag{6}$$

This integration, however, is difficult to perform for complex lattice oscillations. Therefore, the collision term is evaluated for one specific type of oscillation at a time, when approximations can easily be introduced. A simple example is the scattering of

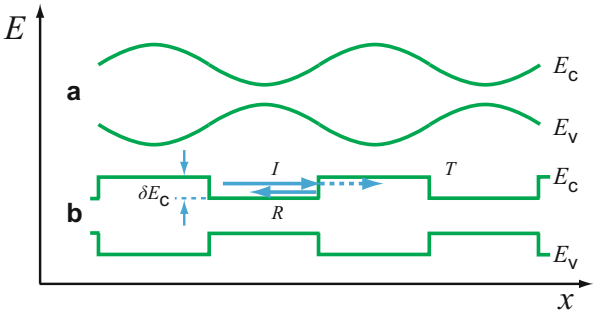


Fig. 2 (**a**) Undulation of the band edges E_c and E_v due to "pressure waves", produced, e.g., in the long-wavelength range of longitudinal acoustic phonons. (**b**) Step-like approximation of these undulations

electrons on longitudinal acoustical phonons, which is discussed in the following section.

2.2 Scattering with Acoustic Phonons

2.2.1 Longitudinal Acoustic Phonon Scattering

The interaction of electrons with longitudinal acoustic phonons can be analyzed in a variety of models; for reviews, see Mitra (1969), Nag (1980), Madelung (1981), Zawadzki (1982), and Seeger (2004). The importance of acoustic-phonon scattering is pointed out by Liu et al. (2017) and also discussed by Protik and Kozinsky (2020).

It is instructive to introduce a classical approach with acoustic waves. In the acoustic branch at longer wavelengths, the lattice is alternatingly compressed and dilated. Consequently, the width of the bandgap is modulated: it widens with compression – see Sect. 2.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids". Electrons are scattered by the modulation of the band edge as illustrated in Fig. 2.

In a further simplified model, the deformation wave is approximated by a potential step in both bands (Fig. 2b; see also McKelvey 1966). An electron wave impinging on such a step is partially transmitted and partially reflected. The reflection probability is estimated from the solution of the Schrödinger equation as the difference between the impinging wave (index I) and the transmitted wave (index T). Their energies are given by

$$E_I = \frac{\hbar^2 k_I^2}{2m_n} \quad \text{and} \quad E_T = \frac{\hbar^2 k_T^2}{2m_n} = E_I - \delta E_c. \tag{7}$$

From continuity of the wavefunction and its derivatives at each step, we obtain for $k_I \cong k_T$ for the reflection probability $R \cong [(k_I - k_T)/(k_I + k_T)]^2$, which yields

$$R \cong \left(\frac{m_n \delta E_c}{2\hbar^2 k_I^2}\right)^2, \tag{8}$$

Table 1 Deformation potentials (eV). The subscripts u and d signify pure shear strain and diagonal component of the deformation-potential tensor, respectively

Material				
Ge	Ξ_u^X	Ξ_d^X	Ξ_u^L	Ξ_d^X
	10.4	0.53	16.4	−6.4
GaAs	$\Xi_u^{\tau(111)}$	$\Xi_d^{\tau(111)}$		$\Xi_d^{\tau(100)}$
	16.5	−8		−11.2

where δE_c is the step-height and k_I is the wavevector of the impinging electron. The step-height is related to the lattice compression by

$$\delta E_c = -\Xi \frac{\delta V}{V} \tag{9}$$

with Ξ as the *deformation potential,*[3] V as the volume, and δV as its change, which is related to the thermal energy by a simple thermodynamic analogy:

$$\frac{1}{2}\delta p\, \delta V = c\, k T, \tag{10}$$

where c is a proportionality factor on the order of 1. Replacing the pressure increment δp from the compressibility (κ) relation

$$\kappa = \frac{1}{V}\frac{\delta V}{\delta p}, \tag{11}$$

we obtain for the probability of reflection from Eqs. 8 to 11:

$$R \cong \left(\frac{m_n}{2\hbar^2 k_I^2}\right)^2 \frac{c\,\kappa\, k T}{V}\Xi^2. \tag{12}$$

The probability of reflection can be connected with a mean free path λ by $\lambda = l/R$, with l as the length of the sample (of volume $V = l^3$). We obtain for $k_I \cong \pi/l$, i.e., for long-wavelength acoustic phonons:

$$\lambda \cong \frac{h^4}{4m_n^2\, c\,\kappa\, k T\, \Xi^2}. \tag{13}$$

[3]The deformation potential is defined as the change in bandgap energy per unit strain and is typically on the order of 10 eV. For a listing, see Table 1. It should be noted that deformation potentials are generally tensors with components Ξ_{ij}, relating the shift of the band edge δE_c to the components of the strain tensor e_{ij} (see ▶ Chap. 4, "Elasticity and Phonons"): $\delta E = \sum_{ij} \Xi_{ij}\, e_{ij}$; Eq. 9 hence provides only an average in anisotropic media.

A more rigorous treatment (Bardeen and Shockley 1950) yields a similar result:

$$\lambda \cong \frac{h^4 c_l}{m_n^2 k T \Xi^2}, \tag{14}$$

where c_l is the elastic constant for longitudinal deformation in cubic crystals: $c_l = c_{11}$ for strain in the $\langle 1\ 0\ 0 \rangle$ direction, $c_l = \frac{1}{2}(c_{11} + c_{12} + c_{44})$ for strain in the $\langle 1\ 1\ 0 \rangle$ direction, and $c_l = \frac{1}{3}(c_{11} + 2c_{12} + 4c_{44})$ for strain in the $\langle 1\ 1\ 1 \rangle$ direction. Here, c_{ij} are components of the elastic stiffness tensor, see Sects. 1.1 and 1.2 of ▶ Chap. 4, "Elasticity and Phonons". Assuming an energy-independent $\lambda = \overline{\lambda}$ using $\overline{\tau} = \overline{\lambda}/v_{\text{rms}}$, and replacing v_{rms} with Eq. 19 of ▶ Chap. 22, "Carrier-Transport Equations", we obtain for the electron mobility due to acoustic phonon scattering

$$\mu_{n,\,\text{ac phon}} = \frac{\sqrt{8\pi}}{3} \frac{e h^4 c_l}{m_n^{5/2} (kT)^{3/2} \Xi^2} \tag{15}$$

$$= 6.1 \times 10^3 \frac{c_l}{10^{12}\,(\text{g cm/s}^2)} \left(\frac{m_0}{m_n}\right)^{5/2} \left(\frac{300\ \text{K}}{T}\right)^{3/2} \left(\frac{\text{eV}}{\Xi}\right)^2 \quad (\text{cm}^2/\text{Vs}).$$

Eq. 15 yields a $T^{-3/2}$ dependence of the mobility at higher temperature, where this type of scattering is dominant. This is generally observed for semiconductors with a direct bandgap, see Fig. 1. In indirect semiconductors intervalley scattering has a significant effect leading to a different dependence,[4] see Sect. 5.2.

The effective mass used in Eqs. 12 to 15 requires the proper mix of density-of-state and mobility effective masses:

$$m_n^{5/2} = m_{n\text{ds}}^{3/2} m_{n\mu}. \tag{16}$$

The deformation potential used here has only slowly varying components in space. Another approach, suggested by Ginter and Mycielski (1970), contains a part of the potential varying with the lattice periodicity, which is more appropriate for shorter-wavelength phonons. This approach is a more general one; still, it gives similar results in a number of examples.

The mobility is determined by the effective mass and the momentum-relaxation time, $\mu_n \propto \overline{\tau}_{\text{m}}/m_n$, see Eq. 2 and Sect. 4.6 of ▶ Chap. 22, "Carrier-Transport Equations". Assuming parameters of a typical semiconductor, Nag (1984) calculated the relaxation time for scattering with acoustic phonons as shown in Fig. 3, indicating a strong dependence on carrier density; comparable dependences are found for alloy scattering discussed in Sect. 4.4. The normalizing constants τ_{ac} and τ_{pz} at the

[4]The experimentally observed exponent of T is -1.67 for Ge (Conwell 1952) and not -1.5. The exponent of T for Si is still larger ($\cong 2.4$). Inserting actual values for Si ($c_l = 15.6{\cdot}10^{10}$ N/m^2, $m_n = 0.2\ m_0$, and $\Xi = 9.5$ eV), one obtains $\mu_n = 5900$ cm^2/Vs, a value that is larger by a factor of ~4 than the measured $\mu_n = 1500$ cm^2/Vs at 300 K.

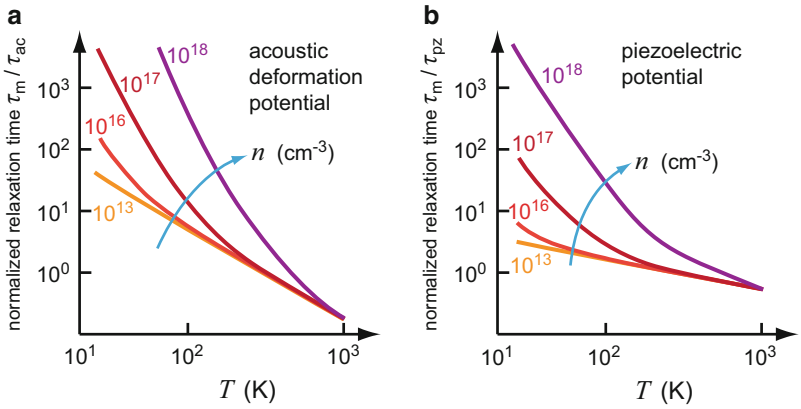


Fig. 3 Calculated momentum-relaxation time for (**a**) scattering by acoustic phonons, and (**b**) scattering by piezoelectric phonons in a semiconductor with $m_n = 0.05\ m_0$, $\varepsilon = 15$, and varied carrier density with n as a function of the lattice temperature. After Nag (1984)

ordinates in Fig. 3 refer to the respective scattering mechanisms and represent the values of τ_m at $T = 300$ K including all controlling physical constants.

2.2.2 Acoustic Phonon Scattering with Piezoelectric Interaction

In piezoelectric crystals, ion oscillations cause a dipole moment, which interacts with carriers rather effectively. A dipole moment can be generated by alternating lattice compression and dilatation, which in turn are caused by longitudinal acoustic phonons. These create an electric field parallel to the propagation direction which has a similar interaction with carriers, although slightly stronger than the acoustic deformation potential discussed above. The resulting mobility shows a somewhat similar behavior (Meyer and Polder 1953):

$$\begin{aligned}\mu_{n,\,\text{pe ac phon}} &= \frac{16\sqrt{2\pi}}{3}\frac{\hbar^2\varepsilon\varepsilon_0}{e\,m_n^{3/2}K^2(kT)^{1/2}} \\ &= 1.5\times10^4\,\frac{\varepsilon}{10}\left(\frac{m_0}{m_n}\right)^{3/2}\left(\frac{10^{-3}}{K}\right)^2\left(\frac{300\ \text{K}}{T}\right)^{1/2}\quad\left(\text{cm}^2/\text{Vs}\right).\end{aligned} \tag{17}$$

Here K is the electromechanical coupling constant[5], which for most semiconductors is on the order of 10^{-3}. At low temperatures we obtain a $T^{-1/2}$ dependence that clearly differs from the $T^{-3/2}$ relation for acoustic deformation potential scattering – see Fig. 7. Piezoelectric phonon scattering can be an important scattering mechanism

[5] K^2 can be expressed as the ratio of the mechanical to the total work in a piezoelectrical material: $K^2 = \left(e_{pz}^2/c_l\right)/\left(\varepsilon\varepsilon_0 + e_{pz}^2/c_l\right)$, with e_{pz} the piezoelectric constant (which is on the order of 10^{-5} As/cm^2), and c_l the longitudinal elastic constant (relating the tension T to the stress S and the electric field F as $T = c_l\,S - e_{pz}\,F$).

in samples with a low density of ionized impurities, which otherwise dominate electron scattering; see Zawadzki (1982), Nag (1984), and Seeger (2004). In GaAs, a comparable contribution of piezoelectric and deformation-potential scattering was found to account for a 15% reduction of the intrinsic electron mobility at 300 K (Liu et al. 2017).

2.3 Scattering with Optical Phonons

2.3.1 Optical Phonon Scattering in Nonpolar Compounds

Low-energy electron scattering with optical phonons is predominantly elastic. This process can be understood as the annihilation of an optical phonon to create a high-energy electron, which in turn immediately creates an optical phonon in a highly probable transition. Therefore, the electron energy is conserved in the turnaround, but not its momentum. When electrons have accumulated sufficient energy to *create* optical phonons, the scattering becomes very effective and is *inelastic*; this will be discussed in ▶ Chap. 24, "Carrier Scattering at High Electric Fields".

Optical phonon scattering in elemental nonpolar semiconductors couples both longitudinal and transverse optical modes with the scattering electron (Boguslawski 1975). It can be estimated by using a deformation-potential formalism for longitudinal optical phonons (see also Conwell 1967). We obtain an electron mobility

$$\mu_{n,\text{opt phon}} = \frac{4\sqrt{2\pi}}{3} \frac{e\hbar^2\rho\sqrt{k\Theta}}{m_n^{5/2} D_\text{o}^2} \phi(T) \tag{18}$$

where ρ is the density of the semiconductor, Θ is the Debye temperature, and D_o is the optical deformation potential (Meyer 1958),

$$\delta E_\text{c} = D_\text{o}\,\delta r, \tag{19}$$

with δr as the change in the interatomic distance, and $\phi(T)$ as a function that contains the temperature dependence of μ and the density of phonons (Seeger 2004). At low temperatures, $\phi(T)$ is large (typically 10^4 to 10^5 at $T = \Theta/10$) and decreases rapidly to a value on the order of 1 near the Debye temperature Θ. The actual form of $\phi(T)$ depends on the approximation used, the function shows a nearly exponential decrease with increasing temperature for $T < \Theta/3$.

Deformation potentials for various semiconductors are listed in Table 2. These deformation potentials of states i are defined by the *hydrostatic strain*:

$$a_i = \frac{dE_i}{d\ln V} = \frac{dE_i}{\frac{1}{V}dV} = B\frac{dE_i}{dp} \tag{20}$$

with B as the bulk modulus, or by *shear strain* in cubic crystals along [111] or [100]

Table 2 Deformation potential at Γ points (in eV)

Solid	$a(\Gamma_1^c)$	$a(\Gamma_1^c) - a(\Gamma_{15}^v)$	b	d	$\lvert d_o \rvert$
C					90
Si	−15.3	−10.0	−2.2	−5.1	40
Ge	−19.6	−12.6	−2.3	−5.0	34
AlSb		−5.9	−1.4	−4.3	37
GaP	−19.9	−9.3	−1.8	−4.5	44
GaAs	−17.5	−9.8	−2.0	−5.4	48
GaSb		−8.3	−1.8	−4.6	32
InP	−18.0	−6.4	−2.0	−5.0	35
InAs		−6.0	−1.8	−3.6	42
InSb	−14.6	−7.7	−2.0	−5.0	39
ZnS	−14.5	−4.0	−0.62	−3.7	4
ZnSe	−11.5	−5.4	−1.2	−4.3	12
ZnTe	−9.5	−5.8	−1.8	−4.6	23
CdS		−3.1			
CdSe		−3.0			
CdTe	−9.5	−3.4	−1.2	−5.4	22
CuCl			−0.7	0.43	7
CuBr			−0.25	−0.65	3.8
CuI			−0.64	−1.4	1.1

Source: Blacha et al. (1984)

$$b = \frac{\delta E}{6\ \varepsilon_{[100]}} \text{ or } d = \frac{\delta E}{2\sqrt{3}\ \varepsilon_{[111]}}, \tag{21}$$

where ε is the strain in the direction indicated by the index. $|d_0|$ is defined by splitting of the Γ_{15} state in the absence of the spin-orbit interaction,

$$d_0 = \frac{\delta E\, a_0}{u} \tag{22}$$

with δE as the observed energy shift, a_0 as the lattice constant, and u as the atomic displacement. The parameter b is related to the optical deformation potential D_o by $D_o = -(3/2)b$.

Numerically, the mobility due to optical phonon scattering (Eq. 18) can be expressed as

$$\mu_{n,\text{opt phon}} = 1.77 \times 10^3 \frac{\left(\dfrac{\rho}{\text{g cm}^{-3}}\right)\left(\dfrac{\Theta}{300\ \text{K}}\right)^{1/2}}{\left(\dfrac{m_n}{m_0}\right)^{5/2}\left(\dfrac{D_o}{10^8\ (\text{eV/cm})}\right)^2}\ \phi(T) \quad \left(\text{cm}^2/\text{Vs}\right). \tag{23}$$

This mobility shows a sufficiently low value near and above the Debye temperature, where $\phi(T) \leq 1$, and becomes the determining factor in high-purity semiconductors.

Nonpolar optical phonon scattering is the only electron-optical phonon interaction in nonpolar semiconductors, such as Si and Ge. It is important for Γ_8 bands (which are usually the valence bands of cubic semiconductors), and, as was pointed out by Harrison (1956), vanishes for the Γ_6 band (usually the conduction band). Therefore, e.g., it is unimportant for *n*-type InSb but important for hole scattering in *p*-type InSb (Bir and Pikus 1974).

2.3.2 Optical Phonon Scattering in Polar Semiconductors

The scattering of carriers with longitudinal optical phonons (*Fröhlich interaction*) in an (at least partially) ionic lattice has a larger influence than the deformation-potential interaction on the carrier mobility due to the larger dipole moment associated with such a lattice vibration. This causes an induced electric field F proportional to the polarization P:

$$F = -\frac{P}{\varepsilon\,\varepsilon_0}. \tag{24}$$

The polarization can be obtained from

$$P = -\frac{e_C\,\delta_a}{V_0} \tag{25}$$

where δ_a is the change in the interatomic distance, V_0 is the volume of the unit lattice cell, and e_C is the Callen effective charge (Eq. 54 in ▶ Chap. 11, "Photon–Phonon Interaction").

The mobility was calculated by Ehrenreich (1961) using a variational method of Howarth and Sondheimer (1953), which accounts for the inelastic scattering. It yields for polar-optical scattering at temperatures well below the Debye temperature ($T \ll \Theta$) an electron mobility (Seeger 2004)

$$\mu_{n,\text{ion,opt phon}} = \frac{e\hbar}{2m_n\,\alpha_c\,k\,\Theta}\ \exp\left(\frac{\Theta}{T}\right) \tag{26}$$

$$= 2.6\times 10^5\ \frac{\exp\left(\dfrac{\Theta}{T}\right)}{\alpha_c\left(\dfrac{m_n}{m_0}\right)\left(\dfrac{\Theta}{\mathrm{K}}\right)}\qquad \left(\mathrm{cm}^2/\mathrm{Vs}\right),$$

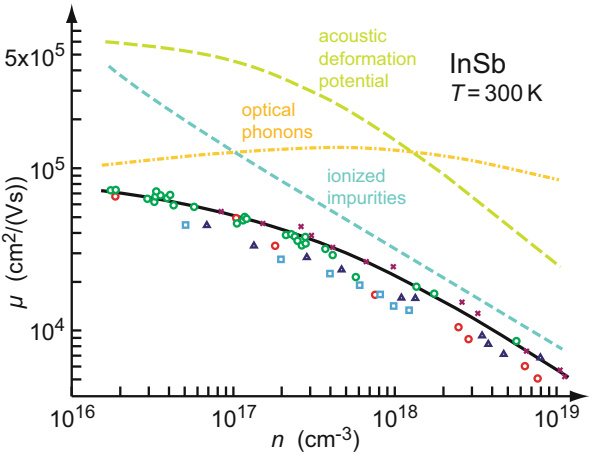


Fig. 4 Electron mobility of *n*-type InSb at 300 K as a function of the free-electron density. Dashed curves indicate the corresponding theoretical contributions. After Zawadzki (1972)

where α_c is the coupling constant given in Eq. 9 of ▶ Chap. 22, "Carrier-Transport Equations". The Debye temperature Θ is used here to account for the LO phonon energy at $\mathbf{q} = 0$ with $\hbar\omega_{LO} = k\,\Theta_{LO} \cong k\,\Theta$.

The mobility given by optical phonon scattering decreases linearly with increasing coupling constant. The mobility increases exponentially with decreasing temperature due to optical phonon freeze-out at lower temperatures. At temperatures $T \geq \Theta$ a different approximation yields:

$$\mu_{n,\,\text{ion, opt phon}} = 1.5\frac{e\hbar}{2m_n\,\alpha_c\,k\,\Theta}\,\sqrt{\frac{T}{\Theta}}\exp\left(\frac{\Theta}{T}\right) \cong 1.6\times 10^5\,\frac{\sqrt{\dfrac{T}{\Theta}}\exp\left(\dfrac{\Theta}{T}\right)}{\alpha_c\left(\dfrac{m_n}{m_0}\right)\left(\dfrac{\Theta}{\text{K}}\right)}. \tag{27}$$

Zawadzki and Szymanska (1971) used a Yukawa-type screened potential that results in a reduced effectiveness of the optical phonon scattering. At higher temperatures they obtained for parabolic bands

$$\mu_{n,\,\text{ion, opt phon}} = \frac{\sqrt{2}}{8\pi}\,\frac{M_r a^3 (k\,\Theta)^2}{e\,e_C^2 m_n^{3/2}\,kT}\,\frac{\sqrt{E-E_c}}{F_{opt}}, \tag{28}$$

where F_{opt} is a screening parameter[6] that depends on the Debye screening length and is on the order of 1 (Zawadzki 1982); M_r, a^3, and e_C are the reduced mass of the ions, the volume of the unit cell, and the Callen effective ionic charge, respectively. F_{opt} results in an increased mobility by a factor of 3.5 for InSb at room temperature and

[6]Here, $F_{opt} = 1 + \frac{2}{\beta}\ln(\beta+1) + 1/(\beta+1)$, With $\beta = (2|\mathbf{k}|L_D)^2$; L_D is the Debye length given in Eqs. 94 of ▶ Chap. 10, "Interaction of Light with Solids", and 49 of ▶ Chap. 16, "Crystal Interfaces".

$n = 10^{19}$ cm^{-3}. When the Fermi level is shifted into the conduction band, the mobility becomes explicitly electron-density-dependent:

$$\mu_{n,\,\mathrm{ion,\,opt\ phon}} = \sqrt{\frac{3}{16\pi}}\,\frac{M_r a^3 (k\Theta)^2 \hbar}{e\, e_C^2 m_n(E_F)\, kT}\;\;\frac{\sqrt[3]{n}}{F_{\mathrm{opt}}}. \tag{29}$$

Figure 4 shows the electron mobility of InSb as a function of the *electron density* at 300 K and identifies the most important branches of scattering by longitudinal optical phonons at low densities, and by ionized impurities at high densities – see Sect. 3.1; scattering by acoustic phonons (Sect. 2.2.1) is of lesser importance. For the dependence on temperature see Fig. 7.

Considering only the Fröhlich interaction in the electron-phonon interaction for computing mobilities around room temperature proved to yield a good estimate for the II-VI semiconductors ZnSe, CdTe, ZnTe, and CdSe; in these direct-bandgap solids the relevant electronic states are near the Γ point and LO phonons provide the dominant scattering mechanism (Vukmirović 2021).

3 Scattering at Coulomb Potentials

When scattering occurs on *charged* defects, the carriers interact with the *long-range Coulomb forces,* resulting in a substantially larger scattering cross-section that is typically on the order of 10^{-13} cm^2.

3.1 Ionized-Impurity Scattering

The original scattering analysis on charged particles was done by Rutherford (1911) for α particles and is easily adapted to carriers scattered by ions of charge z in a solid. This is accomplished after introducing the screening of this potential in the solid by using the dielectric constant $\varepsilon_{\mathrm{stat}}$ and considering Bloch electrons by using the carrier effective mass m_n. A differential cross section for the scattering of an electron of

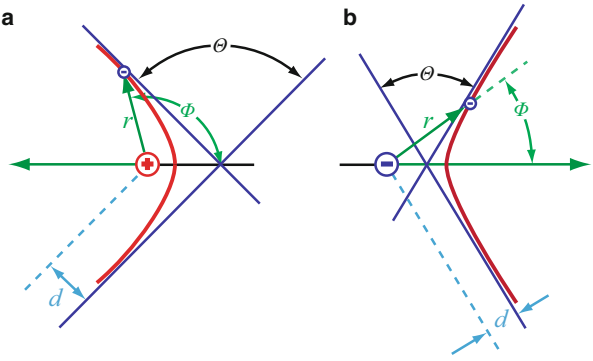


Fig. 5 Electron trajectories for (**a**) a Coulomb-attractive and (**b**) a Coulomb-repulsive scattering center

velocity v_{rms} under an angle θ into an element of solid angle $d\Omega = 2\pi \sin\theta\, d\theta$ (see Fig. 5) is then given (Leighton 1959) by

$$s_n(\theta)d\Omega = \left(\frac{Z\ e^2}{8\pi\varepsilon\varepsilon_0 m_n v_{rms}^2}\right)^2 \sin^{-4}\frac{\theta}{2}d\Omega \tag{30}$$

with

$$\frac{\theta}{2} = \tan^{-1}\ \frac{Z\ e^2}{4\pi\varepsilon\varepsilon_0 d\, m_n v_{rms}^2}, \tag{31}$$

which shows a rather slow decrease of the scattering angle θ with increasing minimum distance d from the center – see Fig. 5. In order to totally randomize the angle after collision, however, only a fraction of $(1 - \cos\theta)$ of all scatterings describes the number of *memory-erasing collisions* (Eq. 125 of ▶ Chap. 22, "Carrier-Transport Equations"):

$$dn = N_{ion} v_{rms} s_n(\theta)\ \ (1 - \cos\theta)d\Omega = d\left(\frac{1}{\tau_m}\right). \tag{32}$$

This number is inversely proportional to the differential momentum-relaxation time.

Integration over all angles θ to obtain the total relaxation time requires a cutoff in order to avoid an infinite result, since $s_n(\theta = 0) = \infty$. Conwell and Weisskopf (1950) assumed that the closest distance d of the trajectory from the center to be considered must be smaller than a maximum distance, $d < d_{max} = \frac{1}{2}N_{ion}^{1/3}$, given by the average distance $2d_{max}$ between nearest-neighbor ionized centers in the crystal. This yields for the time between collisions of carriers with velocity v_{rms}:

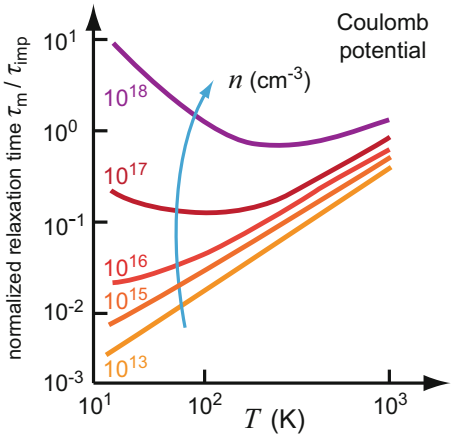


Fig. 6 Momentum-relaxation time for scattering by ionized impurities with the carrier density $n = 10^{13}$ to 10^{18} cm^{-3} for curves 1 to 5 as the family parameter; $m_n = 0.05\ m_0$ and $\varepsilon = 15$ are assumed. The normalizing constant τ_{imp} refers to τ_m at $T = 300$ K. After Nag (1984)

$$\tau = \frac{(4\pi\,\varepsilon_{\text{stat}}\,\varepsilon_0)^2 m_n^2\, v_{\text{rms}}^3}{2\pi Z^2 e^4\;\; N_{\text{ion}}}\;\; \frac{1}{\ln\left[1+\left(\frac{\varepsilon_{\text{stat}}\,\varepsilon_0\, m_n\, v_{\text{rms}}^2}{2\,Z\,e^2\,N_{\text{ion}}^{1/3}}\right)^2\right]}. \tag{33}$$

After averaging, we obtain the momentum-relaxation time τ_{m}. The dependence of τ_{m} on the temperature is given in Fig. 6.

Using $v_{\text{rms}} = \sqrt{3\,kT/m_n}$ and the *averaged* momentum-relaxation time, we obtain for the mobility with $\mu_n = e\overline{\tau}_{\text{m}}/m_n$:

$$\mu_{n,\text{ion imp}} = \frac{8\sqrt{2}(4\pi\,\varepsilon_{\text{stat}}\,\varepsilon_0)^2 (kT)^{3/2}}{\pi^{3/2} Z^2 e^3\, m_n^{1/2}\;\; N_{\text{ion}}}\;\; \frac{1}{\ln\left[1+\left(\frac{3\varepsilon_{\text{stat}}\,\varepsilon_0\,kT}{2\,Z\,e^2\,N_{\text{ion}}^{1/3}}\right)^2\right]} \tag{34}$$

The mobility in this Conwell-Weisskopf dependence increases with temperature $\propto T^{3/2}$, i.e., faster electrons are less effectively scattered; the logarithmic dependence is usually neglected. As expected, μ_n decreases inversely with the density of scattering centers and with the square of their charge $e\,Z$. A conversion of Eq. 34, using the definition of a scattering cross-section, reveals that, when disregarding the logarithmic term, s_n is given by

$$s_n = \frac{\pi^{3/2} Z^2 e^4}{8\sqrt{6}(4\pi\,\varepsilon_{\text{stat}}\,\varepsilon_0)^2 (kT)^2} = \pi r_{\text{ion}}^2, \tag{35}$$

where r_{ion} is the "scattering radius" of the ion. This scattering radius can be compared with r_{C}, the radius of a Coulomb well at a depth of kT:

$$\frac{Z e^2}{4\pi\,\varepsilon_{\text{stat}}\,\varepsilon_0\, r_{\text{C}}} = kT \;\text{ or }\; r_{\text{C}} = \frac{Z e^2}{4\pi\,\varepsilon_{\text{stat}}\,\varepsilon_0\, kT}. \tag{36}$$

We now relate the scattering radius defined in Eq. 35 with the above defined Coulomb radius by

$$r_{\text{ion}} = c_{\text{corr}}\, r_{\text{C}} \tag{37}$$

where c_{corr} is a correction factor. By comparison with Eq. 34, now including the logarithmic term, we obtain:

$$c_{\text{corr}} = 2\left(\frac{2\pi}{3}\right)^{1/4} \frac{1}{\ln\left[1+\left(\frac{3\varepsilon_{\text{stat}}\,\varepsilon_0\,kT}{2\,Z\,e^2\,N_{\text{ion}}^{1/3}}\right)^2\right]} \tag{38}$$

which is on the order of 1. We thereby see that, except for this correction factor, the scattering cross-section is equal to the square of the Coulomb radius at $E = E_{\text{c}} - kT$.

Therefore, the cross section decreases $\propto 1/T^2$, or the increase in mobility is directly related to a decrease of r_C with T. This relationship is closely associated with the trapping of a charge at a Coulomb-attractive center when energy is dissipated – see Sect. 2 of ▶ Chap. 30, "Carrier Recombination and Noise"). For such inelastic events to occur, the carrier must penetrate to $r < r_C$.

With the Coulomb radius introduced above, the Conwell-Weisskopf formula can be rewritten as the classical scattering relation

$$\mu_{n,\,\text{ion imp}} = \frac{e}{m_n}\,\frac{c_{\text{corr}}}{v_n\, s_n\; N_{\text{ion}}}. \tag{39}$$

A somewhat refined approach was suggested by Brooks and Herring (Brooks 1955). They assumed a cut-off in the integration over θ by replacing the Coulomb potential with a screened Yukawa potential and used the Debye length L_D as the screening length. In addition, they replaced the electron density in Eq. 94 of ▶ Chap. 10, "Interaction of Light with Solids", with $n + (n + N_a)[1 - (n + N_a)/N_d]$, considering partial compensation. The ensuing result is similar to Eq. 34:

$$\mu_{n,\text{ion imp}} = \frac{2^{7/2}(4\pi\,\varepsilon_{\text{stat}}\,\varepsilon_0)^2(kT)^{3/2}}{\pi^{3/2}Z^2\,e^3\,m_n^{1/2}\;\;N_{\text{ion}}}\;\;\frac{1}{\ln(1+\beta^2)-\beta^2(1+\beta^2)}, \tag{40}$$

$$\text{with}\;\; \beta = \frac{\sqrt{8m_n}}{\hbar}\sqrt{3\,kT}L_D.$$

Comparing the cut-off by Brooks and Herring with that of Conwell and Weisskopf, we observe a simple relation:

$$r_{\text{corr}}L_D^2 = n^{-1}, \tag{41}$$

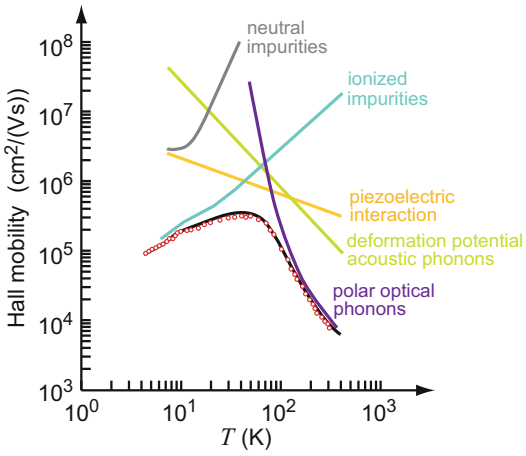


Fig. 7 Electron mobility in *n*-type GaAs as measured by Stillman et al. (1970). Curves without data points indicate the corresponding theoretical contributions. After Fletcher and Butcher (1972)

which introduces the carrier-density dependence into the scattering. For a review, see Zawadzki (1982), and Chattopadhyay and Queisser (1981). For silicon, scattering rates and electron mobilities limited by phonon and ionized-impurity scattering calculated from first principles indicate a dominance of the latter for donor densities above 10^{16} cm^{-3} (Restrepo et al. 2009); for the additional effect of plasmons see Sect. 3.4.2, for neutral defects see Sect. 4.1. An efficient first-principles approach for computing impurity scattering is discussed by Deng et al. (2020). The dependence of the mobility on the electron density is given in Fig. 4 (Sect. 2.3.2).

The effect of various scattering mechanisms on the electron mobility is shown in Fig. 7 as a function of the temperature. It indicates major scattering at ionized impurities (low *T,* see Sect. 3.1) and by longitudinal optical phonons (high *T,* see Sect. 2.3.2); whereas scattering by acoustic phonons, piezoelectric scattering, and scattering at neutral impurities are of lesser importance, in this order.

Ionized-impurity scattering can be suppressed in modulation-doped heterostructures, where the impurities are spatially separated from the mobile electrons; the latter are confined in a high-mobility two-dimensional electron gas (2DEG), see Sect. 1.2 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors".

Quantum Corrections for Ion Scattering Several assumptions of the Brooks-Herring approximation are often not fulfilled (Moore 1967). Corrections obtained by dropping these assumptions may be expressed in a linearized form:

$$\mu = \mu_0 \frac{1}{1 + \delta_\mathrm{B} + \delta_\mathrm{m} + \delta_\mathrm{d}} \tag{42}$$

with three correction contributions:

1. The Born approximation, used to estimate the scattering probability, requires $|\mathbf{k}|\, L_\mathrm{D} \gg 1$, with the average wavevector $|\mathbf{k}| = m_n\, v_\mathrm{rms}/\hbar$. When this condition is not fulfilled, a Born correction component δ_B is introduced with

$$\delta_\mathrm{B} = \frac{2Q(\beta)}{\mathbf{k}^2 L_\mathrm{D}\, a_\mathrm{qH}} \tag{43}$$

with $\beta = 4L_\mathrm{D}^2\, |\mathbf{k}|^2$, and $Q(\beta)$ a slowly varying function: $0.2 < Q < 0.8$. One estimates $0.1 < \delta_\mathrm{B} < 1$ for ion densities between 10^{16} and 10^{19} cm^{-3}. It is a more important correction for lower temperatures ($T < 100$ K, Moore 1967).

2. When coherent scattering occurs from more than one ion, the mean free path becomes comparable to the screening length. The multiple-scattering correction factor δ_m has been estimated by Raymond et al. (1977) and is of minor importance at low temperatures.

3. A dressing effect can be expected to take care of the chemical individuality of the scattering center. The effect of the electron wavefunction of the scattering ion is rather small: δ_d is about 30–50% of δ_m. However, the defect individuality is observed at higher impurity densities for InSb doped with Se or Te (Demchuk and Tsidilkovskii 1977) and for Si or Ge doped with As or Sb (Morimoto and Tani

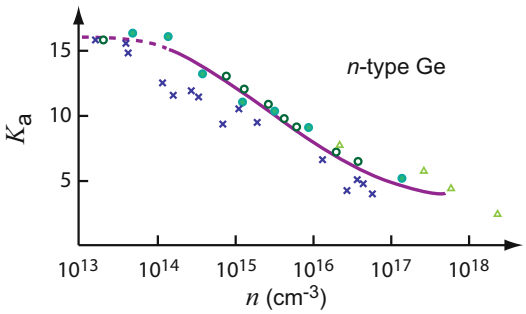


Fig. 8 Mobility-anisotropy factor K_a measured at 77 K in *n*-type Ge; data after Baranskii et al. (1975) calculated curve after Dakhovskii and Mikhai (1965)

1962). It is suggested that the stress field surrounding impurities of different sizes is the reason for the individuality of the scattering probability of such ions rather than central-cell potential corrections (Morgan 1972).

3.2 Coulomb Scattering in Anisotropic Semiconductors

Carrier scattering at an ionized impurity is sensitive to the anisotropy of the band structure, since low-angle scattering events dominate and interfere with randomizing electron velocities (Herring and Vogt 1956). One also requires that the screening length be substantially smaller than the mean free path in order to prevent successive collisions from occurring in the same defect-potential region.

The anisotropy of the effective mass and of the density of states, which are largest along the long axes of the valley ellipsoids, influences the mobility (Boiko 1959). The mobility can be expressed as (Samoilovich et al. 1961)

$$\mu = e\left(\frac{2}{3}\frac{\langle\tau_\perp\rangle}{m_\perp^*} + \frac{1}{3}\frac{\langle\tau_\parallel\rangle}{m_\parallel^*}\right) \tag{44}$$

with $m_\parallel^*/m_\perp^* = 19$ or 5.2 for Ge or Si, respectively.

With a screened Yukawa potential (Sect. 1.3.5 of ▶ Chap. 10, "Interaction of Light with Solids"), the anisotropy factor $K_a = m_\perp^*/m_\parallel^*$ decreases with increasing carrier density as shown in Fig. 8. A smaller cross section of the scattering centers with higher carrier densities is caused by a decreasing Debye length $L_D \propto 1/\sqrt{n}$; it also renders ion scattering more randomizing with increasing carrier density n.

The isotropic matrix-elements for ionized impurity, polar and non-polar electron-phonon, and piezoelectric scattering were recently extended for calculating anisotropic crystals, yielding reasonable agreement of mobility against experiment for many semiconductors (Ganose et al. 2021).

3.3 Scattering at Dislocations

The edge dislocation presents a major disturbance in carrier transport within the lattice because of the charging of the reconstructed core states (see Sect. 4.2 of ▶ Chap. 15, "Crystal Defects") and the surrounding stress field. Another disturbance can be expected from the charges induced in piezoelectric crystals and can be caused by the strain field around edge and screw dislocations (Levinson 1966, Pödör 1970). These core states attract electrons or holes, depending on the position of the Fermi level, and are screened by free carriers surrounding the line as a cylinder of a radius equal to the Debye length $L_D = \sqrt{\varepsilon \varepsilon_0 kT/(e^2 p_0)}$, where p_0 is the hole density in the bulk of the semiconductor, when electrons are attracted to the dislocation core.

Core states in Ge and Si are positively charged at low temperatures and negatively charged at higher temperatures. Bonch-Bruevich and Kogan (1959) discussed the carrier scattering at a charged cylinder. By treating the scattering cylinder similarly to a scattering center of spherical symmetry, Pödör (1966) obtained an expression for the mobility due to dislocation scattering:

$$\mu_{\text{disloc}} = \frac{75\sqrt{p_0}a_l^2(\varepsilon_{\text{stat}}\varepsilon_0)^{3/2}}{e^2 m_n^{1/2} N_{\text{disloc}}} kT, \tag{45}$$

where N_{disloc} (cm^{-2}) is the dislocation density, and a_l is the distance between charged defect centers along the dislocation core-line. Even without charge of the dislocation core at a "neutrality temperature" between negative and positive core charges, a lower mobility in deformed Ge is observed. This is caused by the scattering at the strain field.

A more refined theory takes into consideration the influence of the deformation potential and the anisotropy of the scattering for dislocation fields with a preferred orientation (Düster and Labusch 1973). The mobility, measured parallel to an array of aligned dislocations, is nearly equal to the mobility without dislocations. In contrast, when carriers move perpendicular to the array, the mobility is substantially reduced. This was experimentally confirmed by Schröter (1969). For a short review see Zawadzki (1982).

Carrier scattering at dislocations is important in highly mismatched heterostructures with a high density of dislocations, such as epitaxial layers of III-nitrides. In GaN the influence of scattering at charged dislocation lines, compared to lattice and ionized-impurity scattering, becomes significant at $N_{\text{disloc}} > 10^9$ cm^{-2} (Gurusinghe and Andersson 2003).[7] In this range the mobility is reasonably described by Eq. 45 (Ng et al. 1998); a more rigorous treatment based on the Boltzmann equation including the Bonch-Bruevich potential also accounts for deviations at $T < 30$ K (Look and Sizelove 1999).

[7] In Eq. 45 the fractional filling f of the dislocation line was assumed unity; Gurusinghe and Andersson (2003) computed a slight dependence on n and N_{disloc}, yielding an increase of f from 0.6 for low N_{disloc} and $n = 10^{15}$ cm^{-3} to 1 for $n > 10^{18}$ cm^{-3}. *Uncharged* dislocations scatter via the deformation potential of the surrounding strain field, see Jena and Mishra (2002).

3.4 Carrier-Carrier and Carrier-Plasmon Scattering

3.4.1 Carrier-Carrier Scattering

The scattering of carriers by other carriers within the same band does not change the total momentum of the carrier gas. Therefore, it does not influence the momentum relaxation-time. Combined with other scattering mechanisms, however, it causes an accelerated relaxation – see Sect. 1.2 of ▶ Chap. 32, "Dynamic Processes". This results in a *slight* decrease in mobility, which is usually on the order of only a few percent. Estimates made by Appel (1961) for covalent semiconductors, and by Bate et al. (1965) for ionic semiconductors, indicate a mostly negligible effect of low-field electron-electron scattering. Some larger effects in energy relaxation are discussed by Dienys and Kancleris (1975) and by Nash and Holm-Kennedy (1974). When a subgroup of electrons is excited to higher energies, electron-electron scattering is important, as it tends to restore a thermal electron distribution.

When carriers of different bands are scattering with each other, a more pronounced influence can be observed, especially at higher fields (see ▶ Chap. 24, "Carrier Scattering at High Electric Fields"), where the carriers are heated to differing degrees according to their effective mass. This can be a significant effect for holes in the light-hole and heavy-hole bands. Scattering reduces the difference between the two carrier temperatures.

Electron-Hole Drag Effect At low fields, the electron-hole scattering can be regarded as similar to the scattering of carriers at ionized impurities. It reduces the mobility of minority carriers (Prince 1953, Ehrenreich 1957) and can be described as a drag effect (McLean and Paige 1960).

The exchange of momentum between the two sets of carriers also results in a *carrier drag* caused by the higher-mobility majority carrier that, under extreme conditions, can lead to a drift reversal of the minority carriers. Minority holes are then dragged toward the anode, and minority electrons toward the cathode. Such a *negative absolute mobility* was observed in GaAs quantum wells by Höpfel et al. (1986). General conditions for observing carrier drag are discussed by Breitkreiz et al. (2013).

3.4.2 Carrier Scattering on Plasmons

At higher carrier densities, we must consider the interaction of an electron with the collective electron plasma (see Sect. 1.1 of ▶ Chap. 12, "Photon–Free-Electron Interaction"), characterized by its plasma frequency ω_p as given in Eq. 4 of ▶ Chap. 12, "Photon–Free-Electron Interaction". A substantial energy loss occurs at resonance when the energy of the carrier equals the plasmon energy $\hbar\omega_p$. For thermal electrons, the corresponding carrier density is below 10^{16} cm^{-3} (Eq. 5 of ▶ Chap. 12, "Photon–Free-Electron Interaction"). With faster electrons, resonance occurs with a denser plasma and losses increase. Examples are photo-excited plasmon–LO-phonon modes at ZnSe/GaAs interfaces (Krost et al. 1992), damping of plasmons in semiconductor superlattices (Tripathi and Sharma 1999), surface plasmons in GaN (Polyakov et al. 1998), and plasmon–phonon modes in highly doped *n*-type InAs (Li et al. 1992). See also the plasmon peaks in electron energy-loss spectra shown in Fig. 3 of ▶ Chap. 12, "Photon–Free-Electron Interaction". In *n*-type silicon, electron–plasmon scattering

is the primary mechanism for the cooling of hot carriers; this process also accounts for the anomalous low mobility ($\sim$100 $cm^2V^{-1}s^{-1}$ above $n = 10^{19}$ cm^{-3} (Caruso and Giustino 2016).

In metals, plasmons have an energy of ~10 eV, and fast electrons penetrating a thin metal foil show distinct losses of a multiple of this plasmon energy; see, e.g., León-Monzón et al. (1996). Metal plasmons are observed in reflected-electron energy-loss microscopy (REELM), see Paparazzo and Zema (1997).

4 Scattering by Neutral Defects

The size of neutral defects ranges from the atomic scale, e.g., for neutral lattice defects, to a macroscopic scale, e.g., for boundaries of the crystal. The nature of carrier scattering is determined by the actual size of the scattering object, compared to the wavelength of the scattered electrons.

4.1 Scattering by Neutral Lattice Defects, Intrinsic Point Defects, and Defect Clusters

Scattering by Neutral Lattice Defects Neutral lattice defects have a scattering cross-section which is about the size of the defect atom, typically 10^{-15} cm^2. As described by Erginsoy (1950), they can become important scattering centers at low temperatures (at $T < 100$ K), when the density of ionic impurities has decreased by carrier trapping due to carrier freezeout, and the phonon scattering has decreased due to phonon freeze out. In analogy to the scattering of electrons by hydrogen atoms, the scattering cross-section s_n is estimated as $\propto \pi a_{qH}^2$, where a_{qH} is the equivalent quasi-Bohr radius in the lattice:

$$s_n = \pi a_{qH}^2 \frac{\lambda_{dB}}{a_{qH}} = 2\pi^2 \frac{a_{qH}}{|\mathbf{k}|}; \tag{46}$$

The Bohr radius is modified by a scattering correction factor λ_{dB}/a_{qH} (Seeger 2004), where λ_{dB} is the de Broglie wavelength. Erginsoy estimated a similar relation $s_n \propto 20\ a_{qH}/|\mathbf{k}|$. Using the gas-kinetic estimate for the collision time

$$\tau_n = \frac{1}{N^0 s_n v_{rms}} \tag{47}$$

and $|\mathbf{k}| = 2\pi/\lambda_{dB} = mv/\hbar$, we obtain with $v \cong v_{rms}$ for the mobility due to neutral impurity scattering

$$\mu^0_{n,\mathrm{imp}} = \frac{e}{m_n}\tau_n = \frac{e}{2\pi^2 a_\mathrm{H}\hbar}\,\frac{m_n/m_0}{\varepsilon_\mathrm{stat} N^0}$$
$$= 1.46\times 10^3 \left(\frac{10^{16}}{N^0\,(\mathrm{cm}^{-3})}\right)\left(\frac{10}{\varepsilon_\mathrm{stat}}\right)\left(\frac{m_n}{m_0}\right)\quad (\mathrm{cm}^2/\mathrm{Vs}); \tag{48}$$

here a_H is the true Bohr radius and N^0 is the density of neutral defects. This mobility is independent of the temperature. The Erginsoy approximation is valid for temperatures $T > 20$ K. For lower temperatures, the screening depends on the energy; Blagosklonskaya et al. (1970) obtained

$$\mu^0_{n,\mathrm{imp}} = \frac{e\ m_n^{3/2}}{\sqrt{2}\pi\, a_\mathrm{qH}^2 N^0}\,\frac{1}{\sqrt{E}\Big((E/E_\mathrm{I}+0.0275)^{-1}+10\Big)}, \tag{49}$$

where E_I is the ionization energy of the impurity.

The temperature dependence of neutral-impurity scattering shown in Fig. 7 is due to the variation of N^0 with temperature; there are less neutral impurities at higher temperature due to ionization.

The scattering is different if the spin of the incident electron is parallel or antiparallel to the electrons in the scattering atom (triplet or singlet states, respectively). In considering also the multivalley structure of the conduction bands, Mattis and Sinha (1970) arrived at results similar to those of Blagosklonskaya et al. (1970) at low temperatures, with only a slight mobility reduction at temperatures above 10 K; see also Norton and Levinstein (1972).

Scattering by Intrinsic Point Defects Carrier scattering by intrinsic point defects, such as by interstitials, vacancies, or antisite defects (an A atom on a B site), is caused by deviations from the periodicity of the lattice potential in the equilibrium position of the disordered atoms. These changes can be seen as a local deformation of the bands. The interaction potential responsible for the scattering may be approximated as a Coulomb potential at charged localized centers (Sect. 3.1) or by using the central-core potential if the defect is not charged relative to the lattice (Sect. 4.1). The intrinsic defects are similar to foreign atoms with respect to their behavior as scattering centers.

In first-principles calculations of electron–defect scattering in Si the relaxation times were found to depend strongly on carrier energy and defect type. Assuming one neutral defect (vacancy or interstitial) in 10^6 atoms the calculated defect-limited carrier mobility below 150 K was much above that of ionized-impurity scattering, which has long-range character; interstitials reduce the mobility twice as much as vacancies (Lu et al. 2019).

All such scattering events at atomic point defects are considered *elastic*: the electron momentum is changed but not its energy; the fraction of energy transferred is $\propto m_n/M \ll 1$. When its energy is changed, however, the electron becomes trapped as discussed in ▶ Chap. 30, "Carrier Recombination and Noise".

Scattering at Defect Clusters Larger clusters of defects, such as associates of defects or small inclusions of a different phase, can interfere with the carrier transport beyond their own occupied volume via a space-charge cloud surrounding the defect. Depending on the charge, size, and distribution of these defects, the interaction can be modeled by a simple neutral center of a larger effective diameter or by a charged center with a surrounding space charge extending up to a distance of several Debye lengths.

The charge of the defect associate may be due to carrier trapping. The charge of a different phase inclusion can be estimated from the difference in electron affinity between inclusion and host. In addition, the strain field surrounding such a defect cluster influences the band edge via the deformation potential and, as such, can act as an extended scattering center.

The influence on carrier transport may range from the scattering similar to that at point defects to the carrier repulsion from areas comparable to or larger than the mean free path.

4.2 Influence of External Surfaces

External surfaces can interact with carriers as perfect scattering surfaces, as surfaces for carrier recombination, and via their space charge. The space charge compensates the surface charge and extends into the bulk to a few Debye lengths. These space-charge effects are of technical importance (for example, in field-effect transistors) and are discussed by, e.g., Anderson (1970) and Sze (1981). Surface scattering reduces the mean free path on the order of the crystal dimensions (see below). It also plays a role at low temperatures in high-mobility semiconductors, where the mean free path becomes comparable to at least one of the crystal dimensions – for example, in thin platelets.

In a homogeneous thin semiconductor platelet, we distinguish in a simple approach for surface scattering between specular and nonspecular scattering at both platelet surfaces[8]. Specular scattering has no influence on the mobility, while nonspecular scattering causes a reduction in the average carrier-relaxation time:

$$\frac{1}{\overline{\tau}} = \frac{1}{\tau_{\text{bulk}}} + \frac{1}{\tau_{\text{surf}}}, \tag{50}$$

with the surface-induced relaxation time τ_{surf} given by (see Many et al. 1965)

$$\tau_{\text{surf}} \cong \frac{\delta}{\lambda}\tau_{\text{bulk}}. \tag{51}$$

[8]More sophisticated estimates are given for thin metal layers (e.g., Fuchs 1938, Sondheimer 1952), where surface effects are less complex. These yield results on the same order of magnitude as those given here.

Here, τ_{bulk} is the bulk relaxation time, λ is the mean free path, and δ is the mean carrier distance from the surface. From Eqs. 50 and 51, we obtain for the ratio of actual-to-bulk mobility

$$\frac{\mu}{\mu_{\text{bulk}}} = \frac{\overline{\tau}}{\tau_{\text{bulk}}} = \frac{1}{1 + (\lambda/\delta)}. \tag{52}$$

Setting $2\,\delta = d$, where d is the platelet thickness, and using s as the fraction of the specular scattering events at the platelet surfaces, we obtain

$$\mu = \mu_{\text{bulk}} \frac{d}{d + 2(1-s)\lambda}, \tag{53}$$

which is shown in Fig. 9 with s as the family parameter.

4.3 Influence of Microcrystallite Boundaries

The influence of microcrystallite boundaries is important for carrier transport in most microcrystalline semiconductors. A substantial reduction in mobility is observed when carriers must pass through the interface between crystallites on their way from one electrode to the other. These interfaces contain a high density of lattice defects, such as dislocations along small-angle grain boundaries (Fig. 28 of ▶ Chap. 15, "Crystal Defects") and vacancies or clusters of defects with substantial lattice relaxation along other grain boundaries (Figs. 29 and 30 of ▶ Chap. 15, "Crystal Defects"). Carrier transport through the interfaces is therefore subject to a high degree of scattering.

For a first approximation, we may assume that a free path ends at such interfaces, with a consequent reduction of the effective mobility: The smaller the crystallite grains, the more the mobility is reduced compared with the mobility in large single crystals, consistent with Eq. 51.

Usually, interfaces have a high density of traps, which can become occupied by carriers. As a consequence, interfaces are often charged. Screening charges are located on both sides of the interface, causing a space-charge triple layer (Fig. 10),

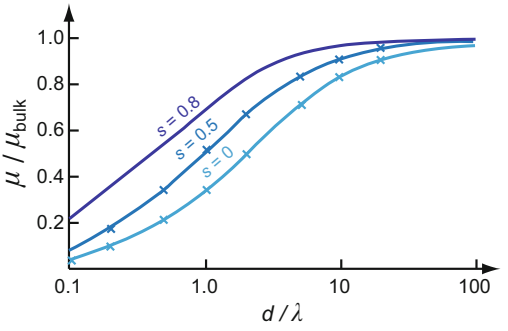


Fig. 9 Carrier mobility in a platelet with respect to the bulk mobility as a function of the platelet thickness d, normalized to the mean free path λ. The specular surface-scattering fraction s is the family parameter. Data points are results of the approximate theory of Anderson (1970), curves are calculated from Sondheimer (1952)

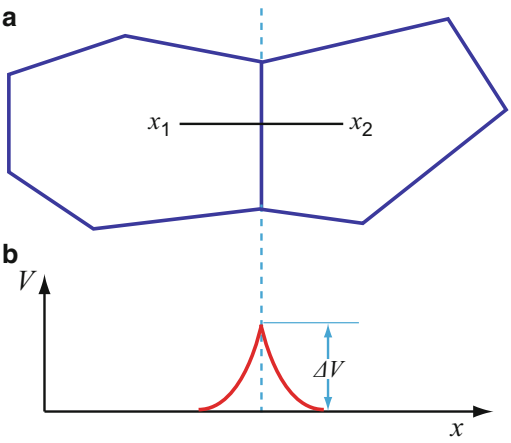


Fig. 10 (**a**) Crystallite boundary and (**b**) potential barrier along line x_1- x_2 as indicated in (*a*)

which produces a potential barrier between each of the crystallites. The height of the barrier can be estimated from the integrated Poisson equation

$$\frac{d^2}{d\mathbf{r}^2} V_{\text{boundary}} = \frac{e\, n_{\text{interface}}\, L_{\text{D}}}{\varepsilon\varepsilon_0}, \tag{54}$$

where $n_{\text{interface}}$ is the surface-charge density at the interface and L_{D} is the Debye length. The actual surface-charge density, however, is usually insufficiently known to make such an estimate meaningful. The barrier height eV_{boundary} is deduced from experimental data. With such a barrier, an exponential dependence of the mobility versus $1/T$ is observed:

$$\mu_{\text{boundary}} = \mu_0 \exp\left(-\frac{e\, V_{\text{boundary}}}{kT}\right), \tag{55}$$

when carriers must pass over these intergrain barriers, that is, when the grain size is smaller than the distance between the electrodes.

The mobility μ_{boundary} is an effective mobility. Within each grain, the carrier mobility μ_0 is larger than the effective mobility. Carrier transport through grain boundaries has received much attention. However, since this topic deals with the inhomogeneous semiconductor, it is not the focus of this book. For a review see Seager (1985).

4.4 Alloy Scattering

Mixing homologous elements forms semiconductor alloys in which – in absence of ordering effects – these elements are statistically distributed. Examples of such random alloys have been mentioned previously, e.g., in Sect. 2.1 of ▶ Chap. 8,

"Bands and Bandgaps in Solids", and 2.5 of ▶ Chap. 19, "Deep-Level Centers", and include $Al_xGa_{1-x}As$, $Al_xGa_{1-x}N$, and $Si_{1-x}Ge_x$. The random distribution of the alloying elements causes a fluctuation in the periodic potential of the lattice and an increased carrier scattering, known as alloy scattering as originally discussed by Wilson (1965); see also Makowski and Glicksman (1973). The alloy-scattering relaxation time was calculated for typical parameters of semiconductors by Nag (1984, see Fig. 11) and for ternary group II-VI and III-V compounds by Auslender and Hava (1993).

In the $Si_{1-x}Ge_x$ alloy, the conduction-band minimum changes for increasing x at $x = 0.85$ from a Δ valley (at $k = 0.8 \times (2\pi/a)$) to an L valley. The band crossing generates an intervalley scattering between the Δ and L band in n-type $Si_{1-x}Ge_x$ in addition to intravalley scattering, producing a dip in $\mu_{alloy}(x)$ (Krishnamurthy et al. 1985, Murphy-Armando and Fahy 2006). In p-type $Si_{1-x}Ge_x$, the substitution of Si by Ge creates a warping of the valence band; this leads to an alloy scattering dominated by heavy holes (Joyce et al. 2007).

In $Al_xGa_{1-x}N$, alloy disorder is a prominent scattering mechanism that limits the carrier mobility in an AlGaN-based two-dimensional electron gas (Simon et al. 2006); more recent first-principles calculations confirm this dominance of alloy scattering at low temperature and its strong contribution at room temperature, with a $T^{-1/2}$ dependence of the mobility (Pant et al. 2020).

The scattering probability, obtained by integrating $S(\mathbf{k}, \mathbf{k}')$ in Eq. 88 of ▶ Chap. 22, "Carrier-Transport Equations", over all $\mathbf{k}'$,

$$S(\mathbf{k}) = \frac{V}{8\pi^3} \int S(\mathbf{k}, \mathbf{k}') \; d\mathbf{k}', \tag{56}$$

can be approximated for alloys by

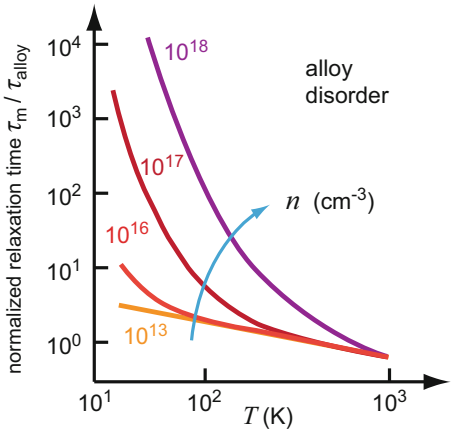


Fig. 11 Momentum-relaxation time for alloy scattering in a semiconductor with $m_n = 0.05\, m_0$ and $\varepsilon = 15$ as a function of the lattice temperature. Family parameter is the carrier density with $n = 10^{13}$ to 10^{18} cm^{-3} for curves 1 to 5, respectively. The normalizing constant τ_{alloy} refers to τ_m at $T = 300$ K. After Nag (1984)

$$S(\mathbf{k}) = \frac{\sqrt{2}}{\pi\hbar^4} N_0 x(1-x) \left(E_{g_1} - E_{g_2}\right)^2 m_n^{3/2} \sqrt{E(\mathbf{k}) - E_c} \tag{57}$$

where E_{g_1} and E_{g_2} are the bandgap energies of the pure constituents of the alloy, and N_0 is the density of atoms in the alloy.

The dependence of the scattering probability on the quantity $x(1-x)$ leads to the rule

$$\mu_{\text{alloy}} \propto \frac{1}{x(1-x)}, \tag{58}$$

which is valid for semiconductors (Slack 1997) as well as for metals (Ziman 1972). Alloy scattering in quantum wires is analyzed by Nag and Gangopadhyay (1998).

5 Multivalley Carrier Transport

Indirect-bandgap semiconductors have several satellite minima at the conduction-band edge in the different crystallographic directions. Scattering of carriers can occur within one of these valleys, *intravalley scattering,* or from one valley to another, *intervalley scattering.* The scattering described in the previous sections is a scattering that leaves each electron within its valley. The carrier distribution in all equivalent valleys remains equal.

5.1 Processes in Intervalley Scattering

Scattering from one valley to another requires a large exchange of momentum, which is provided by phonons of large **q** values, i.e., for optical phonons or for acoustic phonons of large energy – see Fig. 12. Such intervalley phonons with energy $\hbar\omega_i$ may transfer electrons in Si from a valley near $0.8\ \pi/a$ to the equivalent valley at $-0.8\ \pi/a$, requiring a momentum exchange[9] of $q = 0.4\ \pi/a$ (Fig. 12). Another scattering into a nonequivalent valley at $0.8\ \pi/b$ needs an even larger momentum exchange.

Polar and piezoelectric scattering does not provide a large momentum transfer and can be neglected for intervalley scattering. All phonons with sufficient momentum have similar energies, usually within a factor of two, whether they are in the acoustic or the optical branch. Intervalley scattering can therefore be treated similarly to optical phonon scattering by replacing the longitudinal optical phonon energy $\hbar\omega_{\text{LO}} = k\Theta$ in Eq. 18 with $\hbar\omega_{\text{inter,ac}} = \hbar\omega_{\text{inter,opt}} = k\Theta_{\text{inter}}$ which requires typically a 20–40% less-energetic phonons.

[9]See Fig. 8b, which shows two equivalent transitions: one requires $2{\cdot}0.8\ \pi/a = 1.6\ \pi/a$ in the extended $E(k)$ diagram, but actually one needs only a momentum transfer of $2{\cdot}0.2\ \pi/a = 0.4\ \pi/a$.

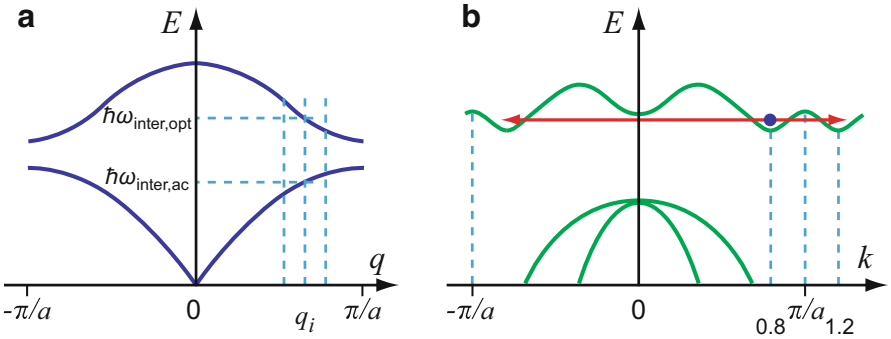


Fig. 12 (**a**) Phonon $E(q)$ and (**b**) electron $E(k)$ diagrams indicating intervalley scattering

The intervalley scattering into a valley on the same axis is referred to as *g-scattering*. It results in a change of the sign of the electron momentum, i.e., "reflecting" the electron path by a sufficiently large angle. These scattering events are also known as *Umklapp processes*, indicating that the momentum has changed its sign without much change in value[10] (in German, *Umklapp* is a mirror operation). The other intervalley scattering processes, which transfer an electron into one of the four valleys on the other axes, are called *f-scatterings*. Selection rules determine which of the valleys can be reached depending on the symmetry of the scattering phonon (Bir and Pikus 1974). Forbidden transitions, however, are also significant for carrier transport (Eaves et al. 1975).

5.2 Mobility for Intervalley Scattering

All intervalley scattering events cause a reduction in the momentum-relaxation time; the mobility is reduced. The Debye temperature Θ_{inter} is large for intervalley phonons (300–720 K) and the function $\phi(T)$ in Eq. 18 can be approximated by an exponential; therefore, Θ_{inter} enters exponentially into a polar-optical scattering, and a reduction from Θ to $\Theta_{\text{inter}} < \Theta$ also reduces the electron mobility:

$$\mu_{\text{inter}} = \frac{e\hbar}{2m_n \alpha_c k \Theta_{\text{inter}}} \sqrt{\frac{T}{\Theta_{\text{inter}}}} \exp\left(\frac{\Theta_{\text{inter}}}{T}\right). \tag{59}$$

[10]Umklapp processes were introduced for scattering in metals, in which changes in the magnitude of the momentum after scattering from one to another side of the near-spherical Fermi surface are even smaller.

Therefore, intervalley scattering becomes the dominant optical phonon-scattering mechanism: the increased phonon interaction with intervalley scattering reduces the relaxation time substantially at higher temperatures.

As shown before, intravalley scattering is insufficient to explain the observed lowering of μ_n with increasing temperatures: acoustic deformation-potential scattering, although frequent enough, yields only a $T^{-3/2}$ decrease of $\mu_n(T)$; optical deformation-potential scattering could give a steeper decrease of $\mu_n(T)$, but does not occur frequently enough to cause a marked reduction in the mobility: there are only few optical phonons at temperatures below the Debye temperature.

The degree of the additional mobility reduction by intervalley scattering depends on the ratio of acoustic to optical deformation-potential interaction rates, ω_{ac} and ω_{opt}, which are described by the ratio of the material-specific coupling constants:

$$\frac{\omega_{\mathrm{opt}}}{\omega_{\mathrm{ac}}} = \frac{1}{2}\left(\frac{D_{\mathrm{inter}}\, v_{\mathrm{s}}}{\Xi_{\mathrm{c}}\, \omega_{\mathrm{inter}}}\right)^2, \tag{60}$$

where D_{inter} is the intervalley optical deformation-potential constant, Ξ_{c} is the acoustic deformation-potential constant, v_{s} is the sound velocity, and ω_{inter} is the phonon frequency for intervalley scattering. Considering the sum of acoustic and intervalley optical scattering relaxation

$$\frac{1}{\tau_{\mathrm{m}}} = \frac{1}{\tau_{\mathrm{ac}}} + \frac{1}{\tau_{\mathrm{inter}}}, \tag{61}$$

we obtain a family of curves shown in Fig. 13. Adjusting the curves by selecting $\Theta_{\mathrm{inter}} = 720$ K and $\omega_{\mathrm{opt}}/\omega_{\mathrm{ac}} = 3$, we obtain for Si a much improved agreement with the experiment (Herring 1955), see Fig. 14.

Rode (1972) computed the electron mobility for *n*-type Si and Ge for a combination of intervalley and intravalley acoustic deformation-potential scattering. Excellent agreement with the experiment was obtained by using two adjustable parameters: the inter- and intravalley deformation potentials. The experimental slope is $T^{-2.42}$ (Putley and Mitchell 1958) and $T^{-1.67}$ (Morin 1954) for Si and Ge, respectively, see Fig. 14. Other intervalley-scattering effects need a larger energy transfer and are discussed in Sect. 3.1 of ▶ Chap. 24, "Carrier Scattering at High Electric Fields".

First-principles calculations for GaAs yield detailed information on the intervalley scattering channels, which contribute substantially to the total electron-scattering rate at the X and L points with ~58% and ~13% at 300 K, respectively (Liu et al 2017). From the X point, phonon wavevectors with $q \cong 0.87$ (in units of $2\pi/a$) induce transitions to the L valley, and vectors with $q \cong 1.0$ induce transitions from the X point to the Γ valley and to other degenerate X valleys. From the L point, $q \cong 0.87$ induces transitions to the Γ and X valleys, and $q \cong 1.0$ transitions to the other degenerate L valleys.

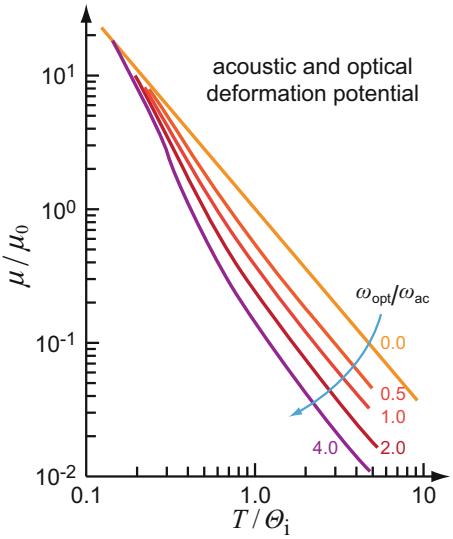


Fig. 13 Intervalley scattering with acoustic and optical deformation potentials as a function of the temperature, and with the coupling-constant ratio as the family parameter. After Herring (1955)

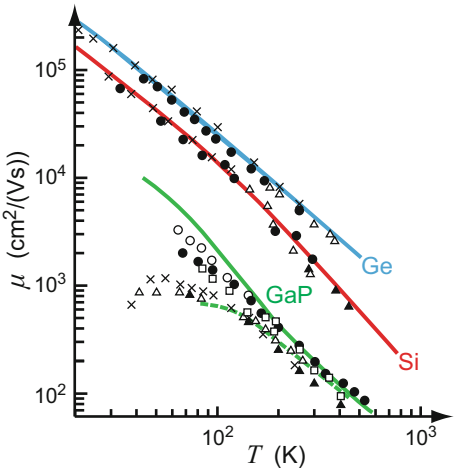


Fig. 14 Temperature dependence of the electron mobility in pure Si, Ge, and GaP. *Solid lines* are calculated for deformation potential inter- and intravalley scattering; ionized-impurity scattering is included in the *dashed curve* for GaP. After Rode (1972)

Resonant Intervalley Scattering A very large additional scattering occurs when nonequivalent valleys coincide in energy. This can be achieved by applying hydrostatic pressure, e.g., 41 kbar for GaAs (see Sect. 3.1 of ▶ Chap. 18, "Shallow-Level Centers"), with the result that the energy for the minima at Γ and X coincide. At this pressure, the mobility decreases markedly because of resonance-intervalley scattering (Kosicki and Paul 1966).

6 Summary

Carriers are scattered by a multitude of deviations from ideal lattice periodicity. Examples include the scattering at neutral or charged impurities as well as intrinsic defects, intravalley scattering of acoustic phonons, and scattering by absorbing an optical phonon followed by immediate reemission. Most of these events are elastic because of the large mass ratio between scattering center and carrier. During elastic scattering, only the momentum of the carrier is changed; its energy is not.

Inelastic scattering causes a change in momentum and energy. Scattering by emitting optical phonons and intervalley scattering are the only inelastic scattering processes during which the carriers can lose a significant amount of their energy to the lattice. This occurs only to a marked degree after the carriers have accumulated a substantial amount of energy from external sources, i.e., a high lattice temperature, electric field, or electromagnetic radiation. Under low-field conditions, many elastic scattering events occur before an inelastic event can follow.

Different types of scattering are dominant for various conditions of temperature, doping, and other material parameters. The type of carrier scattering determines the relaxation time for energy and momentum; both, inversely added, determine the carrier mobility. Usually, ionized-impurity scattering is dominant at low temperatures, and scattering at various phonons at high temperatures. At a given temperature, the ionized-impurity scattering becomes dominant at sufficient doping levels. Usually, the scattering at neutral impurities and with intravalley acoustic phonons are of lesser importance near or below room temperature.

Carrier scattering determines the relaxation time, and with it, the carrier mobility. The carrier mobility is one of the key parameters for transport in semiconductors. By proper treatment, the carrier mobility can be changed within certain limits with concurrent improvement in the performance of devices.

References

Anderson JC (1970) Conduction in thin semiconductor films. Adv Phys 19:311

Appel J (1961) Electron-electron scattering and transport phenomena in nonpolar semiconductors. Phys Rev 122:1760

Auslender M, Hava S (1993) On the calculation of alloy scattering relaxation time for ternary III–V and II–VI semiconductors. Solid State Commun 87:335

Baranskii PI, Klochov WP, Potykievich IV (1975) Semiconductor Electronics. Naukova Dumka, Kiev

Bardeen J, Shockley W (1950) Deformation potentials and mobilities in non-polar crystals. Phys Rev 80:72

Bate RT, Baxter RD, Reid FJ, Beer AC (1965) Conduction electron scattering by ionized donors in InSb at 80°K. J Phys Chem Sol 26:1205

Bir GL, Pikus GE (1974) Symmetry and strain induced effects in semiconductors. Wiley, New York

Blacha A, Presting H, Cardona M (1984) Deformation potentials of $k = 0$ states of tetrahedral semiconductors. Phys Stat Sol B 126:11

Blagosklonskaya LE, Gershenzon EI, Ladyshinskii YP, Popova AP (1970) Scattering of electrons by neutral donors in semiconductors. Sov Phys Sol State 11:2402

Boguslawski P (1975) Nonpolar scattering of electrons by optical phonons in small-gap semiconductors. Phys Stat Sol B 70:53
Boiko II (1959) The theory of the mobility of electrons. Sov Phys Sol State 1:518
Bonch-Bruevich VL, Kogan SM (1959) The theory of electron plasma in semiconductors. Sov Phys Sol State 1:1118
Breitkreiz M, Brydon PMR, Timm C (2013) Transport anomalies due to anisotropic interband scattering. Phys Rev B 88:085103
Brooks H (1955) Theory of the electrical properties of germanium and silicon. Adv Electronics Electron Phys vol 7, pp 85–182. Academic Press, New York
Caruso F, Giustino F (2016) Theory of electron-plasmon coupling in semiconductors. Phys Rev B 94:115208
Chattopadhyay D, Queisser HJ (1981) Electron scattering by ionized impurities in semiconductors. Rev Mod Phys 53:745
Conwell EM (1952) Properties of silicon and germanium. Proc Inst Radio Engrs 40:1327
Conwell EM (1967) High-field transport in semiconductors. Academic Press, New York
Conwell EM, Weisskopf VF (1950) Theory of impurity scattering in semiconductors. Phys Rev 77:388
Dakhovskii IV, Mikhai EF (1965) Calculation of the anisotropy parameter for n-type Si. Sov Phys Sol State 6:2785
Debye PP, Conwell EM (1954) Electrical properties of n-type germanium. Phys Rev 93:693
Demchuk KM, Tsidilkovskii IM (1977) Scattering of electrons by the deformation potential in doped InSb. Phys Stat Sol B 82:59
Deng T, Wu G, Sullivan MB, Wong ZM, Hippalgaonkar K, Wang J-S, Yang S-W (2020) EPIC STAR: a reliable and efficient approach for phonon-and impurity-limited charge transport calculations. npj Comput Mater 6:46
Dienys V, Kancleris Z (1975) Influence of E–E scattering on the phenomenological energy relaxation time in nonpolar semiconductors. Phys Stat Sol B 67:317
Düster F, Labusch R (1973) On the mobility of holes in deformed semiconductors. Phys Stat Sol B 60:161
Eaves L, Hoult RA, Stradling RA, Tidey RJ, Portal JC, Askenazy S (1975) Fourier analysis of magnetophonon and two-dimensional Shubnikov-de Haas magnetoresistance structure. J Phys C 8:1034
Ehrenreich H (1957) Electron scattering in InSb. J Phys Chem Sol 2:131
Ehrenreich H (1961) Band structure and transport properties of some 3-5 compounds. J Appl Phys Suppl 32:2155
Erginsoy C (1950) Neutral impurity scattering in semiconductors. Phys Rev 79:1013
Fletcher K, Butcher PN (1972) An exact solution of the linearized Boltzmann equation with applications to the Hall mobility and Hall factor of *n*-GaAs. J Phys C 5:212
Fuchs K (1938) The conductivity of thin metallic films according to the electron theory of metals. Proc Cambridge Philos Soc 34:100
Ganose AM, Park J, Faghaninia A, Woods-Robinson R, Persson KA, Jai A (2021) Efficient calculation of carrier scattering rates from *first principles*. Nat Commun 12:2222
Ginter J, Mycielski J (1970) Localized potential method in the theory of electron-phonon interaction. J Phys C 3:L1
Giustino F (2017) Electron-phonon interactions from *first principles*. Rev Mod Phys 89:015003
Gurusinghe MN, Andersson TG (2003) Mobility in epitaxial GaN: limitations of free-electron concentration due to dislocations and compensation. Phys Rev B 67:235208
Harrison WA (1956) Scattering of electrons by lattice vibrations in nonpolar crystals. Phys Rev 104:1281
Herring C (1955) Transport properties of a many-valley semiconductor. Bell Sys Tech J 34:237
Herring C, Vogt E (1956) Transport and deformation-potential theory for many-valley semiconductors with anisotropic scattering. Phys Rev 101:944
Höpfel RA, Shah J, Gossard AC (1986) Nonequilibrium electron-hole plasma in GaAs quantum wells. Phys Rev Lett 56:765

Howarth DJ, Sondheimer EH (1953) The theory of electronic conduction in polar semi-conductors. Proc Roy Soc (London) A 219:53
Jena D, Mishra UK (2002) Effect of scattering by strain fields surrounding edge dislocations on electron transport in two-dimensional electron gases. Appl Phys Lett 80:64
Joyce S, Murphy-Armando F, Fahy S (2007) *First-principles* calculation of *p*-type alloy scattering in $Si_{1-x}Ge_x$. Phys Rev B 75:155201
Kosicki BB, Paul W (1966) Evidence for quasilocalized states associated with high-energy conduction-band minima in semiconductors, particularly Se-doped GaSb. Phys Rev Lett 17:246
Krishnamurthy S, Sher A, Chen A-B (1985) Generalized Brooks' formula and the electron mobility in Si_xGe_{1-x} alloys. Appl Phys Lett 47:160
Krost A, Richter W, Zahn DRT (1992) Photoexcited plasmon-LO-phonon modes at the ZnSe/GaAs interface. Appl Surf Sci 56:691
Leighton RB (1959) Principles of modern physics. McGraw-Hill, New York
León-Monzón K, Rodríguez-Coppola H, Velasco VR, García-Moliner F (1996) The inverse dielectric function of a quasi-two-dimensional electron gas in a quantum well: plasmons in a thin metal layer. J Phys Condensed Matter 8:665
Levinson IB (1966) Piezoelectric scattering by uncharged dislocations. Sov Phys Sol State 7:2336
Li YB, Ferguson IT, Stradling RA, Zallen R (1992) Raman scattering by plasmon-phonon modes in highly doped n-InAs grown by molecular beam epitaxy. Semicond Sci Technol 7:1149
Liu TH, Zhou J, Liao B, Singh DJ, Chen G (2017) *First-principles* mode-by-mode analysis for electron-phonon scattering channels and mean free path spectra in GaAs. Phys Rev B 95:075206
Look DC, Sizelove JR (1999) Dislocation scattering in GaN. Phys Rev Lett 82:1237
Lu IT, Zhou JJ, Bernardi M (2019) Efficient *ab initio* calculations of electron-defect scattering and defect-limited carrier mobility. Phys Rev Mater 3:033804
Madelung O (1981) Introduction to solid state theory. Springer, Berlin/New York
Makowski L, Glicksman M (1973) Disorder scattering in solid solutions of III–V semiconducting compounds. J Phys Chem Sol 34:487
Many A, Goldstein Y, Grover NB (1965) Semiconductor surfaces. North Holland, Amsterdam
Mattis D, Sinha O (1970) Impurity scattering in semiconductors. Ann Phys 61:214
McKelvey JP (1966) Solid state and semiconductor physics. Harper & Row, New York
McLean TP, Paige EGS (1960) A theory of the effects of carrier-carrier scattering on mobility in semiconductors. J Phys Chem Sol 16:220
Meyer HJG (1958) Infrared absorption by conduction electrons in germanium. Phys Rev 112:298
Meyer HJG, Polder D (1953) Note on polar scattering of conduction electrons in regular crystals. Physica 19:255
Mitra TK (1969) Electron-phonon interaction in the modified tight-binding approximation. J Phys C 2:52
Moore EJ (1967) Quantum-transport theories and multiple scattering in doped semiconductors. I. Formal theory. Phys Rev 160:607. And: Quantum-transport theories and multiple scattering in doped semiconductors. II. Mobility of n-type gallium arsenide Phys Rev 160:618
Morgan TN (1972) How big is an impurity? – Studies of local strain fields in GaP. In: Proceedings of the international conference on physics of semiconductors, PWN Polish Scietific Publishers, Warsaw, pp 989–1000
Morimoto T, Tani K (1962) Scattering of charge carriers from point imperfections in semiconductors. J Phys Soc Jpn 17:1121
Morin FJ (1954) Lattice-scattering mobility in germanium. Phys Rev 93:62
Murphy-Armando F, Fahy S (2006) First-Principles calculation of alloy scattering in Ge_xSi_{1-x}. Phys Rev Lett 97:096606
Nag BR (1980) Electron transport in compound semiconductors. Springer, Berlin
Nag BR (1984) Relaxation of carriers. In: Alfano RR (ed) Semiconductors probed by ultrafast laser spectroscopy vol I. Academic Press, Orlando, pp 3–44

Nag BR, Gangopadhyay S (1998) Alloy scattering in quantum wires. Semicond Sci Technol 13:417
Nash JG, Holm-Kennedy JW (1974) Experimental determination of highly concentration-sensitive effects of intervalley electron-electron scattering on electric-field-dependent repopulation in *n*-Si at 77 K. Appl Phys Lett 24:139
Ng HM, Doppalapudi D, Moustakas TD, Weimann NG, Eastman LF (1998) The role of dislocation scattering in *n*-type GaN films. Appl Phys Lett 73:821
Norton P, Levinstein H (1972) Determination of compensation density by Hall and mobility analysis in copper-doped germanium. Phys Rev B 6:470
Pant N, Deng Z, Kioupakis E (2020) High electron mobility of $Al_xGa_{1-x}N$ evaluated by unfolding the DFT band structure. Appl Phys Lett 117:242105
Paparazzo E, Zema M (1997) Reflected electron energy loss microscopy and scanning Auger microscopy studies of electron irradiated alkali halide surfaces. Surf Sci 372:L301
Pödör BP (1966) Electron mobility in plastically deformed germanium. Phys Stat Sol B 16:K167
Pödör BP (1970) On the dislocation scattering in silicon-on-insulator films. Phys Stat Sol A 2:K197
Polyakov VM, Tautz FS, Sloboshanin S, Schaefer JA, Usikov AS, Ja Ber B (1998) Surface plasmons at MOCVD-grown GaN(000-1). Semicond Sci Technol 13:1396
Poncé S, Li W, Reichardt S, Giustino F (2020) *First-principles* calculations of charge carrier mobility and conductivity in bulk semiconductors and two-dimensional materials. Rep Prog Phys 83:036501
Poncé S, Macheda F, Margine ER, Marzari N, Bonini N, Giustino F (2021) *First-principles* predictions of Hall and drift mobilities in semiconductors. Phys Rev Res 3:043022
Prince MB (1953) Experimental confirmation of relation between pulse drift mobility and charge carrier drift mobility in germanium. Phys Rev 91:271
Protik NH, Kozinsky B (2020) Electron-phonon drag enhancement of transport properties from a fully coupled ab initio Boltzmann formalism. Phys Rev B 102:245202
Putley EH, Mitchell WH (1958) The electrical conductivity and Hall effect of silicon. Proc Phys Soc (London) A 72:193
Raymond A, Robert JL, Pistoulet B (1977) New method for measuring the compensation and the spatial fluctuations of impurities in n-type III-V compounds - application to bulk and epitaxial layers. In: Hilsum C (ed) Proceeding of the 4th international conference on GaAs and related compounds. Institute of Physics, London
Restrepo OD, Varga K, Pantelides ST (2009) *First-principles* calculations of electron mobilities in silicon: phonon and Coulomb scattering. Appl Phys Lett 94:212103
Rode DL (1972) Electron mobility in Ge, Si, and GaP. Phys Stat Sol B 53:245
Rode DL, Knight S (1971) Electron transport in GaAs. Phys Rev B 3:2534
Rutherford E (1911) The scattering of alpha and beta particles by matter and the structure of the atom. Philos Mag 21:669
Samoilovich AG, Korenblit IYa, Dakhovskii IV, Iskra VD (1961) Solution of the kinetic equation for anisotropic electron scattering. Sov Phys Sol State 3:2148. And: The anisotropy of electron scattering by ionized impurities and acoustic phonons. Sov Phys Sol State 3:2385
Schröter W (1969) Trägerbeweglichkeit in verformtem Germanium. Phys Stat Sol B 31:177. (Carrier mobility in deformed germanium, in German)
Seager CH (1985) Grain boundaries in polycrystalline silicon. Ann Rev Mater Sci 15:271
Seeger K (2004) Semiconductor Physics, 9th edn. Springer, Berlin
Simon J, Wang AK, Xing H, Rajan S, Jena D (2006) Carrier transport and confinement in polarization-induced three-dimensional electron slabs: Importance of alloy scattering in AlGaN. Appl Phys Lett 88:042109
Slack GA (1997) New materials and performance limits for thermoelectric cooling. In: Rowe D (ed) CRC Handbook on thermoelectricity. CRC Press, New York, p 407
Sondheimer EH (1952) The mean free path of electrons in metals. Adv Phys 1:1
Stillman GE, Wolfe CM, Dimmock JO (1970) Hall coefficient factor for polar mode scattering in *n*-type GaAs. J Phys Chem Sol 31:1199

Sze SM (1981) Physics of semiconductor devices. Wiley, New York
Tripathi P, Sharma AC (1999) Plasmons and their damping in a doped semiconductor superlattice. Pramana J Phys 52:101
Vukmirović N (2021) Calculations of electron mobility in II-VI semiconductors. Phys Rev B 104:085203
Wilson AH (1965) Theory of metals. Cambridge University Press, London
Zawadzki W (1972) Electron scattering and transport phenomena in small-gap semiconductors. In: Proceedings of the 11th international conference on physics of semiconductors. PWN Polish Scientific Publishing, Warsaw, pp 87–108
Zawadzki W (1982) Mechanisms of electron scattering in semiconductors. In: Moss TS, Paul W (eds) Handbook on semiconductors, Theory and transport properties, vol 1. North Holland, Amsterdam, pp 713–803
Zawadzki W, Szymanska W (1971) Elastic electron scattering in InSb-type semiconductors. Phys Stat Sol B 45:415
Ziman JM (1972) Principles of the theory of solids, Cambridge University Press, Cambridge, UK

Carrier Scattering at High Electric Fields

24

Karl W. Böer and Udo W. Pohl

Contents

1 Transport-Velocity Saturation 1024
1.1 Drift-Velocity Saturation in External Fields 1025
1.2 Carrier-Diffusion Saturation 1028
1.3 High-Field Carrier Transport in Built-In Electric Fields 1031
2 Distribution Function at Higher Electric Fields 1034
2.1 Warm and Hot Carriers 1034
2.2 Mobility Changes Induced by Optical Excitation 1036
2.3 Numerical Solution of the Boltzmann Equation 1038
3 Elastic and Inelastic Scattering at High Electric Fields 1040
3.1 Intravalley Scattering at High Electric Fields 1041
3.2 Intervalley Scattering at High Electric Fields 1047
4 Summary 1050
References 1051

Abstract

At high electric fields the scattering of carriers is significantly influenced by carrier heating. Usually scattering is increased and counteracts an enhanced accumulation of carrier energy from the field. Such an increased scattering is observed for interaction with phonons, causing a decrease in carrier mobility: first mostly from acoustical scattering and at higher fields from LO-phonon scattering. The drift velocity is thereby limited to values close to the thermal carrier velocity. The deformed carrier distribution can be approximated by a Boltzmann distribution, assuming a carrier temperature elevated above the lattice temperature. Above the energy of longitudinal

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_24

optical phonons the distribution is skewed substantially, and a description of the distribution function and carrier mobility by a Fourier series is more appropriate. Also numerical solutions of transport problems are then advantageous.

Keywords

Carrier distribution · Carrier heating · Dember effect · Diffusion-current saturation · Drift-velocity saturation · Electron temperature · Gunn effect · High-field transport · Hot electrons · Intervalley scattering · Intravalley scattering · Lattice temperature · Saturation velocity · Scattering at high electric fields · Warm electrons

1 Transport-Velocity Saturation

In the preceding ▶ Chap. 23, "Carrier Scattering at Low Electric Fields", low electric fields F were assumed in carrier transport; the energy gain of carriers between scattering events under such conditions is limited to an only small fraction of kT. The incremental speed gained between collisions is then much less than the mean carrier velocity v_{rms}, and the linearization used in the Drude-Sommerfeld theory (Sect. 2 in ▶ Chap. 22, "Carrier-Transport Equations") is justified; this yields Ohm's law $v_{\mathrm{d}} = \mu F$, with a field-*independent* mobility μ and carrier density. In this chapter, changes in the carrier mobility are analyzed for larger electric fields. These changes are caused by a substantial perturbation of the carrier path between scattering events as illustrated in Fig. 1; they involve a significant increase in carrier energy. A discussion of changes in the carrier density induced by high fields will be presented in ▶ Chap. 29, "Carrier Generation".

At sufficiently high fields, the carrier drift-velocity v_{d} increases *sublinearly* with field F (subohmic behavior) and eventually saturates.[1] The reason for a decrease of the transport velocity and its saturation can be seen either from microscopic arguments, dealing with the specific scattering events, or from a more general picture, as discussed below.

Microscopic arguments relate to the scattering events that become more plentiful or more efficient when carriers gain more energy from the field, since the additional energy must be dissipated to the lattice according to detailed-balance considerations. The energy dissipation increases rapidly when the additional energy gained from the field exceeds the LO phonon energy. These phonons interact strongly with electrons and effectively remove the surplus energy. Increased scattering results in a decreased mobility. Such a decrease of the mobility with sufficient electric field F is discussed in ▶ Chap. 23, "Carrier Scattering at Low Electric Fields", Sects. 2.3, 3.4.2, and 5. It leads into a range in which the carrier mobility decreases first proportional to $F^{-1/2}$.

[1]This sublinearity occurs as long as no run-away currents preceding the dielectric breakdown are initiated (Yu and Cardona 1996). The run-away current regime will be discussed in Sect. 2.2 of ▶ Chap. 29, "Carrier Generation".

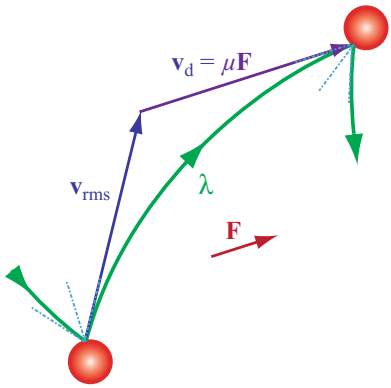


Fig. 1 Carrier path λ between two scattering events at high electric fields **F**, showing the comparable magnitudes of drift velocity $v_d = \mu F$ and thermal velocity v_{rms}

At still higher fields, it decreases proportionally to $1/F$ and results in a *drift-velocity saturation*, yielding $v_d = \mu F = \text{const}$.

General arguments identify a maximum transport velocity that can be deduced easily for a *built-in* electric field: for thermodynamic equilibrium, i.e., for zero bias and no light, the current vanishes at any position within a semiconductor. Therefore, throughout the semiconductor, the drift current is equal and opposite in sign to the diffusion current. This diffusion current (as shown in Sect. 3.2 of ▶ Chap. 22, "Carrier-Transport Equations") can never exceed the thermal velocity of the carriers, thus limiting the drift velocity to the thermal carrier velocity for these conditions. Since the carrier-distribution function in thermal equilibrium is the undeformed Boltzmann distribution (Roberson 1993), this velocity is v_{rms} given by Eq. 19 of ▶ Chap. 22, "Carrier-Transport Equations".

Although such an argument cannot be made for *external* electric fields, the mobility behavior is expected to result in a velocity saturation on the same order of magnitude.

1.1 Drift-Velocity Saturation in External Fields

In an external electric field, the carriers are heated. The drift current is well understood at low fields (Sect. 3.1 of ▶ Chap. 22, "Carrier-Transport Equations", and ▶ Chap. 23, "Carrier Scattering at Low Electric Fields"). At high fields, it requires the selection of the most important scattering mechanisms related to material, doping, and temperature for the computation of the deformed distribution function appropriate for the field range under investigation – see Sects. 3.4.2 and 5 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields".

An example for the relation of the onset of field-induced changes in the mobility and the onset of carrier heating, with consequent changes in the distribution function, can be found in Fig. 2, here obtained for GaAs with

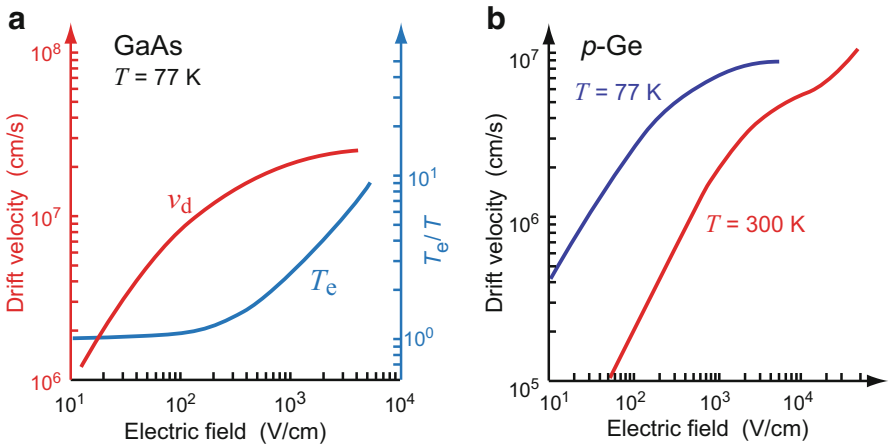


Fig. 2 **(a)** Drift velocity v_d and electron temperature T_e calculated for electrons in the central valley of GaAs at 77 K. **(b)** Saturation of the drift velocity of holes in *p*-type Ge, after Prior (1960) and Bray and Brown (1960)

predominant polar-optical phonon scattering (Price 1977). It is assumed that the electrons are in thermal equilibrium *among themselves* with an electron temperature T_e, which is larger than the lattice temperature T (*hot electrons*, see Sect. 2.1). This assumption is reasonable due to the very fast equilibration of hot-electron ensembles: the ultrafast equilibration in momentum space ($\ll 1$ ps) results in an energy-relaxation with a common energy-relaxation time τ_e at any valleys in a time domain still in the ps range, with a rate ruled by the excess energy only (Tanimura et al. 2016, 2021).

An empirical formula describes the observed high-field behavior of the drift velocity (Jacoboni et al. 1977):

$$v_d = \frac{\mu_0 \; F}{\left(1 + (\mu_0 F / v_{sat})^{\beta}\right)^{1/\beta}}, \tag{1}$$

where v_{sat} is the saturation velocity, β is a temperature-dependent parameter[2] on the order of 1, and μ_0 is the low-field mobility.

This saturation velocity is on the same order of magnitude as the rms velocity of carriers. For a specific scattering mechanism, it can be estimated; it is given in Sect.

[2]The parameters v_{sat} and β are obtained from curve-fitting, depending on materials; for Si they are (Canali et al. 1975) for electrons $v_{sat\ n} = 1.53{\cdot}10^9 \times T^{-0.87}$ (cm/s), $\beta_n = 2.57{\cdot}10^{-2} \times T^{0.66}$ (cm/s), and for holes $v_{sat\ p} = 1.62{\cdot}10^8 \times T^{-0.52}$ (cm/s), $\beta_p = 0.46 \times T^{0.17}$ (cm/s), with the temperature given in K.

3.1.2, Eq. 35 for optical deformation-potential scattering. The drift current, e.g., for electrons, is thus limited to

$$j_{n,\mathrm{drift}} = e n \mu_n(F) \; F \leq e n v_{\mathrm{sat}\,n}. \tag{2}$$

For a review, see Jacoboni and Reggiani (1979). Drift-velocity saturation in AlGaAs/GaAs heterostructures is reported in Wirner et al. (1992); drift velocity in GaN and InN along with a comparison to GaAs is reviewed by Siddiqua and O'Leary (2018).

Mobility Reduction of Holes At high electric fields, the mobility of holes decreases because of the repopulation of the different valence bands. In typical semiconductors with light-hole and heavy-hole bands, the heavy-hole bands become more populated at higher fields because of both the more effective heating of the holes in light-hole bands and the more effective scattering. This causes a mobility reduction. At still higher fields, the more effective optical phonon scattering cuts off the distribution function at $\hbar\omega_{\mathrm{LO}}$ and finally leads to a $1/F$ decrease of the mobility, i.e., to drift-velocity saturation also for holes – see Figs. 2b and 3.

High-Energy Drift Measurements The high-field carrier drift can be measured directly. Well-compensated semiconductors with good ohmic contacts are used at low temperatures to reduce the conductivity and thus Joule's heating. This and a variety of other measurements, including microwave and time-of-flight techniques, are summarized by Jacoboni et al. (1977). In small-size semiconductors a velocity overshoot occurs when the mean free path exceeds the device dimensions (ballistic regime); see Ohno (1990) and Sect. 3.1 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors".

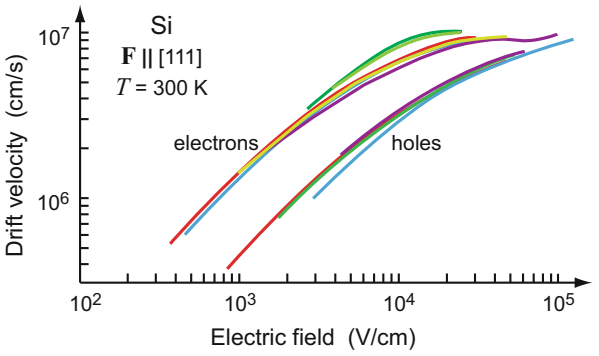


Fig. 3 Drift-velocity limitation in Si at 300 K due to a decrease of the mobility $\propto 1/F$ at fields in the 10^4 V/cm range, applied in the [111] direction. The different curves represent different sets of measurements collected by Jacoboni et al. (1977)

Although each method has certain shortcomings,[3] the resulting $\mu(F)$ dependences are reassuringly similar (see Fig. 3) when care is taken to avoid, or to minimize, modifying effects.

One method for measuring the high-field carrier drift, that avoids contact effects, can be used when stationary high-field domains can be obtained (Böer and Voss 1968). Within such a domain the semiconductor is free of space-charges; the quasi-Fermi levels and bands are tilted parallel to each other, and the domain field is an *external* field.[4] The tilted bands can be imaged by the Franz-Keldysh effect, using light near the fundamental absorption band-edge – thereby distinguishing spatial regions of low and high electric fields. Within the low-field domain, the Hall mobility can be measured and its field dependence can be directly obtained (Böer and Bogus 1968); it shows current saturation near v_{rms}. For a detailed discussion on these high-field domains see the appendix.

1.2 Carrier-Diffusion Saturation

The equivalent of a drift current in an exclusively external field is the diffusion current without a space-charge region. Such a current can be obtained in good approximation as a minority-carrier current with the density gradient created in a homogeneous semiconductor by an inhomogeneous optical excitation, as indicated in Fig. 4. The resulting density gradient of the minority carriers (here n) causes a diffusion current of

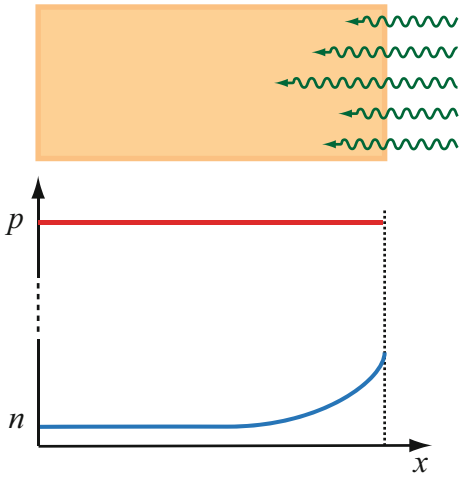


Fig. 4 (**a**) Semiconductor with intrinsic optical excitation absorbed in a surface layer on the right. (**b**) Majority (p) and minority (n) carrier density distributions

[3] For example, conductivity measurements rely on carrier densities obtained from low-field measurements using Hall-effect data; time-of-flight techniques can hardly exclude shallow-level trapping.

[4] Compare to the high-field domains in superlattices discussed in Sect. 2.3.2 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors".

these carriers. This current is called the *Dember current* (Dember 1931) in short-circuit conditions. Photogenerated inhomogeneous carrier distributions (*photo-Dember effect*) can be exploited for creating pulsed THz radiation from semiconductor surfaces (Guzelturk et al. 2018; Kini and Vaisakh 2020; Tonouchi 2020).

The minority-carrier diffusion current is given by the conventional proportionality to the carrier gradient, yielding by use of the Einstein relation

$$j_{n,\mathrm{diff}} = \mu_n kT \frac{dn}{dx}. \tag{3}$$

The diffusion current is illustrated in Fig. 5a. The figure also contains two quasi-Fermi levels; at the two electrodes, the quasi-Fermi levels coincide. The collapse of both quasi-Fermi levels is due to perfect recombination at the metal surface. This diffusion current is also the total electron current as long as the field is small enough to neglect the drift component.

The diffusion current increases proportional to the gradient of the carrier density. However, since it is derived as the difference of two random-walk currents (Eq. 52 of ▶ Chap. 22, "Carrier-Transport Equations") through an arbitrary surface, it is limited, as discussed in Sect. 3.2 of ▶ Chap. 22, "Carrier-Transport Equations", by the Richardson-Dushman current

$$j_{n,\mathrm{diff,max}} = e\, n_0\, v_{\mathrm{rms}}. \tag{4}$$

Therefore, the diffusion current also saturates; the saturation velocity is the thermal electron velocity v_{rms}. A comparison between Fig. 5a and b indicates

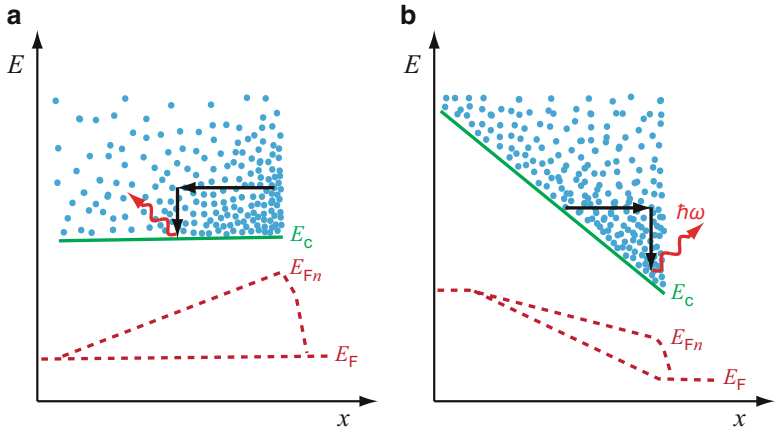


Fig. 5 (**a**) Electron diffusion only, indicated by a sloping (quasi) Fermi level E_{Fn} and a horizontal band. (**b**) Mixed case of built-in and external fields with carrier drift and diffusion, indicated by different slopes of the Fermi level and band. The carrier-density distribution is indicated by the *dot density*

that, for a mixed field condition, carrier heating also occurs and causes a deformed carrier distribution. The root-mean-square velocity v_{rms} must then be calculated from this *deformed* Boltzmann function. Here, the electrochemical field gradient determines carrier heating with a similar result as for an *external* field. In order to identify this averaging with a modified distribution, the rms velocity v_{rms} is now replaced with a modified $\tilde{v}_{\mathrm{rms}}$ (Böer 1985a). Thus, we obtain

$$j_{n,\mathrm{diff}} = \mu_n kT \frac{dn}{dx} \leq e n \tilde{v}_{\mathrm{rms}}. \tag{5}$$

Both drift and diffusion currents are therefore limited at high-electric fields or high-density gradients, respectively, to similar velocities v_{sat} and $\tilde{v}_{\mathrm{rms}}$ (Eqs. 2 and 5).

The saturation diffusion-current can be used to estimate the carrier lifetime in junction devices and provides additional means for analyzing high-injection properties in such devices (Ivanov et al. 2000). A method to extract the diffusion saturation-current in a modified *pn* structure is described by Mimila-Arroyo (2019).

High-Energy Diffusion Measurements The diffusion of carriers, which are excited to higher densities and thus extend within the band to higher energies, can be measured directly. These measurements involve transient effects, such as using field or light pulses, and electro-optical effects for detection. For instance, one uses an optical grating created by electron-induced bleaching and the disappearance of the grating due to carrier diffusion after the initiating light is switched off. The density of carriers that determines the change in the optical absorption (see Sect. 2.2.1 of ▶ Chap. 12, "Photon–Free-Electron Interaction") is given by

$$\frac{dn(x,y,t)}{dt} = g(x,y,t) - c_{\mathrm{cv}}(n(x,y,t))^2 + D\ \partial^2/\partial r^2(n(x,y,t)), \tag{6}$$

with the optical generation rate $g(x, y, t)$ that produces the carrier-density grating by bleaching of a grating design created by the interference pattern of two coherent light beams. The second term describes the recombination of these carriers (Sect. 1.2 of ▶ Chap. 31, "Photoconductivity"). The third term gives the out-diffusion of carriers from the highly illuminated segments of the interference pattern and thereby reduces its contrast.

When probing with a third light beam (Fig. 6), we observe a diffracted beam with an intensity proportional to the contrast of the diffraction pattern, which provides a direct measure of the diffusion constant D in Eq. 6. Such measurements were carried out by Smirl et al. (1982) for Ge, who obtained $D = 142\ \mathrm{cm}^2/\mathrm{s}$ at 135 K and $53\ \mathrm{cm}^2/\mathrm{s}$

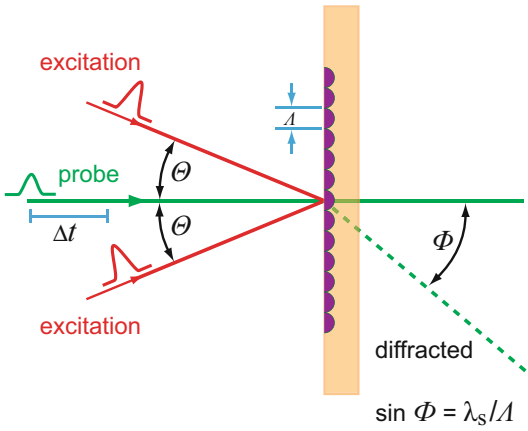


Fig. 6 Schematics of a light-beam arrangement for creating a diffraction grating by induced electron absorption within a semiconductor and consequent light diffraction by a low-intensity probe beam

at 295 K for ambipolar diffusion under the given experimental conditions. For a review on transient grating studies of carrier diffusion, see Miller (1999); various grating techniques for diffusion measurements are compared by Chen et al. (2018).

The results of these diffusion studies are complicated by several overlapping effects:

- The high density of carriers necessary for obtaining measurable optical absorption causes enhanced diffusion because of degeneracy.
- Many-body effects enhance the diffusion due to carrier screening and reduce it due to bandgap narrowing, resulting in selftrapping.
- The increased carrier energy, expressed by an increased carrier temperature, influences diffusion.
- The increased lattice temperature caused by the optical excitation results in a decrease in diffusion; it increases scattering with phonons – see Smirl (1984) and van Driel (1985).

The diffusion coefficient of electrons as a function of an external field, which heates the electrons, measured at different temperatures in Si is shown in Fig. 7. The solid curves are obtained from Monte-Carlo simulation – see Sect. 2.3. The change of the diffusion coefficient at higher fields shows the expected $1/F$ behavior, as observed for the drift mobility in the same field range of up to ~10^4 V/cm. It is caused by a reduction in the carrier mobility.

1.3 High-Field Carrier Transport in Built-In Electric Fields

It is common practice to assume for a built-in electric field (i.e., the field in a space-charge region with negligible external perturbation) the validity of the same basic transport equations as for an external field (Eqs. 61 and 62 of ▶ Chap. 22, "Carrier-

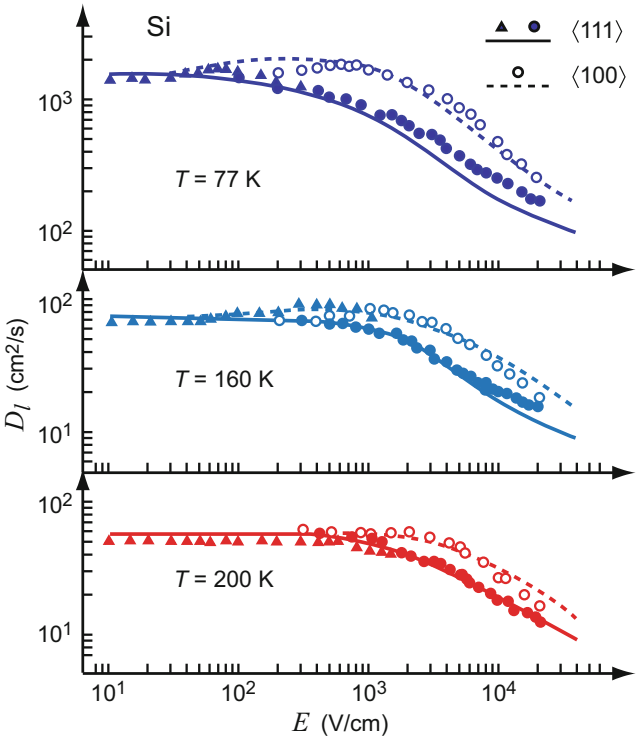


Fig. 7 Longitudinal diffusion coefficient of electrons in Si as a function of an external electric field *F*. *Filled* and *open circles* refer to time-of-flight measurements with field directions along ⟨111⟩ and ⟨100⟩, respectively, *triangles* are electronic-noise data. The curves correspond to Monte-Carlo computation. After Brunetti et al. (1981)

Transport Equations"); this is permissible for low fields. In the absence of an external bias, the carrier equilibrium in a built-in field region is expressed by the balance between drift and diffusion currents of opposite sign – see Sect. 3.4 of ▶ Chap. 22, "Carrier-Transport Equations". This formal treatment is valid until the built-in field exceeds values for which an external field would result in substantial carrier heating, whereas the built-in field does not. In this field-range, the mobility becomes field-dependent and, for reasons indicated in the above mentioned section, the difference between a built-in and an external field must be considered.

Saturation Currents with Built-In Electric Fields The limiting diffusion current is given by the Richardson-Dushman thermal emission current (Sect. 1.2),

$$j_{n,\mathrm{diff},\mathrm{max}} = j_{n,\mathrm{RD}} = e\,n\,v^{*}_{\mathrm{rms}}, \tag{7}$$

with $v^*_{\mathrm{rms}} = v_{\mathrm{rms}}/\sqrt{6\pi}$ for planar geometry and Boltzmann statistics; for built-in fields, the carrier-distribution function remains a Boltzmann distribution at lattice temperature. Within a space-charge region, e.g., a *pn* junction at zero bias and in thermodynamic equilibrium (no external excitation), the drift and diffusion currents must be equal with opposite sign to render the total electron and hole current separately zero in every volume element of the semiconductor. Therefore, even for high built-in fields, the maximum drift velocity cannot exceed the Richardson-Dushman velocity $v^*_{\mathrm{RD,max}} = v^*_{\mathrm{rms}}$.Therefore, the drift current must also be limited to $j_{n,\mathrm{diff,max}} = e\,n\,v^*_{\mathrm{rms}}$. With applied bias, however, the situation becomes more complex and needs additional discussion.

High-Field Carrier Transport in Mixed Electric Fields Space-charge regions with external bias show a mixed, partially built-in, partially external-field behavior, see Fig. 5b. In some regions of the semiconductor the field is predominantly external, whereas in others it is predominantly a built-in field (Böer 1985a). In contrast to a semiconductor with only an external field, the carrier transport controlled by the built-in part of the field is not accessible to direct measurements. Both drift and diffusion current densities are highly inhomogeneous[5]. In a typical *pn* junction, these current densities show a maximum on the order of 10 kA/cm^2 where the carrier gradient peaks, although the net current is usually a very small fraction thereof.

Since within the built-in field region the net current is small compared to the drift or diffusion current, we obtain from

$$j_{n,\,\mathrm{drift}} \cong -j_{n,\mathrm{diff}} \tag{8}$$

an equation that *is independent of the mobility*:

$$\frac{1}{n}\frac{dn}{dx} = \frac{eF}{kT} = -\frac{e}{kT}\frac{d\psi}{dx}. \tag{9}$$

It yields the well-known *Boltzmann condition,* resulting in an exponential dependence of the carrier density on the electrostatic potential:

$$n = n_0 \exp\left(-\frac{e\,(\psi_n - \psi_{n0})}{kT}\right). \tag{10}$$

This relation holds throughout most of the built-in field region in which Eq. 8 is sufficiently well fulfilled. This region is therefore referred to as the *Boltzmann region.* It represents the parts of the space-charge region with a sufficiently high majority-carrier gradient and, at zero bias, comprises the entire space-charge region. With bias, however, there are regions adjacent to the Boltzmann region in which Eq. 8 no longer holds.

[5]In strongly inhomogeneous electric fields, the local carrier mobility was found to depend on the product of the electric field and the gradient of the quasi-Fermi potential; see Zakhleniuk (2006).

These are the regions of predominant drift or diffusion currents – see the quantitative discussion by Böer (1985b); these regions with predominant drift and diffusion currents are referred to as *drift only* or *diffusion only* regions, respectively. In semiconductors within a region of a built-in electric field, they represent well-distinguished regions of major drop in the minority-carrier electrochemical potential and lie adjacent to the part of a junction with dominant space charge.

Substantial carrier heating in space-charge regions (Böer 1985a) is restricted to minority carriers; that is, quasi-Fermi levels of minority carriers show the highest slopes when a bias is applied. Marked majority-carrier heating in devices with space-charge regions will occur only when the external fields become comparable to the built-in field, and this occurs only close to dielectric breakdown.

2 Distribution Function at Higher Electric Fields

We have indicated before that the carrier drift and diffusion currents saturate at a value close to the thermal rms velocity. The reasons for such a saturation will be analyzed in a microscopic model in the following sections.

First, the concept of an electron temperature that is above the lattice temperature when an external field is applied will be introduced. For this purpose we start with the Boltzmann equation. With an electric field F_x applied along the x direction, the electron distribution function $f(E)$ is deformed, as discussed in Sect. 4 of ▶ Chap. 22, "Carrier-Transport Equations"; from Eq. 96 of that chapter we obtain

$$f(v) = f_0(v) - \frac{e}{m_n}\tau(v)\ F_x \frac{\partial f(v)}{\partial v_x}, \tag{11}$$

with the carrier velocity v and the relaxation time τ. For *small fields* $f(v)$ the derivative may be replaced by $f_0(v)$. This can be seen by forming the derivative of $f(v)$ from Eq. 11,

$$\frac{\partial f(v)}{\partial v_x} = \frac{\partial f_0(v)}{\partial v_x} - \frac{e}{m_n}\tau(v)\ F_x \frac{\partial^2 f(v)}{\partial v_x^2}, \tag{12}$$

when introduced into Eq. 11, it gives a third term that is proportional to F_x^2 and can be neglected for small fields.

When *higher fields* are applied, however, we must consider higher-order terms in Eq. 11; this will be discussed in the following section.

2.1 Warm and Hot Carriers

The expression describing the field-induced changes in the distribution function can be extended in two ways (see Seeger 2004). The *first approach* uses an extension of the Fourier development given in Eq. 11:

$$f(v) = f_0(v) - \mu_n F_x \frac{\partial f_0(v)}{\partial v_x} + (\mu_n F_x)^2 \frac{\partial^2 f_0(v)}{\partial v_x^2} - \ldots . \tag{13}$$

This introduces higher terms in the definition equation for the drift velocity (Eq. 1), which can now be expressed as

$$v_{\mathrm{d}\,n} = \mu_{n_0}\left(F_x + \alpha_3 F_x^3 + \alpha_5 F_x^5 + \ldots\right); \tag{14}$$

the even terms vanish because f_0 is an even function. Retaining the definition of the drift velocity as $v_\mathrm{d} = \mu\, F_x$ obtains a field-dependent mobility (Seeger 2004):

$$\mu_n = \mu_{n_0}\left(1 + \alpha_3 F_x^2 + \alpha_5 F_x^4 + \ldots\right); \tag{15}$$

with μ_{n_0} as the field-independent mobility. Carriers are called *warm* when the second term in Eq. 15 is included and *hot* when still higher terms are considered. For warm electrons in GaAs quantum wells see Cross et al. (1999), for hot electrons in superlattices see Burtyka et al. (1991); a general review is given in Ridley (1997).

In a *second approach*, a deformed distribution function $f(E)$ is introduced, which is approximated by the Boltzmann distribution[6] at an elevated *carrier temperature,* $T_\mathrm{e} > T$, with T_e the carrier temperature (here discussed for electrons) and T the lattice temperature:

$$f(E) = f_0(E, T_\mathrm{e}) \propto \exp(-E/(kT_\mathrm{e})). \tag{16}$$

The increase in carrier temperature can be estimated from the incremental carrier energy in field direction – see also ▶ Chap. 29, "Carrier Generation".

$$\frac{m_n}{2}\Delta v^2 = \mu_n e F^2 \tau_e = \frac{3}{2}k\ \ (T_\mathrm{e} - T), \tag{17}$$

where τ_e is the energy-relaxation time. The field-dependent mobility can be expressed in terms of the increased electron temperature. For instance, for *warm electrons*

$$\mu_n = \mu_{n_0}(1 + (T_\mathrm{e} - T)\varphi(T) + \ldots), \tag{18}$$

where $\varphi(T)$ is a function of the lattice temperature alone and depends on the scattering mechanism. *Hot* electrons can leave the semiconductor as shown by Maňkoš et al. (1996) and Iveland et al. (2013).

[6]This approximation is no longer acceptable for electrons exceeding the energy of LO phonons, as the tail of the distribution becomes substantially deformed, see Sect. 2.2.1 in ▶ Chap. 29, "Carrier Generation".

At sufficiently high external electric fields, the carrier-distribution function is substantially deformed. It may be approximated by a Boltzmann distribution with a carrier temperature that significantly exceeds the lattice temperature ($T_e > T$):

$$f = \frac{n}{I_n} \exp\left(-\frac{E}{kT_e}\right), \tag{19}$$

where I_n is a normalization integral, given by

$$I_n = \frac{1}{4\pi^3} \int_{BZ} \left(-\frac{E(\mathbf{k})}{kT_e}\right) d\mathbf{k}, \tag{20}$$

with the integration extending over the Brillouin zone BZ. At high electron densities, scattering between carriers is more frequent than with lattice defects. Here, a *displaced Maxwellian distribution* (Nag 1980) describes the electron ensemble:

$$f(\mathbf{k}) = c \ \exp\left(-\frac{\hbar^2 |\mathbf{k} - \mathbf{k}_0|^2}{2 m_n k T_e}\right) = \ c \ \exp\left(-\frac{E - \hbar \mathbf{v}_d \cdot \mathbf{k}}{kT_e}\right). \tag{21}$$

Here $\mathbf{k}_0$ is the displacement vector of the momentum in field direction and $\hbar^2 (k - k_0)^2 / m_n^2$ is the square of the drift velocity; c is a constant.

Analytical expressions require a consideration of the different scattering mechanisms given in Sect. 3. Since serious simplifications are introduced in such treatments also numerical solutions of the Boltzmann equation considered in Sect. 2.3 are interesting to provide an insight into high-field transport.

2.2 Mobility Changes Induced by Optical Excitation

When a semiconductor is illuminated, carriers are optically excited and a number of defects will change their charge state. Charged centers may become neutral (e.g., a donor-type center after trapping an electron or an acceptor-type center after trapping a hole), while other centers may become negatively or positively charged when trapping a photoexcited carrier or when ionized by light; see, e.g., Haug and Koch (1990) and Henneberger et al. (1993).

The redistribution of carriers, indicated by the split of the two quasi-Fermi levels under illumination, can be used to identify the degree of recharging: centers with their energy levels between the two quasi-Fermi levels are preferentially recharged. Centers just above E_{F_p} tend to be depleted, whereas centers just below E_{F_n} tend to be filled by the optical excitation. During illumination the filling of centers between these quasi-Fermi levels is inverted from thermodynamic equilibrium – see Fig. 8.

The changes in the densities of neutral and charged centers consequently cause changes in the relative magnitude of scattering on these two types of defects, as described in Sects. 4.1 and 3.1 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields". In addition, an optical generation initially creates carriers which are not in

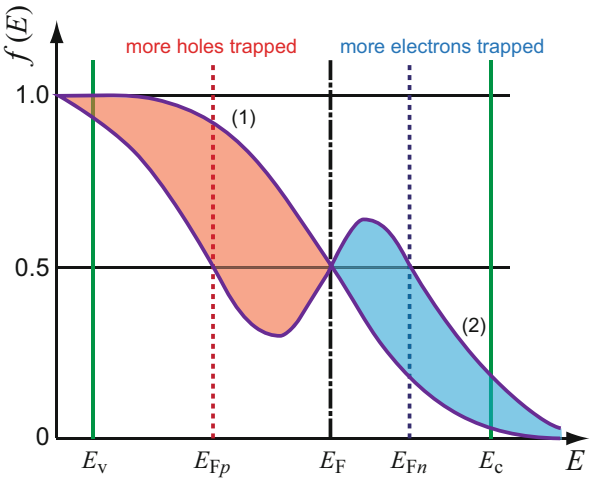


Fig. 8 (*1*) Fermi distribution in thermal equilibrium and (*2*) quasi-Fermi distribution in steady state with light. Ranges of redistributed carriers are *shaded*; the states of higher energy are more negatively charged, those of lower energy more positively. The densities of recharged levels under both must be equal to maintain quasineutrality

thermal equilibrium. Usually, they are excited to levels substantially beyond the band edge. Between continuous excitation and relaxation due to scattering, the resulting distribution function becomes severely deformed.

Both the defect-center recharging and the change in the distribution function will cause the mobility to become a function of the optical generation rate g_o:

$$\mu = \mu(g_o). \tag{22}$$

Energy Relaxation of Optically Excited Carriers The absorption into states higher in the conduction bands produces high-energy electrons which will relax into lower states according to their energy-relaxation time. As a consequence, the electron gas is heated. Scattering with longitudinal optical phonons is the main energy-dissipation mechanism to the lattice.

On the other hand, if the optical excitation proceeds within less than kT of the band edge, the produced electron distribution is compressed and may be approximated by a Boltzmann distribution of a lower-than-lattice temperature. This means that the *electron gas is cooled*, a phenomenon referred to as *optical cooling*. This effect is accompanied by a heating of the *lattice*, caused by the fraction of nonradiatively recombining electrons, which produce large numbers of phonons – see Budkin and Tarasenko (2011).

Further inside the band, scattering probabilities are different between the different branches and vary widely within each branch. Again, scattering with LO phonons is the predominant energy-relaxation mechanism. Fröhlich and Paranjape (1956) estimate the energy-relaxation time for such a scattering:

$$\tau_{e,\mathrm{LO}} \cong \frac{1}{\alpha_c\,\omega_{\mathrm{LO}}}\,\frac{k}{\cosh^{-1}(k\,a)}, \tag{23}$$

with $k = \sqrt{2\,m_n\,\omega_{\mathrm{LO}}/\hbar}$. For further estimations on energy relaxation of optically generated carriers, see in ▶ Chap. 32, "Dynamic Processes" the Sects. 1.2.2 and 1.2.3. Using the Boltzmann equation, the surplus energy $E - E_c$, given by the optical excitation to an electron in the conduction band, can be equated to the surplus energy provided by an external electric field (Ferry 1980). However, the influence of optical excitation on the carrier distribution is distinctly different; it can severely skew the distribution. Kinetic effects are helpful in distinguishing the different relaxation mechanisms, as discussed in the section mentioned above.

An example of optical carrier heating is the oscillatory photoconductivity observed in InSb and GaSb, when monotonically increasing the photon energy of the exciting light. Such nonmonotonic changes in photoconductivity can be explained by a nonmonotonic carrier density caused by a preferred recombination from the band edge. When electrons are excited to an energy above the band edge equal to an integer of the LO phonon energy, the LO scattering brings the excited electrons close to the bottom of the band, i.e., to $k \cong 0$, from which preferred recombination with holes near $k = 0$ occurs. Consequently, this reduces the steady-state carrier density more readily than recombination from higher states in the conduction band (Habegger and Fan 1964).

2.3 Numerical Solution of the Boltzmann Equation

The analytical methods discussed in Sect. 3 for solving the Boltzmann equation require drastic simplifications and permit only one type of interaction, i.e., a specific scattering mechanism to be discussed at a time. Often the approximations used are justified only in a very limited parameter range. This becomes critical at higher electric fields, when major deviations from the Boltzmann distribution are observed. Here, the truncation of an expansion of the distribution function, as given in Eq. 103 of ▶ Chap. 22, "Carrier-Transport Equations", is no longer an acceptable approximation. Numerical solutions of the specific transport problems are then advantageous.

Iterative Method An *iterative technique* proposed by Budd (1966), extending the *variable path method* of Chambers (1952), yields solutions of the Boltzmann equation by stepwise processing the evolving carrier-distribution function (Rode 1970). A modification of this method, involving a fictitious *self-scattering* for the purpose of mathematical simplification, was proposed by Rees (1969, 1972).

Monte Carlo Method This method simulates the motion of a carrier under the influence of an electric field and the different scattering mechanisms by applying basic kinetic laws. Here, the scattering process (phonons, impurities, carrier-to-carrier, etc.) and the final states are stochastically selected, with a probability

distribution given by the density of states times the squared matrix elements, determined from a microscopic theory of the different scattering centers (Kurosawa 1966).

When successive scattering events of one electron are followed long enough, its behavior is equivalent to that of the average behavior of the entire electron ensemble (*ergodicity*). Figure 9 illustrates the principle of the Monte Carlo technique. Subfigure *a* shows the actual electron path in two dimensions under the influence of a large external field. Subfigure *b* shows the same eight events as line segments in momentum space. These segments are connected by gray lines representing the momentum changes in each of the scattering events. Subfigure *c* gives the velocity of the carrier averaged at the *n*th point over all previous $(n - 1)$ paths. This average velocity approaches the drift velocity (dashdotted line) when enough paths are taken; 90 paths are shown in subfigure *c*. The drift velocity is a direct measure of the mobility, which then can be obtained by using the relation $v_d = \mu_d F$.

The Monte Carlo method permits the extraction of derived physical information from simulated experiments and is a powerful tool for the discussion of stationary (Jacoboni and Reggiani 1983) or transient (Oh et al. 1992; Reklaitis 2012) transport effects in semiconductors. It also has become the preferred technique to analyze the carrier transport under nonhomogeneous conditions, such as in device simulation (Jacoboni and Lugli 1989); for GaN-based high electron-mobility transistors see, e.g., Fang et al. (2019), for III-N LEDs see Kivisaari et al. (2015).

Ab Initio Approaches The development of the density-functional perturbation theory (DFPT) and its application to lattice-dynamics in crystals allows for calculating phonon dispersions (Baroni et al. 2001), which can be combined with electronic band structures and electron-electron scattering within a hot-electron

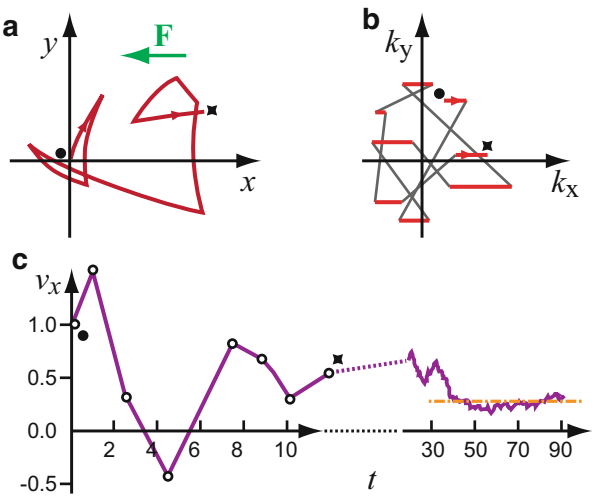


Fig. 9 Schematics for Monte Carlo method. (**a**) Real space path of an electron with large electric field **F** in the *x* direction; (**b**) same path as in (*a*) but in momentum space; (**c**) carrier velocity averaged over all steps starting from step 1 up to the running step number as a function of simulation time. After Brunetti and Jacoboni (1984)

ensemble computed using the *GW* method.[7] Such an ab initio approach yields results free of empirical parameters. Additionally calculating electron-phonon scattering-matrix elements using a Wannier-function formalism (Giustino et al. 2007), hot-carrier dynamics was more recently calculated for GaAs yielding electron-phonon relaxation times in agreement with ultrafast optical experiments (Bernardi et al. 2015). The study showed that the ultrafast (tens of femtoseconds) hot-electron relaxation time originates from electron-phonon scattering and that all optical and acoustic modes contribute substantially to electron-phonon scattering, with a dominant contribution from transverse acoustic modes. While the (doping dependent) equilibration in momentum space proceeds in the range of 10^1 to 10^2 fs, energy relaxation occurs on longer time scales and depends on the excess energy (Tanimura et al. 2016, 2021); a study of Si is reported by Minamitani (2021); see also Fischetti et al. (2019).

For a detailed description of the field dependence, the collision term in the Boltzmann equation must be evaluated analytically. This will be done for a few examples in the following sections.

3 Elastic and Inelastic Scattering at High Electric Fields

At low electric fields, the energy gained from the field is transmitted to the lattice predominantly via scattering with longitudinal optical phonons, see Sect. 4.6.2 of ▶ Chap. 22, "Carrier-Transport Equations", and Sect. 2.3 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields". Intermediate scattering with *acoustic* phonons is substantially elastic; this means that the electron gains more energy from a sufficiently high electric field than it can dissipate by generating acoustic phonons.

At higher fields the carrier temperature is increased markedly above the lattice temperature. However, when the carrier temperature approaches the Debye temperature, at which optical phonons can be generated in large quantities, a further rise in electron temperature is slowed down; that is, scattering increases substantially and the mobility decreases with increasing field.

A measure of the interaction between electrons and acoustic or optical phonons was obtained in Sect. 4.6.2 of ▶ Chap. 22, "Carrier-Transport Equations", from the ratio between momentum and energy relaxation. This ratio is changed at high fields according to the electron temperature, indicating the average energy gain of the electron ensemble from the field. For nondegeneracy, it can be estimated from (Seeger 2004)

$$\frac{\langle \tau_e \rangle}{\langle \tau_m \rangle} \simeq \frac{3k(T_e - T)}{8 m_n v_s^2}, \tag{24}$$

[7] G is the Green's function and W is the screened Coulomb potential; GW is employed for calculating the electron exchange-correlation interactions – see also Sects. 2.1.7 and 2.2.3 of ▶ Chap. 7, "Quantum Mechanics of Electrons in Crystals".

where v_s is the sound velocity, and T is the lattice temperature. The ratio of elastic to inelastic scattering events is not significantly changed for the *average* electron, since T_e usually remains close to T, while the *fast* electrons in the high-energy tail of the distribution are affected drastically (Brunetti and Jacoboni 1984). The dependence of τ_m on T for significantly varying electron temperature T_e is illustrated for optical nonpolar phonon scattering in Fig. 12.

3.1 Intravalley Scattering at High Electric Fields

When electrons are heated only slightly, their main scattering will remain with the abundant low-energy acoustic phonons (Srivastava 1990) determining the momentum relaxation.

3.1.1 Scattering with Acoustic Phonons

The dependence of the mobility on the electric field F can be estimated from the energy-balance equation. In steady state, the collision term must equal the incremental electron energy between collisions:

$$\left\langle -\frac{\partial E}{\partial t} \right\rangle_{\text{coll}} = \frac{m_n}{2} v_d^2 \left\langle \tau_m^{-1} \right\rangle = e\mu F^2. \tag{25}$$

When this energy is dissipated with acoustic phonons, the collision term can be estimated (Seeger 2004) as

$$\left\langle -\frac{\partial E}{\partial t} \right\rangle_{\text{coll}} = \frac{m_n}{2} v_{s,l}^2 \left\langle \tau_m^{-1} \right\rangle c_a \frac{T_e - T}{T}, \tag{26}$$

where $v_{s,l}$ is the velocity of longitudinal acoustic phonons, and c_a is a proportionality factor ($\cong 32/(3\pi)$). The factor $(T_e - T)/T$ is introduced to account for the increased average energy of the phonons created by collisions with warm electrons at energy $(3/2)kT_e$. Combining Eqs. 25 and 26, we obtain

$$\frac{T_e - T}{T} = \frac{2}{c_a} \left(\frac{\mu F}{v_{s,l}} \right)^2. \tag{27}$$

Introducing the approximation

$$\mu = \mu_0 \sqrt{\frac{T}{T_e}}, \tag{28}$$

yields for the field dependence of the mobility for *warm electrons* ($\mu_0 F \ll v_{s,l}$)

$$\mu \cong \mu_0 \left[1 + \frac{2}{c_a}\left(\frac{\mu_0 F}{v_{s,l}}\right)^2\right]^{-1/2}. \tag{29}$$

At higher fields ($\mu_0 F \gg v_{s,l}$) we obtain for the relative electron temperature

$$\frac{T_e}{T} \cong \frac{1}{2}\left[1 + \sqrt{1 + \frac{4}{c_a}\left(\frac{\mu_0 F}{v_{s,l}}\right)^2}\right], \tag{30}$$

which is shown in Fig. 10 as a function of the low-field drift velocity $\mu_0 F$, with the sound velocity as the family parameter. Eliminating T_e/T from Eqs. 28 and 30, we obtain the Shockley approximation (Shockley 1951)

$$\mu \cong \mu_0 \; c_a^{1/4} \sqrt{v_{s,l}/(\mu_0 F)}, \tag{31}$$

i.e., a square-root branch of the field dependence of the drift velocity v_d in the high-field regime

$$v_d = \mu F = c_a^{1/4} \sqrt{v_{s,l} \; \mu_0 F}. \tag{32}$$

The field dependence of the mobility for scattering with *acoustic* phonons is given by Eq. 29 and shown in Fig. 11

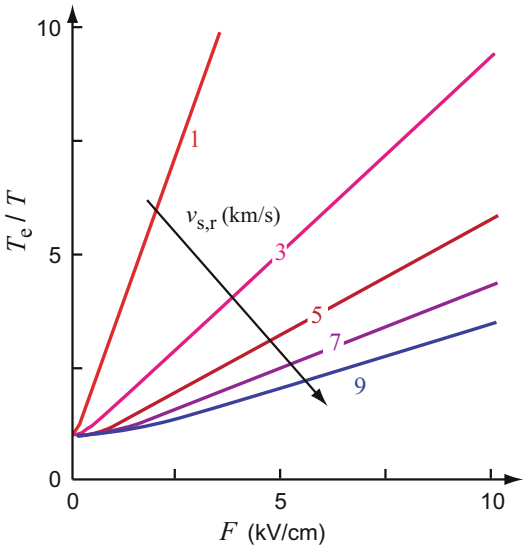


Fig. 10 Electron temperature for acoustic deformation-potential scattering as a function of the electric field with the sound velocity as the *family parameter*, according to Eq. 30

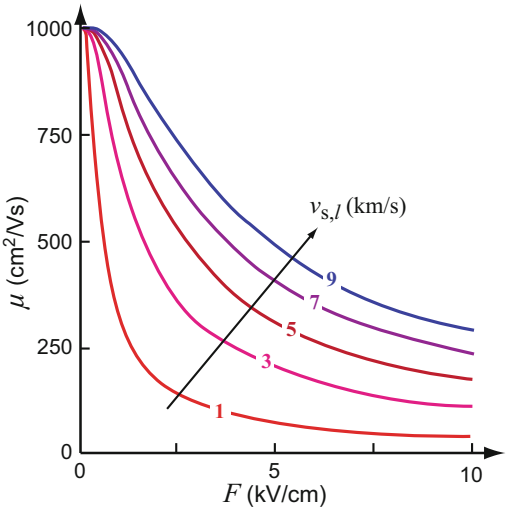


Fig. 11 Field dependence of the electron mobility for acoustic deformation-potential scattering according to Eq. 29. The sound velocity $v_{s,l}$ is the *family parameter*

3.1.2 Scattering with Optical Phonons

At sufficiently high electric fields a large number of electrons have enough energy to dissipate this surplus energy through the creation of longitudinal optical phonons.

Scattering with Optical Deformation Potential The scattering mechanism is similar to that described before, except that the fraction of electrons involved in this type of scattering is larger. Therefore, a more substantial reduction of the average relaxation time $\overline{\tau}$ results, and $\mu = (e/m_n)\overline{\tau}$ decreases more rapidly with the electron temperature. The dependence of the momentum-relaxation time τ_m on the ratio of lattice versus electron temperature T/T_e is illustrated in Fig. 12, calculated applying various approaches. The normalization factor τ_{op} represents the value of τ_m at $T = 300$ K. It decreases significantly for a given lattice temperature as T_e increases.

Applying an average relaxation time $\overline{\tau}$ yields a mobility (Seeger 2004):

$$\mu(T_e) = \frac{3\pi^{3/2}\hbar^2\rho\sqrt{k\Theta}}{2m_n^{3/2}D_o^2\Theta^{3/2}}\ \varphi_e(T_e) \tag{33}$$

with Θ the Debye temperature, D_o the optical deformation-potential constant of the band edge, ρ the mass density, and

$$\varphi_e(T_e) = \frac{\left(\frac{T}{2\Theta}\right)^{3/2}\sinh\left(\frac{\Theta}{2T}\right)}{\cosh\left(\frac{T_e - T}{T_e}\frac{\Theta}{2T}\right)\ K_2\left(\frac{\Theta}{2T_e}\right) + \sinh\left(\frac{T_e - T}{T_e}\frac{\Theta}{2T}\right)\ K_1\left(\frac{\Theta}{2T_e}\right)}. \tag{34}$$

For the modified Bessel functions K_1 and K_2 see Abramowitz and Stegun (1968). With increasing electron temperature, the electron mobility decreases, see also

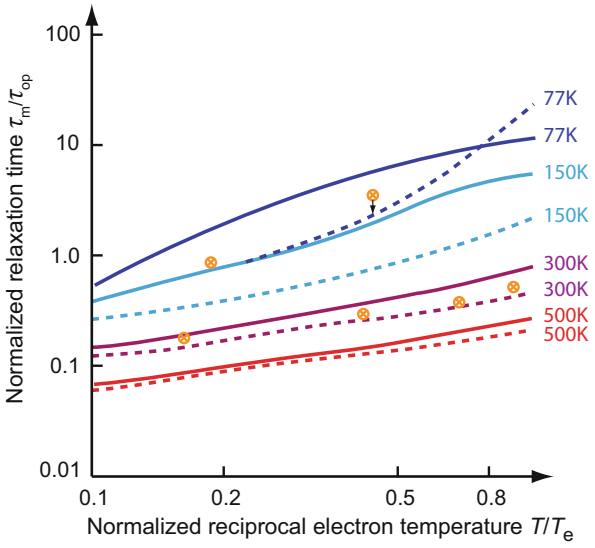


Fig. 12 Momentum-relaxation time τ_m for high electric fields, calculated as a function of electron temperature T_e with the lattice temperature T as *family parameter* for nonpolar optical phonon scattering. *Solid* and *dashed curves* refer to a Maxwell or a displaced Maxwell distribution function, respectively, *circles* are results of Monte-Carlo simulations. After Nag (1984)

Fig. 12. Introducing the field dependence of the electron temperature from Eq. 27, we obtain a drift velocity which first increases linearly with the field and then levels off at the saturation velocity

$$v_{d,sat} = \sqrt{3\,k\,\Theta\ \tanh(\Theta/(2T))/(4\,m_n)}. \tag{35}$$

The saturation drift-velocity can be obtained by eliminating the electric field F from the energy balance $e\ v_d\ F = (3\ k\ \Theta/2)\ (\exp(x) - 1)$ and the momentum balance $e\ F = 2 m_n v_d\ (\exp(x) + 1)$ with $x = \Theta/T$. It is close to the thermal velocity at the Debye temperature: $(m_n/2) v_{d,sat}^2 = (3/2)\, k\, \Theta\, c_{corr}$, where $c_{corr} = 1/\{4\ \coth\ [\Theta/(2T)]\}$ is a correction factor on the order of one. The temperature dependence of $v_{d,\ sat}$ is shown in Fig. 13.

The dependence of the drift velocity on the electric field F is shown in Fig. 14 for various values of the lattice temperature T. It indicates that, above the Debye temperature, saturation is approached for $\mu_0\ F \cong 2\ v_{d,sat}$. At lower temperatures, drift-velocity saturation occurs more gradually.

The saturation velocities for different materials are given in Fig. 15, as compiled by Ferry (1975), and indicate a satisfactory agreement with Eq. 35.

Polar Optical Scattering at High Fields With polar optical scattering, the mobility as a function of the electron temperature at high fields is similar to the optical deformation-potential scattering in the warm-electron range. An equation similar to Eq. 33 is obtained, except that the order of both modified Bessel functions in the φ_e dependence is reduced by one (Seeger 2004).

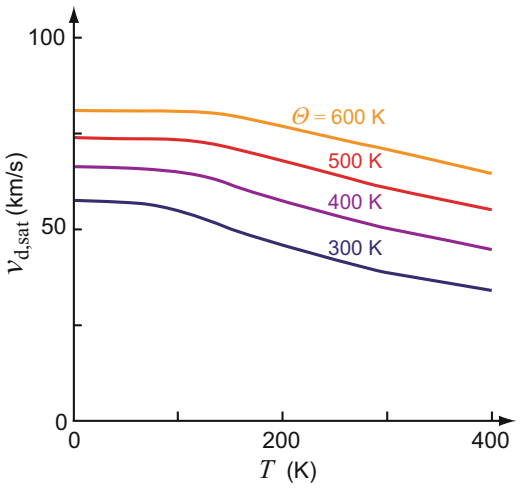


Fig. 13 Saturation drift-velocity as a function of the lattice temperature according to Eq. 35, for $m_n = m_0$, with the Debye temperature Θ as *family parameter*

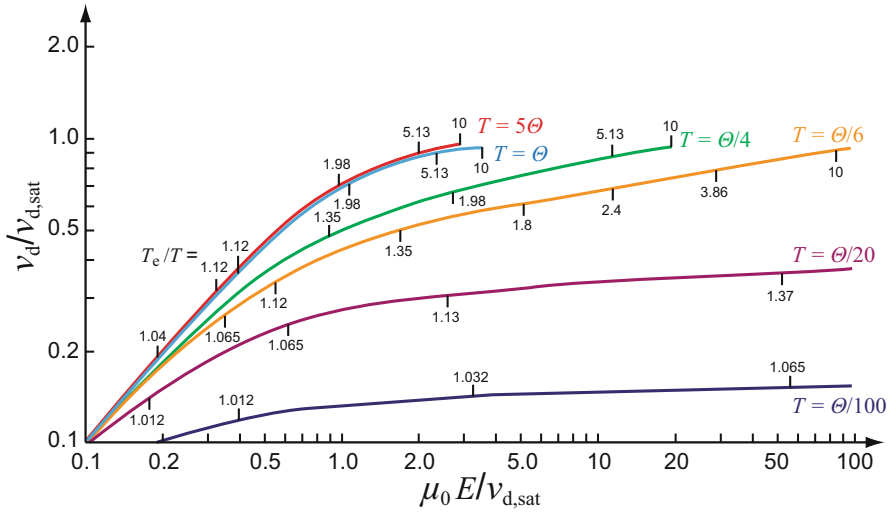


Fig. 14 Drift velocity for optical deformation-potential scattering as a function of the electric field F, with T/Θ as family parameter and the relative electron temperature T_e/T given along the various curves. After Seeger (2004)

The drift velocity does not saturate at high fields but shows an increase above the threshold field before the hot-electron range is reached. Here the onset of dielectric breakdown effects begins (Conwell 1967), see Fig. 16.

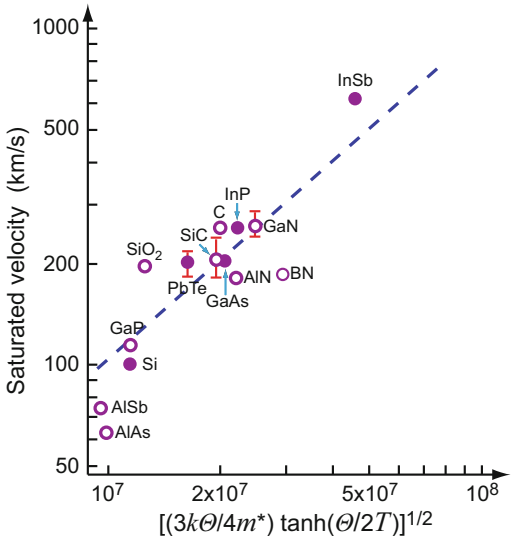


Fig. 15 Saturation drift-velocity $v_{d,sat}$ determined from experiment (*solid symbols*) or detailed calculation (*open circles*) including data from Ferry (1975), compared with the simple Eq. 35 (*dashed line*)

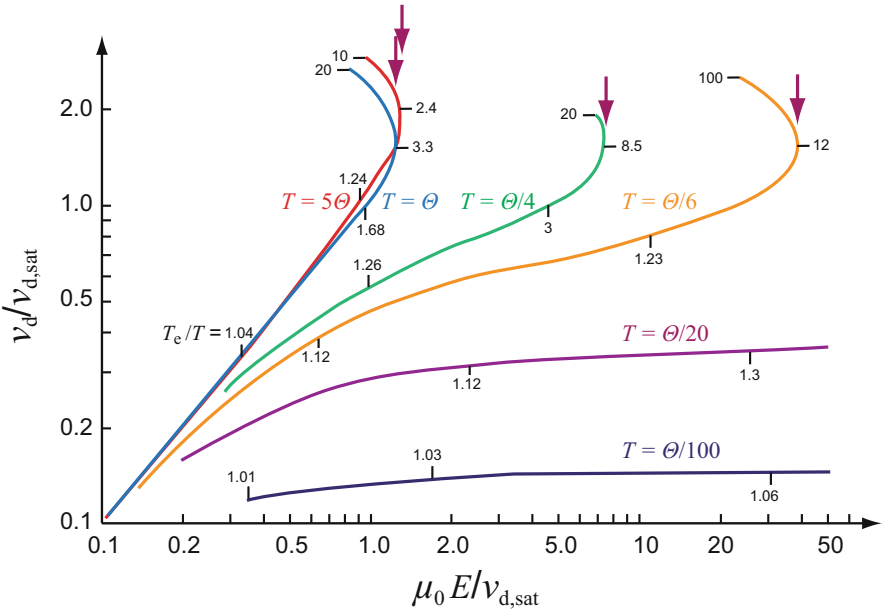


Fig. 16 Drift velocity as a function of the electric field for polar optical scattering with the lattice temperature as the family parameter, indicating the onset of dielectric breakdown effects (*arrows*). Values of the relative electron temperature are indicated along each of the curves. After Seeger (2004)

3.2 Intervalley Scattering at High Electric Fields

Similar to carrier scattering al low fields (Sect. 5 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields") we have to distinguish between intervalley scattering for equivalent and nonequivalent valleys (Kim and Yu 1991). Equivalent valleys have their minima at the same energy, the conduction-band edge. Nonequivalent valleys have a slightly higher energy and hence need additional carrier energy to become populated. For a review on intervalley scattering of electrons in GaAs see Mickevicius and Reklaitis (1990).

In the presence of intervalley scattering, carriers are transferred to satellite valleys where they are stored for some time. Scattering from the Γ valley to the L or X valleys does not arise from the polar LO scattering in compound semiconductors, but from phonons near the Brillouin zone boundary, and this is a very rapid process (Ferry 2021). The scattering to and from the satellite valleys can follow a variety of paths, for example, from Γ to X, and from here they can either return to Γ, or scatter to L. In any case, the electrons will after some time delay end up back in the Γ valley.

3.2.1 Equivalent-Intervalley Scattering

Within the satellite valleys, the dominant scattering is the intervalley scattering between equivalent valleys. Intervalley scattering with optical phonons can be evaluated for warm electrons using a deformation-potential approximation. The resulting mobility is given by

$$\mu_i = \frac{4\alpha_c\, k\, \Theta_i}{3\hbar}\sqrt{\frac{T_e}{\Theta_i}}\frac{\sinh\left(\dfrac{\Theta_i}{2T}\right)}{\cosh\left(\dfrac{T_e - T}{T_e}\dfrac{\Theta}{2T}\right)\; K_1\left(\dfrac{\Theta_i}{2T_e}\right)}, \tag{36}$$

where K_1 is the modified first-order Bessel function.

At higher electric fields, the anisotropy of the valleys must be taken into consideration. The direction of the field relative to the valley axis identifies the relevant effective mass, which changes with different alignment. In the **k**-space representation of Fig. 17 the effective mass is proportional to the square of the projection of the constant-energy surfaces in field direction (green arrows in the figure). It is smallest in the direction of the short axis and largest in the direction of the long axis of the rotational ellipsoid. Hence, electron heating, which is proportional to $\mu\, F = (e/m_n)\tau_m\, F$, is most effective when the electric field F is aligned with the short axis of the ellipsoid. This ellipsoid is called the *hot ellipsoid* and the corresponding valley the *hot valley*; the other with the long axis in field direction is called the *cool ellipsoid* and the corresponding valley the *cool valley.* Scattering proceeds preferentially from the hot to the cool valley, since hot electrons have a higher average energy and consequently generate more intervalley phonons. This influences the average electron mobility.

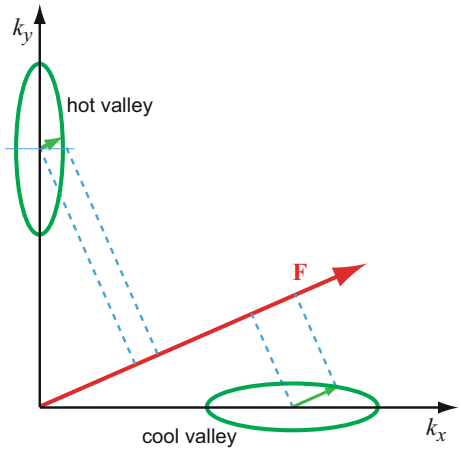


Fig. 17 Intervalley scattering into nonequivalent valleys from hot to cool valleys; *ellipses* represent constant-energy surfaces. Hot valleys have their short axis aligned closer to the field direction

The largest repopulation is observed when the field is applied in the direction of one of the main axes of the ellipsoids. No change in the population occurs when the field direction is 45° off the direction of the main axes.

The Sasaki-Shibuya Effect The anisotropy of the intervalley scattering at high fields, which causes a repopulation from hot to cool valleys, also causes the effective conductivity to become anisotropic. The current prefers to flow in a direction closer to an alignment with the long axis of the ellipsoid; that is, the current may deviate from the direction of the applied field. As a result, the surfaces of the semiconductor, having a component in this preferred direction, become charged. This produces a field component perpendicular to the applied electric field. It is similar to the Hall field, which is caused by surface charging in a magnetic field. Applied and induced field vectors define an angle ϑ, called the *Sasaki angle* (Fig. 18):

$$\frac{F_t}{F_l} = \frac{j_t}{j_l} = \tan\vartheta \cong \vartheta. \tag{37}$$

Subscripts t and l stand for *transverse* and *longitudinal.* The effect, called the *Sasaki-Shibuya effect*, can be used to obtain information about the anisotropy of the equi-energy surfaces in $E(\mathbf{k})$ (Shibuya 1955; Sasaki et al. 1959).

3.2.2 Intervalley Scattering into Nonequivalent Valleys

A rather large effect is observed when electrons are scattered into higher valleys in which they have a substantially higher effective mass – as, e.g., in GaAs.

In an external field of sufficient magnitude, electrons in a valley with a small effective mass are heated very efficiently. When scattered into higher satellite valleys (see Fig. 19) with a higher effective mass, heating is reduced and backscattering is lowered. Therefore, a substantial fraction of the conduction electrons can be pumped

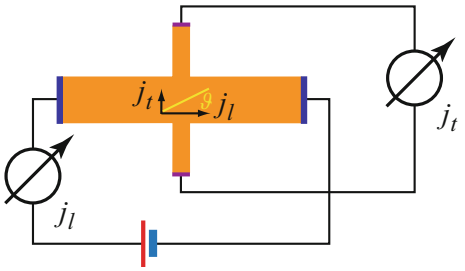


Fig. 18 Sasaki-Shibuya measurement of longitudinal and transverse currents in a semiconductor with elliptical valleys

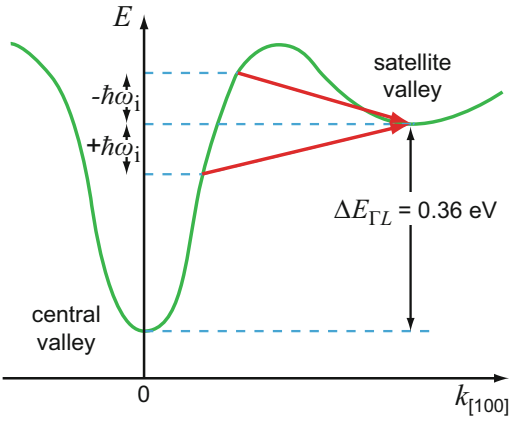


Fig. 19 Electron scattering (*arrows*) into a higher satellite valley with higher effective mass

into the higher valley. As a result, the average electron mobility is reduced, and the current therefore increases less than ohmically with increased field (see Jacoboni et al. 1981 for Ge and Alberigi-Quaranta et al. 1971 for compound semiconductors).

Negative Differential Conductivity Effects If the ratio of mobilities is high enough, one observes a *negative differential conductivity* (Singh 1993) as shown in curve 2 of Fig. 20a. Corresponding experimental results are shown in Fig. 20b for several III-V compounds. Here a field range exists in which the current decreases with increasing field. In this range, stationary solutions of the transport equation may not exist, and high-field domains may develop which can move through the semiconductor and cause current oscillations. These have been observed by Gunn (1963) and are termed *Gunn-effect* oscillations (Dalven 1990).

If such a repopulation, with an increasing electric field, competes with thermal excitation at higher temperatures, only a change in the slope of the current-voltage characteristic is observed, without going through a maximum (curve 1 in Fig. 20a). Increased doping in the high-doping range causes a decrease in the threshold field and finally the disappearance of the negative differential conductivity (Seeger 2004).

In GaAs, the L satellite minimum lies 0.36 eV above the Γ point $E(\mathbf{k})$ minimum at $\mathbf{k} = 0$ – see Fig. 19. The effective mass in the central minimum is $m_n = 0.07\, m_0$, and

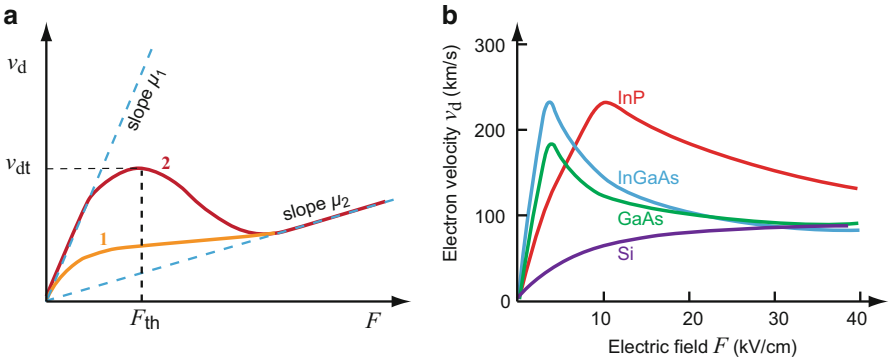


Fig. 20 (**a**) Drift velocity as a function of the field when major repopulation of the higher satellite valley with substantially lower effective mass becomes effective at the threshold field F_{th} (schematic). (**b**) Measured field dependence of the drift velocity for several semiconductors. After Evans and Robson (1974)

$m_n \cong m_0$ in the satellite minima.[8] This causes a reduction of the mobility in *n*-type GaAs from ~8000 cm^2/Vs at low fields to ~200 cm^2/Vs at high fields, when most of the electrons are pumped into the satellite valley. Fields of 3 kV/cm are sufficient for achieving the necessary pumping. A similar effect is observed when strong electron-electron scattering is present, see Asche (1989) and Ohno (1994).

In group-III nitrides, both intervalley electron-transfer to satellite valleys with heavier effective mass or the negative-mass region beyond the inflection point of $m_n(k)$ may generate a negative differential resistance (NDR). Hybrid functional DFT calculations by Yan et al. (2014) indicate that electron-transfer NDR will be dominant in AlN, while negative-mass NDR can be realized in GaN; in $Al_xGa_{1-x}N$, the change of dominant character is expected at $x = 0.64$. For GaN, experiment and simulation show a negative differential mobility above an electron-velocity peak of 2.4×10^7 cm/s at 180 kV Schwierz (2005), and Gunn-effect oscillations were reported by Yilmazoglu et al. (2008).

4 Summary

High external electric fields cause a deformation of the carrier distribution with more carriers at higher energies. In a first approximation, such a deformed distribution can be approximated by a Boltzmann distribution, assuming an equilibrated carrier ensemble at a carrier temperature T_e above lattice temperature T, with acceptable errors when $T_e - T \ll T$ (warm carriers). When carriers are accelerated further, the distribution is skewed substantially near and above an energy $\hbar\omega_{LO}$, where strong

[8]The effective mass in one of the four satellite minima is 0.4 m_0; the density-of-state mass is hence $(0.4^{3/2} \times 4)^{2/3}\, m_0 = 1.01\, m_0$).

inelastic scattering occurs, and the concept of carrier heating with a further elevated T_e (hot carriers) becomes less satisfactory. A better description includes higher terms in the Fourier development of the distribution function or mobility. Also a numerical solution of the Boltzmann equation yields better results.

With increasing electric field F, an increased scattering counteracts an increased accumulation of energy from the field. Such increased scattering is observed for interaction with phonons, causing a decrease in mobility: first $\propto 1/\sqrt{F}$ from mostly acoustical scattering, and at higher fields $\propto 1/F$ from LO-phonon scattering. Scattering with ionized impurities, however, decreases with increasing field, as carriers can penetrate closer to the center before being scattered.

The dominating effect of LO-phonon scattering causes a saturation of the drift velocity close to the rms velocity of carriers. Similarly, a saturation of the diffusion current is observed at high carrier-density gradients, with a maximum effective diffusion velocity also close to the rms velocity. With additional thermal excitation, the saturation branch of the drift velocity may be hidden by dielectric breakdown effects.

References

Abramowitz M, Stegun IA (eds) (1968) Handbook of mathematical functions. Dover, New York

Alberigi-Quaranta A, Jacoboni C, Ottaviani G (1971) Negative differential mobility in III-V and II-VI semiconducting compounds. Rivista del Nuovo Comento 1:445

Asche M (1989) Multivalued hot electron distributions as spontaneous symmetry breaking. Solid State Electron 32:1633

Baroni S, de Gironcoli S, Dal Corso A, Paolo G (2001) Phonons and related crystal properties from density-functional perturbation theory. Rev Mod Phys 73:515

Bernardi M, Vigil-Fowler D, Ong CS, Neaton JB, Louie SG (2015) Ab initio study of hot electrons in GaAs. Proc Natl Acad Sci U S A 112:5291

Böer KW (1985a) High-field carrier transport in inhomogeneous semiconductors. Ann Phys 497:371

Böer KW (1985b) Current-voltage characteristics of diodes with and without light. Phys Stat Sol A 87:719

Böer KW, Bogus K (1968) Electron mobility in CdS at high electric fields. Phys Rev 176:899

Böer KW, Voss P (1968) Stationary high-field domains in the range of negative differential conductivity in CdS single crystals. Phys Rev 171:899

Bray R, Brown DM (1960) Lattice scattering mechanisms in *p*-type germanium. In: Proceedings of the 5th international conference on the physics of semiconductors, Prague, Academic, New York, pp 82–85

Brunetti R, Jacoboni C (1984) Transient and stationary properties of hot-carrier diffusivity in semiconductors. In: Alfano RR (ed) Semiconductors probed by ultrafast laser spectroscopy vol I. Academic, Orlando, pp 367–412

Brunetti R, Jacoboni C, Nava F, Reggiani L, Bosman G, Zijlstra RJJ (1981) Diffusion coefficient of electrons in silicon. J Appl Phys 52:6713

Budd HF (1966) Hot carriers and the variable path method. J Phys Soc Jpn Suppl 21:420

Budkin GV, Tarasenko SA (2011) Heating and cooling of a two-dimensional electron gas by terahertz radiation. J Exp Theor Phys 112: 656

Burtyka MV, Glukhov OV, Yakovenko VM (1991) Interaction of hot electrons with two-dimensional gas in semiconductor superlattices. Solid State Electron 34:559

Canali C, Majni G, Minder R, Ottaviani G (1975) Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature. IEEE Trans Electron Devices 22:1045

Chambers RG (1952) The kinetic formulation of conduction problems. Proc Phys Soc Lond, Sect A 65:458

Chen K, Meng X, He F, Zhou Y, Jeong J, Sheehan S, Bank SR, Wang Y (2018) Comparison between grating imaging and transient grating techniques on measuring carrier diffusion in semiconductor. Nanosc Microsc Thermophys Eng 22(4): 348

Conwell EM (1967) High-field transport in semiconductors. Academic, New York

Cross AJ, Kent AJ, Hawker P, Lehmann D, Cz J, Henini M (1999) Phonon emission by warm electrons in GaAs quantum wells: the effect of well width on the acoustic-optic crossover. Physica B 263:526

Dalven R (1990) Introduction to applied solid state physics, 2nd edn. Plenum Press, New York

Dember H (1931) Über eine photoelektronische Kraft in Kupferoxydul-Kristallen. Phys Z 32:554. (On a photoelectric E.M.F. in cuprous oxide crystals, in German)

Evans AGR, Robson PN (1974) Drift mobility measurements in thin epitaxial semiconductor layers using time-of-flight techniques. Solid State Electron 17:805

Fang J, Fischetti MV, Schrimpf RD, Reed RA, Bellotti E, Pantelides ST (2019) Electron transport properties of $Al_xGa_{1-x}N$/GaN transistors based on first-principles calculations and Boltzmann-equation Monte Carlo simulations. Phys Rev Appl 11: 044045

Ferry DK (1975) High-field transport in wide-band-gap semiconductors. Phys Rev B 12:2361

Ferry DK (1980) Modeling of carrier transport in the finite collision duration regime: effects in submicron semiconductor devices. In: Ferry DK, Barker JR, Jacoboni C (eds) Physics of nonlinear transport in semiconductors. Plenum Press, New York, pp 577–588

Ferry DK (2021) Non-equilibrium longitudinal optical phonons and their lifetimes. Appl Phys Rev 8: 021324

Fischetti MV, Yoder PD, Khatami MM, Gaddemane G, Van de Put ML (2019) "Hot electrons in Si lose energy mostly to optical phonons": truth or myth? Appl Phys Lett 114: 222104

Fröhlich H, Paranjape BV (1956) Dielectric breakdown in solids. Proc Phys Soc Lond, Sect B 69:21

Giustino F, Cohen ML, Louie SG (2007) Electron-phonon interaction using Wannier functions. Phys Rev B 76:165108

Gunn JB (1963) Microwave oscillations of current in III–V semiconductors. Solid State Commun 1:88

Guzelturk B, Belisle RA, Smith MD, Bruening K, Prasanna R, Yuan Y, Gopalan V, Tassone CJ, Karunadasa HI, McGehee MD, Lindenberg AM (2018) Terahertz emission from hybrid perovskites driven by ultrafast charge separation and strong electron-phonon coupling. Adv Mater 30: 1704737

Habegger MA, Fan HY (1964) Oscillatory intrinsic photoconductivity of GaSb and InSb. Phys Rev Lett 12:99

Haug H, Koch SW (1990) Quantum theory of optical and electronic properties of semiconductors. World Scientific, Singapore

Henneberger F, Schmidt-Rink S, Göbel EO (eds) (1993) Optics of semiconductor nanostructures. Akademie-Verlag, Berlin

Ivanov PA, Levinshtein ME, Palmour JW, Rumyantsev SL, Singh R (2000) 'Classical' current-voltage characteristics of 4H-SiC p^+-n junction diodes. Semicond Sci Technol 15:908

Iveland J, Martinelli L, Peretti J, Speck JS, Weisbuch C (2013) Direct measurement of Auger electrons emitted from a semiconductor light-emitting diode under electrical injection: identification of the dominant mechanism for efficiency droop. Phys Rev Lett 110: 177406

Jacoboni C, Reggiani L (1979) Bulk hot-electron properties of cubic semiconductors. Adv Phys 28:493

Jacoboni C, Reggiani L (1983) The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. Rev Mod Phys 55:645

Jacoboni C, Canali C, Ottaviani G, Alberigi Quaranta A (1977) A review of some charge transport properties of silicon. Solid State Electron 20:77
Jacoboni C, Lugli P (1989) The Monte Carlo method for semiconductor device simulation. Springer, Wien
Jacoboni C, Nava F, Canali C, Ottaviani G (1981) Electron drift velocity and diffusivity in germanium. Phys Rev B 24:1014
Kim D-S, Yu PY (1991) Hot-electron relaxations and hot phonons in GaAs studied by sub-picosecond Raman scattering. Phys Rev B 43:4158
Kini RN, Vaisakh CP (2020) Terahertz emission mechanisms in III–V semiconductors: The influence of isoelectronic dopants. In: Biswas A, Banerjee A, Acharyya A, Inokawa H, Roy J (eds) Emerging trends in terahertz solid-state physics and devices. Springer, Singapore, pp 169-187
Kivisaari P, Oksanen J, Tulkki J, Sadi T (2015) Monte Carlo simulation of hot carrier transport in III-N LEDs. J Comput Electron 14: 382
Kurosawa T (1966) Monte Carlo simulation of hot electron problems. J Phys Soc Jpn Suppl 21:424
Maňkoš M, Tromp RM, Reuter MC, Cartier E (1996) Imaging hot-electron emission from metal-oxide-semiconductor structures. Phys Rev Lett 76: 3200
Mickevicius R, Reklaitis A (1990) Electron intervalley scattering in gallium arsenide. Semicond Sci Technol 5:805
Miller A (1999) Transient grating studies of carrier diffusion and mobility in semiconductors. In Garmire E, Kost A (eds) Nonlinear Optics in Semiconductors II. Academic, San Diego, pp 287–312
Mimila-Arroyo J (2019) On the importance of the junction temperature on obtaining the charge transport parameters of a solar cell. Semicond Sci Technol 34: 065019
Minamitani E (2021) *Ab initio* analysis for the initial process of Joule heating in semiconductors. Phys Rev B 104: 085202
Nag BR (1980) Electron transport in compound semiconductors. Springer, Berlin
Nag BR (1984) Relaxation of carriers. In: Alfano RR (ed) Semiconductors probed by ultrafast laser spectroscopy vol I. Academic, Orlando, pp 3–44
Oh KH, Ong CK, Tan BTG (1992) Field dependence of the overshoot phenomena in InP. J Phys Chem Solids 53:555
Ohno Y (1990) Estimation of velocity-overshoot in small size semiconductors. Solid State Electron 33:935
Ohno Y (1994) A new bulk negative differential conductance mechanism with multiple steady states. Semicond Sci Technol 9:615
Price PJ (1977) Calculation of hot electron phenomena. Solid State Electron 21:9
Prior AC (1960) A reversed carrier transport effect in germanium. Proc Phys Soc 76:465
Rees HD (1969) Calculation of distribution functions by exploiting the stability of the steady state. J Phys Chem Solids 30:643
Rees HD (1972) Numerical solution of electron motion in solids. J Phys C 5:641
Reklaitis A (2012) High field electron and hole transport in wurtzite InN. Phys Stat Sol (b) 249:1566
Ridley BK (1997) Electrons and phonons in semiconductor multilayers. Cambridge University Press, Cambridge/New York
Roberson HS (1993) Statistical thermophysics. Prentice-Hall, Englewood Cliffs
Rode DL (1970) Electron mobility in direct-gap polar semiconductors. Phys Rev B 2:1012
Sasaki W, Shibuya M, Mizuguchi K, Hatoyama G (1959) Anisotropy of hot electrons in germanium. J Phys Chem Solids 8:250
Schwierz F (2005) An electron mobility model for wurtzite GaN. Solid State Electron 49:889
Seeger K (2004) Semiconductor physics, 9th edn. Springer, Berlin
Shibuya M (1955) Hot electron problem in semiconductors with spheroidal energy surfaces. Phys Rev 99:1189
Shockley W (1951) Hot electrons in germanium and Ohm's law. Bell Syst Tech J 30:990

Siddiqua P, O'Leary SK (2018) Electron transport within the wurtzite and zinc-blende phases of gallium nitride and indium nitride. J Mater Sci Mater Electron 29: 3511

Singh J (1993) Physics of semiconductors and their heterostructures. McGraw-Hill, New York

Smirl AL (1984) Dynamics of high-density transient electron-hole plasmas in germanium. In: Alfano RR (ed) Semiconductors probed by ultrafast laser spectroscopy vol I. Academic, Orlando, pp 198–273

Smirl AL, Moss SC, Lindle JR (1982) Picosecond dynamics of high-density laser-induced transient plasma gratings in germanium. Phys Rev B 25:264

Srivastava GP (1990) The physics of phonons. Hilger, Bristol

Tanimura H, Kanasaki J, Tanimura K, Sjakste J, Vast N, Calandra M, Mauri F (2016) Formation of hot-electron ensembles quasiequilibrated in momentum space by ultrafast momentum scattering of highly excited hot electrons photoinjected into the Γ valley of GaAs. Phys Rev B 93:161203

Tanimura H, Tanimura K, Kanasaki J (2021) Ultrafast relaxation of photoinjected nonthermal electrons in the Γ valley of GaAs studied by time- and angle-resolved photoemission spectroscopy. Phys Rev B 104: 245201

Tonouchi M (2020) Simplified formulas for the generation of terahertz waves from semiconductor surfaces excited with a femtosecond laser. J Appl Phys 127: 245703

van Driel HM (1985) Physics of pulsed laser processing of semiconductors. In: Alfano RR (ed) Semiconductors probed by ultrafast laser spectroscopy vol II. Academic, Orlando, pp 57–94

Wirner C, Witzany M, Kiener C, Zandler G, Bohm G, Gornik E, Vogl P, Weimann G (1992) Experimental and theoretical investigation of the drift velocity and velocity distribution function in GaAs/AlGaAs heterostructures. Semicond Sci Technol 7:B267

Yan Q, Kioupakis E, Jena D, Van de Walle CG (2014) *First-principles* study of high-field-related electronic behavior of group-III nitrides. Phys Rev B 90: 121201

Yilmazoglu O, Mutamba K, Pavlidis D, Karaduman T (2008) First observation of bias oscillations in GaN Gunn diodes on GaN substrate. IEEE Transact Electron Devices 55: 1563

Yu PY, Cardona M (1996) Fundamentals of semiconductors, ch 5. Springer, Berlin

Zakhleniuk NA (2006) Nonequilibrium drift-diffusion transport in semiconductors in presence of strong inhomogeneous electric fields. Appl Phys Lett 89: 252112

Carriers in Magnetic Fields and Temperature Gradients

25

Karl W. Böer and Udo W. Pohl

Contents

1 Transport Equations and Thermoelectric Effects 1056
1.1 The Boltzmann Equation in Magnetic Fields and Temperature Gradients 1056
1.2 Transport Equations 1058
1.3 Cyclotron Resonance 1070
2 Quantum Effects in a Strong Magnetic Field 1070
2.1 Quasi-Free Carriers in a Strong Magnetic Field 1070
2.2 Diamagnetic and Paramagnetic Electron Resonance 1072
2.3 Density of States and DeHaas-Type Effects 1074
3 Ballistic Transport in Strong Magnetic Fields 1077
3.1 The Integer Quantum-Hall Effect 1077
3.2 The Fractional Quantum Hall Effect 1081
4 Summary 1083
References 1084

Abstract

The application of a magnetic field in addition to an electric field yields significant information on carrier polarity and mobility, on the effective mass, and on the origin of energy levels in paramagnetic centers. If a temperature gradient exists in addition to an electric field, thermoelectric effects occur with useful applications, such as the Seebeck effect rendering thermoelectricity used in thermocouples and the Peltier effect applied for cooling. If a magnetic field is added to the temperature gradient and to the electric field, several galvanomagnetic and thermomagnetic effects are observed.

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_25

In strong magnetic fields, the electronic density of states is changed: energy levels condense on quantized Landau levels with cylindrical equi-energy surfaces in **k** space. Quantities controlled by their vicinity to the Fermi energy then show an oscillatory dependence on the magnetic field, such as the DeHaas-van Alphen oscillations of the magnetic susceptibility and Shubnikov-DeHaas oscillations of the resistivity.

If scattering is suppressed in highly pure samples at very low temperature, a strong magnetic field forces carriers to propagate on edge states at the sample surface, creating a topological insulator with no conductance in the bulk. In a two-dimensional electron gas, this leads to the quantum Hall effect, which established an international metrological standard for the electrical resistance. The related fractional quantum Hall effect lead to the discovery of composite fermions, quasi-particles composed of an electron and flux quanta, which conjointly carry a fractional charge. The quantum spin Hall phase represents a third type of topological insulators, which require no external magnetic field.

Keywords

DeHaas-van Alphen effect · Fractional quantum Hall effect · Galvanomagnetic effects · Hall effect · Hall mobility · Landau levels · Magnetoresistance · Peltier effect · Quantum Hall effect · Quantum spin Hall phase · Seebeck effect · Shubnikov-DeHaas effect · Thermoelectric effects · Thermomagnetic effects · Topological insulator

1 Transport Equations and Thermoelectric Effects

The transport of carriers is described by the Boltzmann equation introduced in Sect. 4 of ▶ Chap. 22, "Carrier-Transport Equations". We now include an additional magnetic field or a temperature gradient into this equation and analyze their influences on the electron transport.

1.1 The Boltzmann Equation in Magnetic Fields and Temperature Gradients

The steady-state Boltzmann equation

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \dot{\mathbf{k}} \cdot \frac{\partial f}{\partial k} + \dot{\mathbf{r}} \cdot \frac{\partial f}{\partial r}, \tag{1}$$

which describes the transport of electrons under the influence of external fields, can readily be expanded to include the *magnetic field*[1] when expressing the forces acting on the electrons (see Eq. 85 of ▶ Chap. 22, "Carrier-Transport Equations") by the sum of the electric field **F** and Lorentz forces exerted by the magnetic field:

$$\dot{\mathbf{k}} = \frac{e}{\hbar}(\mathbf{F} + \dot{\mathbf{r}} \times \mathbf{B}). \tag{2}$$

Here $\dot{\mathbf{r}}$ is the group velocity of the electron wave-packet and **B** is the magnetic induction. The distribution function $f = f(\mathbf{r}, \mathbf{k}, T, t)$ contains the temperature, which also can include temperature gradients.

Using the relaxation-time approximation and setting $f = f_0 + \delta f$, where δf is a small perturbation of the Fermi distribution-function f_0, we have

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{\delta f}{\tau(E)}. \tag{3}$$

Assuming that each collision probability is independent of the collision angle, the Boltzmann equation can be integrated. This can be done in a closed form when $B = 0$:

$$f = \frac{f_0(1-f_0)\tau}{kT}\left(\mathbf{v} \cdot \left[e\mathbf{F} - \frac{\partial E_{\text{F}}}{\partial \mathbf{r}} - \frac{E - E_{\text{F}}}{T}\frac{\partial T}{\partial \mathbf{r}}\right]\right). \tag{4}$$

Here and in the following equations, the Fermi energy E_{F} is used, which includes the potential energy as well as changes in the carrier density, and permits a simplified expression. When applying the equation to a deviation from thermal equilibrium, E_{F} must be replaced by the quasi-Fermi energies $E_{\text{F}n}$ or $E_{\text{F}p}$ for electrons or holes, respectively – see Sect. 3.3 of ▶ Chap. 22, "Carrier-Transport Equations".

When the magnetic induction **B** is included, its influence can no longer be treated as a small perturbation. In contrast to the electrical and thermal conductivities, which are observed at small fields, typical magnetical effects, such as the Hall effect and magneto-resistance, require rather large fields to become observable. Mathematically, this means that the term proportional to the *gradient* of δf must also be taken into consideration. This yields

$$\delta f = \frac{f_0(1-f_0)\tau}{kT}\left(\mathbf{v} \cdot \left[e\mathbf{F} - \frac{\partial E_{\text{F}}}{\partial \mathbf{r}} - \frac{E - E_{\text{F}}}{T}\frac{\partial T}{\partial \mathbf{r}}\right]\right) + \frac{e}{\hbar}\tau(\mathbf{v} \times \mathbf{B})\frac{\partial \delta f}{\partial \mathbf{k}}, \tag{5}$$

[1] In the following sections the *magnetic induction* **B** is used, which is connected to the *magnetic field* **H** by $\mathbf{B} = \mu\mu_0\mathbf{H}$, with μ_0 the permeability of free space and μ the relative permeability. Occasionally, the *magnetization* **M** is used, defined by $\mathbf{B} = \mu_0(\mathbf{H} + \mathbf{M})$ with $\mathbf{M} = \chi_{\text{mag}}\mathbf{H}$, $\mu = 1 + \chi_{\text{mag}}$, and with χ_{mag} the *magnetic susceptibility*, see ▶ Chap. 9, "Magnetic Semiconductors".

which can be solved by iteration. Evaluating the solution near the bottom of the conduction band, where, for spherical equi-energy surfaces, the electron velocity can be expressed as $\mathbf{v} = \hbar\mathbf{k}/m_n$, we obtain

$$\delta f = \frac{f_0(1-f_0)}{kT} \frac{\tau[\mathbf{v}\cdot\mathbf{f} + \mathbf{v}\cdot(\mathbf{b}\times\mathbf{f}) + (\mathbf{v}\cdot\mathbf{b})(\mathbf{b}\cdot\mathbf{f})]}{1+b^2}, \tag{6}$$

with the abbreviations

$$\mathbf{b} = \frac{e}{m_n}\tau\mathbf{B} \quad \text{and} \quad \mathbf{f} = e\mathbf{F} + \frac{\partial E_F}{\partial \mathbf{r}} + \frac{E-E_F}{T}\frac{\partial T}{\partial \mathbf{r}}. \tag{7}$$

The distribution function $f = f[\mathbf{r}, \mathbf{k}, T(\mathbf{r}), \mathbf{B}, t]$ now contains the influence of electric, thermal, and magnetic fields; for more detail, see Haug (1972) and Madelung (1981).

The deformed distribution function causes changes in the transport properties, i.e., changes in the electrical or thermal currents as a result of the interacting fields. Rather than following a stringent development of the transport from the Boltzmann equation, a task first solved for carrier conduction by Bloch, the following section will take an alternative, semiempirical approach by describing the different currents with proportionality constants. These constants are later interpreted by a microscopic model.

1.2 Transport Equations

The two governing transport equations, dealing a with carrier current **j** and an energy (heat) current per unit area **w** [with $|\mathbf{w}| = \partial q/(A\partial t)$ and q the heat], are given in their general form as

$$\mathbf{j} = -\frac{e}{\hbar}\int \frac{\partial E(\mathbf{k})}{\partial \mathbf{k}}\; g(\mathbf{k})\; f[\mathbf{r},\mathbf{k},T(\mathbf{r}),\mathbf{B},t]\,d\mathbf{k}, \tag{8}$$

$$\mathbf{w} = \frac{1}{\hbar}\int \frac{\partial E(\mathbf{k})}{\partial \mathbf{k}}\; E(\mathbf{k})\; g(\mathbf{k})\; f[\mathbf{r},\mathbf{k},T(\mathbf{r}),\mathbf{B},t]\,d\mathbf{k}; \tag{9}$$

they contain the density of states $g(\mathbf{k})$ and the distribution functions developed in the previous section.

In addition, we need two *conservation laws* to describe the transport behavior in a homogeneous semiconductor: the conservation of the *number of carriers*

$$e\frac{\partial n}{\partial t} + \frac{\partial}{\partial \mathbf{r}}\cdot\mathbf{j} = 0, \tag{10}$$

and the conservation of *energy*[2]

$$\rho \frac{\partial u}{\partial t} + \frac{\partial}{\partial \mathbf{r}} \mathbf{w} = -\mathbf{F} \cdot \mathbf{j}, \tag{11}$$

where ρ is the mass density and u is the specific internal energy.

The solution of these transport equations in steady state can be expressed as a linear combination of transport parameters and driving forces. For example, when only electric fields act we obtain

$$\begin{aligned} \mathbf{j} &= \alpha_{11} \frac{\partial \varphi}{\partial \mathbf{r}} \\ \mathbf{w} &= \alpha_{21} \frac{\partial \varphi}{\partial \mathbf{r}}; \end{aligned} \tag{12}$$

when electric and thermal fields act we have

$$\begin{aligned} \mathbf{j} &= \alpha_{11} \frac{\partial \varphi}{\partial \mathbf{r}} + \alpha_{12} \frac{\partial T}{\partial \mathbf{r}} \\ \mathbf{w} &= \alpha_{21} \frac{\partial \varphi}{\partial \mathbf{r}} + \alpha_{22} \frac{\partial T}{\partial \mathbf{r}}; \end{aligned} \tag{13}$$

and when incorporating in addition a magnetic field we obtain

$$\begin{aligned} \mathbf{j} = {} & \alpha_{11} \frac{\partial \varphi}{\partial \mathbf{r}} + \alpha_{12} \frac{\partial T}{\partial \mathbf{r}} + \beta_{11} \left(\mathbf{B} \times \frac{\partial \varphi}{\partial \mathbf{r}} \right) + \beta_{12} \left(\mathbf{B} \times \frac{\partial T}{\partial \mathbf{r}} \right) \\ & + \gamma_{11} \mathbf{B} \cdot \left(\mathbf{B} \cdot \frac{\partial \varphi}{\partial \mathbf{r}} \right) + \gamma_{12} \mathbf{B} \cdot \left(\mathbf{B} \cdot \frac{\partial T}{\partial \mathbf{r}} \right) \\ \mathbf{w} = {} & \alpha_{21} \frac{\partial \varphi}{\partial \mathbf{r}} + \alpha_{22} \frac{\partial T}{\partial \mathbf{r}} + \beta_{21} \left(\mathbf{B} \times \frac{\partial \varphi}{\partial \mathbf{r}} \right) + \beta_{22} \left(\mathbf{B} \times \frac{\partial T}{\partial \mathbf{r}} \right) \\ & + \gamma_{21} \mathbf{B} \cdot \left(\mathbf{B} \cdot \frac{\partial \varphi}{\partial \mathbf{r}} \right) + \gamma_{22} \mathbf{B} \cdot \left(\mathbf{B} \cdot \frac{\partial T}{\partial \mathbf{r}} \right). \end{aligned} \tag{14}$$

Here φ is the electrochemical *potential,* distinguished from the electrochemical *energy* $e\varphi \equiv E_F$, the Fermi energy. At a deviation from thermodynamic equilibrium, we have to replace φ with φ_n, the quasi-Fermi potential for electrons and φ_p for holes (see Sect. 3.3 of ▶ Chap. 22, "Carrier-Transport Equations"). The coefficients α_{ik}, β_{ik}, and γ_{ik} are the *transport coefficients*: the electrical conductivity $\alpha_{11} = \sigma_c$, the Seebeck coefficient

[2]Although the electric and magnetic fields act as external forces, and we have e ($\mathbf{F} + \mathbf{v} \times \mathbf{B}$) as total force, the scalar product of ($\mathbf{v} \times \mathbf{B}$) · $\mathbf{j}$ is zero since the vectors $\mathbf{v} \times \mathbf{B}$ and $\mathbf{j}$ are perpendicular to each other; in first approximation, there is no energy input into the carrier gas from a magnetic field.

$-\alpha_{11}^{-1}\alpha_{12} = S$, the Peltier coefficient $\alpha_{21}\alpha_{11}^{-1} = \Pi$, and the electronic contribution of the thermal conductivity involving the respective carriers (with subscript $c = n$ or p) $\alpha_{21}\alpha_{11}^{-1}\alpha_{12} - \alpha_{22} = \kappa_c$; for details, see Beer 1963; Madelung 1981.

In anisotropic semiconductors, each of the transport parameters is a *tensor,* e.g., $\sigma_n = \sigma_{ik}^{(n)} = e\,n\,\mu_{ik}$. Important relations connect the different transport coefficients, such as the *Onsager relations,* obtained from the reciprocity of the effects

$$\alpha_{ik}(\mathbf{B}) = \overline{\alpha}_{ki}(-\mathbf{B}), \tag{15}$$

where $\overline{\alpha}_{ki}$ is the transposed tensor of α_{ik}.

The transport coefficients are directly accessible through experimental observation. Their magnitudes depend on the relative orientation of the different fields and, for an anisotropic semiconductor, also on the relative crystallographic orientation. In samples that permit electrical currents in the x direction only, one also distinguishes *isothermal* and *adiabatic galvanomagnetic effects*,[3] depending on whether $\partial T/\partial y = 0$ or $w_y = 0$ (see Madelung 1981).

An overview of the different possibilities in an isotropic semiconductor is given in Tables 1 and 2. Some of the effects listed have gained technical interest or are used extensively for analytical purposes; the Hall effect and the Peltier effect are examples.

In the following sections, a few of the more important effects will be discussed in some detail and the corresponding transport parameters will be analyzed in a microscopic model to yield information about the basic transport properties.

1.2.1 Thermoelectric Effects

There are four experimentally accessible constants that describe the relations between the electric and thermal fields and the electric and thermal currents given in Eq. 13. It is convenient to invert these equations, which directly yields the four conventional parameters (electrical resistivity $\rho = 1/\sigma$, Seebeck coefficient S, Peltier coefficient Π, and thermal conductivity κ):

$$\begin{aligned} \mathbf{F}^* &= \rho\,\mathbf{j} + S\frac{\partial T}{\partial \mathbf{r}}, \\ \mathbf{w}^* &= \Pi\,\mathbf{j} - \kappa\frac{\partial T}{\partial \mathbf{r}}, \end{aligned} \tag{16}$$

with $\mathbf{F}^* = \mathbf{F} - \partial\varphi/\partial\mathbf{r}$ and $\mathbf{w}^* = \mathbf{w} - \mathbf{j}\ \varphi/e$. The coefficients can be obtained by solving the Boltzmann equation for a small perturbation (Conwell 1982). The results are listed in Table 1.

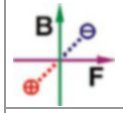
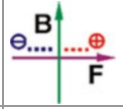
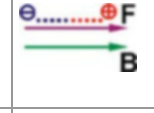
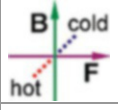
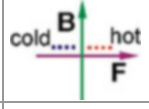
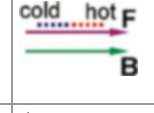
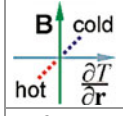
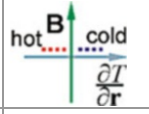
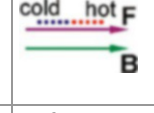
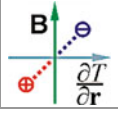
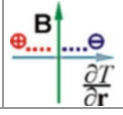
One relation between the Peltier coefficient and the Seebeck coefficient, called the *Kelvin relation*, is often useful:

[3] Galvanomagnetic effects signify electrical and thermal phenomena occurring when a current passes through a solid placed in a magnetic field, see Table 2. It should be noted that often different sign conventions in defining the tensor coefficients are used in literature.

Table 1 Electric and thermoelectric effects. *emf* denotes electromotive force, Δp in Dember's effect is the carrier excess over equilibrium values at illuminated surface (0) or backside of sample (d)

		Applied quantity	Result	Coefficient	Name of effect
Homogeneous effects	Electric	Electric field $\mathbf{F}$	Current density $\mathbf{j}$	$\mathbf{j} = \sigma\mathbf{F}$ Conductivity σ	Conductivity
		Gradient of carrier generation $\Delta p(\mathbf{r})$	Photo emf V_{emf}	$V_{\text{emf}} = (e/\sigma)\,(D_n - D_p)\times\{\Delta p(0) - \Delta p(d)\}$ Dember voltage V_{emf}	Dember effect
	Thermoelectric	Gradient of temperature $\partial T/\partial\mathbf{r}$	Heat flow $\mathbf{w}$ Thermo emf V_{emf}	$\mathbf{w} = -\kappa\partial T/\partial\mathbf{r}$ Heat conductivity κ $-\partial V_{\text{emf}}/\partial\mathbf{r} = S\partial T/\partial\mathbf{r}$ Thermopower or Seebeck coefficient S	Seebeck effect
		Gradient of temperature parallel to $\mathbf{j}$	Heat production $\partial q/\partial t$	$\partial q/\partial t = K\,\mathbf{j}\,\partial T/\partial\mathbf{r}$ Thomson coefficient K $K \cong T\,\partial S/\partial T$ (S = Seebeck coefficient)	Thomson effect
		Changing gradient of temperature $\partial T(t)/\partial\mathbf{r}$	Thermo emf		Inverse Thomson effect
Inhomogeneous effects	Electric	Optical carrier generation between two materials	Photo emf		Photovoltaic effect
	Thermoelectric	Difference of temperature between two materials	Thermo emf V_{emf}	$V_{\text{emf}} = (S_2 - S_1)\,\Delta T$ Seebeck coefficient S	Seebeck effect
		Current I through sequence of two materials	Heat production $\partial q/\partial t$, temperature difference between contact points	$\mathbf{w} = \Pi\mathbf{j}$, or $\partial q/\partial t = (\Pi_2 - \Pi_1)\,I$ Peltier coefficient Π	Peltier effect

Table 2 Galvanomagnetic and thermomagnetic effects. Orthogonal axes refer to magnetic induction **B**, electric field **F**, and temperature field given by the gradient $\partial T/\partial \mathbf{r}$. The electric field **F*** relates to the field **F** by $\mathbf{F}^* = \mathbf{F} - (1/e)\,\partial E_F/\partial \mathbf{r}$ and includes spatial changes of the Fermi energy; **w*** is the heat current related to **F***

	Applied quantity	Transversal effects				Longitudinal effects	
		B⊥ (**j** or $\partial T/\partial \mathbf{r}$)				**B** ∥ (**j** or $\partial T/\partial \mathbf{r}$)	
		Result	Name	Result	Name	Result	Name
Galvanomagnetic effects	Electric field **F**	emf 	Hall effect $R_H = \frac{F_y}{j_x B_z} = \frac{1}{ne}$	$\Delta\sigma$ 	Magneto ρ_{trans} $\rho_{xx} = \frac{F_x^*}{j_x}$	$\Delta\sigma$ 	Magneto ρ_{long} $\rho_{xx} = \frac{F_z^*}{j_z}$
		ΔT 	Ettingshausen effect $\frac{\partial T/\partial y}{j_x B_z} = \frac{\Pi_{yx}}{\kappa_{xx} B_z}$	ΔT 	Nernst effect $\frac{\partial T}{\partial x} = \frac{j_x \Pi_{xx}}{\kappa_{xx}}$	ΔT 	Longitudinal Nernst effect $\frac{\partial T}{\partial z} = \frac{j_x \Pi_{zz}}{\kappa_{zz}}$
Thermomagnetic effects	Temperature field $\partial T/\partial \mathbf{r}$	ΔT 	Righi-Leduc effect $S = \frac{\partial T/\partial y}{\partial T/\partial z \cdot B_z}$	$\Delta\kappa$ 	Maggi-Righi-Leduc effect $\kappa_{xx} = \frac{w_x^*}{\partial T/\partial x}$	$\Delta\kappa$ 	Longitudinal Maggi-Righi-Leduc effect $\kappa_{zz} = \frac{F_x^*}{\partial T/\partial z}$
		emf 	1.Ettingshausen-Nernst effect $Q = \frac{F_y^*}{\partial T/\partial z \cdot B_z}$	emf 	2.Ettingshausen-Nernst effect $S_{xx} = \frac{w_x^*}{\partial T/\partial x}$	emf	Longitudinal 2. Ettingshausen-Nernst effect $S_{zz} = \frac{F_z^*}{\partial T/\partial z}$

$$S = \Pi/T. \tag{17}$$

Another relation is the *Wiedemann-Franz law* Eq. 40 of ▶ Chap. 5, "Phonon-Induced Thermal Properties", which holds for metals, i.e., as long as the thermal conductivity is determined by the electron gas alone and the lattice conductivity is negligible:

$$\kappa = L\sigma T, \quad \text{with } L = \frac{1}{3}\left(\frac{\pi\kappa}{e}\right)^2 = 2.45 \times 10^{-8} \left(\frac{\mathrm{V}^2}{\mathrm{K}^2}\right), \tag{18}$$

where L is the Lorentz number for a strongly degenerate electron gas, see Sect. 3.3.2 of ▶ Chap. 5, "Phonon-Induced Thermal Properties". For a comprehensive review, see Beer (1963).

When exposed to a temperature gradient, the electron gas at the hotter end obtains a higher kinetic energy. Therefore, some of these electrons in a "simple metal", i.e., an alkali metal, move preferentially to the cooler end, charging it negatively. The Seebeck coefficient, also referred to as *thermoelectric power*, can be obtained classically by equating the currents caused by an electric field and by a thermal gradient, yielding with Drude's model $S = c_V^{(e)}/(3ne)$. When replacing the specific heat of the electron gas with $c_V^{(e)} = (\pi^2/2)/(kT/E_\mathrm{F})nk$, we obtain

$$S = -\frac{\pi^2}{3}\frac{k}{e}\frac{kT}{E_\mathrm{F}}, \tag{19}$$

except for a factor of 2 due to insufficient consideration of scattering (see below). With $k/e \cong 86\ \mu\mathrm{eV/K}$, we expect S to be typically on the order of 1 μeV per degree. The Seebeck coefficient S observed for metals at $T = 300$ K is on the same order of magnitude[4] and can be positive or negative.

For semiconductors, the thermoelectric power is usually much larger and is approximated by

$$S_n = -(k/e)\ [r - \ln(N_\mathrm{c}/n)] \ \text{ or } \ S_p = (k/e)\ [r - \ln(N_\mathrm{v}/p)] \tag{20}$$

for n-type or p-type semiconductors, respectively. Here r is a parameter depending on the scattering mechanism:

$r = 1$ for amorphous semiconductors (Friedman 1971)
$r = 2$ for acoustic phonon scattering
$r = 3$ for (polar) optical phonon scattering
$r = 4$ for ionized impurity scattering
$r = 2.5$ for neutral impurity scattering.

For a cubic ambipolar semiconductor, we have

[4]Experimental Seebeck coefficients S (in μeV/K) are −8.3 (Na), −15.6 (K), −4.4 (Pt), +1.7 (Au), +11.5 (Li), and +0.2 (W).

$$S = \frac{S_n \sigma_n + S_p \sigma_p}{\sigma_n + \sigma_p}. \tag{21}$$

See, for example, Smith (1952) and Tauc (1954). Under some experimental conditions, the thermoelectric effect can be observed as a change of the refractive index, as described by Xu and Shen (1994).

Thermoelectricity, the direct conversion of thermal to electrical energy, is an active field of research to harvest unused waste heat. The suitability of a solid can be assessed by its figure of merit ZT, defined as $ZT = (S^2\sigma\, T)/(\kappa_L + \kappa_{el})$, with κ_L and κ_{el} the lattice and electronic contributions to the thermal conductivity. Z (in units K^{-1}) determines the maximum conversion efficiency; a practical material should meet $Z \times T \gtrsim 3$. Highest values are found for doped semiconductors, since insulators have poor electrical conductivity σ, and metals have low Seebeck coefficients S and significant electronic thermal conductivity. The search for optimum solids focuses on low thermal conductivity (to avoid thermal bypass, see Sect. 4 of ▶ Chap. 5, "Phonon-Induced Thermal Properties") and a suitable band structure to obtain a high thermoelectric power factor $S^2\sigma$ (Park et al. 2021). Studies indicate a benefit of a narrow energy distribution in the band structure and a low effective mass (Bilc et al. 2015); a high power factor may be obtained for low-dimensional semiconductors (Hung et al. 2016) or with half-Heusler compounds ABX with X = Sn, Sb, or Bi, (Zeier et al. 2016). Strategies and achievements are comprehensively reviewed by Mao et al. (2018); for the measurement of the thermoelectric power factor, see Haupt et al. (2020).

The Inhomogeneous Thermoelectric Effect The thermo emf is usually measured between two endpoints of a metal wire that is connected to a second metal wire. The two connecting points are kept at different temperatures. The resulting *thermo emf* is then given by the difference of the thermoelectric power of the two metals times the temperature difference between the two connecting points

$$\Delta\varphi = (S_2 - S_1)\ \ (T_2 - T_1), \tag{22}$$

where S_i are the two respective Seebeck coefficients. When measured against a metal with exceptionally small S, e.g., lead, one obtains the tabulated value for an *absolute thermoelectric power*. For a review, see Pollock (1985).

1.2.2 Magneto-Electric Effects

With a magnetic induction, the Lorentz force results in a curving of the electron path. When the magnetic induction is small enough, so that between scattering events only a small deviation from the straight path occurs, the superposition of electric field **F** and magnetic induction **B** results in a bending of the electron path independently. This means that in the relaxation-time approach of the Boltzmann equation, two components must be distinguished. From

$$e\,(\mathbf{F}+\mathbf{v}\times\mathbf{B})\frac{\partial f}{\partial \mathbf{p}}=-\frac{\delta f}{\tau_{\mathrm{m}}}\quad \text{with}\,\delta f=\delta f_1(\mathbf{F})+\delta f_2(\mathbf{F},\mathbf{B}) \tag{23}$$

we obtain

$$\delta f_1(\mathbf{F})\cong -e\,\tau_{\mathrm{m}}\mathbf{F}\;\;\partial f/\partial\mathbf{p}=\frac{e}{kT}\tau_{\mathrm{m}}\mathbf{F}\cdot\mathbf{v}f_0, \tag{24}$$

as discussed in Sect. 4.3 (Eq. 97) of ▶ Chap. 22, "Carrier-Transport Equations", and

$$\delta f_2(\mathbf{F},\mathbf{B})\cong -e\,\tau_{\mathrm{m}}(\mathbf{v}\times\mathbf{B})\;\;\partial f/\partial\mathbf{p}=-\frac{e^2\tau_{\mathrm{m}}^2}{m_n kT}(\mathbf{v}\times\mathbf{B})\cdot\mathbf{F}f_0. \tag{25}$$

From Eq. 25 it follows that a magnetic induction parallel to the electric field has no effect ($\delta f_2=0$), whereas with a magnetic induction component perpendicular to **F** the contribution of δf_2 becomes finite. This contribution determines the Hall effect. Effects occurring in solids with a simultaneously applied magnetic field **B** and an electric field **F** or a temperature gradient $\partial/\partial\mathbf{r}$ are listed in Table 2.

The Hall Effect

For definiteness, we assume $\mathbf{F}=(F_x,0,0)$ and $\mathbf{B}=(0,0,B_z)$ for the relative orientations of the electric field **F** and the magnetic induction **B**. The current density is given by

$$\mathbf{j}=\frac{e^2}{kT}\left(\sum_{\mathbf{v}}\tau f_0\mathbf{F}\cdot\mathbf{v}\;\;\mathbf{v}-\frac{e}{m_n}\sum_{\mathbf{v}}\tau^2 f_0(\mathbf{v}\times\mathbf{B}\cdot\mathbf{F})\;\;\mathbf{v}\right), \tag{26}$$

which for the chosen fields has the components

$$j_x=\frac{e^2}{kT}\left(\sum_{\mathbf{v}}\tau f_0\langle v_x^2\rangle F_x+\frac{e}{m_n}\sum_{\mathbf{v}}\tau^2 f_0\langle v_x^2\rangle B_z F_y\right)=\sigma_{xx}F_x+\sigma_{xy}F_y, \tag{27}$$

$$j_y=\frac{e^2}{kT}\left(-\frac{e}{m_n}\sum_{\mathbf{v}}\tau^2 f_0\left\langle v_y^2\right\rangle B_z F_x+\sum_{\mathbf{v}}\tau f_0\left\langle v_y^2\right\rangle F_y\right)=\sigma_{yx}F_x+\sigma_{yy}F_y. \tag{28}$$

The components of the *magneto-conductivity tensor* are

$$\sigma_{xx}=\sigma_{yy}=\frac{ne^2}{m_n}\frac{\langle E\tau\rangle}{\langle E\rangle}\quad and\quad -\sigma_{yx}=\sigma_{xy}=\frac{ne^2}{m_n}\frac{eB_z}{m_n}\frac{\langle E\tau^2\rangle}{\langle E\rangle}. \tag{29}$$

A more general expression of the average is used here in terms of the energy-distribution function, which is equivalent to the relation $\langle v^2\tau\rangle/\langle v^2\rangle$ for quasi-free

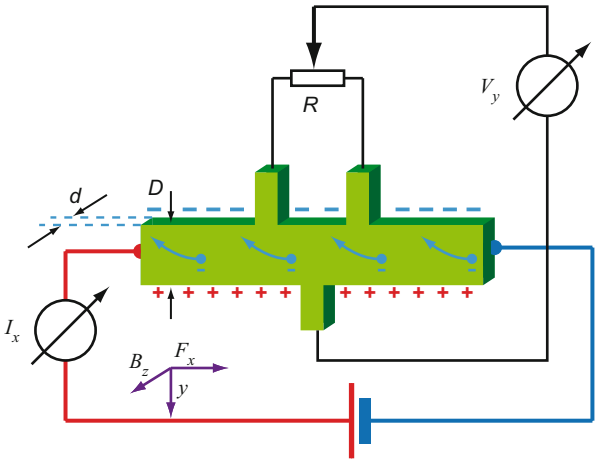


Fig. 1 Experimental setup for Hall-effect measurements in a long two-dimensional sample of thickness d and width D. The Hall angle is determined by a setting of the rheostat R, which renders the current component $j_y = 0$

electrons with $\langle E \rangle = m_n \langle v^2 \rangle / 2$. When more complex equi-energy surfaces are involved, the anisotropy of the effective carrier mass must be considered; for more detail see Conwell (1982).

For a two-dimensional semiconductor in the shape of the platelet shown in Fig. 1, the initial **B**-induced current in y direction causes a charging of the corresponding surfaces until the polarization field forces j_y to vanish. From Eqs. 28 and 29 we obtain

$$\frac{F_y}{F_x} = -\frac{\sigma_{yx}}{\sigma_{xx}} = \frac{e}{m_n} B_z \frac{\langle E\tau^2 \rangle}{\langle E\tau \rangle}. \tag{30}$$

The ratio of the resulting fields determines the *Hall angle*

$$\theta_{\mathrm{H}} = \tan^{-1}\left(\frac{F_y}{F_x}\right) \cong \frac{e}{m_n} B_z \langle \tau_{\mathrm{m}} \rangle = B_z \mu_{\mathrm{H}}. \tag{31}$$

This permits a direct measurement of the *Hall mobility* μ_{H}. The subscript H is used to distinguish the Hall mobility from the carrier mobility μ; the latter is usually slightly smaller than μ_{H}, namely $\mu_{\mathrm{H}}/\mu = \langle E\tau^2 \rangle \langle E \rangle / \langle E\tau \rangle^2 \cong 3\pi/8$ for acoustic mode scattering, $\mu_{\mathrm{H}}/\mu \cong 1.7$ for ionized-impurity scattering, and $\mu_{\mathrm{H}}/\mu \cong 1$ for higher defect densities and temperatures.

The electric field F_y corresponds to the Hall voltage $V_y = F_y D = \langle v_x \rangle B_z D$, where D is the sample width (see Fig. 1). Expressing this voltage by the current $I_x = j_x D \cdot d$ with $j_x = e n \langle v_x \rangle$ and d the sample thickness, we obtain

$$V_y = R_{\mathrm{H}} \frac{I_x B_z}{d}. \tag{32}$$

R_{H} is the *Hall coefficient* given by

$$R_{\mathrm{H}} = \frac{F_y}{j_x B_z} = -\frac{1}{B_z}\frac{\sigma_{yx}}{\sigma_{xx}\sigma_{yy}} = \frac{1}{e\,n}\frac{\langle E\tau^2\rangle\langle E\rangle}{\langle E\tau\rangle^2}. \tag{33}$$

The Hall constant is proportional to $1/(en)$, except for a numerical factor that depends on the scattering mechanism and is on the order of 1; this factor applies also for the ratio μ_{H}/μ as noted above.

For ellipsoidal equi-energy surfaces, the Hall constant is given by (Herring 1955)

$$R_{\mathrm{H}} = \frac{1}{e\,n}\frac{\langle E\tau^2\rangle\langle E\rangle}{\langle E\tau\rangle^2}\frac{3\left\{\left(m_x m_y\right)^{-1} + \left(m_y m_z\right)^{-1} + \left(m_z m_x\right)^{-1}\right\}}{\left(m_x^{-1} + m_y^{-1} + m_z^{-1}\right)^2}. \tag{34}$$

When electrons and holes are present in comparable densities (compensated semiconductors) or two types of carriers (electrons or holes in different bands, or polarons) are present, both types contribute to the Hall constant:

$$R_{\mathrm{H}} = \frac{n_1\, e_1 \mu_1 \mu_{\mathrm{H}_1} + n_2\, e_2 \mu_2 \mu_{\mathrm{H}_2}}{\left(n_1\, e_1 \mu_1 + n_2\, e_2 \mu_2\right)^2}. \tag{35}$$

With $(e_1,\ e_2) = (-e,\ +e)$ for electrons and holes, respectively, the sign of the Hall constant indicates the *type* of majority carrier: it is negative for *n*-type and positive for *p*-type conduction. Here, the signs of e and μ are carried in accordance with the sign convention, see Sect. 2 of ▶ Chap. 22, "Carrier-Transport Equations".

Transverse Magnetoresistance

For higher magnetic induction, we can no longer ignore second-order terms ($\propto B^2$). These terms cause a reduction in the conductivity with increased magnetic induction. This results from the fact that the Hall field compensates only for the deflection of electrons with average velocity, while slower or faster electrons of the distribution are more or less deflected, resulting in a less favorable path average for the carrier conductivity. Scattering itself, however, is not influenced by magnetic induction. This *magnetoresistance* effect, discovered by Thomson (1857), yields information about the anisotropy of the effective mass (Glicksman 1958).

For a quantitative treatment, an alternative method to the evaluation of the Boltzmann equation will be used (Seeger 2004). It is based on the equation of motion for quasi-free electrons (Brooks 1955):

$$m_n\ d\mathbf{v}/dt = e\,(\mathbf{F} + \mathbf{v} \times \mathbf{B}). \tag{36}$$

With $\mathbf{B} = (0,0,B_z)$ and $\mathbf{F}$ the electric field in Eq. 36, we obtain two components

$$\left.\begin{aligned} dv_x/dt &= eF_x/m_n + \omega_{\mathrm{c}}\ v_y \\ dv_y/dt &= eF_y/m_n - \omega_{\mathrm{c}}\ v_x \end{aligned}\right\} \quad \text{with}\ \ \omega_{\mathrm{c}} = \frac{eB_z}{m_n}, \tag{37}$$

where ω_{c} is the *cyclotron frequency*. These two components can be discussed in a complex plane:

$$v = v_x + i\,v_y \quad \text{and} \quad F = F_x + i F_y, \tag{38}$$

yielding from Eq. 37

$$dv/dt = e\,F/m_n - i\,\omega_c\ \ v. \tag{39}$$

This equation can be integrated after both sides are multiplied by $\exp(i\omega_c t)$, yielding for the drift velocity (McKelvey 1966)

$$v_d = v_0 \exp(-i\omega_c t) + e\,F/(i\omega_c m_n)\ \ \{1 - \exp(-i\omega_c t)\}, \tag{40}$$

which shows oscillatory behavior. Due to interfering scattering, however, only a fraction of a cycle about **B** in the carrier path is completed for $\omega_c\ \tau_m < 1$. Considering a *distribution* of relaxation times τ_m, we obtain for the average drift velocity

$$\langle v_d \rangle = \frac{\int_0^\infty v_d(t)\ \exp(-t/\tau_m)\ dt}{\int_0^\infty \exp(-t/\tau_m)\ dt} = \frac{1}{1 + i\omega_c\,\tau_m}\left(v_0 + \frac{e\,F\,\tau_m}{m_n}\right). \tag{41}$$

The first term (v_0) of Eq. 40 drops out when averaging over all angles. Separating the real and imaginary parts of v and F, we obtain

$$\begin{aligned} v_x &= \frac{e}{m_n}\left\{\left(\tau_m - \omega_c^2 \frac{\tau_m^3}{1+\omega_c^2\,\tau_m^2}\right)\ F_x + \omega_c \frac{\tau_m^2}{1+\omega_c^2\,\tau_m^2} F_y\right\} \\ v_y &= \frac{e}{m_n}\left\{-\omega_c \frac{\tau_m^2}{1+\omega_c^2\,\tau_m^2} F_x + \frac{\tau_m}{1+\omega_c^2\,\tau_m^2} F_y\right\}. \end{aligned} \tag{42}$$

For the current densities $j = e\,n\overline{v}_d$ in the x and y directions, we must average these velocities, yielding

$$\begin{aligned} j_x &= e\,n\frac{e}{m_n}\left\{\left(\langle\tau_m\rangle - \omega_c^2\left\langle\frac{\tau_m^3}{1+\omega_c^2\,\tau_m^2}\right\rangle\right)\ F_x + \omega_c\left\langle\frac{\tau_m^2}{1+\omega_c^2\,\tau_m^2}\right\rangle F_y\right\} \\ j_y &= e\,n\frac{e}{m_n}\left\{-\omega_c\left\langle\frac{\tau_m^2}{1+\omega_c^2\,\tau_m^2}\right\rangle F_x + \left\langle\frac{\tau_m}{1+\omega_c^2\,\tau_m^2}\right\rangle F_y\right\}. \end{aligned} \tag{43}$$

For $\omega_c\ \tau_m \ll 1$, which is generally fulfilled, we can neglect the frequency dependence in the denominators. With $j_y = 0$, we then obtain from Eq. 43 by eliminating F_y

$$j_x = e\,n\frac{e}{m_n}\langle\tau_m\rangle F_x\left\{1 - \frac{e^2 B_z^2}{m_n^2}\,\frac{\langle\tau_m^3\rangle\langle\tau_m\rangle - \langle\tau_m^2\rangle^2}{\langle\tau_m\rangle^2}\right\}. \tag{44}$$

This expression contains a second-order term that causes a decrease of the current j_x with increasing magnetic induction. With $\rho = 1/\sigma$ and $e\,\langle\tau_m\rangle/m_n = \mu_n$, and thus $j_x = \sigma F_x\{1 - f(B_z^2)\}$, we obtain for the *magnetoresistance coefficient*

$$\frac{\Delta\rho}{\rho B_z^2} = \mu_n^2 \frac{\langle\tau_m^3\rangle\langle\tau_m\rangle - \langle\tau_m^2\rangle^2}{\langle\tau_m\rangle^4}; \tag{45}$$

that is, the coefficient is essentially equal to μ_n^2 except for the term containing the relaxation-time averages. This term represents a numerical factor that depends on the scattering mechanism and lies between 0.38 and 2.15 (Seeger 2004).

The case of magnetoresistance with *two* carriers is straightforward (McKelvey 1966) and is additive for both carriers, even though they may be of opposite sign. The case of nonspherical equi-energy surfaces is rather involved and is summarized by Conwell (1982), see also Beer (1963).

As an example, the magnetoresistance of a two-carrier semiconductor, *p*-type Ge, is given in Fig. 2. The two carrier types are light hole (p_{lh}, μ_{lh}) and heavy hole (p_{hh}, μ_{hh}). Predominant carrier scattering is assumed to be due to acoustic phonons. The magnetoresistance coefficient is then given by (Seeger 2004)

$$\frac{\Delta\rho}{\rho B_z^2} = \mu_{hh}^2 \frac{9\pi}{16}\left\{\frac{1+\eta\beta^3}{1+\eta\beta} - \frac{\pi}{4}\left(\frac{1+\eta\beta^2}{1+\eta\beta}\right)^2\right\}, \tag{46}$$

with the carrier-density ratio $\eta = p_{lh}/p_{hh}$ and their corresponding mobility ratio $\beta = \mu_{lh}/\mu_{hh}$. Although in *p*-type Ge at 205 K only 4% of the holes are in the light-hole band ($\eta = 0.04$), the large ratio of hole mobilities ($\beta = 8$) renders the numerical factor in Eq. 46 greater by a factor of 24 than for a single-carrier model. As a result,

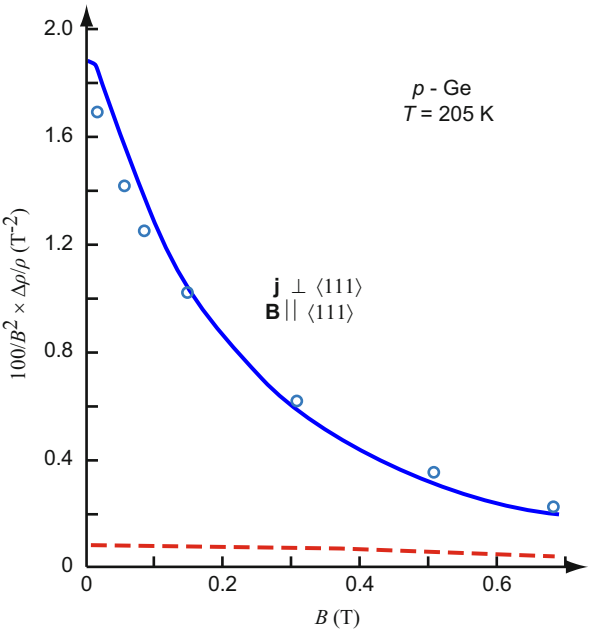


Fig. 2 Transverse magnetoresistance of *p*-type Ge at 205 K as observed (*blue solid curve*) and calculated for a single-carrier heavy-hole model (*red dashed curve*) (After Willardson et al. 1954)

the magnetoresistance is substantially enhanced by the light carriers; this is illustrated in Fig. 2, where the dashed curve is a calculation neglecting the high mobility of light holes.

1.3 Cyclotron Resonance

When the magnetic field is strong enough, and the mean free path is long enough for carriers to complete cyclic paths about the applied magnetic induction, strong resonances in an oscillating electromagnetic probing field are observed at the cyclotron frequency (Eq. 24 of ▶ Chap. 8, "Bands and Bandgaps in Solids")

$$\omega_c = \frac{eB}{m^*} = 178.4 \frac{m_0}{m^*} \; B \;\; (\mathrm{GHz/T}). \tag{47}$$

For the derivation of resonance conditions, see McKelvey (1966). The more cycles are completed before scattering occurs the sharper is the cyclotron-resonance line width. Scattering, with its limiting relaxation time, acts as the damping parameter in the resonance equation (Eq. 41), and with $\tau_m = 1/\gamma$ in Eq. 34 of ▶ Chap. 17, "Optical Properties of Defects", determines the resulting line shape.

Cyclotron-resonance measurements are well suited for determining the effective mass in different crystallographic directions as pointed out in Sect. 1.2.5 of ▶ Chap. 8, "Bands and Bandgaps in Solids".

2 Quantum Effects in a Strong Magnetic Field

When the magnetic induction becomes large enough so that $\hbar\omega_c$ is no longer much smaller than kT, quantum-mechanical effects must be considered: the splitting of electron energies into Landau levels. The influence of a strong magnetic induction will be discussed in two steps: first including scattering to obtain information on the density of states, and, as in Sect. 3, excluding scattering.

2.1 Quasi-Free Carriers in a Strong Magnetic Field

Assuming that the magnetic induction is given by $\mathbf{B} = (0,0,B_z)$ acting in the z direction, the electron motion is described by the Schrödinger equation (Landau 1933)

$$\left\{ -\frac{\hbar^2}{2m_n}\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + \frac{m_n \omega_c^2}{8}\left(x^2 + y^2\right) - i\hbar\omega_c\left(x\frac{\partial}{\partial y} + y\frac{\partial}{\partial x}\right)\right\} \psi = E \; \psi, \tag{48}$$

including two additional terms caused by the magnetic induction and depending on the cyclotron frequency ω_c. These terms impose a constraint on the electron motion in the xy plane due to the Lorentz force. The electron motion is given semiclassically by

$$\begin{aligned} x &= x_0 + \sqrt{\frac{2\hbar}{m_n \omega_c}\left(\nu + \frac{1}{2}\right)} \cos(\omega_c t) = x_0 + r_\nu \cos(\omega_c t) \\ y &= y_0 + \sqrt{\frac{2\hbar}{m_n \omega_c}\left(\nu + \frac{1}{2}\right)} \sin(\omega_c t) = y_0 + r_\nu \sin(\omega_c t), \end{aligned} \tag{49}$$

which are circles with a radius r_ν (Eq. 62), determined by the magnetic induction and the quantum number ν. In k space, we consequently obtain, using $k_x = (m_n/\hbar) \times (dx/dt)$ and $k_y = (m_n/\hbar)(dy/dt)$,

$$k_x^2 + k_y^2 = k_\nu^2 = (2 m_n \omega_c/\hbar)\left(\nu + \frac{1}{2}\right). \tag{50}$$

Equation 50 describes in **k** space a set of cylinder surfaces determined by the quantum number ν with a radius r_ν inverse to B_z, shown in Fig. 3.

The energy of the electrons on these surfaces, obtained as eigenvalues of the Schrödinger equation (48), is given by

$$E = \hbar^2 k_z^2/(2 m_n) + \left(\nu + \frac{1}{2}\right)\hbar\omega_c. \tag{51}$$

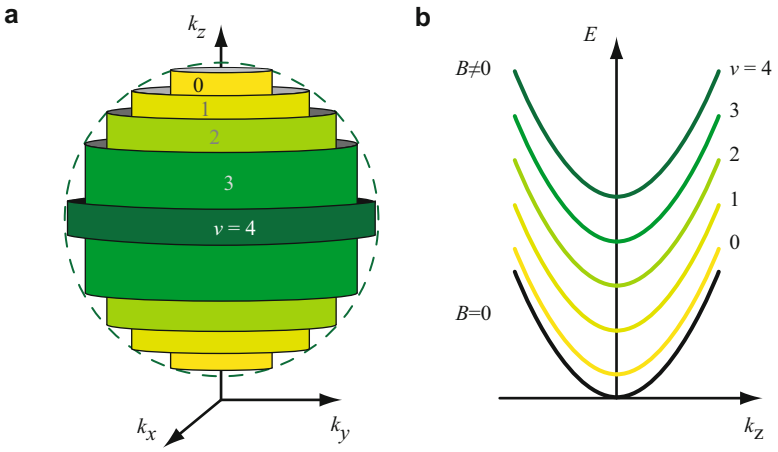


Fig. 3 (**a**) Constant-energy surfaces in **k** space for a given magnetic induction $B \neq 0$, resulting in concentric cylinders for each of the Landau levels, which are labeled by their quantum number ν. (**b**) $E(k_z)$ dispersion relation for zero-magnetic induction (*lowest curve*) and for a constant magnetic induction $B \neq 0$ showing the split into a sequence of Landau bands

In anisotropic semiconductors with anisotropic effective mass, the relative direction of the magnetic field and the crystal orientation must be considered and are included in the cyclotron frequency, as given in Eq. 25 of ▶ Chap. 8, "Bands and Bandgaps in Solids".

The effect described above shows that the application of a strong magnetic induction substantially changes the behavior of Bloch electrons from being quasi-free to being confined in the *xy* plane. This results in a splitting into magnetic subbands[5] referred to as *Landau levels* at a given magnetic field according to the quantum number ν; whereas in the k_z direction, although the $E \propto k_z^2$ relation known for free electrons holds, it is offset by steps of the height of the cyclotron energy.

2.2 Diamagnetic and Paramagnetic Electron Resonance

The interactions between free electrons and a magnetic field due to the Lorentz force, leading to cyclotron resonances, are *diamagnetic interactions.* With a sufficient density of *free electrons*, the semiconductor becomes diamagnetic, i.e., its magnetic moment becomes negative: an oblong probe of the semiconductor turns perpendicular to the magnetic flux when suspended from a filament to permit free rotation. The induced magnetic momentum opposes its inducing force.

When including the *electron spin* in this discussion, we must consider an additional *paramagnetic interaction.* This interaction produces a positive contribution to the magnetic moment. The eigenvalues of the Schrödinger equation including spin interaction (last term) are

$$E = \hbar^2 k_z^2/(2m_n) + \left(\nu + \frac{1}{2}\right)\hbar\omega_c \pm \frac{1}{2} g\mu_B B, \tag{52}$$

with – or + dependent on parallel or antiparallel spin, respectively. $\mu_B = e\hbar/(2m^*)$ is the *Bohr magneton*, and g is the *Landé g factor*. Figure 4 illustrates the additional splitting of the Landau levels due to the spin. For free electrons in vacuum, the cyclotron frequency is $\omega_c = eB/m_0$, which can also be expressed with the Bohr magneton $\mu_B = e\hbar^2/(2m_0)$ as $\omega_c = 2\mu_B B/\hbar$. The corresponding frequency in an

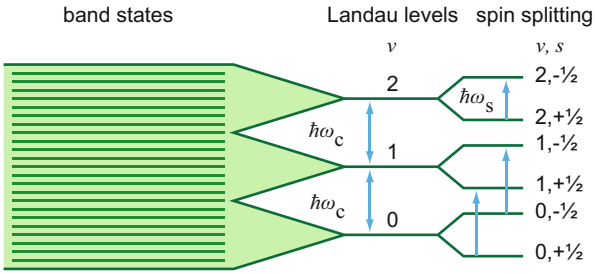


Fig. 4 Splitting of band states into Landau levels, considering diamagnetic interaction by orbiting electrons and additional paramagnetic interaction with the electron spin resulting in further splitting

[5] In the k_z direction there are subbands; in the k_x and k_y directions there are discrete levels in $E(k)$.

atom is the *Larmor frequency* $\omega_L = g\mu_B B/\hbar$, which is equal to ω_c for $g = 2$. For electrons orbiting within a semiconductor, g can deviate substantially from 2, depending on the effective mass and the spin-orbit splitting energy Δ_0 (Lax et al. 1959):

$$g \cong 2\left\{1 + \frac{m_n - m_0}{m_n}\,\frac{\Delta_0}{E_g + 2\Delta_0}\right\}, \tag{53}$$

and may even become negative ($g \cong -50$ for InSb). Since the effective mass has tensor properties, g is also a tensor. Some values of g are listed in Roth and Lax (1959). Care must be taken not to measure the superimposed response of defects in the semiconductor. Furthermore, Landé factors depend on temperature, see Eq. 53 containing E_g and m_n; e.g., in GaAs the g factor of free conduction-band electrons varies from -0.48 at $T \to 0$ to -0.33 at 300 K (Hübner et al. 2009).

With a sufficient density of *impurities* with uncompensated spins, we can observe resonant absorption when flipping the spin by electromagnetic radiation of a frequency corresponding to the energy difference between states with parallel and antiparallel spin. This is an *electron paramagnetic resonance* (EPR), also called *electron spin resonance* (ESR), and occurs at

$$\hbar\omega_S = g\mu_B B, \tag{54}$$

see also Sect. 3.4.2 in ▶ Chap. 18, "Shallow-Level Centers". These resonances provide information about the symmetry of the impurity (by the angle dependence of **B** with respect to the crystallographic axes), the density of uncompensated spins (by the strength of the resonance), and about the g factor and hence the chemical identity (by the resonance frequency).

The resonance absorption can be measured directly by interacting with an electromagnetic field of appropriate frequency (ω_S, typically in the GHz range) or optically by observing changes in the intensity or polarization of laser-excited luminescence – see Baranov et al. (2017) and a review by Cavenett (1981), see also optically induced electron spin resonance of A centers reported by Allen (1995).

Further information about defect centers can be obtained from the paramagnetic interaction with nuclear spins, which can be measured by inducing spin-flipping by absorption of electromagnetic radiation (*nuclear spin resonance*), see Sect. 3.4.2 of ▶ Chap. 18, "Shallow-Level Centers". There is a wide variety of interactions involving the nuclear spin of defects that can be used for analyzing certain defect properties – see Bagraev and Mashkov (1986).

In addition to the paramagnetic interaction of electron spins, the *Pauli spin paramagnetism* at an impurity center, there is the diamagnetic part due to the orbital quantization: for bound electrons, this is the *Landau diamagnetism* that is 1/3 the magnitude of the Pauli contribution. For more detail, see Wilson (1954) and Raoux et al. (2015); the large diamagnetism of bismuth and the role of Landau contributions are reviewed in Fuseya et al. (2015). Landau diamagnetism plays also an important role in two- and one-dimensional quantum structures, see Koyano and Kurita (1998) and Tamura and Ueda (1996).

2.3 Density of States and DeHaas-Type Effects

Density of States in Magnetic Fields The modified $E(k)$ relation described in the previous section is shown in Fig. 3a, b with an applied magnetic field. The density of states depends on the Landau quantum number ν, and is given for each subband by

$$g(k_z)\,dk_z = \frac{2}{(2\pi)^2}\,\frac{m_n\,\omega_c}{\hbar}\,dk_z. \tag{55}$$

Using Eq. 51, we obtain the density of states as a function of the energy

$$g_\nu(E)\,dE = \frac{1}{(2\pi)^2}\left(\frac{2m_n}{\hbar^2}\right)^{3/2}\frac{\hbar\,\omega_c}{\sqrt{E-(\nu+\frac{1}{2})\,\hbar\,\omega_c}}\,dE, \tag{56}$$

which shows an inverse square-root dependence. The *total* density of states is obtained by summation over all possible quantum numbers ν within the band: near the bottom of it, lifted to $E_c + \hbar\,\omega_c/2$, for $E_c + \hbar\,\omega_c/2 < E < E_c + 3\hbar\,\omega_c/2$ with a summation over only one subband, for $E_c + 3\hbar\,\omega_c/2 < E < E_c + 5\hbar\,\omega_c/2$ with summation over two subbands, and so on for the conduction band; this yields the total density of states

$$g(E) = \frac{1}{(2\pi)^2}\left(\frac{2m_n}{\hbar^2}\right)^{3/2}\sum_{\nu=0}^{\infty}\frac{\hbar\,\omega_c}{\sqrt{E-(\nu+\frac{1}{2})\,\hbar\,\omega_c}}. \tag{57}$$

This density-of-state distribution is compared in Fig. 5 with the undisturbed distribution for vanishing magnetic induction $g_0(E)$ showing a square-root dependence for parabolic bands.

DeHaas-Type Effects In a metal or a degenerate semiconductor, the Fermi level lies within the conduction band. With increasing magnetic field, the spacing of the Landau levels $\Delta E = \hbar\,\omega_c = \hbar\,eB/m^*$ increases and causes one after the other of these levels to cross E_F, see Fig. 5. This leads to a periodically changing the density of states $g(E,B)$ at E_F, and in turn causes the amplitude of certain properties which are determined near the Fermi surface to change periodically.

A prominent effect is the oscillatory dependence of the magnetic susceptibility on the inverse magnetic field, referred to as the *DeHaas-van Alphen effect* shown in Fig. 6a (DeHaas and van Alphen 1930). The magnetization M is given by

$$M = -\frac{\partial U}{\partial B}, \tag{58}$$

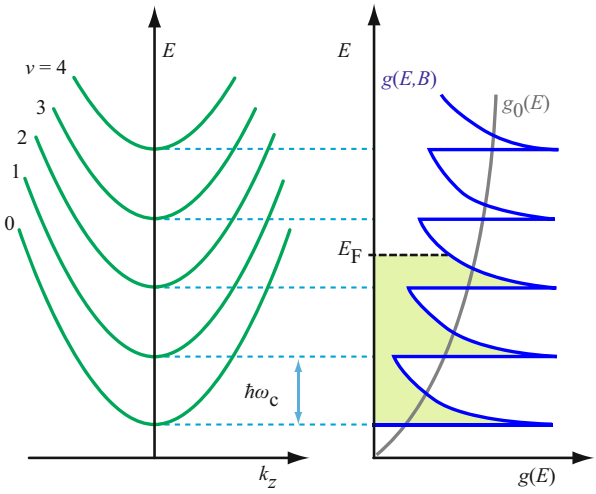


Fig. 5 Landau levels (*left*) and corresponding density of state (*right*) with magnetic induction g (E,B), neglecting spin and lifetime broadening. The *gray curve* is the density of state $g_0(E)$ without magnetic induction; the total areas under $g(E,B)$ and $g_0(E)$ up to any given energy are the same. The shaded area represents the filling of levels up to the Fermi energy E_F

where U is the internal energy.[6] Since the energy of the electron system changes periodically with B^{-1} due to the changes of $g(E,B)$ by level crossing at E_F, so does the magnetization. The period is given by

$$\Delta\left(B^{-1}\right) = \frac{2\pi e}{\hbar A_{\text{orbit}}}, \tag{59}$$

where A_{orbit} is the extremal cross-section area[7] included by the orbit on the Fermi surface perpendicular to **B**. Using this relation the shape of the Fermi surface can be measured.

Another effect related to the changing the density of states $g(E,B)$ at E_F is the oscillatory dependence of the electrical conductivity on the inverse magnetic field, referred to as the *Shubnikov-DeHaas effect* (Fig. 6b).

In actual materials, the orbits occur only rarely at a *spherical* Fermi surface. In semiconductors, the orbits may be elliptical (in sidevalleys); the orbits can also have quite complicated shapes (see Sects. 1.1.4 and 1.2.5 of ▶ Chap. 8, "Bands and Bandgaps in Solids") that make the oscillations more complex (see Fig. 6a for a metal, Fig. 18 of ▶ Chap. 8, "Bands and Bandgaps in Solids", for Ge) and in turn provide information about the shapes of the Fermi surfaces (first suggested by Onsager 1952, see also Shoenberg 1969; Ziman 1972). As an example, the ratio of the wavelength of oscillation shown in Fig. 6a gives the ratio of the area of belly and neck (here 9); see also Fig. 9 of ▶ Chap. 8, "Bands and Bandgaps in Solids".

[6] At finite temperature U must be replaced by $U - TS$, where S is the entropy of the system.

[7] The area in **k** space has a unit of *length*$^{-2}$.

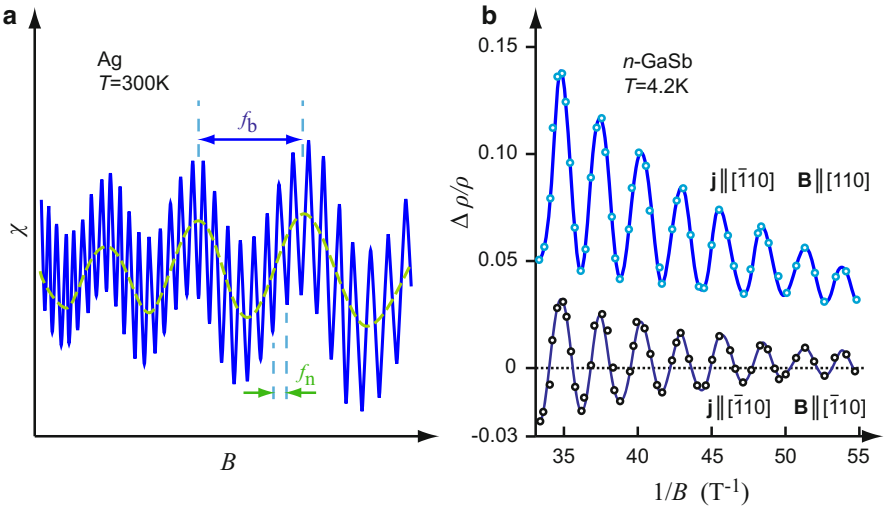


Fig. 6 (**a**) DeHaas-van Alphen oscillations of the magnetic susceptibility in silver with neck (high-frequency f_n) and belly (low-frequency f_b) oscillations when a Landau level passes through the Fermi surface (After Joseph and Thorsen 1965). (**b**) Shubnikov and DeHaas (1930) oscillations of the relative resistivity as a function of the magnetic field in GaSb at 4.2 K (After Becker and Fan 1964)

Magneto-Phonon Effects When the spacing between the Landau levels coincides with the energy of longitudinal optical phonons,

$$\hbar\omega_{LO} = i\ \hbar\omega_c, \quad i = 1,\ 2,\ 3,\ \ldots, \tag{60}$$

electrons can be transferred more easily by scattering with these phonons between different Landau levels. This causes a more pronounced change in the magnetoresistance, with a period length

$$\Delta\left(B^{-1}\right) = \frac{e}{m_n\,\omega_{LO}}, \tag{61}$$

which was first observed by Firsov et al. (1964) in InSb. The effect is small but observable at an intermediate range of the magnetic induction (Eq. 60), temperature (to have sufficient optical phonons), and doping (the effect is sensitive to changes in scattering, see Gurevich and Firsov 1964). It can be used to obtain information about the effective carrier mass.

When higher electric fields are applied in addition to the magnetic field, carrier heating takes place, and distinct multiphonon transitions can be observed – see the review by Stradling (1984).

3 Ballistic Transport in Strong Magnetic Fields

New transport phenomena occur in a strong magnetic field (above ~10 T) if the motion of carriers is not disturbed by scattering. Such a situation can be achieved experimentally by confining electrons in a two-dimensional (2D) electron gas (2DEG, see Sect. 4 of ▶ Chap. 12, "Photon–Free-Electron Interaction"), using samples of extraordinary high purity and applying very low temperatures (below ~1 K). In a 2DEG scattering by impurities is suppressed by *modulation doping*, which provides a spatial separation of donors and the free electrons[8]; these can then move free in the 2D quantum well or in the inversion layer of the 2D channel in a field-effect transistor (Fig. 11 of ▶ Chap. 12, "Photon–Free-Electron Interaction"). If also the scattering at phonons is suppressed by applying low temperatures, electron mobilities exceeding 10^7 $cm^2/(Vs)$ have been demonstrated in a modulation-doped $Al_xGa_{1-x}As/GaAs$ structure (Pfeiffer et al. 1989).

3.1 The Integer Quantum-Hall Effect

If a magnetic field B_z is applied and no scattering occurs, electrons move in circles when exposed to Lorentz forces ($e\,\mathbf{v} \times \mathbf{B}$), with the radius, frequency, and energy given by

$$r_c = \frac{m_n v}{e B_z}, \quad \omega_c = \frac{e B_z}{m_n}, \quad E = \frac{m_n}{2}\,\omega_c^2 r_c^2. \tag{62}$$

With the addition of an electric field F_x in the x direction, the electrons move perpendicularly to B_z and F_x *in the y direction* with constant velocity[9] of the center of each circle, forming *trochoids* (flat spirals) as shown in Fig. 7:

$$v_y = \frac{F_x}{B_z} \quad \text{or} \quad \sigma_{xy} = \frac{e\,n}{B_z}, \tag{63}$$

while

$$v_x = 0 \quad \text{or} \quad \sigma_{xx} = 0. \tag{64}$$

With substantial scattering, σ_{xy} decreases and σ_{xx} increases; σ_{xy} is responsible for the Hall voltage, while σ_{xx} is related to the longitudinal magnetoresistance.

[8]This concept applies also for a two-dimensional hole gas (2DHG). However, the mobility of holes is usually much lower due to a larger effective mass.

[9]In contrast to the case of vanishing magnetic induction where the motion proceeds in the x direction and, without scattering, is accelerated (ballistic transport, see Sect. 3.1 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors").

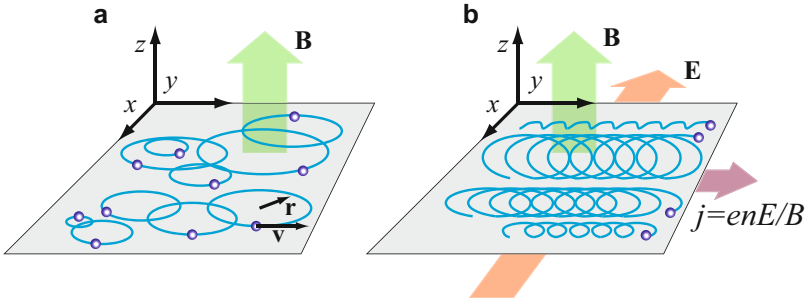


Fig. 7 (**a**) Two-dimensional electron gas with magnetic field **B** only. (**b**) Movement of these electrons in the *y* direction when an additional electric field **F** acts in the *x* direction

When the magnetic field is large enough to cause significant Landau-level splitting ($\hbar\omega_c \geq kT$), instead of a continuum of states in the band we obtain a set of discrete energy levels at

$$E_\nu = E_c + \left(\nu + \frac{1}{2}\right)\hbar\omega_c, \quad \nu = 0,\ 1,\ 2,\ \ldots, \tag{65}$$

with radii

$$r_\nu = \sqrt{\frac{2\hbar}{eB_z}\left(\nu + \frac{1}{2}\right)}. \tag{66}$$

Each of these Landau levels is degenerate, permitting occupation by

$$D = \frac{1}{2\pi r_{\nu=0}^2} = \frac{eB_z}{h} \quad \left(\mathrm{cm}^{-2}\right) \tag{67}$$

electrons; e.g., at $B = 10$ T each Landau level can be occupied by $D \cong 2.4 \times 10^{11}$ electrons per cm^2 in the 2D conductive layer. As B is gradually increased, only D out of n electrons in the conduction band can fill the first Landau level; then D will fill the second level and so on, until all electrons are distributed, with the highest level at $T = 0$ partially filled, as long as the electron density n is not accidentally an integer multiple of D. The ratio n/D is referred to as the filling factor ν. The filling of the Landau levels can be done by electron injection[10] with increasing bias (V_{SD} in Fig. 8) at constant magnetic induction, or by constant bias and increasing magnetic induction, which results in fewer Landau levels below E_F. This causes carrier rearrangement

[10]Electron injection relates to electrode properties not discussed in this book. It provides an experimental means of increasing the carrier density by simply increasing the bias, thereby injecting more carriers from an appropriate electrode. For a review, see Rose (1978).

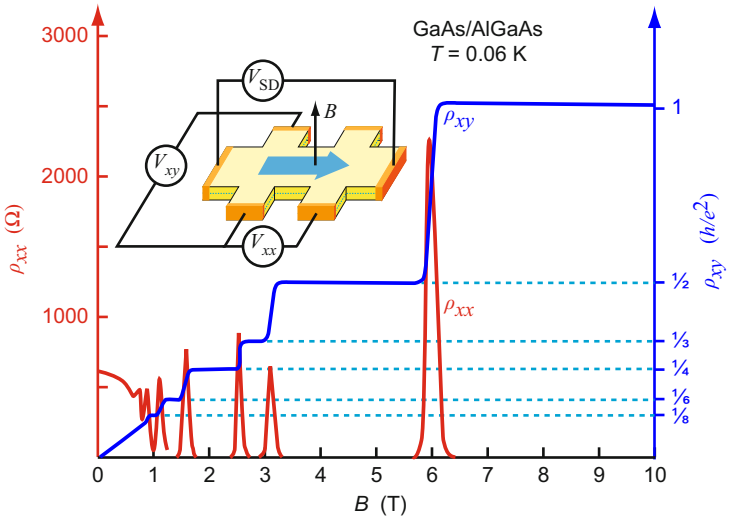


Fig. 8 Measured Hall resistance ρ_{xy} and longitudinal magnetoresistance ρ_{xx} in a GaAs/Al_xGa_{1-x} As heterojunction at $T = 0.06$ K as a function of the magnetic induction, with Landau-level steps $1/\nu$ ($\nu = 1, 2, 3 \ldots$) indicated at the right axis; after Tsui et al. 1994). *Inset*: Experimental setup to measure the quantized Hall effect

whenever a Landau level passes over the Fermi level, which tends to increase E_F and thereby changes the injection.

Scattering can only occur for electrons in the highest Landau level, and only if this level is incompletely filled or $kT > \hbar\omega_c$. Consequently, the electron ensemble will follow unperturbed trochoids as shown in Fig. 7b. When the Fermi level coincides with a Landau level, the magnetoresistivity $\rho_{xx} = 1/\sigma_{xx}$ vanishes, and the Hall resistance $\rho_{xy} = 1/\sigma_{xy}$ shows a pronounced step at $(h/e^2)/\nu$ (von Klitzing et al. 1980), see Fig. 8.

From Eqs. 63 and 67 we can eliminate the incremental electron density per Landau step (where $n = D$) and obtain a step distance, referred to as the *von-Klitzing constant* R_K, of

$$\frac{1}{\Delta\sigma_{xy}} = \frac{B_z}{eD} = \frac{h}{e^2} \equiv R_K = 25,812.807 \ \Omega. \tag{68}$$

It must be noted that the measured Hall-resistance steps in $\rho_{xy} = V_{xy}/I_{SD}$ are not influenced by layer material, layer geometry, defects, or by the carrier effective-mass. They are observed in entirely different structures based on Si, III-V compounds, and even graphene and oxide (Tsukazaki et al. 2010).They are measured precisely to within 1 part in 10^8. Since 1990 the step distance has been defined to $R_K = 25.812807$ kΩ (exact) and has established a new international standard for the electrical resistance, i.e.,

for the unit Ohm. The step distance can also be used to measure the fine-structure constant[11] $\alpha = \mu_0 c e^2/(2h)$ (Tsui et al. 1982; von Klitzing 1981, 1986).

The rounding of the step edges and the step width is due to localized defect states in the gap between the Landau levels, and the broadening of the Landau levels into narrow Landau bands. The persistence of the plateau for a substantial width is more difficult to understand (Stormer and Tsui 1983); carrier transport responsible for the resistance values of ρ_{xy} and ρ_{xx} occurs almost exclusively via small paths at the sample edges illustrated in Fig. 9: since virtually no scattering occurs within the sample, only collisions at the edges remain, yielding carrier motion along segments of circles. Carriers in these spatially separated *edge states* of the two opposite sample sides propagate in opposite directions.

Topological Insulators Structures that are insulating in the bulk but conducting at their boundaries represent a new state of matter referred to as *topological insulator*. The surface states of these topological insulators are symmetry protected by particle-number conservation and time-reversal symmetry (Gu and Wen 2009; Pollmann et al. 2012). The integer and fractional quantum Hall states in a 2D electron gas were the first realizations of topological phases in matter. Later the quantum spin Hall insulator was predicted (Kane and Mele 2005; Bernevig and Zhang 2006) and eventually demonstrated in $Hg_xCd_{1-x}Te$ quantum wells at mK temperatures (König et al. 2007) as a third type of 2D topological insulator. The quantum spin Hall phase requires no external magnetic field; it can be described as a parallel pair of quantum-Hall-like edge states (*helical edge states*) with opposite spins propagating in opposite directions (*spin-momentum locking*). There exists also a *three-dimensional* topological-insulator phase, which can be realized at room temperature without magnetic field (Hasan and Kane 2010), particularly in semiconductors with Heusler structure. For more information, see Sect. 4 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors"

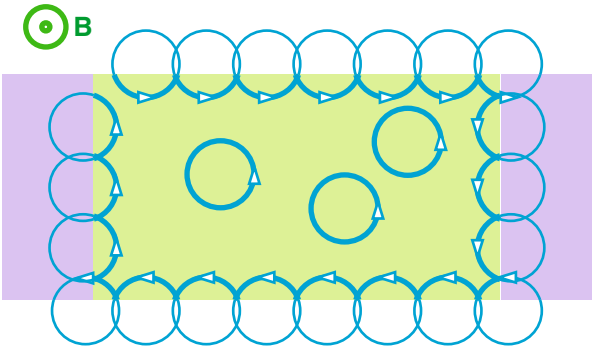


Fig. 9 Motion of electrons in a two-dimensional electron gas with solely scattering at the sample edges, subjected to a strong magnetic field. Conduction between contact pads (*light blue*) occurs via these edge states, where carriers move in opposite directions along the two opposite side edges

[11]The velocity of light contained in α is the best known of the three constants.

3.2 The Fractional Quantum Hall Effect

At magnetic fields above the Landau plateau of the filling factor $\nu = 1$, more such plateaus were observed in high-quality GaAs/$Al_{1-x}Ga_xAs$ heterostructure samples at very low temperatures (Tsui et al. 1982). These steps refer to *fractional* quantum numbers. The first pronounced fractional filling factor appears at $\nu = 1/3$ and is identified at 0.09 K to better than 3 parts in 10^5 (Chang et al. 1984). This step is shown in Fig. 10; note the disappearance of the features as the temperature approaches $T = 4$ K.

Later a rich structure with more and more fractional steps has been identified by plateaus in the Hall resistance ρ_{xy} and peaks in the longitudinal magnetoresistance ρ_{xx}, see Fig. 11. The additional plateaus refer to fractional filling factors for $\nu = p/q$, where q must be an odd integer for the initially discovered states (3, 5, 7, 9),[12] while p is a positive integer (1 . . . 13), see Tsui and Stormer (1986), White et al. (1990).

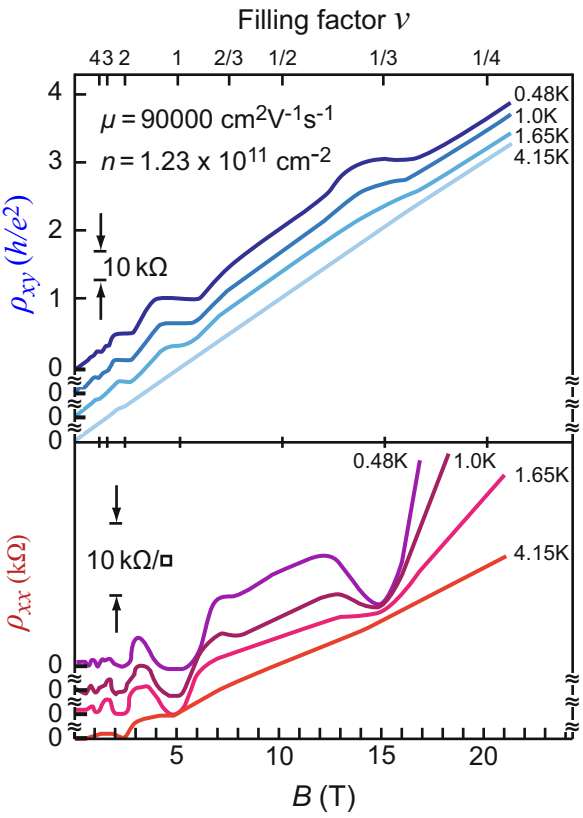


Fig. 10 Appearance of an additional step in ρ_{xy} and a dip in ρ_{xx} at high magnetic field B and temperatures well below 4 K (After Tsui et al. 1982)

[12]Later also even denominators were observed, see Willett et al. (1987).

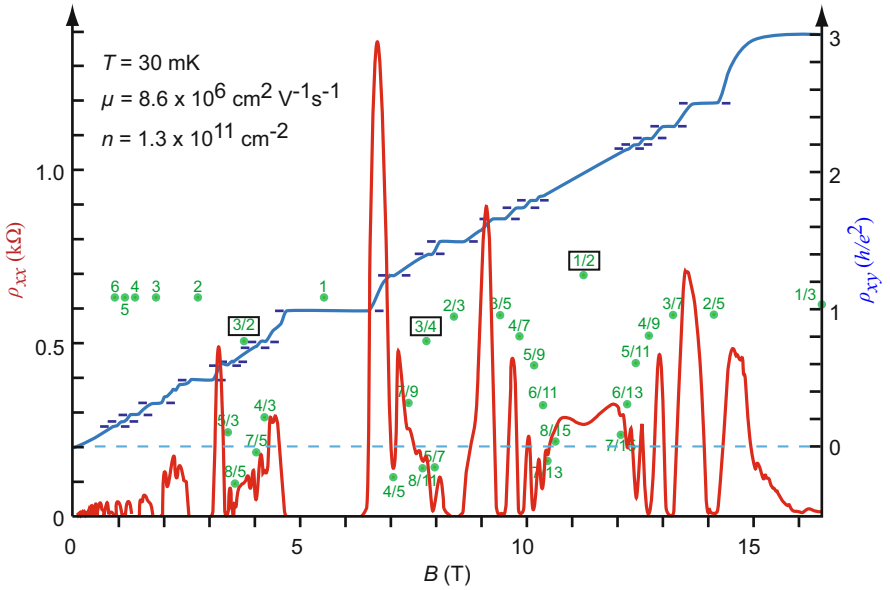


Fig. 11 Appearance of an additional step in ρ_{xy} and a dip in ρ_{xx} at a fractional filling factor $\nu = 1/3$ at high magnetic field B and temperatures well below 4 K (After Abstreiter 1998)

The dominance of the effect at fractions of 1/3 and 2/3 indicates the existence of quasi particles carrying a fractional charge (1/3) e. If we assume an energy gap to appear in the lowest Landau level at $\nu = 1/3$, we can apply the model of the integer quantum-Hall effect to these quasi particles for describing the effects at such fractional filling factors.

The explanation of the fractional quantum Hall effect requires the additional consideration of the electron–electron interaction. If the position of the jth electron at (x_j, y_j) in the two-dimensional electron gas is denoted by a complex number $z_j = x_j - i\,y_j$, the many-electron wavefunction of the $1/\nu$-states can be written as a product over all differences between particle positions $(z_j - z_k)$,

$$\Psi_{1/\nu}(z_1, z_2, \ldots z_N) = (z_1 - z_2)^{1/\nu}(z_1 - z_3)^{1/\nu}(z_2 - z_3)^{1/\nu} \ldots \quad (z_{N-1} - z_N)^{1/\nu} \quad (69)$$

(Laughlin 1981, 1983); this wavefunction obeys Pauli's principle, since it becomes 0 for two electrons at the same position. Furthermore, ν must be odd to ensure that Ψ changes sign when two electrons are exchanged. In the ground state, the electrons have an optimum correlation with minimum Coulomb repulsion. Ψ must have a finite density of zeros equal to the density of magnetic flux quanta $\phi_0 = h/(2e)$ (vortices) penetrating the sample. This means each electron experiences ν zeros of Ψ at the positions of the adjacent electrons, or, ν vortices are *bound* to each electron: electron plus ν vortices represent a new quasi particle. The addition or subtraction of a single electron or flux quantum to the ensemble disturbs the ordering at the expense of a substantial energy cost. The Laughlin state hence represents a *condensed* many-

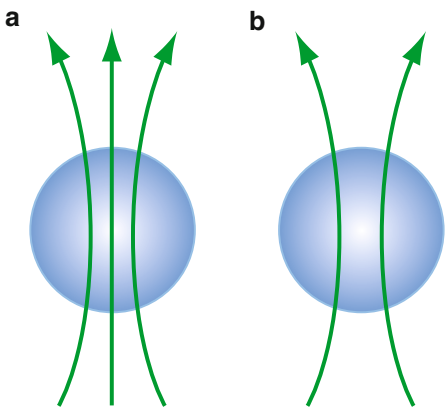


Fig. 12 Schematic of composite particles consisting of an electron (*blue circle*) and (**a**) three or (**b**) two magnetic flux quanta (*green arrows*)

particle ground state, separated by an energy gap to an unbound electron state. Since the position of an electron is not fixed, the ensemble is referred to as a *quantum liquid*.

The new quasi particle is called a *composite Fermion* and is illustrated in Fig. 12. An electron added to the ensemble immediately splits up into a number of the composite particles with fractional charge each, where charges $1/m$ (m integer) occur with m the eigenvalue of the angular momentum of the many-electron wavefunction in the lowest Landau level (Laughlin 1983). Besides states of filling 1/3, 1/5, . . . also states of fractions $1 - 1/\nu$ appear due to the electron–hole symmetry. Plateaus at filling factors 2/5, 3/7, 4/9, etc., may be explained by a hierarchy of condensed states, where p/q-states are assigned to daughter states of $1/q$-states. Near the filling factor $1/q$ then many excited quasi particles correlate to optimize the total energy and condense into a state with filling factor p/q (Jain 1989, 1990).

The striking symmetry of the magnetoresistivity (ρ_{xx}) oscillations around the filling factor ½ (near $B_{1/2} = 11$ T in Fig. 11) is related to composite Fermions composed of one electron and *two* flux quanta. For these quasi particles, the magnetic field $B_{1/2}$ corresponds to an effective field $B^* = 0$; $B_{1/2}$ may be considered a field to provide the two flux quanta per electron for creating this quasi-particle. The filling factor 1/3 for these composite fermions then corresponds to a filling factor of 1 for normal electrons, and the filling factors 2/5, 3/7, 4/9 (at 13–14 T in Fig. 11) correspond to $\nu = 2$, 3, 4 of the integer Hall effect. For more information, see Halperin and Jain (2020).

4 Summary

The application of a magnetic field or of a temperature gradient in addition to an electric field is an important tool to analyze properties of a semiconductor. The most significant information is obtained for the mobility by the Hall-effect, the effective mass by the cyclotron resonance, the origin of levels in a complex spectrum by spin-flip resonance, and the type of conductivity by the Hall effect or the thermo-

electromotive force (thermo-emf). In addition, a number of thermoelectric effects occur with useful applications, such as the Seebeck effect rendering thermoelectricity used in thermocouples or in radioisotope thermal generators (nuclear batteries) for remote power generation, and the Peltier effect applied for cooling; moreover, several galvanomagnetic and thermomagnetic effects are observed.

In strong magnetic fields, the density of states of quasi-free electrons is substantially altered; energy levels condense on quantized Landau levels with cylindrical equi-energy surfaces in **k** space. The shape of the density of states gives rise to an oscillatory behavior of quantities controlled by their vicinity to the Fermi energy when the magnetic field is increased. Prominent examples are DeHaas-van Alphen oscillations of the magnetic susceptibility and Shubnikov-DeHaas oscillations of the resistivity.

Additional transport phenomena are observed if scattering is suppressed in pure samples at very low temperatures. The strong magnetic induction forces carriers to propagate on edge states near the sample surface, creating a topological insulator with no conductance in the bulk. The effect was first observed in the quantum-Hall effect, which provides a means for high-precision measurement of elementary constants (h/e^2), and since 1990 became an international metrological standard for the electrical resistance. New insight into quantum effects of a confined two-dimensional electron gas was provided by the fractional quantum Hall effect, which lead to the discovery of composite fermions composed of an electron and some flux quanta, conjointly carrying a fractional charge $(1/m)e$, with m an integer. The more recently discovered quantum spin Hall phase represents a third type of topological insulators, which require no external magnetic field.

References

Abstreiter G (1998) Die Entdeckung des fraktionalen Quanten-Hall-Effekts. Phys Bl 54:1098 (The discovery of the fractional quantum Hall effect, in German)

Allen JW (1995) Spectroscopy of lattice defects in tetrahedral II-VI compounds. Semicond Sci Technol 10:1049

Bagraev NT, Mashkov VA (1986) Optical nuclear polarization and spin-dependent reactions in semiconductors. In: von Bardeleben HJ (ed) Defects in semiconductors, Mater Sci Forum, vol 10–12, pp 435–443. Trans Tech Publ, Switzerland

Baranov PG, von Bardeleben HJ, Jelezko F, Wrachtrup J (2017) Magnetic resonance of semiconductors and their nanostructures. Springer, Wien

Becker WM, Fan HY (1964) Magnetoresistance oscillations in *n*-GaSb. In: Hulin M (ed) Proc 7th international conference physics of semiconductors, Paris. Dumond, Paris, pp 663–667

Beer AC (1963) Galvanomagnetic effects in semiconductors, vol Supplement 4, Solid state physics. Academic Press, New York

Bernevig BA, Zhang S-C (2006) Quantum spin Hall effect. Phys Rev Lett 96:106802

Bilc DI, Hautier G, Waroquiers D, Rignanese GM, Ghosez P (2015) Low-dimensional transport and large thermoelectric power factors in bulk semiconductors by band engineering of highly directional electronic states. Phys Rev Lett 114:136601

Brooks H (1955) Theory of the electrical properties of germanium and silicon. Adv Electron Electron Phys 7:85

Cavenett BC (1981) Optically detected magnetic resonance (O.D.M.R.) investigations of recombination processes in semiconductors. Adv Phys 30:475

Chang AM, Berglund P, Tsui DC, Stormer HL, Hwang JCM (1984) Higher-order states in the multiple-series, fractional, quantum Hall effect. Phys Rev Lett 53:997
Conwell E (1982) Transport: the Boltzmann equation. In: Paul W, Moss TS (eds) Handbook on semiconductors, vol. 1. Band theory and transport properties. North Holland, Amsterdam, pp 513–561
DeHaas WJ, van Alphen PM (1930) Note on the dependence of the susceptibility of diamagnetic metal on the field. Leiden Commun 208d and Leiden Commun 212a
Firsov YA, Gurevich VL, Parfeniev RV, Shalyt SS (1964) Investigation of a new type of oscillations in the magnetoresistance. Phys Rev Lett 12:660
Friedman L (1971) Hall conductivity of amorphous semiconductors in the random phase model. J Non-Cryst Solids 6:329
Fuseya Y, Ogata M, Fukuyama H (2015) Transport properties and diamagnetism of dirac electrons in bismuth. J Phys Soc Jpn 84:012001
Glicksman M (1958) The magnetoresistivity of germanium and silicon. In: Gibson AF (ed) Progress in semiconductors, vol 3. Wiley, New York, pp 1–26
Gu Z-C, Wen X-G (2009) Tensor-entanglement-filtering renormalization approach and symmetry protected topological order. Phys Rev B 80:155131
Gurevich VL, Firsov YA (1964) A new oscillation mode of the longitudinal magnetoresistance of semiconductors. Sov Phys JETP 20:489
Halperin BI, Jain JK (2020) Fractional quantum Hall effects – new developments. World Scientific, Publ Co, Singapore
Hasan MZ, Kane CL (2010) Topological insulators. Rev Mod Phys 82:3045
Haug A (1972) Theoretical solid state physics. Pergamon Press, Oxford
Haupt S, Edler F, Bartel M, Pernau HF (2020) Van der Pauw device used to investigate the thermoelectric power factor. Rev Sci Instrum 91:115102
Herring C (1955) Transport properties of a many-valley semiconductor. Bell Syst Tech J 34:237
Hübner J, Döhrmann S, Hägele D, Oestreich M (2009) Temperature-dependent electron Landé g factor and the interband matrix element of GaAs. Phys Rev B 79:193307
Hung NT, Hasdeo EH, Nugraha ART, Dresselhaus MS, Saito R (2016) Quantum effects in the thermoelectric power factor of low-dimensional semiconductors. Phys Rev Lett 117:036602
Jain JK (1989) Composite-fermion approach for the fractional quantum Hall effect. Phys Rev Lett 63:199
Jain JK (1990) Theory of the fractional quantum Hall effect. Phys Rev B 41:7653
Joseph AS, Thorsen AC (1965) Low-field de Haas-van Alphen effect in Ag. Phys Rev 138:A1159
Kane CL, Mele EJ (2005) Z_2 topological order and the quantum spin Hall effect. Phys Rev Lett 95:146802
König M, Wiedmann S, Brüne C, Roth A, Buhmann H, Molenkamp LW, Qi X-L, Zhang S-C (2007) Quantum spin Hall insulator state in HgTe quantum wells. Science 318:766
Koyano M, Kurita R (1998) Magnetization of quasi-two-dimensional conductor η-Mo_4O_{11}. Solid State Commun 105:743
Landau LD (1933) Motion of electrons in a crystal lattice. Phys Z Sowjetunion 3:664
Laughlin RB (1981) Quantized Hall conductivity in two dimensions. Phys Rev B 23:5632
Laughlin RB (1983) Anomalous quantum Hall effect: an incompressible quantum fluid with fractionally charged excitations. Phys Rev Lett 50:1395
Lax B, Roth LM, Zwerdling S (1959) Quantum magneto-absorption phenomena in semiconductors. J Phys Chem Solid 8:311
Madelung O (1981) Introduction to solid state theory, vol 2. Springer, Berlin
Mao J, Liu Z, Zhou J, Zhu H, Zhang Q, Chen G, Ren Z (2018) Advances in thermoelectrics. Adv Phys 67(2):69
McKelvey JP (1966) Solid state and semiconductor physics. Harper & Row, New York
Onsager L (1952) Interpretation of the de Haas-van Alphen effect. Philos Mag 43:1006
Park J, Xia Y, Ozoliņš V, Jain A (2021) Optimal band structure for thermoelectrics with realistic scattering and bands. npj Comput Mater 7:43
Pfeiffer LN, West KW, Störmer HL, Baldwin KW (1989) Electron mobilities exceeding 10^7 cm^2/Vs in modulation-doped GaAs. Appl Phys Lett 55:1888

Pollmann F, Berg E, Turner AM, Oshikawa M (2012) Symmetry protection of topological phases in one-dimensional quantum spin systems. Phys Rev B 85:075125
Pollock DD (1985) Thermoelectricity: theory, thermometry, tool. ASTM, Philadelphia
Raoux A, Piéchon F, Fuchs JN, Montambaux G (2015) Orbital magnetism in coupled-bands models. Phys Rev B 91:085120
Rose A (1978) Concepts in photoconductivity and allied problems. RE Krieger Publication Co, New York
Roth LM, Lax B (1959) *g* factor of electrons in germanium. Phys Rev Lett 3:217
Seeger K (2004) Semiconductor physics, 9th edn. Springer, Berlin
Shoenberg D (1969) Electronic structure: the experimental results. In: Ziman JM (ed) Physics of metals, vol 1. Cambridge University Press, Cambridge, UK, pp 62–112
Shubnikov L, DeHaas WJ (1930) Magnetische Widerstandsvergrößerung in Einkristallen von Wismut bei tiefen Temperaturen. Leiden Commun 207a,c,d; and: Leiden Commun 210a (Magnetic increase of the resistance of bismuth at low temperatures, in German)
Smith RA (1952) The physical properties of thermodynamics. Chapman & Hall, London
Stormer HL, Tsui DC (1983) The quantized Hall effect. Science 220:1241
Stradling RA (1984) Studies of the free and bound magneto-polaron and associated transport experiments in *n*-InSb and other semiconductors. In: Devreese JT, Peeters FM (eds) Polarons and excitons in polar semiconductors and ionic crystals. Plenum Press, New York
Tamura H, Ueda M (1996) Effects of disorder and electron–electron interactions on orbital magnetism in quantum dots. Physica B 227:21
Tauc J (1954) Theory of thermoelectric power in semiconductors. Phys Rev 95:1394
Thomson W (1857) On the electro-dynamic qualities of metals: effects of magnetization on the electric conductivity of nickel and of iron. Proc Roy Soc London 8:546
Tsui DC, Stormer HL (1986) The fractional quantum Hall effect. IEEE J Quantum Electron QE 22:1711
Tsui DC, Stormer HL, Gossard AC (1982) Two-dimensional magnetotransport in the extreme quantum limit. Phys Rev Lett 48:1559
Tsui DC, Janssen M, Viehweger O, Fastenrath U, Hajdu J (1994) Introduction to the theory of the integer quantum hall effect. VCH, Weinheim
Tsukazaki A, Akasaka S, Nakahara K, Ohno Y, Ohno H, Maryenko D, Ohtomo A, Kawasaki M (2010) Observation of the fractional quantum Hall effect in an oxide. Nat Mater 9:889
von Klitzing K (1981) Two-dimensional systems: a method for the determination of the fine structure constant. Surf Sci 113:1
von Klitzing K (1986) The quantized Hall effect. Rev Mod Phys 58:519
von Klitzing K, Dorda G, Pepper M (1980) New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. Phys Rev Lett 45:494
White CRH, Davies M, Henini M, Davidson BR, Main PC, Owers-Bradley JR, Eaves L, Hughes OH, Heath M, Skolnick MS (1990) The observation of the fractional quantum Hall effect in a single (AlGa)As/GaAs/(AlGa)As quantum well. Semicond Sci Technol 5:792
Willardson RK, Harman TC, Beer AC (1954) Transverse Hall and magnetoresistance effects in *p*-type germanium. Phys Rev 96:1512
Willett R, Eisenstein JP, Störmer HL, Tsui DC, Gossard AC, English JH (1987) Observation of an even-denominator quantum number in the fractional quantum Hall effect. Phys Rev Lett 59:1776
Wilson AH (1954) The theory of metals. Cambridge University Press, Cambridge, UK
Xu HF, Shen ZZ (1994) Mechanism of refractive-index change in a Ce:SBN single crystal illuminated by nominal homogeneous light. Opt Lett 19:2092
Ziman JM (1972) Principles of the theory of solids. Cambridge University Press, Cambridge, UK
Zeier WG, Schmitt J, Hautier G, Aydemir U, Gibbs ZM, Felser C, Snyder GJ (2016) Engineering half-Heusler thermoelectric materials using Zintl chemistry. Nat Rev Mater 1:16032

Superconductivity

26

Karl W. Böer and Udo W. Pohl

Contents

1 Low-Temperature Superconductors 1088
1.1 Superconductive Solids 1088
1.2 Cooper Pairs and Condensation 1090
1.3 The Critical Temperature 1092
1.4 Meissner-Ochsenfeld Effect and Type I or Type II Superconductors 1094
1.5 Josephson Tunneling and SQUID 1097
2 High-Temperature Superconductors 1099
2.1 Ceramic High-T_c Superconductors 1099
2.2 Normal-State Properties of High-T_c Superconductors 1102
2.3 The Superconductive State of High-T_c Superconductors 1104
2.4 Mediating-Partner Models for High-T_c Superconductors 1106
3 Summary 1109
References 1109

Abstract

There exists a large diversity of superconductors following different mechanisms to achieve the superconducting phase. Low-temperature superconductivity appears in metals and degenerate semiconductors; it is induced by the formation of electron pairs in a bipolaron state referred to as Cooper pair. The superconductive state is separated from the normal-conductivity state by an energy gap below the Fermi energy. This gap appears at the critical temperature and widens as the temperature decreases. In type I low-temperature superconductors an external magnetic field is expelled from the bulk up to an upper value, which

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_26

eliminates superconductivity. In type II superconductors an array of flux lines penetrates into the bulk above a lower critical field, creating a mixed normal and superconductive phase up to the upper critical field.

High-temperature superconductivity of type II is observed mostly in layered compounds such as cuprates, iron pnictides, and nickelates, with critical temperatures exceeding 100 K. Superconductivity in these materials is usually carried by hole pairs and requires sufficient doping. The mechanism of pair formation differs from that in metals and involves an interaction with spin fluctuations. The symmetry of the layered superconductive system and of the superconductive gap is lower than in the basically isotropic metals; in cuprates pairs with a lateral *d* symmetry are found.

Keywords

Anderson RVB model · BCS theory · Ceramic superconductor · Cooper pair · Critical temperature · Cuprates · Flux-line lattice · High-T_c superconductor · Isotope effect · Jospehson tunneling · London penetration depth · Magnetic ordering · Meissner phase · Meissner-Ochsenfeld effect · Nickelates · Organic superconductor · Pnictides · Superconduction energy-gap · SQUID · Two-fluid model · Type I and II superconductors · Vortices

1 Low-Temperature Superconductors

1.1 Superconductive Solids

Superconductivity denotes the occurrence of zero electrical resistance and the expulsion of an external magnetic field below a characteristic critical temperature. There exist two classes of semiconductors that show superconductivity, low-temperature superconductors and high-temperature superconductors, which follow different paths to become superconductive.

Inorganic Superconductors The *low-temperature superconductors* are metals or highly doped semiconductors with their superconductivity carried by electron-related quasiparticles. These materials become superconductive at very low temperatures – almost exclusively well below 30 K. Most *high-temperature superconductors* have a semiconducting antiferromagnetic ground state and can be doped with sufficient carrier density to become superconductors at higher temperatures (albeit still low, typically 40 … 150 K). They have highly anisotropic lattices, and their conduction mechanism is linear or layer-like with a superconductivity usually carried by hole-related quasiparticles. A list of some superconductive materials is given in Table 1. For a complete list of classic low-temperature superconductors see Harshman and Mills (1992). High-temperature superconductors are discussed in Sect. 2.

Low-temperature superconductivity in *semiconductors* was predicted by Gurevich et al. (1962) and by Cohen (1964), and in 1964 it was observed in $Ge_{1-x}Te$, SnTe, and $SrTiO_3$ (Hein et al. 1964; Schooley et al. 1964). Transition

Table 1 Various superconducting materials and their transition temperature T_c

Type	Material	T_c (K)
High T_c	$HgBa_2Ca_2Cu_3O_{8+\delta}$	135
	$Tl_2Ba_2Ca_2Cu_3O_{10+\delta}$	125
	$YBa_2Cu_3O_7$	92
	$Bi_2Sr_2CaCu_2O_8$	89
	$La_{1.85}Sr_{0.15}CuO_4$	39
	$Nd_{1.85}Ce_{0.15}CuO_4$	24
	$RbCs_2C_{60}$	33
Low T_c	Nb_3Ge	23.2
	Nb	9.25
	Pb	7.20
	Ta	4.39
	Hg	4.15
	Sn	3.72

temperatures are typically below 1 K, and depend on the stoichiometry, which influences the carrier density. The critical temperature increases with increasing carrier density, i.e., decreasing stoichiometry. When the electron density exceeds 10^{20} cm^{-3} in $SrTiO_3$, the transition temperature again decreases (Appel 1966); the superconducting state of this most dilute bulk superconductor is discussed by Gastiasoro et al. (2020). Many-valley semiconductors have a higher critical temperature as they have a larger density of states, since the density-of-state mass is increased. A review of superconductivity in the elements, alloys and simple compounds is given by Webb et al. (2015).

Organic Superconductors A variety of *organic semiconductors* has been observed to become superconductive at low temperatures (typically below 10 K) – for reviews, see Ishiguro et al. (1998), Ardavan et al. (2012). The superconductive phase was first discovered in compounds based on quasi-one-dimensional conducting radical salts, including the $TMTSF_2X$ salts (Jérome 1994), where TMTSF denotes tetramethyl-tetraselenafulvalene (Fig. 1a) and X is an electron acceptor such as, e.g., AsF_6, TaF_6, or ClO_4. The molecules arrange in a stacked order with a relative good, nearly one-dimensional normal conductivity along the stacking direction. Most of these compounds become superconductive only at high hydrostatic pressure, see Table 2; an exception is the similar structure $TMTTF_2ClO_4$, where TMTTF denotes tetramethyl-tetrathiafulvalene with Se of TMTSF being replaced by S. The chemical structure of the molecules in two superconducting compounds is depicted in Fig. 1.

Most attention is devoted to two new classes with $(ET)_2X$ structure, having a critical temperature of 12 K, and of fullerites with T_c up to 33 K. In superconductors based on the BEDT-TTF compound usually the abbreviation ET is used for this molecule. The molecules form planes which are separated by anions like, e.g., I_3 or $Cu[N(CN)_2]Br$. The molecules arrange in various phases with orderings in fishbone

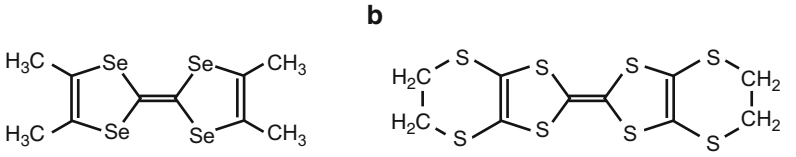


Fig. 1 Chemical structure of molecules in organic superconductors. (**a**) tetramethyl-tetraselenafulvalene (TMTSF), (**b**) bisethylenedithio-tetrathiafulvalene (BEDT-TTF)

Table 2 Organic superconducting materials with their transition temperature T_c and ambient pressure P

Compound	T_c (K)	P (bar)
$(TMTTF)_2ClO_4$	1.4	1
$(TMTSF)_2PF_4$	1.1	6500
$(TMTSF)_2AsF_6$	1.1	9500
$(TMTSF)_2TaF_6$	1.35	11000
β-$(ET)_2I_3$	1.5	1
κ-$(ET)_2I_3$	3.6	1
κ-$(ET)_2Cu(NCS)_2$	10.4	1
κ-$(ET)_2Cu[N(CN)_2]Cl$	12.8	1.3
κ-$(ET)_2Cu[N(CN)_2]Br$	11.6	1
Sr_6C_{60}	6.8	1
K_3C_{60}	19.3	1
K_2RbC_{60}	23	1
Rb_3C_{60}	29.4	1
$RbCs_2C_{60}$	33	1

or checkerboard structures, yielding two-dimensional conductivity. Similar to the $TMTSF_2X$ salts they are type II superconductors (Sect. 1.4).

Superconductors based on the insulator fullerene C_{60} form a three-dimensional isotropic solid crystallizing in an fcc arrangement. Conductivity is achieved by doping with alkali dopants such as Rb to form, e.g., Rb_3C_{60} molecules.

1.2 Cooper Pairs and Condensation

We will first review basic properties of the conventional low-temperature superconductors, although this type of superconductivity is mostly observed in metals and thus beyond the scope of this book; for a general overview, see Buckel and Kleiner (2004); for non-centrosymmetric superconductors see Smidman et al. (2017). According to the Bardeen–Cooper–Shrieffer (*BCS*) theory (Bardeen et al. 1957), superconductivity originates from the formation of electron pairs; for reviews, see Schrieffer (1964) and Allen and Mitrovic (1982). Such Cooper pairs can undergo Bose–Einstein condensation; superconductivity is assigned to this condensation (Blatt 1961). The BCS theory predicts that the lowest energy state of such electron pairs is separated from higher energy states of unpaired electrons by an energy gap,

which prevents any damping of the electron transport at low temperatures. For this damping, scattering events must occur with energy transfer. However, since all states below the bandgap are occupied, no such scattering is possible for the Cooper pairs. They can move through the lattice as condensed bosons without energy loss, similarly to superfluid helium.

Electrons are fermions with energies up to some eV due to the Pauli principle, i.e., with thermal energies corresponding to very high temperatures. Since the transition to superconductivity occurs at quite low temperatures, an interaction operating at such energies is required. The model of superconductivity is based on the idea of forming electron *pairs* which are *bosons*; these quasiparticles follow Bose–Einstein statistics and can condensate into a common ground state. Electrons can form such pairs despite the strong Coulomb repulsion if they have opposite spin; they are assisted in pair formation by lattice polarization as evidenced by the isotope effect (Sect. 1.3). This assistance can be imagined by recognizing the polarization well created by an electron,[1] which at sufficiently low temperatures traps another electron in the same well, forming an electron pair. This pair is sometimes referred to as a *bipolaron*, see Sect. 1.2.2 of ▶ Chap. 22, "Carrier-Transport Equations".

The electron interaction with the lattice, necessary for pair formation, can be modeled by assuming emission and absorption of phonons from paired electrons, which results in eigenstates below those of two independent electrons. Pairs formed between two electrons with opposite spin and wavevector directly at the Fermi surface of a metal or a highly degenerate semiconductor at $T \to 0$ K are called *Cooper pairs*. With pair formation, the electrons can lower their energy to occupy a state at an energy Δ *below* the Fermi energy E_F,

$$E_\mathrm{Cp} = E_\mathrm{F} - \Delta, \text{ with } \Delta \cong \hbar\omega_\mathrm{D} \exp\left[-\frac{1}{g(E_\mathrm{F})V}\right]. \tag{1}$$

Here ω_D is the Debye frequency, $g(E_\mathrm{F})$ is the density of states at E_F, and $-|V|$ is the interaction energy, which is of the form

$$V = \left|M_{kq}\right|^2 \frac{2\hbar\omega_q}{\left[E(k+q) - E(k)\right]^2 - \left(\hbar\omega_q\right)^2}, \tag{2}$$

with M_{kq} as the electron–phonon matrix element; ω_q is an average phonon frequency typical of the lattice, with $\omega_q \cong \omega_\mathrm{D}/2$. A reduction in energy is possible despite full occupation of every single energy state below E_F by electrons, since Cooper pairs are bosons and as such can condense at a lower ground-state energy.

An energy Δ per electron is gained when Cooper pairs are formed. This renders the filled Fermi sphere unstable, and provides the driving force for Bose–Einstein

[1] The polaron state introduced in Sect. 1.2 of ▶ Chap. 22, "Carrier-Transport Equations".

condensation of such pairs. When condensation of Cooper pairs takes place, the excited states for individual electrons are given by

$$E(k) = \sqrt{(E - E_F)^2 + \Delta^2}, \tag{3}$$

which has a minimum at $E = E_F$, and creates a gap of 2Δ between the ground and excited states for two paired electrons. This energy gap makes superconductivity plausible: it does not permit any scattering with an energy exchange of less than 2Δ. An electron can only be scattered when the displacement of the Fermi sphere in an external field exceeds 2Δ:

$$\Delta E = \frac{\hbar^2}{2m}\left[(k_F + \delta k)^2 - (k_F - \delta k)^2\right] = \frac{2\hbar^2 k_F \delta k}{m} \geq 2\Delta. \tag{4}$$

Therefore, the current density $\mathbf{j} = e\, n\mathbf{v} = e\, n\hbar\mathbf{k}/m$ must remain below a *critical current density* j_{crit}, above which superconductivity vanishes:

$$j \leq \frac{2\, e\, n\, \Delta}{\hbar\, k_F} = j_{crit}. \tag{5}$$

With typical values for $\Delta \cong 1$ meV, $n \cong 10^{22}$ cm^{-3}, and $k_F \cong 10^8$ cm^{-1}, we obtain $j_{crit} \cong 10^7$ A/cm^2 in an order-of-magnitude agreement with the experiment. In a classical sense these high current densities indicate a very high drift velocity v_d of superconducting electrons: $j = env_d$ yields v_d on the order of 10^4 cm/s at j_{crit}.

1.3 The Critical Temperature

Cooper-pair formation requires pairs of electrons with $\mathbf{k}\uparrow$ and $-\mathbf{k}\downarrow$ at the Fermi surface. The density of such pairs decreases with increasing temperature, following Fermi-Dirac statistics. Since the gap energy Δ is a function of available electrons,[2] Δ shrinks with increasing temperature and vanishes at a critical temperature T_c, which can be estimated from a numerical integration of Eq. 1, yielding

$$kT_c \cong (2/3.5)\,\Delta_{T=0}. \tag{6}$$

Since $T_c \propto \Delta \propto \omega_D \propto 1/\sqrt{M}$, we expect an increase in the critical temperature with decreasing mass M of the oscillating atoms. With different isotopes of a given element, only the mass changes while all other parameters are left unchanged.

[2] $g(E_F)$ was used in Eq. 1 at $T = 0$ K.

Such an *isotope effect* of T_c is indeed observed[3] and shown for Hg (a solid below −39 °C) in Fig. 2.

The superconductive state can be approximately described as consisting of a superfluid liquid of Cooper pairs, which is mixed with a liquid of normal electrons; the density ratio of both is determined by the temperature: the density of Cooper pairs decreases as the temperature increases. Above T_c no Cooper pair can exist and the gap vanishes. Therefore, all carriers above T_c are normal electrons. Below T_c, more and more electrons condense as Cooper pairs. Evidence for such a *two-fluid model* can be obtained from optical absorption and ultrasound attenuation. Only the Cooper pairs have an energy gap, yielding a far-IR absorption edge. Only the free electrons provide ultrasound attenuation; Cooper pairs cannot be split by the low energy of ultrasonic phonons. This model yields a temperature-dependent density ratio between electrons and Cooper pairs and provides evidence for a temperature-dependent gap $\Delta(T)$, as shown in Fig. 3.

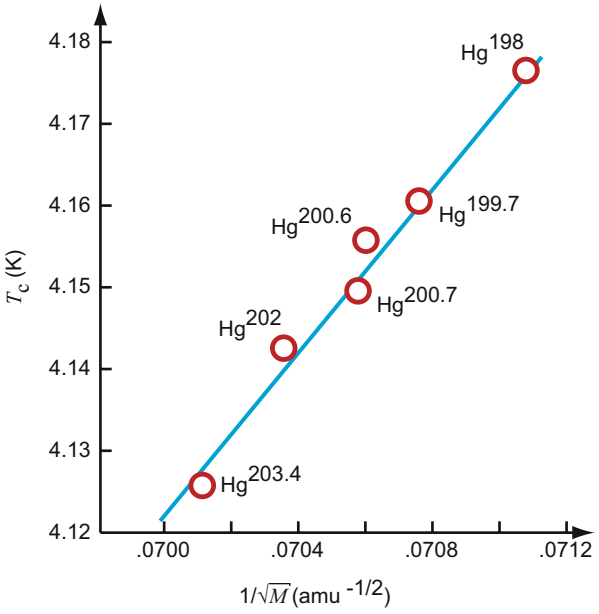


Fig. 2 Isotope effect of the critical temperature T_c for samples of Hg isotopes with different mass M, according to data from Reynolds et al. (1950)

[3]Exponents β in $T_c \propto M^{-\beta}$ are 0.5 for isotopes of Hg, Cd, and Tl, 0.48 for Pb, 0.47 for Sn; also smaller effects are observed, e.g., $\beta = 0.33$ for Mo and 0.2 for Os.

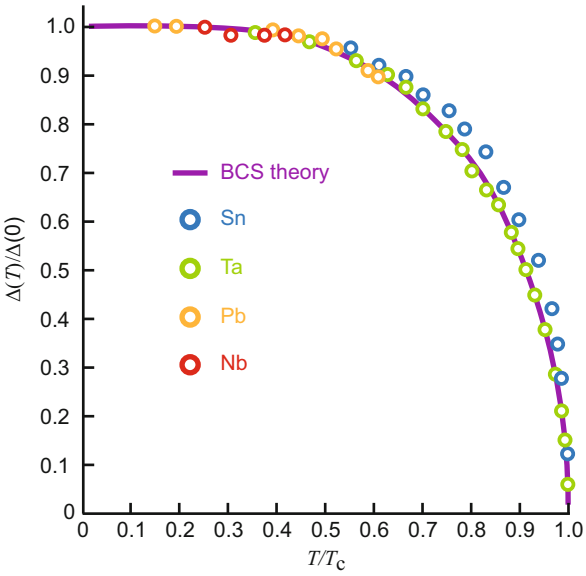


Fig. 3 Temperature dependence of the energy gap Δ, depicted in reduced units; for absolute values of T_c see Table 1. Symbols are experimental results, the solid curve is calculated from the BCS theory (After Townsend and Sutton 1962)

1.4 Meissner-Ochsenfeld Effect and Type I or Type II Superconductors

Meissner-Ochsenfeld Effect A superconductor is characterized by a second effect in addition to the vanishing resistance: an externally applied magnetic field is expelled from a superconductor when cooled below T_c; this phenomenon is referred to as the *Meissner-Ochsenfeld effect* (Meissner and Ochsenfeld 1933). Such a perfect diamagnetic behavior cannot be explained solely by a vanishing resistivity. For this, Maxwell's equations of electrodynamics must be modified, as performed by the phenomenological London theory for superconductors with the London equations

$$\frac{\partial}{\partial \mathbf{r}} \times \mathbf{j}_s = -\frac{\mathbf{B}}{\mu_0 \lambda^2} \quad \text{and} \quad \frac{\partial \mathbf{j}_s}{\partial t} = \frac{\mathbf{E}}{\mu_0 \lambda^2}, \tag{7}$$

where $\mathbf{j}_s$ is the superconducting current density and λ is the *London penetration depth*[4] of the magnetic field (London and London 1935). The penetration depth denotes the exponential decrease of an externally applied magnetic field **B** inside the superconductor according to $B = B_{surface}\exp(-z/\lambda)$. It is on the order of 500 Å in

[4] λ is related to the density of superconducting carriers n_s by $\lambda = \sqrt{m_{Cp}/\left(\mu_0 q_{Cp}^2 n_s\right)}$, where $m_{Cp} = 2m_n$ and $q_{Cp} = -2e$.

typical metals. With increasing temperature the penetration depth increases according to

$$\lambda(T) = \lambda(T=0)\frac{1}{\sqrt{1-(T/T_c)^4}} \tag{8}$$

as shown in Fig. 4. When the penetration depth exceeds the thickness of the wire carrying the current, superconductivity has vanished.

Type I or Type II Superconductors Superconductors that expel the magnetic flux as described in the previous section are called *type I superconductors*. They show diamagnetic behavior with $\chi_{\text{mag}} = -1$, created by surface currents within the penetration depth λ that induce a magnetization $\mathbf{M}$ which perfectly cancels the external field $\mathbf{H}$ inside the superconductor according to $\mathbf{M} = \chi_{\text{mag}}\ \mathbf{H} = -\mathbf{H}$. Above a critical magnetic field H_c, the entire superconductor becomes normal conductive. The phase boundary between the normal and superconductive states can empirically be described by

$$H(T) = H(T=0)\ \left[1-(T/T_c)^2\right], \tag{9}$$

where $B(T=0) = \mu_0\, H(T=0)$ is usually on the order of 0.01 . . . 0.1 T for Al and V, respectively, as examples. All superconductive metals are type I, except for Nb.

Type II superconductors have two critical magnetic fields, H_{c1} and H_{c2}. Below H_{c1}, the superconductor displays type I behavior. Above H_{c1}, the magnetic field penetrates into the superconductor, and microscopic flux lines, also called *vortices*,

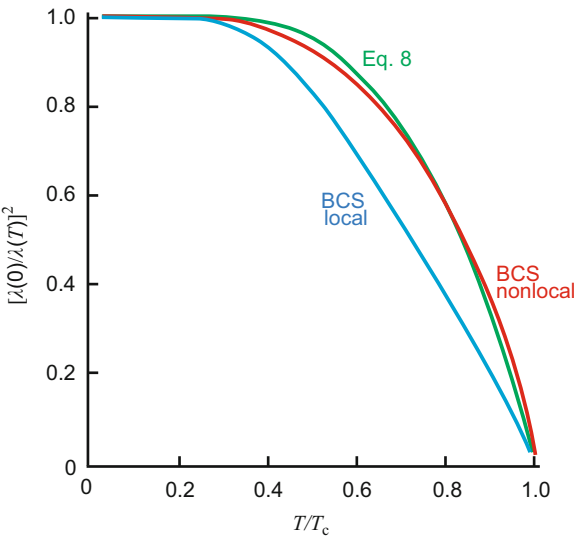


Fig. 4 Temperature-dependent penetration depth of an external magnetic field according to the BCS theory or nonlocal approximations as indicated; the *green curve* shows the dependence according to Eq. 8 (After Fetter and Walecka 1971)

are formed in a regular trigonal close-packed pattern: the flux-line lattice imaged in Fig. 5. The period of the pattern grows with decreasing magnetic field, and the flux of the magnetic field through one elementary vortex cell is a constant, referred to as the magnetic flux quantum $\Phi_0 = h/(2e)$. The area of a vortex is partially normal-conductive and increases at higher temperatures (for constant field) on the expense of the superconductive area outside the vortices (Fig. 5b, c), until above T_c no superconductive area remains.

Above H_{c2}, the type II superconductor becomes normal-conductive. Typical fields H_{c2} are an order of magnitude higher than H_{c1} with the maximum known value at ~60 T for $PbMo_6S_8$, a *Chevrel-phase* superconductor. A typical phase diagram for type I and type II superconductors is shown in Fig. 6.

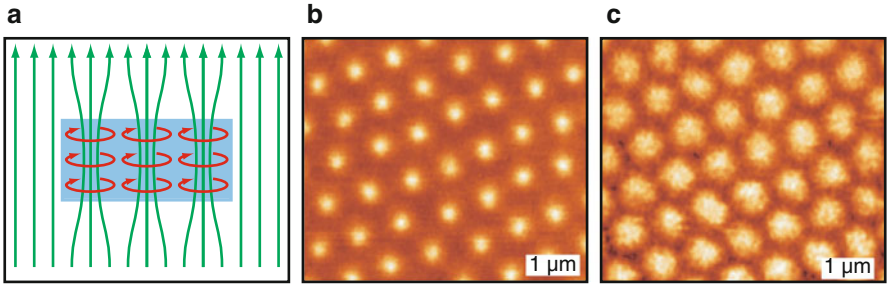


Fig. 5 (**a**) Schematic of magnetic vortices (*flux lines*) in a type II superconductor (*blue rectangle*). *Red circles* indicate currents shielding areas outside the penetrating magnetic field. (**b**) Plane-view imaging of flux lines (*bright spots*) emerging from the surface of the high-temperature superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$ using magnetic force microscopy with a magnetic-tip cantilever. A constant external field $B \cong 2$ mT is applied at $T = 5.1$ K. (**c**) Same as panel (**b**) at $T = 38.1$ K (After Schwarz et al. 2010)

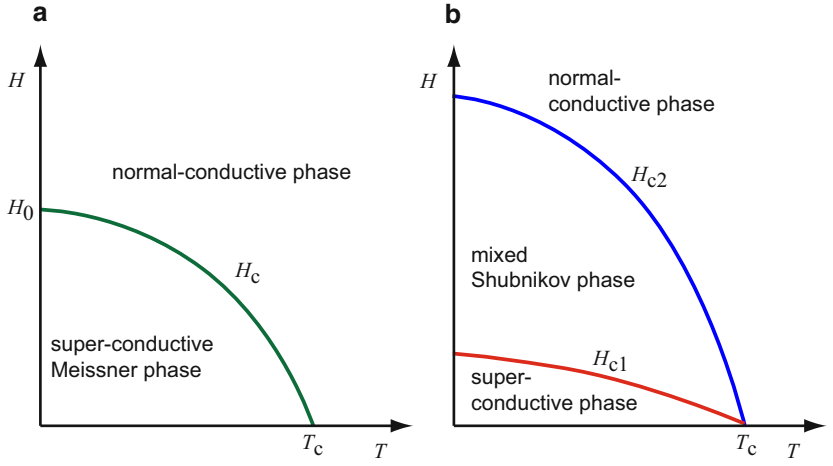


Fig. 6 Phase boundaries between normal and superconductive states (**a**) for type I and (**b**) for type II superconductors, with critical field boundaries indicated

The flux-line pattern is influenced by crystal defects. Flux lines may move in the presence of a current that creates Lorentz forces normal to these lines. Consequently, some minor power dissipation is observed in such type II superconductors above H_{c1}. Some crystal defects are known to pin flux lines and thereby create a more stable superconductor at high fields – see Matsushita (2014), Feighan et al. (2017). This is important for NMR imaging, where high and extremely constant fields over long periods are required.

1.5 Josephson Tunneling and SQUID

Tunneling between two superconductors through a very thin insulating layer provides the most direct evidence for the energy gap. If both superconductors are metals in their normal-conductive phase, tunneling proceeds at $T \geq 0$ K without threshold, see Fig. 7a.

If one of the metals is a superconductor, no current flows near zero bias. Only when electrons are elevated above the gap does a small current flow (*Giaever tunneling* – Fig. 7c light green curve for $0 < T < T_c$). With further increased bias, electrons from below the Fermi surface can be drawn across the barrier (*Josephson tunneling*, Fig. 7c green curve), and the tunneling current increases steeply.

Superconducting Quantum-Interference Device The flux of an externally applied magnetic field enters a type II superconductor in integer multiples of flux

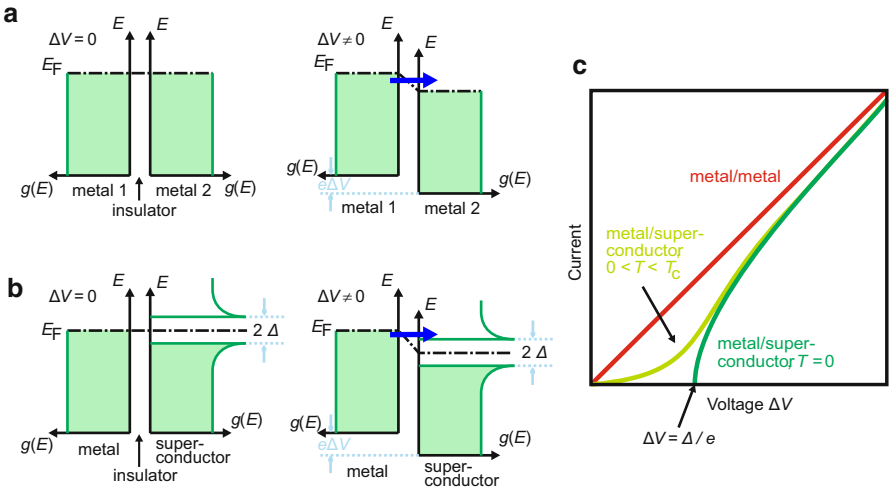


Fig. 7 Tunneling through a thin insulating layer between two metals: (**a**) both metals are normal conductors; (**b**) one metal is a superconductor: current flow requires at $T = 0$ a voltage exceeding Δ/e. (**c**) *I-V* characteristics of the tunnel current for a metal-metal junction (*straight red curve*) and a metal-superconductor junction at $T = 0$ (*green curve*) and at $T > 0$ (*light green*)

quanta $\Phi_0 = h/(2e)$ (Sect. 1.4). This applies also for the magnetic flux through a superconductive ring, which is given by

$$\Phi_{\text{ring}} = \Phi_{\text{ext}} + L j_s = i\,\Phi_0; \tag{10}$$

here L is the inductance of the ring, j_s is the superconducting current induced by the magnetic field, and i is an integer. Thus, although the externally applied flux Φ_{ext} and the current j_s are continuous quantities, the magnetic flux inside the ring is not; this can only be fulfilled if j_s oscillates with an amplitude $|Lj_s| \leq \Phi_0/2$, when an externally applied magnetic field is continuously increased: the current in the ring j_s completes the total flux in the ring to an integer multiple of Φ_0.

The induced superconducting current j_s can be measured using a superconducting quantum-interference device (SQUID); a schematic of the setup is given in Fig. 8. Two very thin insulators are introduced into the superconductive ring considered above. They constitute tunnel junctions (*Josephson junctions*) between two superconductors similar to the junction depicted in Fig. 7b. At a Josephson junction the current j_s tunnels through the barrier, and the two collective wavefunctions of the Cooper pairs on both side of a junction experience a phase difference γ. The phase difference is related to the maximum critical current density j_c through such a junction by $j_s = j_c \sin\gamma$. Using this relation, the current in the ring with two Josephson contacts can be written

$$j_s = j_c/2 \times (\sin\gamma_1 - \sin\gamma_2), \tag{11}$$

with phase differences γ_1 and γ_2 at the two junctions.

If this ring is connected to an external current source, the current entering the ring splits into two parts. If a small external field **B** applied, a screening current j_s begins to circulate in the ring to cancel the applied external flux. In one branch of the ring j_s

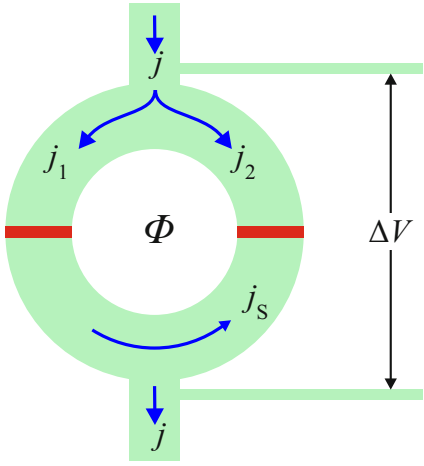


Fig. 8 Superconducting ring with two Josephson junctions (*red* thin insulators), building a superconducting quantum-interference device (SQUID)

is in the same direction as the applied current, $j_1 = j/2 + j_s$, and it is opposite in the other branch, yielding $j_2 = j/2 - j_s$. When the current in either branch exceeds the critical current density j_c of the Josephson junction, a voltage ΔV appears across the junction. If the external flux is further increased and eventually exceeds $\Phi_0/2$, an increase of the flux to $1 \times \Phi_0$ is energetically more favorable than a stronger shielding current to prevent a flux from entering into the ring. The screening current thus flows in the opposite direction. Every time the flux increases by multiples of $\Phi_0/2$, j_s changes direction and hence oscillates as a function of the applied external magnetic field. The period corresponds to the addition of a single flux quantum Φ_0 to the flux already penetrating the ring.

SQUIDs are very sensitive magnetometers and are widely used for measuring magnetic properties of materials, magnetic resonance imaging (MRI), or in biology for studies such as magnetoencephalography (MEG); for reviews see Sarracanie and Salameh (2020), Geethanath and Vaughan (2019).

2 High-Temperature Superconductors

2.1 Ceramic High-T_c Superconductors

Transition temperatures exceeding the highest observed in metals have been reported for the class of high-temperature superconductors (also termed high-T_c superconductors), which are all type II superconductors. These are usually layered compounds that are semiconductors at or above room temperature. In 1986 the perovskite-type ceramic $La_{2-x}Ba_xCuO_4$ was found to become superconductive near 30 K (Bednorz and Müller 1986; T_c was later specified to 35 K), substantially above the highest T_c value of a conventional superconductor (Nb_3Ge, 23.2 K). Shortly after this discovery further ceramic compounds with even much higher T_c values surpassing the boiling point of liquid nitrogen (77 K) were found: Y-Ba-Cu-O, Bi-Sr-Ca-Cu-O, and Hg-Ba-Ca-Cu-O. Later also other structures such as the pnictides were discovered, see Fig. 9. Over 150 high-temperature superconducting compounds with T_c exceeding that of Nb_3Ge were found; they may be roughly divided into cuprates, the related nickelates, bismuthades, and iron pnictides – in addition to the fullerenes mentioned with organic superconductors in Sect. 1.1, carbon nanotubes, and various other inorganic compounds such as hydrogen compounds (e.g., H_2S, LaH_{10}) – the latter at extremely high pressure. Some ceramic superconductors are listed in Table 1.

A common feature of the ceramic superconductors is the layer structure. A prominent large family are the superconductive *cuprates*; they comprise La-*M*-Cu-O-type compounds (M = Ba, Sr, Ca), Y-Ba-Cu-O-type compounds like $RBa_2Cu_3O_{6+x}$ (R = Y, La, Ca, or a rare-earth element), and Bi-, Tl-, Hg-type compounds like, e.g., $Tl_mBa_2Ca_{n-1}Cu_nO_{2n+m+2+\delta}$ (m = 1,2; n = 1−4). They all have CuO_2 layers, which carry the superconductive electron system. The other layers serve as dopants to provide the carriers ("charge reservoir") and couple the stacked layers. At least two of such weakly

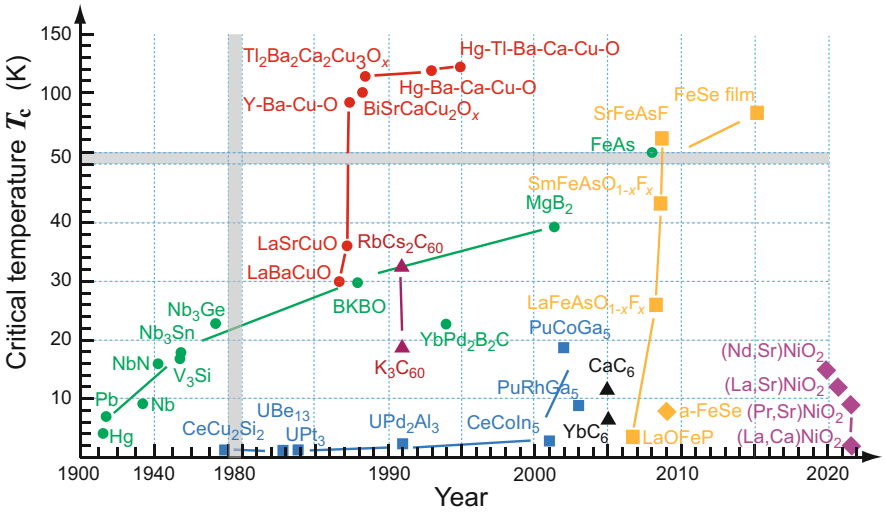


Fig. 9 History of discovered superconductors. High-temperature cuprate superconductors are indicated by *red circles*, iron pnictides by *orange squares*, and nickelates by *purple diamonds*; conventional low-temperature superconductors are marked by *green circles*. Note the change in abscissa scaling at the year 1980 and the change in ordinate scaling at $T_c = 50$ K

coupled layers are needed to create superconductivity, and T_c increases with the number of these layers. The cuprates are reviewed in detail by Plakida (2010).

Another prominent family are the more recently found *iron pnictides* such as, e.g., *M*FFeAs (M = Ca, Sr, Ba, Eu) with doping by La^{3+} ions. The pnictides have conductive FeAs layers playing the role of the CuO_2 layers in the cuprates. The lattice structure of some cuprate superconductors is shown in Fig. 10. Their structure is usually closely related to the perovskite structure, with alternating layers of CuO_2 planes and an oxygen deficiency.

The *nickelates* such as $Nd_{1-x}Sr_xNiO_2$ are another quite recently found family (Li et al. 2019). The (unusual) valence Ni^{1+} is isoelectronic (d^9) with Cu^{2+}; the parent nickelate $RNiO_2$ (R= La, Nd) has the same crystalline structure and a similar half-filling $d_{x^2-y^2}$ orbital configuration as the cuprate parent compound $CaCuO_2$. Calculations confirm in cuprates and nickelates a dominant pairing tendency in the $d_{x^2-y^2}$ band (Wu et al. 2020). In both families, the superconductivity is induced by charge doping in the parent compound,[5] and a "superconducting dome" of comparable width for the composition parameter x is found in the phase diagram (Zeng et al. (2020), cf. also Fig. 11). Despite the similarities in the crystalline and electronic structures, the nickelates feature striking differences. The parent compound $NdNiO_2$ shows no magnetic order and is (weakly) conducting, while $CaCuO_2$ is an insulating antiferromagnet. Moreover, the hybridization between Ni $3d_{x^2-y^2}$ and O $2p$ is weak,

[5]For hole doping of $NdNiO_2$, trivalent Nd^{3+} is replaced by divalent Sr^{2+}.

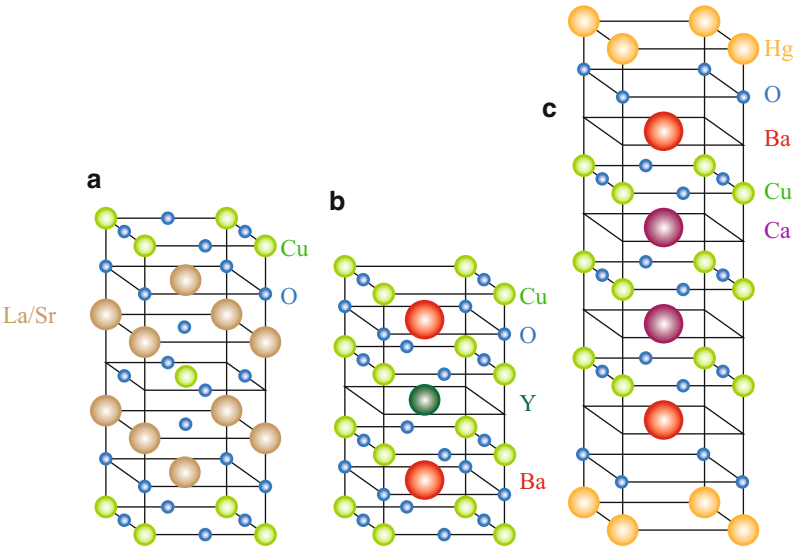


Fig. 10 Unit cells of superconductive cuprates with CuO_2 layers. (**a**) $La_{2-x}Sr_xCuO_4$, (**b**) $YBa_2Cu_3O_6$, (**c**) $HgBa_2Ca_2Cu_3O_8$

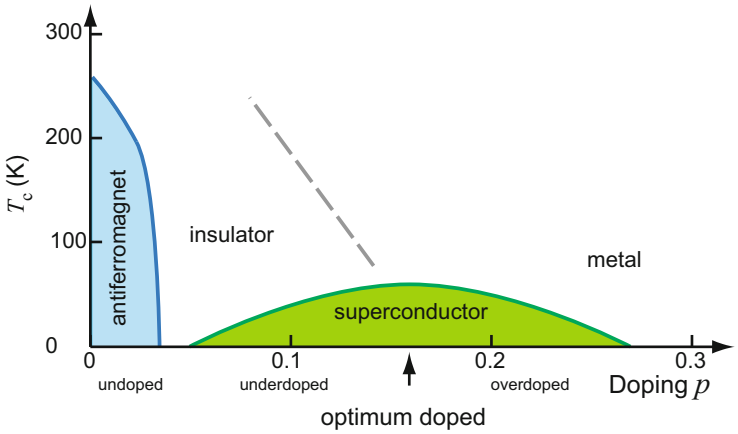


Fig. 11 Schematic phase diagram of a cuprate superconductor, indicating the dependence of T_c on doping; p is given in units of holes per Cu atom. An antiferromagnetic phase occurs at low doping, superconductivity appears in an intermediate doping range

while it is strong between Cu $3d_{x^2-y^2}$ and O $2p$ in cuprates; the doped holes in cuprates reside therefore at the oxygen sites, but they are at the Ni sites in $Nd_{1-x}Sr_xNiO_2$. In addition, $5d$ orbitals of the interlayer Nd^{3+}ions participate in the Fermi surface and hence contribute to transport. This indicates that

superconductivity of nickelates requires consideration of more orbitals than for cuprates (Nomura et al. 2019). An early review is given by Ji et al. (2021).

The critical temperature of high-T_c superconductors increases with increasing carrier density, similar to semiconducting low-temperature superconductors. The carrier density is generally low (typ. 5×10^{21} cm^{-3}) compared to conventional superconductors ($5 \times 10^{22} \ldots 10^{23}$ cm^{-3}). The doping properties are outlined below for cuprates; similar properties are also found for the iron pnictides (Ishida et al. 2009).

Undoped cuprates are often Mott insulators with antiferromagnetic ordering of the spins of the Cu^{2+} ions, which have an incomplete d shell ($3d^9$) with one hole spin $S = \frac{1}{2}\hbar$. The antiferromagnetic correlations in cuprates create a pseudogap, which is detrimental to superconductivity and is of a different origin than the superconducting gap. Superconductivity appears upon doping, as depicted in the universal phase diagram of a cuprate with p-type conductivity in Fig. 11. Doping with holes is achieved,[6] e.g., in $La_{2-x}M_xCuO_4$, by substituting La^{3+} ions with divalent ions such as Ba^{2+}, Sr^{2+}, or Ca^{2+}. The critical temperature $T_c/T_{c,max}$ is an approximately parabolic function of the hole concentration, with a maximum near 0.16 holes per Cu atom in the superconducting layer (Tallon et al. 1995). A similar phase diagram applies for cuprates with n-type conductivity, which are a minority of the ceramic copper-oxide-based superconductors. They differ in their magnetic texture and doping dependence (Zhang and Bennemann 1995), and their critical temperature is generally somewhat lower (~30 K).

The critical temperature of superconductivity can often be raised when hydrostatic pressure is applied, as noted for organic superconductors in Table 2 and as earlier found for $La_{2-x}Ba_xCuO_4$ by Chu et al. (1987). This effect can be mimicked in ceramics by replacing some atoms apart from the superconductive layer with smaller atoms, e.g., Ba with Sr.

It is instructive to discuss first the properties of these materials at temperatures above the superconductive transition temperature.

2.2 Normal-State Properties of High-T_c Superconductors

Experiments on high-temperature superconductors yield complex and often irregular results, although some generic trends are found, and a couple of general conclusions are drawn in Sect. 2.4. Here we point out a few prominent findings.

The conductivity in ceramic superconductors is highly anisotropic, with $\sigma_\parallel/\sigma_\perp = 10^2$–$10^5$ near the transition temperature. The conductivity within the Cu-O layers of cuprates is metallic,[7] whereas the current perpendicular to these layers is

[6]Doping can also be performed by changing the composition with oxygen. An excess of oxygen extracts electrons from the environment to form O^{2-} ions; these electrons partly come from the CuO_2 layers, leaving holes.

[7]Metallic conductivity is found above the Néel temperature, which, e.g., for undoped $YBa_2Cu_3O_{6+x}$, is quite high (500 K for $x = 0$, decreasing with doping). Below this temperature antiferromagnetic ordering leads to an energy gap and consequently to insolating behavior.

substantially reduced, probably due to tunneling through barriers. Rather unusual is the ubiquitous temperature-linear increase of the resistivity above the transition temperature depicted in Fig. 12 for $La_{2-x}Sr_xCuO_4$ with compositions $0.075 \leq x \leq 0.15$ that show superconductivity; in contrast, at compositions $x \leq 0.07$ that become not superconductive, the more common behavior of the resistivity is observed: it changes $\propto T^2$, indicating electron–electron scattering. A further peculiarity of the linear dependence is the wide range without saturation at higher temperatures (Gurvitch and Fiory 1987). Such a saturation is commonly found in transition metals, when the mean free path approaches the interatomic distance (Ioffe-Regel limit). A more recent study of the resistivity curvature indicates a gradual change in carrier density with doping and temperature – rather than an abrupt change of the Fermi surface in a scenario of a quantum-phase transition (Pelc et al. 2020). Up to now there exists no microscopic theory to describe such transport properties along with the observed pseudogap behavior and the pairing mechanism. For reviews see Hussey (2007), Ji et al. (2021).

The Hall coefficient of the high-T_c cuprates shows a strong temperature dependence instead of an expected near-constant behavior that prevails for non-superconducting compositions (Hwang et al. 1994; Segawa and Ando 2004). As an example the $1/T$ dependence of YBa_2CuO_7 above the transition temperature is shown in Fig. 13. The temperature and doping dependence of the Hall effect is related to complex topological changes of the Fermi surface, which are not yet fully understood.

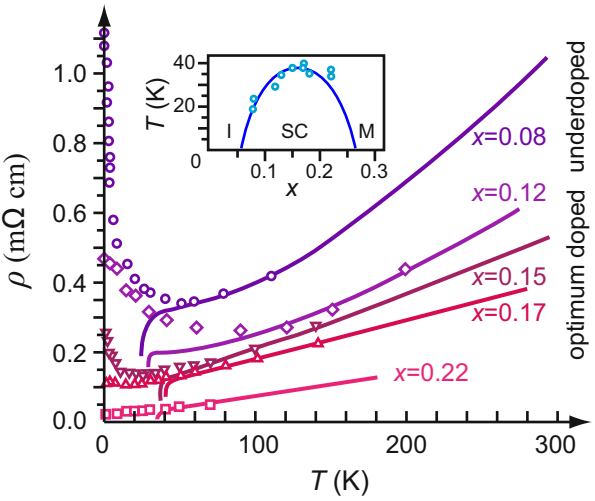


Fig. 12 Resistivity ρ of $La_{2-x}Sr_xCuO_4$ single crystals within the CuO_4 plane as a function of temperature; the composition x is the family parameter. *Solid lines* are measurements at zero magnetic field; at low T (symbols) superconductivity is suppressed by applying a high field (60 T) which exceeds the upper critical field. *Inset*: dependence of the critical temperature T_c on the Sr composition x; *I*, *SC*, and *M* denote insulating, superconducting, and metallic, respectively (After Boebinger et al. 1996)

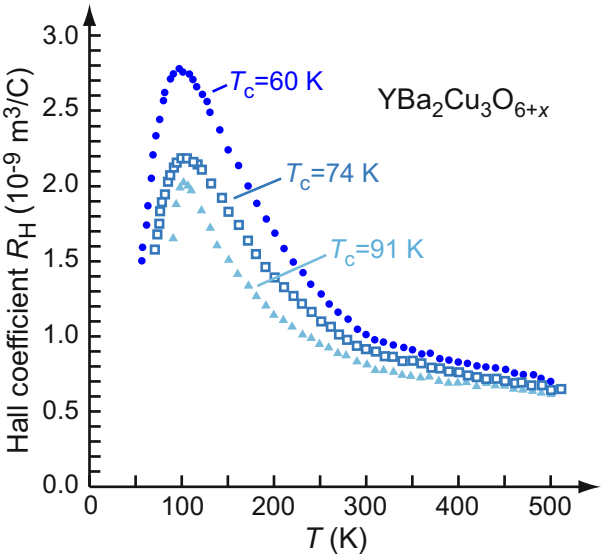


Fig. 13 Hall coefficient R_H of differently doped $YBa_2Cu_3O_{6+x}$ single crystals with **B** ∥ **c** as a function of temperature for samples with various critical temperatures T_c (After Harris et al. 1992)

2.3 The Superconductive State of High-T_c Superconductors

In conventional low-temperature superconductors the wavefunction of cooper pairs is generally isotropic and shows virtually no dependence on the orientation of **k**; it has *s* symmetry. This applies also for the Fermi sphere and the superconduction energy-gap (Fig. 7), which appears below T_c at the Fermi surface.

Ceramic superconductors are different. The model system of the cuprates is characterized by the superconducting CuO_2 layers illustrated in Fig. 14. Without doping there is one electron (correspondingly one hole) in the $d_{x^2-y^2}$ orbital of each copper ion; in absence of magnetic ordering this yields metallic conductivity of a half-filled band. Doping with holes disturbs the magnetic order (lowers $T_{Néel}$), until eventually superconductance occurs. Holes coupled to a pair repel each other and hence do not form a wavefunction with *s* symmetry, where the probability for both holes is maximum at the same place. Since the angular momentum of the pair must be even, $L = 2\hbar$ is the smallest possible angular momentum, yielding *d* symmetry for a Cooper pair in cuprates and hence also for the symmetry of the superconductive system.

The *d*-wave nature of cuprate superconductors was confirmed by many experiments, most directly by nodes observed in the excitation spectrum using angle-resolved photoemission spectroscopy (ARPES) for mapping the Fermi surface in momentum space (Platé et al. 2005, Horio et al. 2018, Sobota and Shen 2021). The results clearly show that the density of carriers in the superconducting state and also the energy gap to normal-state carriers has four lateral nodes in **k** space with a

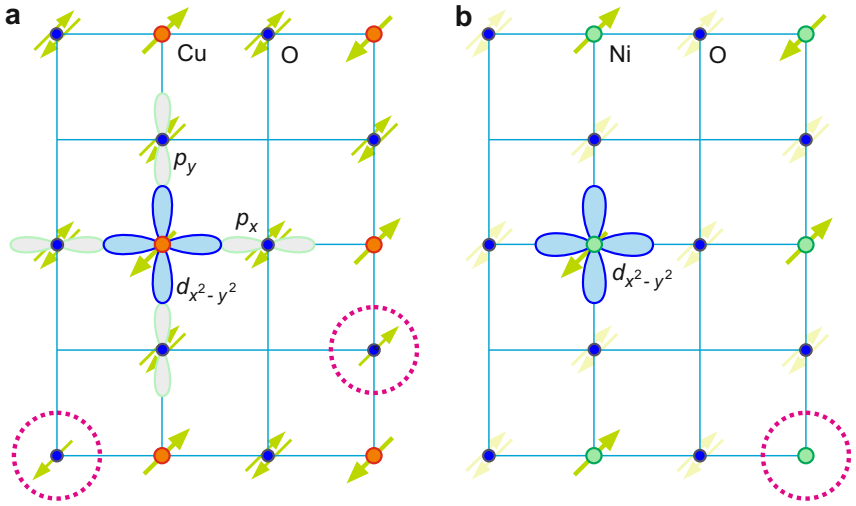


Fig. 14 (**a**) Schematic of the CuO_2 layer in a cuprate superconductor, showing the half-filled $d_{x^2-y^2}$ orbital of a Cu ion and the p_x, p_y orbitals of adjacent O ions. *Dotted circles* indicate holes in p orbitals of oxygen. (**b**) NiO_2 layer in a nickelate superconductor with Ni $d_{x^2-y^2}$ orbitals indicated. The *dotted circle* denotes a hole in a $d_{x^2-y^2}$ orbital of nickel.

symmetry corresponding to d character. Such a symmetry is also found in experiments on tunneling currents.

The superconduction gap can directly be measured using scanning tunneling spectroscopy (STS). Measurement of the tunneling conductance at one point of the sample yields the surface-near local density of states (LDOS), averaged over the in-plane angles of $\mathbf{k}$. In low-temperature superconductors the conductance and correspondingly the DOS at E_F vanish below T_c (Fig. 7b); above T_c the conductance has a U shape centered at zero bias in STS measurements with peaks at $\pm\,\Delta$. In high-temperature superconductors the nodes in $\Delta(\mathbf{k})$ result in a linear increase of the DOS, yielding a V shaped conductance centered at zero bias in STS measurements with a peak separation of $2\Delta_p$ depending on doping, see Fig. 15.

In the superconductive state, the critical current density is also anisotropic ($j_{c\parallel}/j_{c\perp} \cong 10^3$). This provides a further indication that superconductivity occurs within the plane of such layered crystals and may be interrupted by interfacing normal conducting layers. In $Bi_{1.6}Pb_{0.6}Sr_{1.8}CaCu_2O_x$ also an in-plane anisotropy of the critical current density was measured (Nakayama et al. 2000).

All high-T_c superconductors show type-II behavior, with low critical fields and substantial flux-line movements, causing some residual power dissipation. The flux-line lattice "melts" at temperatures[8] well below T_c, when thermal fluctuations of the

[8]The flux-line melting temperature is about 75 K for a material with $T_c = 93$ K, and is somewhat lower for $T_c = 125$ K material.

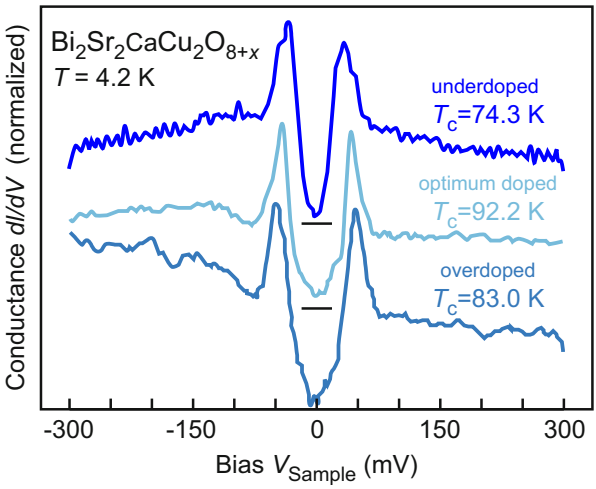


Fig. 15 Differential conductance of differently doped $Bi_2Sr_2CaCu_2O_{8+x}$ single crystals measured using scanning tunneling spectroscopy. *Horizontal lines* indicate zero conductance for the respective vertically offset spectra. Positive and negative sample bias results in electron tunneling into unoccupied and out of occupied electronic states, respectively (After Renner et al. 1998)

vortex positions become comparable to the generally small coherence length of the Cooper pairs. The critical current density may be increased if the flux lines are pinned by, e.g., point defects (Theuss 1993), or radiation damage induced by neutron bombardment (van Dover et al. 1989; Terai et al. 1997). The pinning mechanism is discussed by Matsushita (2014), see also Xu et al. (1995) and Schwarz et al. (2010); general pinning strategies are discussed by Feighan et al. (2017).

The upper critical magnetic field $H_{c_2}(T)$ has a large value and a large slope $\partial H_{c_2}/\partial T$. From critical-field measurements, the size of the quasiparticles can be determined (Tajima et al. 1988). In $EuBa_2Cu_3O_7$, they display a pancake shape, with coherent lengths of 35 Å diameter and 3.8 Å height lying in the Cu-O plane.

2.4 Mediating-Partner Models for High-T_c Superconductors

Even after about four decades of intensive research there exists no all-encompassing microscopic model on the origin of high-T_c superconductivity, in contrast to the BCS theory valid for low-temperature superconductors. Still a number of conclusions are generally agreed (Leggett 2006):

- High-T_c superconductivity originates from the formation of Cooper-pair-like carrier pairs.
- In the cuprates the superconductive electron system is located in the CuO_2 layers, in the nickelates correspondingly in the NiO_2 layers, and in the iron pnictides in the FeAs layers.
- In the two-dimensionally layered high-T_c superconductors the Cooper pairs form independently in the different layers, i.e., there is hardly hopping from one layer to another.

- Cooper pairs are not coupled by an electron–phonon interaction like those in low-temperature superconductors. There are indications for an at least predominant *electronic* pairing of wavefunctions.

Superconductance in ceramic compounds appears at sufficient doping (Fig. 11). Doping with holes generally leads to higher critical temperatures as doping with electrons. The mediating partner required to provide the coupling of the quasi-Cooper pair also needs to be a boson. As such, one may consider resonance valence bands (Anderson 1987), magnons (Chen and Goddard 1988), spin fluctuations (Schrieffer et al. 1988; Schrieffer 1991), plasmons (Tajima et al. 1991; Bill et al. 2003), and excitons.

Electron interaction with spin fluctuations are prominent in high-temperature superconductors. There are essentially two mechanisms creating hole (or electron) pairs. Intersubband hopping in a lattice with antiferromagnetic order lowers the electronic kinetic energy and may induce pairing by antiferromagnetic exchange. This mechanism is proposed by Anderson (1987) and is outlined in the following.

The Anderson RVB Theory The formation of Cooper pairs by phonon interaction in low-temperature superconductors can be imagined as trapping of an electron in the lattice-deformation funnel created by another electron (Sect. 1.2). In a high-T_c superconductor, the role of phonons is in Anderson's RVB model replaced by spin-density waves[9] (Anderson 1987): the spin of a moving electron creates a spin-density wave around it, causing a nearby electron to fall into the spin depression created by the first electron. Since the Coulomb repulsion prevents pairing of the two electrons on the same lattice site, the pairing occurs at near-neighbor lattice sites. The symmetry of the Cooper pair is hence not *s*-like as in the classical BCS theory. In the cuprates the pair wavefunction has $d_{x^2-y^2}$ symmetry instead[10] (Tsuei et al. 1997; Kirtley et al. 2006); this reflects the symmetry of the Cu orbitals in the Cu-O layers as indicated in Fig. 14.

In Anderson's resonating-valence-bond (RVB) theory there exist magnetic singlet pairs in the undoped insulating magnetic phase of cuprates (Fig. 16). These pairs become Cooper pairs when the insulator is sufficiently strongly doped (Fig. 11), providing a predominantly electronic and magnetic mechanism for superconductivity (Lee et al. 2006). Doping with holes in cuprates destroys the antiferromagnetic order of the two-dimensional Mott insulator and creates a disordered state like a spin liquid of resonating valence bonds. Spin fluctuations play also a major role in the pairing of the iron pnictides (Hosono and Ren 2009), while it is less pronounced in the more recently found nickelates (Lu et al. 2021).

[9]All high-T_c superconductors are strong spin-density wave systems; the recently found nickelates might be an exception – for a discussion see Lu et al. (2021).

[10]In La-based cuprates some hybridization with d_{z^2} was found (Matt et al. 2018).

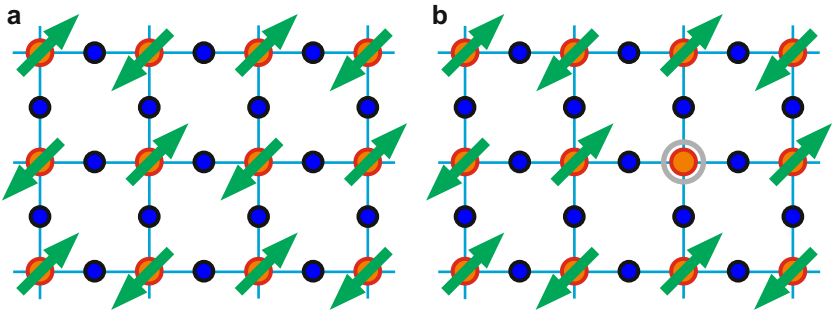


Fig. 16 Schematic of the two-dimensional Cu-O layer in a cuprate superconductor (**a**) in the insulating undoped case showing antiferromagnetic orientation of the spins, and (**b**) doped with a hole located at a Cu ion

Charge-Fluctuation Models Within the classical BSC theory a high critical temperature T_c is connected to a large energy of the mediating boson, which may basically be an exciton or a plasmon. Considering an exciton as mediating partner, two groups of electrons are required: more localized electrons providing the coupling and the coupled electrons in the superconducting band. The excitons in cuprates may result from a charge transfer according to $Cu^{2+}O^{2-} \rightarrow Cu^{1+}O^{1-}$ (Varma et al. 1987, Varma 1997). Such a charge-transfer excitation can induce pairing with *s* or d_{xy} symmetry. For a review, see Little (1992).

The two-dimensionality and van Hove singularities close to the Fermi energy favor charge instabilities and thereby the possibility of pairing by charge fluctuations. Also the occurrence of specific plasma excitations is enhanced by such conditions, basically allowing for plasmon-induced pairing. The two-dimensionality[11] and increasing ionicity[12] supports the formation of tightly bound excitons, which can survive higher temperatures.

Résumé on Mediating-Partners The present status indicates that the pair formation in high-temperature superconductors could actually comprise more than one singular mechanism. Spin fluctuation certainly contributes to pairing as an important part, which may be substantially assisted by charge fluctuations that in turn enhance electron–phonon interaction.

[11]Confinement of three-dimensional excitons to two dimensions leads to a substantial increase of the exciton binding-energy see Sect. 2.1 of ▶ Chap. 14, "Excitons". The superlattice-like (2D) structure of the ceramic superconductors could provide such a confinement.

[12]Here excitons, or polarons, become more tightly bound (Frenkel excitons). Earlier observation of heavy-fermion superconductors (Stewart 1984; Joynt and Taillefer 2002) may point toward the assistance of more tightly bound polarons (with $m_n \cong 200\ m_0$) in forming superconducting compounds, (e.g., $CeCu_2Si_2$ or UPt_3).

3 Summary

Superconductivity was in early days believed to be restricted to metals and to temperatures below 25 K. It involves Cooper pairs of electrons, which can be formed due to phonon interaction. Bose–Einstein condensation of such pairs creates a bandgap that eliminates damping of the transport, which still occurs for the remaining normal electrons. An injection of electrons above the gap, via Giaever tunneling, shows conductivity of normal electrons in superconductors. The density of superconducting Cooper pairs increases with decreasing temperature at the expense of normal electrons, causing an increasing superconduction bandgap for Cooper pairs, an increasing critical current (above which superconductivity vanishes), and a decreasing penetration depth for a magnetic field – the Meissner field expulsion. Above the temperature-dependent critical field, these type I superconductors become normal conductive. Type II superconductors enter a mixed Shubnikov phase above a lower critical magnetic field, with a regular pattern of flux lines penetrating the superconductor. The density of these (partially normal-conductive) vortices increases for increasing field, until above the upper critical field superconductivity vanishes.

High-temperature superconductivity is found mostly in perovskite-type solids, with transition temperatures exceeding 100 K. In the large family of cuprates superconductivity occurs in the CuO_2 layers and is usually carried by quasiparticles related to pairs of holes. Generally undoped compounds are insulating due to magnetic ordering of the spins located at the Cu ions; these cuprates require doping of the layers beside the CuO_2 layers with sufficient holes to generate superconductivity. Similar conditions apply for the related nickelates with superconducting NiO_2 layers. Unlike in low-temperature superconductors the wavefunction of Cooper pairs is anisotropic with in-plane *d* symmetry, yielding a corresponding symmetry of the Fermi surface and the superconduction energy-gap. Similar conditions are found for the family of iron pnictides, where the superconductive electron system is carried in FeAs layers. The microscopic mechanism for pair formation in high-temperature superconductors differs from that found in typical metals. While spin fluctuations certainly play an important role, a generally accepted model is still to be developed.

References

Allen PB, Mitrović B (1982) Theory of superconducting T_c. Solid State Phys 37:1

Anderson P (1987) The resonating valence bond state in La_2CuO_4 and superconductivity. Science 235:1196

Appel J (1966) Superconductivity in pseudoferroelectrics. Phys Rev Lett 17:1045

Ardavan A, Brown S, Kagoshima S, Kanoda K, Kuroki K, Mori H, Ogata M, Uji S, Wosnitza J (2012) Recent topics of organic superconductors. J Phys Soc Jpn 81:011004

Bardeen J, Cooper LN, Schrieffer JR (1957) Theory of superconductivity. Phys Rev 108:1175.

Batlogg B (1991) Physical properties of high-T_c superconductors. Phys Today 44:44

Bednorz JG, Müller KA (1986) Possible high-T_c superconductivity in the Ba-La-Cu-O system. Z Phys B 64:189

Bill A, Morawitz H, Kresin VZ (2003) Electronic collective modes and superconductivity in layered conductors. Phys Rev B 68:144519

Blatt JM (1961) Persistent ring currents in an ideal Bose gas. Prog Theor Phys 26:761

Boebinger GS, Ando A, Passner A, Kimura T, Okuya M, Shimoyama J, Kishio K, Tamasaku K, Ichikawa N, Uchida S (1996) Insulator-to-metal crossover in the normal state of $La_{2-x}Sr_xCuO_4$ near optimum doping. Phys Rev Lett 77:5417

Buckel W, Kleiner R (2004) Superconductivity: fundamentals and applications. Wiley-VCH, Weinheim

Chen G, Goddard WA (1988) The magnon pairing mechanism of superconductivity in cuprate ceramics. Science 239:899

Chu CW (1997) High-temperature superconducting materials: a decade of impressive advancement of T_c. IEEE Trans Appl Supercond 7:80

Chu CW, Hor PH, Meng RL, Gao L, Huang ZJ, Wang YQ (1987) Evidence for superconductivity above 40 K in the La-Ba-Cu-O compound system. Phys Rev Lett 58:405

Cohen ML (1964) Superconductivity in many-valley semiconductors and in semimetals. Phys Rev 134:A511

Feighan JPF, Kursumovic A, MacManus-Driscoll JL (2017) Materials design for artificial pinning centres in superconductor PLD coated conductors. Supercond Sci Technol 30:123001

Fetter AL, Walecka JD (1971) Quantum theory of many-particle systems. McGraw-Hill, New York

Gastiasoro MN, Ruhman J, Fernandes RM (2020) Superconductivity in dilute $SrTiO_3$: a review. Ann Phys 417:168107

Geethanath S, Vaughan Jr JT (2019) Accessible magnetic resonance imaging: a review. J Magn Reson Imaging 49:e65

Gurevich VL, Larkin AI, Firsov YuA (1962) On the possibility of superconductivity in semiconductors. Sov Phys Sol State 4:131

Gurvitch M, Fiory AT (1987) Resistivity of $La_{1.825}Sr_{0.175}CuO_4$ and $YBa_2Cu_3O_7$ to 1100 K: absence of saturation and its implications. Phys Rev Lett 59:1337

Harris JM, Yan YF, Ong NP (1992) Experimental test of the T^2 law for the Hall angle from T_c to 500 K in oxygen-reduced $YBa_2Cu_3O_{6+x}$ crystals. Phys Rev B 46:14293

Harshman DR, Mills AP (1992) Concerning the nature of high-Tc superconductivity: Survey of experimental properties and implications for interlayer coupling. Phys Rev B 45:10684

Hein RA , Gibson JW, Mazelsky R, Miller RC, Hulm JK (1964) Superconductivity in germanium telluride. Phys Rev Lett 12:320

Horio M, Hauser K, Sassa Y, Mingazheva Z, Sutter D, Kramer K, Cook A, Nocerino E, Forslund OK, Tjernberg O, Kobayashi M, Chikina A, Schröter NBM, Krieger JA, Schmitt T, Strocov VN, Pyon S, Takayama T, Takagi H, Lipscombe OJ, Hayden SM, Ishikado M, Eisaki H, Neupert T, Månsson M, Matt CE, Chang J (2018) Three-dimensional Fermi surface of overdoped La-based cuprates. Phys Rev Lett 121:077004

Hosono H, Ren Z-A (2009) Focus on iron-based superconductors. New J Phys 11:025003

Hussey NE (2007) Normal state transport properties. In: Schrieffer JR, Brooks JS (eds) Handbook of high-temperature superconductivity. Theory and experiment. Springer, New York, pp 399–425

Hwang HY, Batlogg B, Takagi H, Kao HL, Kwo J, Cava RJ, Krajewski JJ, Peck WF Jr (1994) Scaling of the temperature dependent Hall effect in $La_{2-x}Sr_xCuO_4$. Phys Rev Lett 72:2636

Ishida K, Nakai Y, Hosono H (2009) To what extent iron-pnictide new superconductors have been clarified: a progress report. J Phys Soc Jpn 78:062001

Ishiguro T, Yamaji K, Saito G (1998) Organic superconductors. Springer, Berlin

Jérome D (1994) Organic superconductors. Solid State Commun 92:89

Ji Y, Liu J, Li L, Liao Z (2021) Superconductivity in infinite layer nickelates. J Appl Phys 130:060901

Joynt R, Taillefer L (2002) The superconducting phases of UPt_3. Rev Mod Phys 74:235

Kirtley JR, Tsuei CC, Ariando A, Verwijs CJM, Harkema S, Hilgenkamp H (2006) Angle-resolved phase-sensitive determination of the in-plane gap symmetry in $YBa_2Cu_3O_{7-\delta}$. Nat Phys 2:190

Lee PA, Nagaosa N, Wen XG (2006) Doping a Mott insulator: physics of high-temperature superconductivity. Rev Mod Phys 78:17

Leggett AJ (2006) What DO we know about high T_c? Nat Phys 2:134

Li D, Lee K, Wang BY, Osada M, Crossley S, Lee HR, Cui Y, Hikita Y, Hwang HY (2019) Superconductivity in an infinite-layer nickelate. Nature 572:624

Little WA (1992) Generalization of BCS superconductivity to non-phonon mediated interactions: the excitonic interaction. In: Maekawa S, Sato M (eds) Physics of high-temperature superconductors. Springer, Heidelberg, pp 113–124

London F, London H (1935) The electromagnetic equations of the supraconductor. Proc Roy Soc (London) A 149:72

Lu H, Rossi M, Nag A, Osada M, DF Li, Lee K, Wang BY, Garcia-Fernandez M, Agrestini S, Shen ZX, Been EM, Moritz B, Devereaux TP, Zaanen J, Hwang HY, Zhou K-J, Lee WS (2021) Magnetic excitations in infinite-layer nickelates. Science 373:213

Matsushita T (2014) Flux pinning in superconductors. Springer, Berlin

Matt CE, Sutter D, Cook AM, Sassa Y, Månsson M, Tjernberg O, Das L, Horio M, Destraz D, Fatuzzo CG, Hauser K, Shi M, Kobayashi M, Strocov VN, Schmitt T, Dudin P, Hoesch M, Pyon S, Takayama T, Takagi H, Lipscombe OJ, Hayden SM, Kurosawa T, Momono N, Oda M, Neupert T, Chang J (2018). Direct observation of orbital hybridisation in a cuprate superconductor. Nature Commun 9:972

Meissner W, Ochsenfeld R (1933) Ein neuer Effekt bei Eintritt der Supraleitfähigkeit. Naturwiss 21:787 (A new effect at the onset of superconductivity, in German)

Nakayama Y, Motohashi T, Otzschi K, Shimoyama J, Kishio K (2000) In-plane anisotropy of critical current density in $Bi_{1.6}Pb_{0.6}Sr_{1.8}CaCu_{2.0}O_y$. Physica C 341:1477

Nomura Y, Hirayama M, Tadano T, Yoshimoto Y, Nakamura K, Arita R (2019) Formation of a two-dimensional single-component correlated electron system and band engineering in the nickelate superconductor $NdNiO_2$. Phys Rev B 100:205138

Pelc D, Veit MJ, Dorow CJ, Ge Y, Barišić N, Greven M (2020) Resistivity phase diagram of cuprates revisited. Phys Rev B 102:075114

Plakida N (2010) High-temperature cuprate superconductors. Springer, Berlin

Platé M, Mottershead JDF, Elfimov IS, Peets DC, Liang R, Bonn DA, Hardy WN, Chiuzbaian S, Falub M, Shi M, Patthey L, Damascelli A (2005) Fermi surface and quasiparticle excitations of overdoped $Tl_2Ba_2CuO_{6+\delta}$. Phys Rev Lett 95:077001

Renner C, Revaz B, Genoud J-Y, Kadowaki K, Fischer Ø (1998) Pseudogap precursor of the superconducting gap in under- and overdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$. Phys Rev Lett 80:149

Reynolds CA, Serin B, Wright WH, Nesbitt LB (1950) Superconductivity of isotopes of mercury. Phys Rev 78:487

Sarracanie M, Salameh N (2020) Low-field MRI: how low can we go? A fresh view on an old debate. Front Phys 8:172

Schooley JF,Hosler WR, Cohen ML (1964) Superconductivity in semiconducting $SrTiO_3$. Phys Rev Let. 12:474

Schrieffer JR (1964) Theory of superconductivity. Benjamin, New York

Schrieffer JR (1991) The influence of spin fluctuations on the physical properties of high T_c materials. Physica C 185:17

Schrieffer JR, Wen X-G, Zhang S-C (1988) Spin-bag mechanism of high-temperature superconductivity. Phys Rev Lett 60:944

Schwarz A, Liebmann M, Pi UH, Wiesendanger R (2010) Real space visualization of thermal fluctuations in a triangular flux-line lattice. New J Phys 12:033022

Segawa K, Ando Y (2004) Intrinsic Hall response of the CuO_2 planes in a chain-plane composite system of $YBa_2Cu_3O_y$. Phys Rev B 69:104521

Smidman M, Salamon MB, Yuan HQ, Agterberg DF (2017) Superconductivity and spin-orbit coupling in non-centrosymmetric materials: a review. Rep Prog Phys 80:036501

Sobota JA, Shen Z-X (2021) Angle-resolved photoemission studies of quantum materials. Rev Mod Phys 93:025006

Stewart GR (1984) Heavy-fermion systems. Rev Mod Phys 56:755
Tajima Y, Hikita M, Ishii T, Fuke H, Sugiyama K, Date M, Yamagishi A, Katsui A, Hidaka Y, Iwata T, Tsurumi S (1988) Upper critical field and resistivity of single-crystal $EuBa_2Cu_3O_y$: direct measurements under high field up to 50 T. Phys Rev B 37:7956
Tajima S, Kaneko T, Wada T, Tomeno I, Tai K, Mizuo Y, Yamauchi H, Koshizuka K, Ido T, Uchida S (1991) Plasmons in high-T_c superconducting cuprates. Physica C 185:1013
Tallon JL, Bernhard C, Shaked H, Hitterman RL, Jorgensen JD (1995) Generic superconducting phase behavior in high-T_c cuprates: T_c variation with hole concentration in $YBa_2Cu_3O_{7-\delta}$. Phys Rev B 51:12911
Terai T, Kobayashi T, Ito Y, Kishio K, Shimoyama J (1997) Change in superconducting properties of Bi-2212 single crystal due to fast neutron irradiation followed by thermal annealing. Physica C 282:2285
Theuss H (1993) Flux line pinning by point defects in single crystalline high-T_c. Physica C 208:155
Townsend P, Sutton J (1962) Investigation by electron tunneling of the superconducting energy gaps in Nb, Ta, Sn and Pb. Phys Rev 128:591
Tsuei CC, Kirtley JR, Ren ZF, Wang JH, Raffy H, Li ZZ (1997) Pure d_{x2-y2} order-parameter symmetry in the tetragonal superconductor $Tl_2Ba_2CuO_{6+\delta}$. Nature 387:481
van Dover RB, Gyorgy EM, Schneemeyer LF, Mitchell JW, Rao KV, Puzniak R, Waszczak JV (1989) Critical currents near 10^6 A cm^{-2} at 77 K in neutron-irradiated single-crystal $YBa_2Cu_3O_7$. Nature 342:55
Varma CM (1997) Non-fermi-liquid states and pairing instability of a general model of copper oxide metals. Phys Rev B 55:14554
Varma CM, Schmitt-Rink S, Abrahams E (1987) Charge transfer excitations and superconductivity in "ionic" metals. Solid State Commun 62:681
Webb GW, Marsiglio F, Hirsch JE (2015) Superconductivity in the elements, alloys and simple compounds. Physica C 514:17
Wu X, Di Sante D, Schwemmer T, Hanke W, Hwang HY, Raghu S, Thomale R (2020) Robust $d_{x^2-y^2}$-wave superconductivity of infinite-layer nickelates. Phys Rev B 101:060504
Xu X, Fang J, Cao X, Li K (1995) The critical current density and the pinning mechanism of epitaxial $YBa_2Cu_3O_{7-\delta}$ thin films. Solid State Commun 93:291
Zeng S, Tang CS, Yin X, Li C, Li M, Huang Z, Hu J, Liu W, Omar GJ, Jani H, Lim ZS, Han K, Wan D, Yang P, Pennycook SJ, Wee ATS, Ariando A (2020) Phase diagram and superconducting dome of infinite-layer $Nd_{1-x}Sr_xNiO_2$ thin films. Phys Rev Lett 125:147003
Zhang W, Bennemann KH (1995) Theory of magnetic textures in high-T_c superconductors: electron versus hole doping dependence. J Phys Condens Matter 7:1335

Carrier Transport in Low-Dimensional Semiconductors

27

Karl W. Böer and Udo W. Pohl

Contents

1 In-Plane Transport in Two-Dimensional Structures 1114
1.1 The Two-Dimensional Electron Gas (2DEG) 1115
1.2 Carrier Mobility in a 2D Electron Gas 1119
2 Perpendicular Carrier Transport in 2D Structures 1122
2.1 Tunneling Through a Planar Barrier 1122
2.2 Tunneling Through a Double-Barrier Quantum Well 1123
2.3 Perpendicular Transport in Superlattices 1125
3 Transport in 1D Structures 1132
3.1 Diffusive and Ballistic Transport 1134
3.2 Quantization of Conductance 1135
3.3 Landauer-Büttiker Formalism 1138
4 Topological Insulators 1140
4.1 Topological Band Theory 1141
4.2 Quantum Spin Hall Insulator 1144
4.3 3D Topological Insulators 1148
5 Zero-Dimensional Transport 1152
5.1 Single-Electron Tunneling 1153
5.2 Coulomb Blockade in Few-Electron Dots 1156
6 Summary 1157
References 1157

Abstract

Carrier transport in semiconductors with reduced dimensions is determined by the low-dimensional density of states. In two-dimensional systems such as quantum

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_27

wells and superlattices, the carrier mobility is highly anisotropic. Parallel to the barriers it may exceed the bulk value by far in a two-dimensional electron gas at low temperature. Perpendicular to the interfaces, carriers have to penetrate the barriers and the mobility is low. Tunneling through thin barriers is an important process; it is enhanced when matched with quantized energy levels and leads to negative differential resistance. In one-dimensional quantum wires, ballistic transport occurs and the conductance gets quantized. Transport through a zero-dimensional quantum dot is affected by charging with single electrons, giving rise to a Coulomb blockade with zero conduction at certain bias values. Low-dimensional carrier transport occurs also in topological insulators. These solids are insulators in the bulk but have conducting states at their edge (two-dimensional quantum spin Hall insulator) or surface (three-dimensional topological insulator). The surface states are protected by symmetry and robust against disruption.

Keywords

Ballistic transport · Bloch oscillations · Carrier mobility · Chiral state · Coherent tunneling · Conductance quantization · Coulomb blockade · Coulomb diamond · Helical state · High-field domain · Landauer-Büttiker formalism · Landauer formula · Negative differential resistance · One-dimensional transport · Quantum Hall state · Quantum spin Hall · Resonant tunneling diode · Sequential tunneling · Spin-momentum locking · Topological insulator · Transmission coefficient · Tunneling · Quantum-cascade laser · Quantum wire · Single-electron tunneling · Subband · Time-reversal symmetry · Two-dimensional electron gas · Wannier–Stark ladder · Zero-dimensional transport

1 In-Plane Transport in Two-Dimensional Structures

The carrier transport in two-dimensional semiconductor structures is highly anisotropic. It is rather low *perpendicular* to the interfaces, where electrons have to penetrate the barriers. Transport through these barriers is controlled by quantum-mechanical tunneling. Furthermore, the potential in double-barrier and superlattice structures leads to electron energies quantized in discrete levels or minibands affecting the tunneling process. Transport perpendicular to barriers is discussed in Sect. 2

In contrast, carrier mobility *parallel* to the interfaces may be very high and is often much higher than in the bulk material. This applies particularly for modulation-doped structures, where a high density of carriers can drift within layers of high purity and lattice perfection, while dopants are confined within barrier layers. At low temperatures, these structures form an ideal two-dimensional electron gas (2DEG), which is applied, e.g., to define the standard for the electrical resistance via the quantum-Hall effect (Sect. 3.1 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients"). The 2DEG is a convenient model system to study quantum transport, because the electron density and hence the Fermi wavelength are adjustable by external potentials; the mean free path of carriers may exceed device dimensions, and lateral

patterning allows for further reducing the dimensionality to planar quantum wires or quantum dots. We therefore consider the properties of such a 2DEG in more detail.

1.1 The Two-Dimensional Electron Gas (2DEG)

In a two-dimensional electron gas (2DEG)[1] the electrons are free to move in two lateral dimensions (x, y) and are confined in the third perpendicular dimension (z). The lateral mobility of the electrons is basically limited by phonon scattering at higher temperatures and by scattering at ionized impurities at low temperatures (Sects. 2 and 3 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields"). The limitation at low temperatures can effectively be circumvented by separating the ionized donor dopants from the free electrons using *modulation doping*. This term indicates that doping is not homogeneous; in the 2D electron gas of a modulation-doped heterostructure, the doping range is restricted to the barrier material, and the 2DEG is created by a band-bending effect at the interface of the heterojunction. Another approach to create a 2DEG is the formation of an inversion layer formed at the gate of a field-effect transistor (FET). In both cases a narrow, highly conductive layer is generated parallel to the interface in an approximately triangular vertical potential $V(z)$. Since usually an additional gate potential is applied to the hetero-junction, both approaches allow to adjust the Fermi level with respect to the band edge and consequently to vary the electron density of the 2DEG.

2D Electron Gas at an Inversion Layer A common device creating a 2D electron gas is the metal-oxide semiconductor FET (MOSFET) based on p-type silicon; for a comprehensive review see Ando et al. (1982); recent progress is briefly reviewed by Alam et al. (2019). The native SiO_2 oxide has a large (~3 eV) conduction-band offset to Si; application of a positive gate voltage bends the conduction-band edge of the p-type Si electrostatically below the Fermi energy E_F, creating an inversion layer with electron conductance at the SiO_2/Si interface (Fig. 1).

The sheet carrier density n_{2D} (in units of m^{-2}) depends on the areal capacitance C_{2D} (in units of F/m^2) formed by the 2DEG and the gate electrode, and on the difference of the threshold voltage V_t and the gate voltage V_g indicated in Fig. 1; the inversion layer is created for $V_g > V_t$. From $e\, n_{2D} = C_{2D}\,(V_g - V_t)$ and $C_{2D} = \varepsilon_{oxide} \times \varepsilon_0 / d_{oxide}$ we obtain

$$n_{2D} = \frac{\varepsilon_{oxide}\,\varepsilon_0}{e\, d_{oxide}} \left(V_g - V_t\right). \tag{1}$$

The electron mobility in the 2DEG of the inversion layer in Si is typically about 10^3 $cm^2V^{-1}s^{-1}$ at 300 K and an order of magnitude higher at low T. It is limited by various effects; the effective electron mass of Si ($m_n = 0.19\ m_0$) is higher than that of direct semiconductors, and there occurs intervalley scattering due to the

[1]The analogous system with *holes* as free carriers is called two-dimensional hole gas (2DHG). Due to their smaller effective mass usually electrons are preferred as carriers in transport devices.

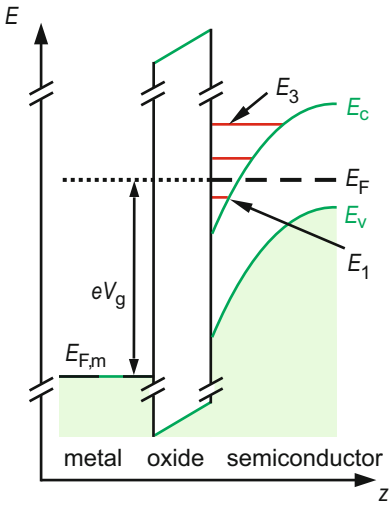


Fig. 1 Energy-band diagram of a metal-oxide-semiconductor (MOS) structure with a 2DEG created at the interface between the oxide and the *p*-type Si, indicating the subbands (*red lines*) created at the interface with the lowest subband E_1 occupied

degeneracy of conduction-band minima, and surface scattering due to the microscopically rough interface of Si to the amorphous SiO_2.

2D Electron Gas at a Heterointerface This approach usually applies a GaAs/AlGaAs interface with a metal gate at the wide-bandgap AlGaAs, forming a modulation-doped FET (MODFET),[2] see Störmer et al. (1979). The AlGaAs is doped with donors, except for a narrow undoped region at the interface to the undoped GaAs. Electrons diffusing from the doped AlGaAs barrier to the interface are confined in the potential well illustrated in Fig. 2 and form a conductive channel. The confinement perpendicular to the interface creates subbands, which can be populated depending on the applied gate voltage; usually only the lowest subband is occupied. The dependence of n_{2D} on gate bias in GaAs-based HEMTs is analyzed by Khandelwal et al. (2013). For a general review see Harris et al. (1989).

The sheet carrier-density n_{2D} of the MODFET is also controlled by the gate potential V_g similar to the Si MOSFET and likewise described by Eq. 1, see also Fig. 5. The threshold voltage V_t is not fixed; it can be adjusted by the Shottky barrier ϕ_{Bn} at the metal/AlGaAs interface (see Fig. 17 of ▶ Chap. 16, "Crystal Interfaces") and the conduction-band offset at the $Al_xGa_{1-x}As/GaAs$ interface.

The 2DEG of the AlGaAs-based MODFET benefits from the low effective mass of electrons in GaAs ($m_n = 0.067\ m_0$) and the continuous crystalline periodicity across the interface of this well lattice-matched material combination, avoiding interface scattering; furthermore, no intervalley scattering occurs in the single-

[2]Such a device is also referred to as high-electron-mobility transistor (HEMT) or two-dimensional electron gas field-effect transitor (TEGFET). Typical Al mole fraction x of $Al_xGa_{1-x}As$ and donor-doping level are 0.30 and 10^{18} cm^{-3}, respectively. Donor doping stops at a spacer distance d (typically on the order of 10 nm, see Fig. 5) away from the interface to GaAs.

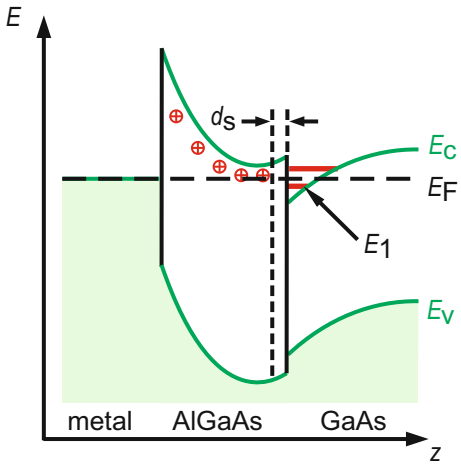


Fig. 2 Energy-band diagram of a modulation-doped AlGaAs/GaAs heterostructure with a metal gate-contact and a 2DEG located at the interface; the ⊕ symbols denote ionized donor atoms. An undoped spacer layer of thickness d_s is indicated

conduction-band GaAs. As a result, electron mobilities are much higher than in Si-based inversion layers, yielding values up to 10^7 $cm^2V^{-1}s^{-1}$ (Pfeiffer et al. 1989).

Energy Levels in a 2D Electron Gas Various quantities change abruptly across the interface of the two-dimensional electron gases described above, particularly the band edge, the effective mass, and the dielectric constant. The resulting energies of the 2DEG can only be evaluated numerically (Stern and Das Sarma 1984). For practical electron gases some approximations are introduced yielding also analytical solutions. The conduction-band offset at the heterojunction is often assumed infinitely high; this approach is reasonable, since the ground-state energy in the confining potential $V(z)$ is in the range of some tens of meV, compared to the very large offset at the SiO_2/Si interface and still ~300 meV at the AlGaAs/GaAs junction. This assumption leads to the boundary condition $\varphi(z = 0) = 0$ for the wavefunction at the interface located at $z = 0$; consequently the variations of the effective mass $m_n(z)$ and the dielectric constant $\varepsilon_r(z)$ are neglected and set to the values of the channel material.

For low carrier densities n_{2D} the potential can then be well approximated by the triangular dependence

$$V(z) = \begin{cases} \infty & z = 0 \\ eFz & z > 0 \end{cases}, \tag{2}$$

where in an AlGaAs/GaAs heterostructure the electric field F is given by the sum of the fields produced by the ionized donors in the barrier, F_0, and by that of the electrons in the 2DEG:

$$F = F_0 + \frac{1}{2}\frac{en_{2D}}{\varepsilon_r\varepsilon_0}; \tag{3}$$

the factor ½ is introduced to obtain an average of the field over the space-charge layer of the electrons. Using the potential of Eq. 2, the analytic solution of the Schrödinger equation yields Airy functions for the electron wavefunctions with energy levels

$$E_i = \left(\frac{\hbar^2}{2m_n}\right)^{1/3} \left[\frac{3\pi}{2}\left(i - \frac{1}{4}\right) eF\right]^{2/3}, \quad i = 1,\ 2,\ 3,\ \ldots. \tag{4}$$

The solutions are illustrated in Fig. 3.

The approximations lead to a ground-state energy E_1 being typically 20% too high, while the energy separation to the first excited level E_2 is less affected.

For larger carrier densities $n_{2D} > 2\,\varepsilon_r\varepsilon_0 F_0/e$, the approximation of Eq. 2 is poor. If only the ground state is occupied, a parameterized wavefunction yields more realistic solutions (Stern 1972).

The energy $E_i(z)$ represents the bottom of the two-dimensional subband i with free carrier motion in the (x, y) plane, where no restriction exists. The carrier density $n_{2D,i}$ in each subband depends on the energy separation to the Fermi level E_F and is given by

$$n_{2D,i} = \frac{m_n kT}{\pi\hbar^2} \ln\left[1 + \exp\left(\frac{E_F - E_i}{kT}\right)\right]. \tag{5}$$

For practical applications usually only the ground state E_0 is occupied to avoid intersubband scattering between occupied bands (Störmer et al. 1982; Piazza et al. 1998). The mobility of the excited subband is substantially lower than that of the ground subband. This behavior originates from stronger ionized-impurity Coulomb scattering for carriers with small wavevectors and the smaller Fermi wavevector of carriers in the higher subband.

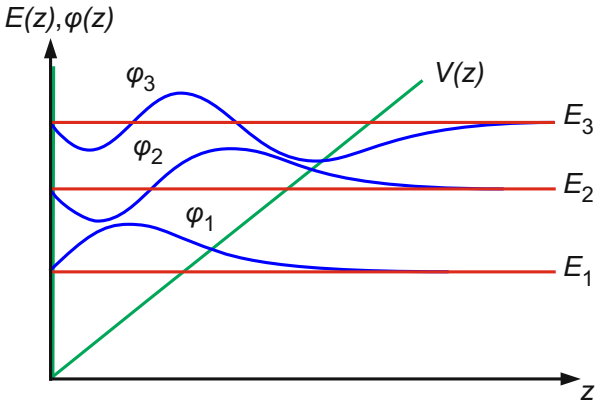


Fig. 3 Wavefunctions $\varphi(z)$ and energy levels E_i in a triangular potential $V(z)$ for typical values of an AlGaAs/GaAs heterointerface; the ground state is located at $E_1 \cong 50$ meV

1.2 Carrier Mobility in a 2D Electron Gas

The mobility of carriers in a 2D electron gas depends on temperature, carrier density, and, in modulation-doped heterostructures, on the thickness of the undoped spacer layer. The temperature dependence of the mobility in an AlGaAs/GaAs heterostructure is shown in Fig. 4, along with limits imposed by various scattering processes. Major effects are scattering at ionized impurities at low temperatures and scattering at optical and acoustical phonons at high temperatures; at low temperature a limit is also given by piezoelectric scattering for very pure samples. In the 2D electron gas the mobility continues to increase for decreasing temperature beyond the expected maximum at $\mu \cong 10^5$ cm^2/(Vs) for a residual donor density of 4×10^{13} cm^{-3} and approaches 5×10^6 cm^2/(Vs) below 10 K.

The high mobility of carriers at low temperature results from the reduced ionized-donor scattering achieved by the spatial separation between the electrons in the two-dimensional channel and the ionized donors in the adjacent barrier region. The Coulomb scattering is further reduced by the thickness of the undoped spacer layer in the barrier between these two regions, which also affects the carrier density. The effect of a varied undoped spacer thickness on mobility and carrier density n_{2D} is shown in Fig. 5: the 40 nm thick n-type AlGaAs barrier with 1.3×10^{18} cm^{-3} Si donors is separated from the conductive GaAs channel by undoped AlGaAs barriers with thicknesses d_s varied from 0.9 to 320 nm. The mobility is maximum in an intermediate range, partly due to the dependence on the carrier density n_{2D}.

For a fixed thickness of the undoped spacer layer, the low-temperature mobility increases with increasing carrier density n_{2D}. The effect is also controlled by the temperature dependence of the scattering processes as illustrated in Fig. 6 for a

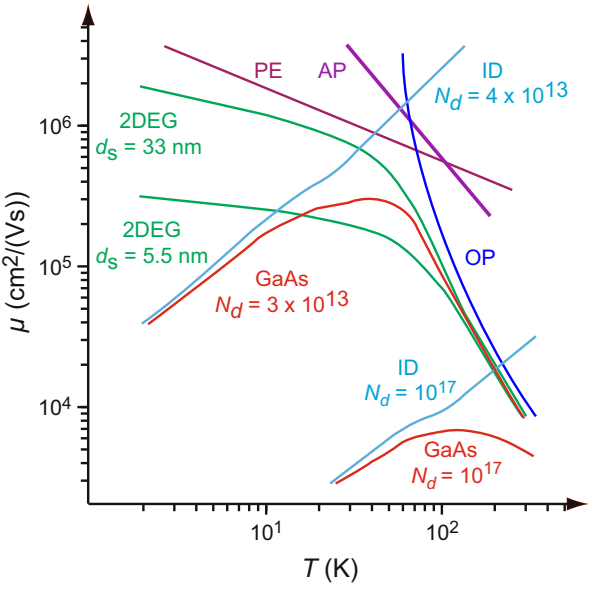


Fig. 4 Temperature dependence of the electron mobility in the 2D electron gas of modulation-doped AlGaAs/GaAs heterostructures (*2DEG*, *green curves*) with different undoped spacer thicknesses d_s, compared to homogeneously doped bulk GaAs (*red curves*) with different donor concentrations N_d. *Other lines* indicate limits by scattering at ionized donors (*ID*), optical phonons (*OP*), acoustic phonons (*AP*), and by piezoelectric scattering (*PE*) (After Morkoç 1985)

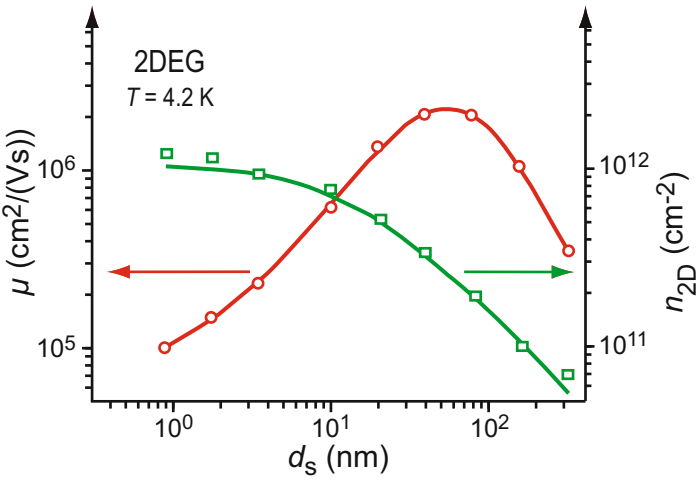


Fig. 5 Dependence of mobility μ and carrier density n_{2D} at low temperature in a 2D electron gas of a modulation-doped AlGaAs/GaAs heterostructure with various thicknesses d_s of the undoped AlGaAs spacer layer (After Harris et al. 1986)

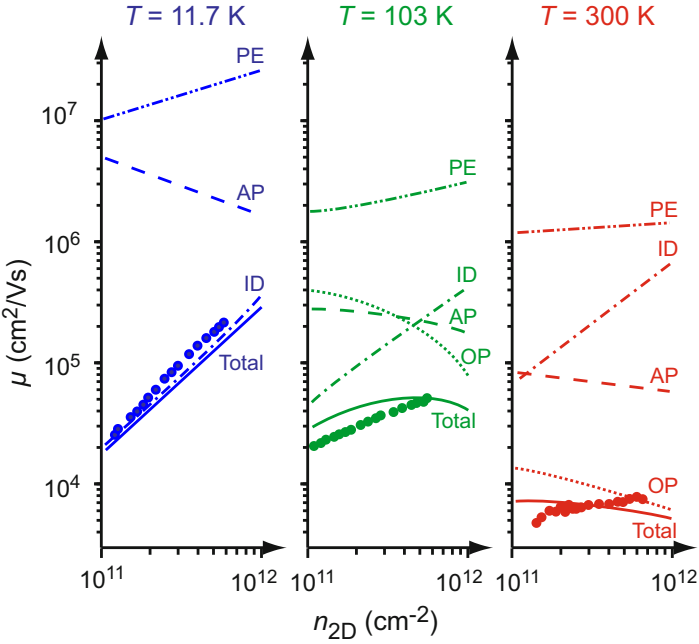


Fig. 6 Dependence of the mobility μ in a 2D electron gas on the carrier density n_{2D} at different temperatures. Calculated limits imposed by various scattering processes are indicated by *broken lines* with labels as in Fig. 4, *circles* are measured data points (After Hirakawa and Sakaki 1986)

modulation-doped AlGaAs/GaAs heterostructure with a 4.5 nm thick spacer layer. The increase of μ at larger carrier density observed at low temperature is less pronounced in the intermediate temperature range due to the contribution of acoustic-phonon scattering; at even higher temperature the polar-optical phonon dominates due to its increased exponential rise in T.

Limiting scattering processes of 2DEGs in hexagonal AlGaN/GaN structures are analyzed in Gurusinghe et al. (2005) and Ahmadi et al. (2016).

Hot-Electron Mobility The carrier mobility in a 2D electron gas considered above refers to low electric fields. The changes of scattering processes at high electric fields discussed for bulk semiconductors in ▶ Chap. 24, "Carrier Scattering at High Electric Fields" also affect the mobility in a 2DEG. The electron mobility at low temperatures decreases rapidly with increasing electric field, as shown in Fig. 7 for a modulation-doped AlGaAs/GaAs heterostructure. In general, the peak drift-velocities do not exceed the bulk velocities; typical values of $1 \ldots 3 \times 10^7$ cm/s are observed, the latter at 4 K (Inoue 1985).

An interesting effect was pointed out by Hess (1981), when with high electric fields carrier acceleration is large enough that electrons can be transferred above the barrier and diffuse into the wide-gap material, which contains most of the donors. Here, the electrons turn from a 2D into a 3D continuum, and the scattering increases substantially: a range of negative differential conductivity can be observed at high donor densities within the barrier layer.

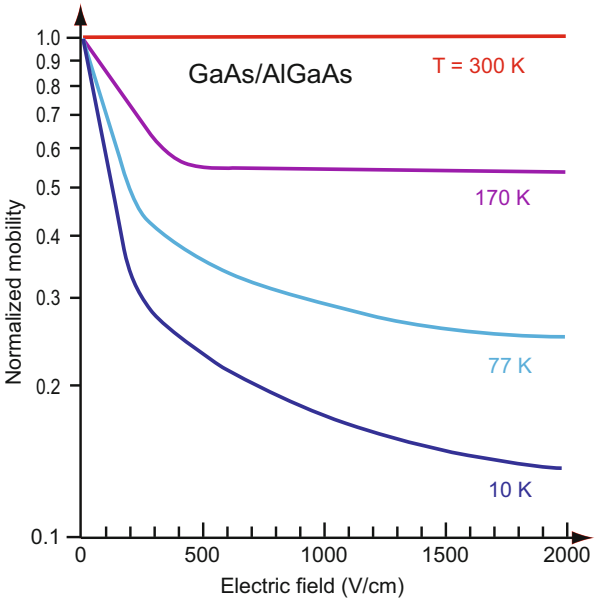


Fig. 7 Normalized electron mobility in a modulation-doped AlGaAs/GaAs structure as a function of the applied electric field for various temperatures (After Keever et al. 1982)

2 Perpendicular Carrier Transport in 2D Structures

2.1 Tunneling Through a Planar Barrier

In a classical consideration a carrier which is incident on an energy barrier with a height eV_0 is always reflected, if eV_0 exceeds the total energy of the carrier. However, if we consider the wavefunction of the carrier, there are also nonzero solutions inside and behind the barrier if the barrier height is finite. The square of the wavefunction behind the barrier represents the probability density for finding the carrier; this probability is hence finite, i.e., the carrier has – with some probability – penetrated the barrier despite a lack of energy. This phenomenon is referred to as a *tunneling process*. Such a tunneling is illustrated in Fig. 8.

The wavefunction of the carrier with mass m_n can in one dimension along z be described by

$$\begin{aligned} \varphi(z) &= A_0 e^{ikz} + A_1 e^{-ikz} \quad z \leq 0 \\ \varphi(z) &= A_2 e^{ikz} \qquad\qquad\quad z \geq d, \end{aligned} \tag{6}$$

where the carrier meets the barrier at $z = 0$; the wavevector is given by

$$k = \sqrt{2m_n E/\hbar^2}. \tag{7}$$

The amplitudes A_0 and A_1 represent the incident and reflected parts of the wavefunctions, and A_2 is the amplitude of the transmitted part behind the barrier with thickness d. Inside the barrier the wavefunction reads

$$\varphi(z) = A_3 e^{i\beta z} + A_4 e^{-i\beta z} \quad 0 < z < d, \tag{8}$$

with

$$\beta = \sqrt{2m_n (eV_0 - E)/\hbar^2}. \tag{9}$$

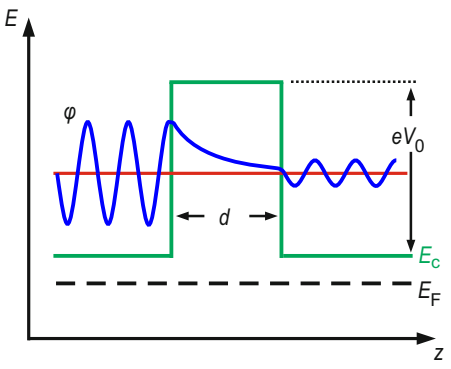


Fig. 8 Energy-band diagram of two semiconductors separated by a barrier of thickness *d*. *Green lines* represent conduction-band edges, *blue lines* illustrate the incident and transmitted part of the wavefunction of a carrier impinging on the barrier from the *left*

The amplitudes are determined by the boundary conditions: The wavefunction and its derivatives must be continuous at the interfaces. The probability of finding the carrier behind the barrier is then given by the transmission coefficient T:

$$T = \left(\frac{A_2}{A_0}\right)^2 = \left[1 + \frac{(eV_0 \sinh \beta d)^2}{4E(eV_0 - E)}\right]^{-1}. \tag{10}$$

For a small transmission probability $\beta d \ll 1$ applies and Eq. 10 can be approximated as

$$T \cong \exp(-2\beta d) = \exp\left(-2d\sqrt{2m_n(eV_0 - E)/\hbar^2}\right). \tag{11}$$

According to Eq. 11 a finite transmission coefficient T requires a small thickness d of the barrier, a low barrier height eV_0, and a small effective mass m_n. For a more detailed discussion of tunneling, see Sect. 2.3 of ▶ Chap. 29, "Carrier Generation".

2.2 Tunneling Through a Double-Barrier Quantum Well

Adding a second barrier with a small distance d_{QW} to the first creates an interesting device: a *resonant tunneling diode*. It may be readily fabricated by a GaAs quantum well (QW) layer sandwiched by two thin AlGaAs barrier layers and further embedded in degenerately n-type doped GaAs. The band diagram of such a double-barrier QW structure is shown in Fig. 9.

Discrete energy levels exist between the two barriers if their distance d_{QW} is small ($d_{QW} \lesssim 10$ nm). Tunneling across the barriers is possible for sufficiently thin barrier

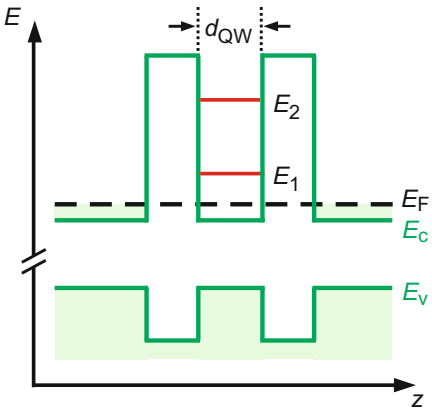


Fig. 9 Energy diagram of a double-barrier resonant tunneling device at zero bias. *Green shading* indicates electron occupation in the bands, *red lines* indicate electron states in the quantum well; respective hole states are not shown

layers (on the order of some nm). However, the transmission coefficient is only large if the energy of an incident carrier equals one of the energy levels in the well.

We focus on the conduction-band edge and on electrons.[3] If a bias is applied across the device, a voltage drop occurs in the barriers and the well, but not in the highly conductive degenerate semiconductors beyond the barriers. The discrete energies in the QW are hence shifted with respect to the conduction bands outside the barriers, such that the energy of incidents electrons can be matched to the energy in the well for an appropriate bias. At this resonance electrons near the Fermi level tunnel through the first barrier into the discrete state of the quantum well, and subsequently through the second barrier into unoccupied states beyond. If the bias is further increased, the conduction-band edge shifts above the energy of the QW and the tunnel current decreases.

The energy diagrams for zero bias, resonant condition and further increased bias are illustrated in Fig. 10, along with the resulting tunneling current of an AlAs/GaAs mesa device with 6 μm diameter. We observe a strong increase of the current up to a peak occurring when the bias establishes resonant conditions at $V_{\text{bias}} = 2E_1/e$ as depicted in Fig. 10b, followed by a decrease down to a minimum at I_{valley}. The range between I_{peak} and I_{valley} has a negative differential resistance (NDR): $dI/dV < 0$. This NDR can, e.g., be applied to overcompensate the ohmic resistance yielding a

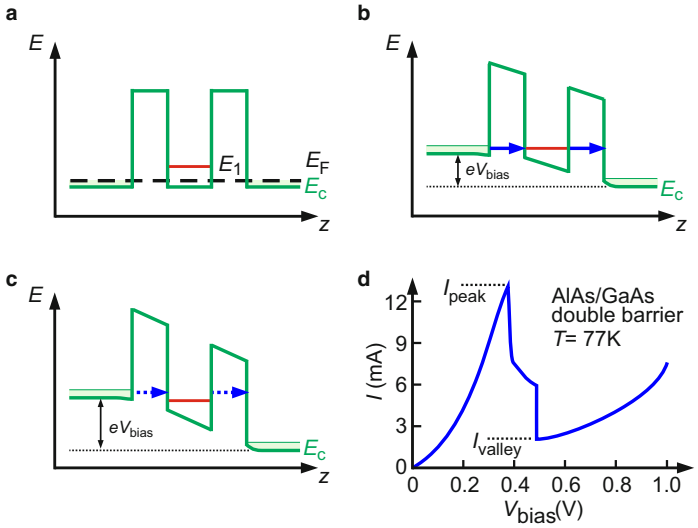


Fig. 10 Energy diagram of the double-barrier device for different bias conditions: (**a**) zero bias, (**b**) resonant bias at $V = 2E_1/e$, and (**c**) $V > E_1/e$. (**d**) Measured current–voltage characteristic of a device with two 2.5 nm thick AlAs barriers cladding a 5 nm thick GaAs quantum well (After Morkoç et al. 1986)

[3]For resonant tunneling of holes, see, e.g., Hayden et al. (1991).

simple oscillator; for applications see Capasso et al. (1990). The peak-to-valley ratio I_{peak}/I_{valley} is relevant for device applications; for the device of Fig. 10d it is 9:1 at $T = 77$ K and 3:1 at room temperature (Morkoç et al. 1986). Its magnitude is limited by the spreading of the distribution function around the resonance energy, scattering of the tunneling electrons by phonons, scattering at defects, and interface roughness. Moreover, thermally assisted tunneling through higher confined levels and thermionic emission over the barriers occur at higher temperatures; such processes reduce the peak-to-valley ratio and should be minimized for device applications.

The wavefunction of the carriers may either maintain quantum-mechanical coherence or, in case of scattering, loose coherence in the tunneling process. These cases are referred to as coherent or sequential tunneling and are discussed in the following section.

2.3 Perpendicular Transport in Superlattices

An equally spaced stacking of resonant-tunneling structures like that described above yields a superlattice, which allows for improved current-peak magnitudes. If the wells are placed sufficiently close together, the wavefunctions in the wells overlap and narrow minibands are formed from the otherwise degenerate energy levels of different wells, see Sect. 3.1.2 of ▶ Chap. 8, "Bands and Bandgaps in Solids". Transport of carriers through the superlattice occurs in these minibands, where they have a large effective mass. The carriers also have to tunnel through the barriers. The effective carrier mobility is therefore rather small.

The minibands are shifted with respect to each other with increasing applied total bias. Consequently, the current will vary according to the density-of-states product in adjacent wells until the minibands in adjacent wells no longer join. Similar to the effect in a double-barrier QW discussed above, a sharp reduction in the current (a negative differential resistivity regime) results, as shown in Fig. 11. When, with further increased bias, a match with the next higher miniband is reached (Fig. 11a), the current increases again until these bands are shifted away from each other and a second range of negative differential resistance appears as illustrated in Fig. 11b, and so on.

The sharpness of the current maxima and their positions are a direct measure of miniband width and energy and are in reasonable agreement with theoretical estimation using a Kronig-Penney potential and tunneling. With broader minibands, occurring in superlattices with narrow barrier layers, one estimates the electronic conduction at very low fields similarly to the classical Drude theory. Because of the relatively small band width, however, the assumption of a constant effective mass is no longer justified.

Coherent and Sequential Tunneling At low bias the tunneling from well to well can proceed within the same miniband over several barriers without losing its phase relation, if there is negligible scattering. This process is referred to as *coherent* or *resonance tunneling*. In coherent resonant tunneling, the effect of negative differential resistance is accompanied by a coherent enhancement of transmission, analogous to the transmission of an optical wave in a Fabry-Pérot filter (Choi et al. 1987a).

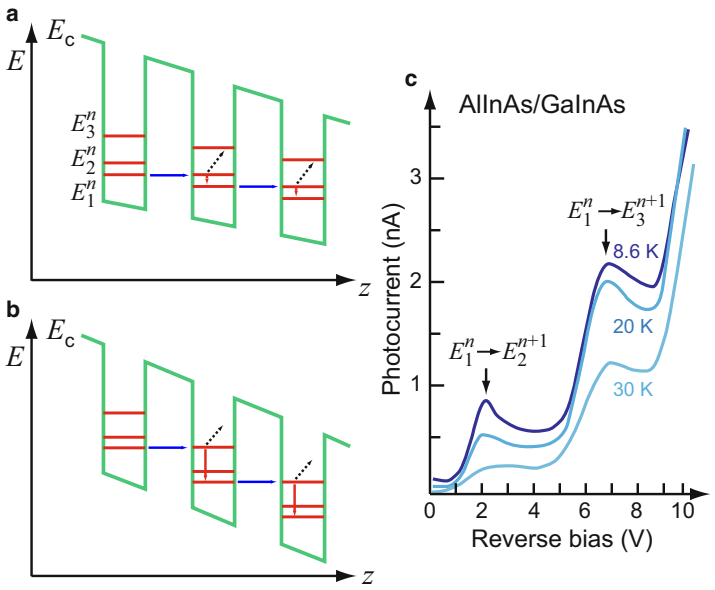


Fig. 11 (**a, b**) Tunneling between quantum wells of an $Al_{0.48}In_{0.52}As/Ga_{0.47}In_{0.53}As$ superlattice including phonon emission and (**c**) corresponding current–voltage characteristics of a 35 period $Al_{0.48}In_{0.52}As/Ga_{0.47}In_{0.53}As$ superlattice with 13.8 nm thick wells and barriers; uppercase indices designate QW numbers (After Capasso et al. 1990)

At increased bias tunneling occurs into an excited state of the subsequent well as illustrated in Fig. 11a, b. A carrier relaxation into the ground state is then required before tunneling to the next well can proceed. Such an inelastic scattering is accompanied by radiative or phonon emission; phonon emission may act as the bottleneck for carrier transport through the superlattice when the gap between the minibands is larger than the energy of an optical phonon $\hbar\omega_{LO}$ (Capasso et al. 1986; Capasso 1987). The coherence of the electron wave is lost in this process, and the tunneling to the next well is described as *sequential tunneling*. Since wavefunction coherence is usually not required, most work focuses on the sequential tunneling. Coherence proved important, however, in quantum-cascade lasers (Sect. 2.3.2); see Callebaut and Hu (2005).

Tunneling Through Wells with Variable Width A special case of resonance tunneling through a superlattice can be achieved when the thickness of consecutive wells decreases in such a way that each of the resulting first levels lines up across the entire superlattice at a certain bias, as shown in Fig. 12. Here, the current through the superlattice is much increased, while it is very small at lower or higher bias values (Brennan and Summers 1987a). The energy of electrons in thinner wells can exceed the impact-ionization energy, rendering the behavior of such structures similar to that of a photomultiplier (Brennan and Summers 1987b).

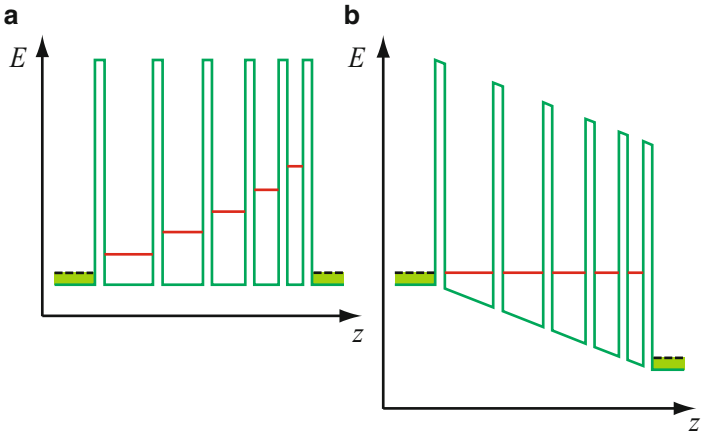


Fig. 12 (**a**) Superlattice with decreasing well width in thermal equilibrium. (**b**) Superlattice with states aligned for maximum tunneling current at a certain applied bias

2.3.1 Wannier-Stark Ladder and Bloch Oscillations

An electric field F_z applied along the growth direction z of the superlattice tilts the band edges; if the related energy shift between adjacent wells is on the order of or larger than the zero-field width of the minibands, the degeneracy of the energies in the individual wells is lifted, see Fig. 13. The formation of minibands is hence largely inhibited, and the wavefunctions along z get localized: they extend over only a few adjacent wells.[4] The discrete energies around a well with eigenenergy E_0 are then given by

$$E_i = E_0 + i e F_z L, \quad i = 0, \quad \pm 1, \quad \pm 2, \quad \ldots, \tag{12}$$

where L is the superlattice period (Mendez et al. 1988). The equally spaced series of energies with differences eF_zL according to Eq. 12 is referred to as *Wannier–Stark ladder*. The transitions indicated in Fig. 13a are measured in optical spectra (Voisin et al. 1988; Leisching et al. 1994).

An additional effect occurs due to the motion of an electron through the periodic potential of the superlattice: the *Bloch oscillations*. The electric field F_z leads to an increase of the carrier momentum k_z. If we neglect scattering the carrier gains energy, its effective mass increases, and eventually changes sign. The carrier velocity hence decreases near the edge of the Brillouin zone, and at further increased momentum the edge of the Brillouin zone with zero group velocity is reached at $k_z = \pi/L$ (L = superlattice period). Bragg reflection then transfers the momentum to $k_z = -\pi/L$, where the carrier is accelerated again. As a consequence, the carrier velocity and thus the macroscopically observed current through the superlattice are expected to oscillate with a *Bloch frequency* ω_B given by

[4]Heavy holes in the valence band get localized much stronger due to their larger effective mass.

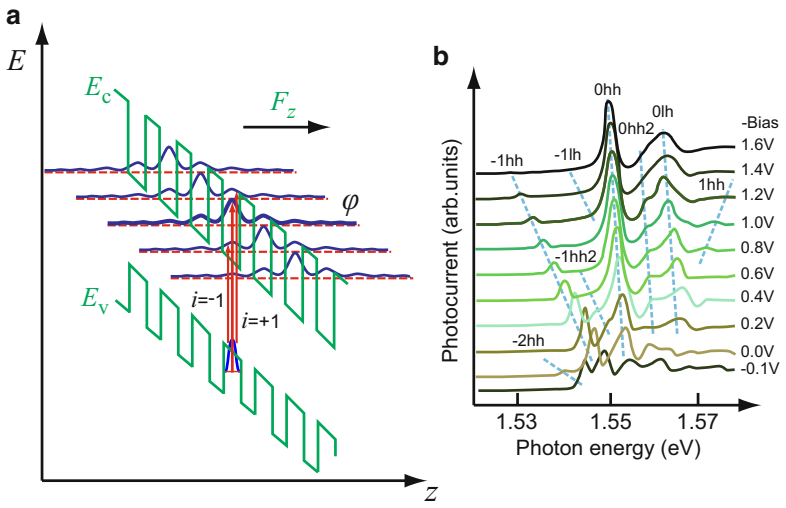


Fig. 13 (**a**) Band diagram of a Wannier–Stark ladder in a superlattice in an electric field F_z along the growth direction. *Vertical arrows* indicate optical transitions between states localized in the wells. (**b**) Photocurrent excitation spectrum of a GaAs/AlGaAs superlattice embedded in a *pin* junction, measured at varied bias; *dashed lines* mark transitions between localized heavy-hole (*hh*) or light-hole (*lh*) states and electron states of a Wannier–Stark ladder (After Leisching et al. 1994)

$$\omega_B = eF_zL/\hbar. \tag{13}$$

This frequency lies in the terahertz (THz) range; Bloch oscillations in superlattices are hence a means to fabricate tunable emitter devices for coherent radiation in the middle to far IR range (Lyssenko and Leo 2011).

Bloch oscillations should in principle already occur in the periodic potential of a crystal (i.e., for L in Eq. 13 replaced by the lattice constant a), but in crystals the oscillation period, which is inverse to the periodicity of the potential, is much longer than the mean scattering time. In a superlattice the period L is substantially larger than the lattice constant of a crystal, and the conditions are more favorable to observe these oscillations (Esaki and Tsu 1970); see also Sect. 2.1 of ▶ Chap. 22, "Carrier-Transport Equations".

2.3.2 Quantum-Cascade Laser and High-Field Domains

The principles of size quantization and tunneling outlined above are exploited in the quantum-cascade laser (QCL). While in conventional semiconductor lasers light is generated by radiative recombination of electrons in the conduction band with holes in the valence band, light in a QCL is generated by transitions between subbands of the conduction band (Faist et al. 1994). The QCL is hence a unipolar device operating solely with electrons. The emission wavelength of a QCL can be tailored from the mid-infrared to the submillimeter-wave region without changing the semiconductor material – just by changing the thickness of the quantum wells.

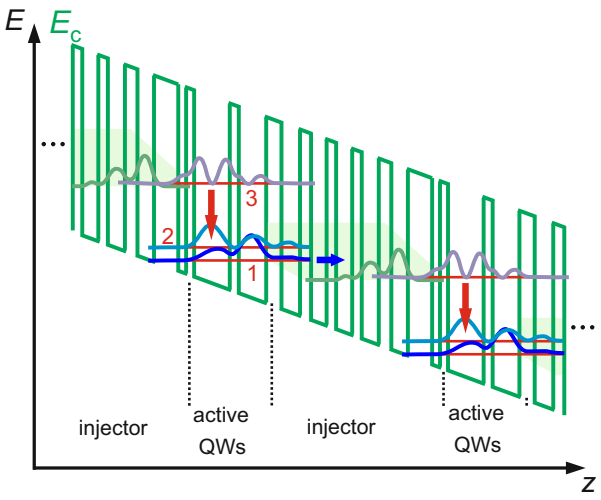


Fig. 14 Conduction-band edge E_c of a quantum cascade laser (*green lines*) with alternating injector and QW sections. *Red horizontal lines* are energy states of the coupled QWs, square of the wavefunctions in the QWs and the injectors are drawn *bluish and green*, respectively, *green shadings* in the injector sections indicate miniband regions. *Red arrows* signify lasing transitions

The quantum-cascade laser consists of many layers with different bandgap energies, divided into injector and active QW sections, see Fig. 14. The transfer rate between two subbands depends upon the overlap of their wavefunctions and their energy spacing. Layer thicknesses are tailored such that electrons are efficiently injected from an injector section into an excited state of a QW (Köhler et al. 2002). QC lasers are typically based on a three-level system; as in any such laser a long lifetime of the upper laser level and an efficient depletion of the lower laser level are needed to obtain population inversion.[5] In the example of Fig. 14 the energy match of E_3 in the two coupled QWs and the miniband in the injector provides a strong injection, and the node of φ_3 placed at the peaks of φ_2 and φ_1 provides a longer lifetime; the absolute maximum of φ_1 in the right QW and its overlap with φ_2 in the left QW produces a fast depopulation[6] of E_2 and electron transfer to the (energy-matched) subsequent injector section after the lasing transition. A QCL may have over 50 active regions and injector sections; thereby each electron generates that many photons when it traverses the structure.

There are various scattering processes operating in the carrier relaxation. Generally intrasubband and intrawell relaxations are faster than intersubband and interwell relaxations (Ferreira and Bastard 1989). The thermalization of the electron gas is provided by highly efficient intrasubband LO-phonon and electron–electron scattering. Intersubband relaxation is dominated by electron–LO-phonon scattering for subband separations exceeding the LO phonon energy (Freeman 2020), but even

[5]A nonthermal electron distribution in the active region makes laser action possible even in the absence of a global population inversion, see Faist et al. (1996).

[6]The transfer from E_2 to E_1 gets particularly efficient if their difference equals the energy of an LO phonon, thereby inducing resonant electron-LO-phonon scattering.

at smaller energy spacings relaxation is fast due to carrier-carrier scattering (Hartig et al. 1998). Besides electron–electron scattering intersubband transport is controlled by scattering at ionized impurities.

A QCL is a complex structure comprising a large number of layers; its fabrication requires an extraordinary precision of growth with respect to layer thickness and homogeneity. QCLs were realized with many materials such as, e.g., InGaAs/InAlAs on InP substrate for the mid-IR range, GaAs/AlGaAs on GaAs for the terahertz region, and InGaAs/AlAsSb on InP for short-wavelengths in the μm range; the short- wavelength limit is determined by the conduction-band offset between wells and barriers. A survey of fabricated quantum-cascade lasers and their temperature performance is given in Fig. 15. The achievements in the development of devices are impressive; continuous-wave operation is produced at room temperature with watt-level output power and power conversion (wall-plug) efficiencies above 20% (over 50% at 77 K). For reviews see Capasso (2010), Faist (2013), Vitiello et al. (2015), and Vitiello and Tredicucci (2021).

QC lasers find numerous applications. The high atmospheric transmission in the 3–5 and 8–12 μm wavelength ranges allows for long-distance remote chemical sensing using laser-absorption spectroscopy, since most chemicals have distinctive absorption lines in the 3–16 μm wavelength range; further applications include optical free-space communication, laser radar, and medical imaging.

Without application of a bias a QCL can also be operated as a photovoltaic detector. The performance can, however, be substantially improved by specifically designing the quantum-cascade structure for detection, because the dependence of

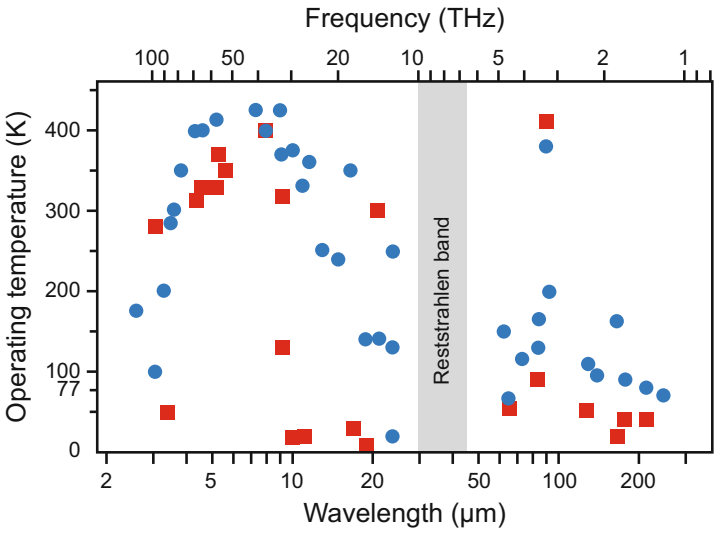


Fig. 15 Emission wavelength of various quantum-cascade lasers, showing room-temperature operation in the mid-IR range and low-temperature operation in the THz range. *Squares* and *circles* denote continuous-wave and pulsed operation, respectively (After Vitiello et al. 2015)

energy levels and wavefunctions on the bias is taken into account in the laser design. An advantage of quantum-cascade detectors over well-established mid-IR detector types is their designable operation wavelength, low-noise photovoltaic operation mode, and the ability of room-temperature operation. For more information see Giorgetta et al. (2009) and Hillbrand et al. (2021).

High-Field Domains In a superlattice with weakly coupled quantum wells a high-field domain may appear at high bias. Such stable stationary domains lead to a characteristic sawtooth pattern in the current–voltage characteristics[7] as shown in Fig. 16a (Esaki and Chang 1974; Kwok et al. 1995). They were also found in quantum-cascade lasers (Lu et al. 2006). The domain appears if carriers cannot tunnel resonantly through the entire superlattice due to a misalignment of the QWs by the applied bias; the resistance then increases at one of the QWs, where most of the applied bias drops at his leading barrier. Now tunneling may occur from the ground state E_1 to an excited state E_2 with a subsequent relaxation by phonon emission as illustrated in Fig. 16b. The voltage drop at the QW is then $V \cong (E_2 - E_1)/e$. A subsequent bias increase will appear across this domain until the ground level rises to level E_2 of the next well. The first domain forms at the anode and then extends one after the other towards the cathode (Choi et al. 1987b). Consequently 48 negative

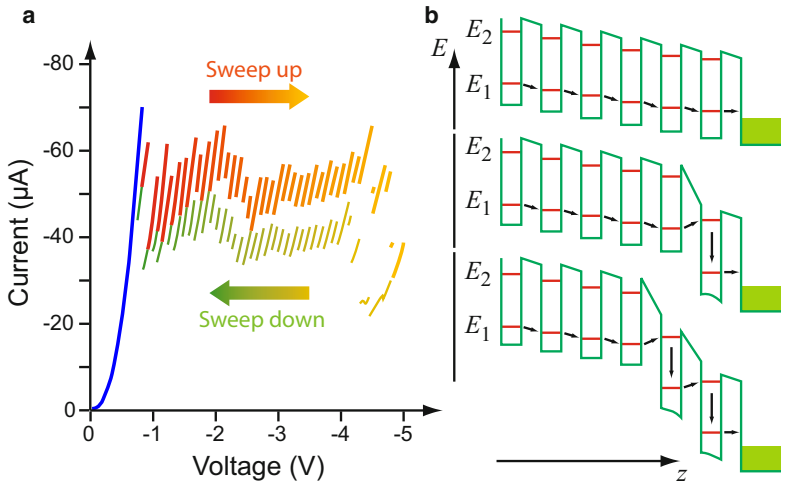


Fig. 16 (**a**) Measured current–voltage characteristic in a superlattice with 40 periods of 9 nm GaAs wells and 4 nm AlAs barriers. The sweep-up was measured from 0 to −5 V and the sweep-down vice versa (After Kastrup et al. 1994). (**b**) Band diagram of the superlattice for several values of the average potential drop per period

[7]The current is usually measured by applying a slow continuous sweep of the bias. As the branches in the *I-V* characteristics overlap, different parts of the branches are observed for sweep-up and sweep-down.

conductance peaks were observed in a superlattice with 49 periods. From optical measurements an energy difference in reasonable agreement with the voltage drop per superlattice period measured in transport was determined.

The instability of a homogeneous field distribution in the superlattice also gives rise to a self-sustained oscillatory behavior at fixed bias in the range of negative differential conduction. The oscillations originate from the formation, motion, and recycling of the domain boundary inside the superlattice (Kastrup et al. 1997) and allow for tunable generation of frequencies over 100 GHz (Schomburg et al. 1999). Also bistability between stationary and oscillatory behavior has been observed at fixed bias within the first plateau of the current–voltage characteristics (Zhang et al. 1997); for a review see Wacker (2002).

3 Transport in 1D Structures

Fabrication of 1D and 0D structures requires confining barriers in addition to those obtained by the epitaxy of 2D layers discussed in the previous sections. Most progress for transport structures was achieved by imposing such a confinement laterally on two-dimensional electron gases[8] (2DEGs) using lithography techniques: the definition of negatively biased gates depletes the underlying 2DEG to the region between the gates. This approach creates smooth lateral confinements with designed shapes and tuning ability. Compared to the confinement of a heterostructure the potential is quite shallow, yielding subband splittings in the range of only some meV in the transverse direction. On the other hand, the straightforward approach of creating high lateral barriers by etching a 2D system is often hampered by the high defect density introduced at the sidewall interfaces for the required small scale; these defects may give rise to pinning of the Fermi level and a high density of centers for nonradiative recombination. Growth of nanowires (Sect. 2.2.2 of ▶ Chap. 3, "The Structure of Semiconductors") is a further technique for fabricating quantum wires with high lateral barriers. For a brief survey on fabrication techniques see Thornton (1994), Ahmed and Nakazato (1996).

Split-Gate Lateral Potential A widely used method for fabricating a one-dimensional channel, referred to as *quantum wire* (QWR), is the split-gate technique using two parallel gate electrodes with a narrow channel in between on top of a 2DEG. At negative bias, the 2DEG is depleted underneath these electrodes due to the electrostatic potential, leaving an undepleted 1D channel between the electrodes.

The negative bias applied to the split gate creates an approximately parabolic potential, which laterally confines the 2DEG to a small region as illustrated in Fig. 17a. When charge accumulates in the 2DEG for less negative gate bias, the potential approximates a truncated parabola with a flat bottom (Laux et al. 1988). The calculated

[8]We focus again on *electron* gases; one-dimensional transport structures made using 2DHGs (two-dimensional hole gases) were studied as well, see, e.g., Danneau et al. (2006).

potential of a GaAs/$Al_{0.26}Ga_{0.74}As$ 2DEG shown in Fig. 17b for a 400 nm wide slit between the gates yields nearly equidistant 1D subbands with ~5 meV spacing. As the effective channel width increases at less negative bias, the level spacing decreases.

The density n_{2D} of the electrons in the channel and the effective width W of the channel depend on the applied bias voltage V_g, see Fig. 17b. Both quantities were found to decrease linearly with V_g as illustrated for a GaAs/AlGaAs structure in Fig. 18; the magnitude of the effect depends on the device geometry. An independent control of n_{2D} and W can be achieved by placing an additional top gate on top of the slit between the split-gate electrodes, vertically separated by an insulating layer (Tkachenko et al. 2001); usually, however, the more simple assembly illustrated in Fig. 17a is used.

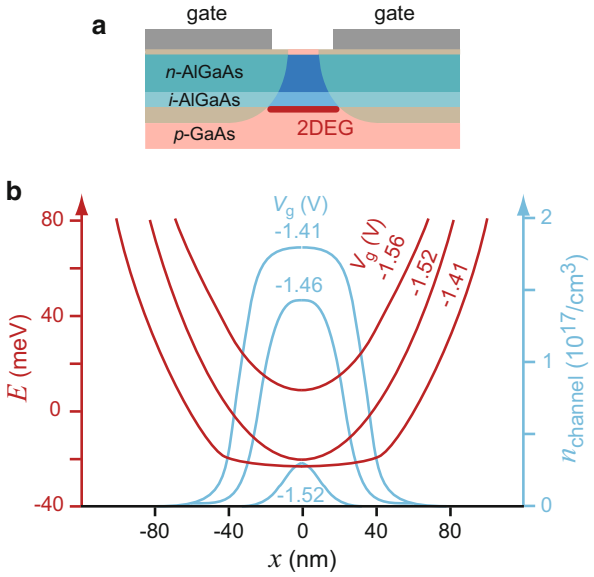


Fig. 17 (**a**) Cross-sectional schematic of split-gate electrodes on top of a two-dimensional electron gas producing a one-dimensional undepleted channel, when a negative potential is applied. (**b**) Calculated potential of a split-gate GaAs/AlGaAs structure for various gate voltages V_g (After Laux et al. 1988)

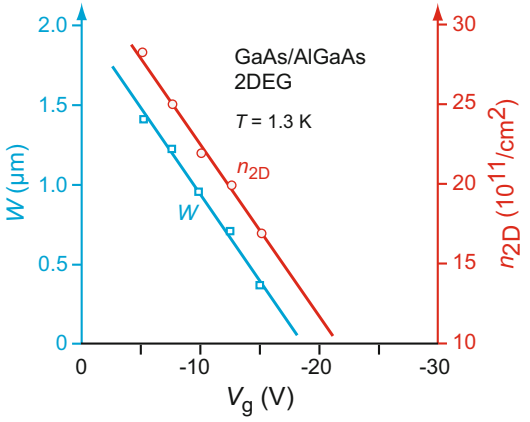


Fig. 18 Electron density n_{2D} and effective width W of a 300 μm long 1D channel imposed on a two-dimensional electron gas (*2DEG*) for varied negative split-gate voltage V_g (After Zheng et al. 1986)

3.1 Diffusive and Ballistic Transport

In the discussion of conductivity in previous chapters we used a classical description of the electron dynamics (▶ Chap. 22, "Carrier-Transport Equations"), including the quantum-mechanical Fermi-Dirac statistics (▶ Chap. 21, "Equilibrium Statistics of Carriers"). For a conductor we derived Ohm's law

$$j = \sigma \; F, \tag{14}$$

which relates the local current density j to the applied electric field F. σ is the conductivity, which, for a 3D conductor of width W, thickness W, and length L, is related to the conductance G by

$$G = \left(W^2/L\right) \; \sigma. \tag{15}$$

In two dimensions Eq. 15 reads

$$G = (W/L) \; \sigma, \tag{16}$$

and G and σ have the same unit. Equations 14, 15, and 16 hold for large homogeneous conductors, where the current is given by a diffusion process characterized by many scattering events as illustrated in Fig. 19a; the mean free path λ of a carrier is much smaller than the sample dimensions W and L. This regime is referred to as *diffusive transport*. Diffusive scattering increases the resistivity $\rho = \sigma^{-1}$.

In a quantum wire, the sample dimensions W and L get smaller than the mean free path λ. This is the *ballistic transport* regime depicted in Fig. 19b. There may occur scattering at boundaries, but – for smooth confinement potentials – such scattering is found to be predominantly specular with a probability close to unity (Thornton et al. 1989). In contrast to diffusive scattering, specular scattering at a boundary with a roughness on a length scale smaller than the Fermi wavelength does not increase the resistivity. In this case the electron motion along the wire is not affected by the lateral confinement; a quantum wire defined by a 2DEG confined along the z direction and by an additional lateral smooth confinement then maintains the resistivity value of the 2DEG. The resistivity refers in the classical description to the mean time τ

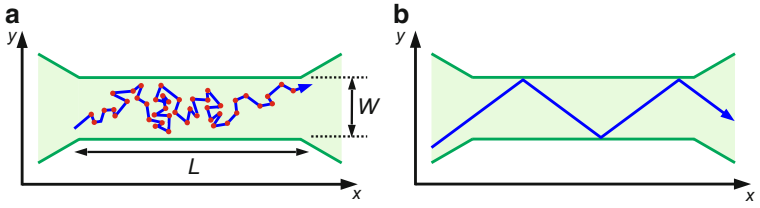


Fig. 19 Trajectories of carriers in (**a**) the diffusive and (**b**) the ballistic transport regime in a conductive wire. *Red circles* indicate scattering at defects

between two scattering events[9] and yields $\rho = m_n/(e^2 n_{2D}\tau(T))$. At zero temperature a residual resisitivity remains due to elastic scattering at defects.

Between the ballistic and the diffusive regimes there is an intermediate range referred to as *quasi-ballistic*. It is characterized by a mean free path λ larger than the wire width W but smaller than the wire length L. For a review on the transport regimes and scattering mechanisms see Beenakker and van Houten (1991).

3.2 Quantization of Conductance

Evidence for ballistic transport is found in the conductance of a quantum wire fabricated by the split-gate technique described above. At sufficiently high negative bias the depletion regions of both gates overlap and the 2DEG gets completely depleted; the conductance G hence drops to zero. As the negative gate voltage is reduced the number of occupied channels below the Fermi level increases. In samples with negligible scattering the conduction then increases in steps as shown in Fig. 20 (Wharam et al. 1988; van Wees et al. 1988). The step height[10] is given by

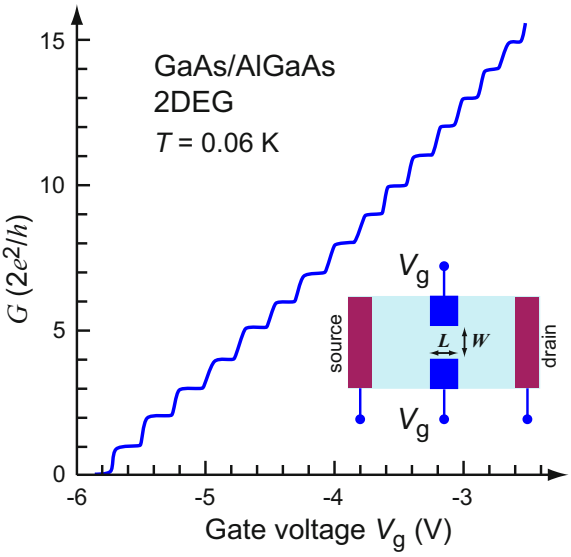


Fig. 20 Differential conductance $G(V_g)$ of a quantum wire after correction for a series resistance of 703Ω. *Inset*: Plan view schematic of the split-gate device (After Thomas et al. 1998)

[9]In the classical description the phase correlation of the carrier wavefunctions before and after scattering is destroyed. In the description of quantum diffusion the phase correlation is limited by *inelastic* scattering events only; the corresponding mean scattering time τ_ϕ is at low temperatures significantly larger than τ.

[10]The conduction step $2e^2/h$ corresponds to a resistance of 12907 Ω.

$$\Delta G = \frac{2e^2}{h}; \tag{17}$$

The quantization of the conductance G in integer multiples of $2e^2/h$ originates from ballistic transport in the 1D quantum wire (Thornton et al. 1986). The factor 2 is a consequence of the spin degeneracy of the transported electrons. If the degeneracy is lifted, the electrons become spin-polarized and an additional conductance plateau is observed at $0.5(2e^2/h)$; this occurs in a large external magnetic field.[11]

The importance of low scattering is illustrated in the simulation shown in Fig. 21. Ionized donors are assumed to be distributed at random in a δ-doping layer separated by a 42 nm spacer from the 2DEG with $n_{2D} = 2.5 \times 10^{12}$ cm^{-2}, and negatively biased split-gate contacts define a 300 nm wide and 600 nm long channel (Nixon et al. 1991). For a uniform (averaged) donor potential the calculated conductance shows pronounced quantized steps, while the accuracy of quantization is reduced by scattering from the random potential of the donors. The quantization eventually disappears at sufficiently strong disruption of ballistic transport, yielding diffusive conduction.

The occurrence of the conductance steps may be understood within the simple model for noninteracting electrons. We consider a 2DEG created by a heterostructure

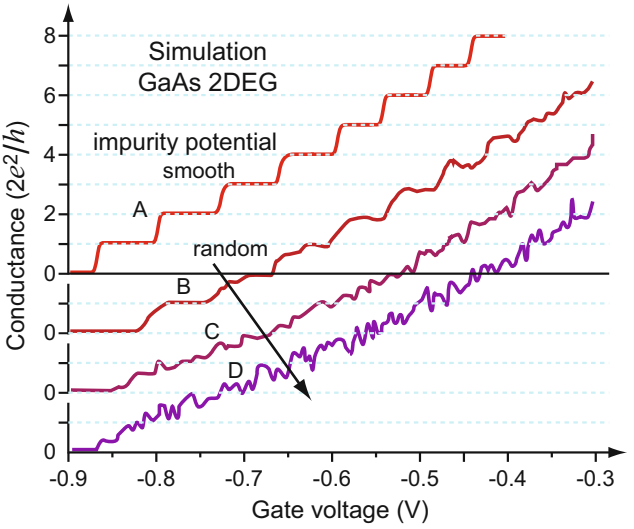


Fig. 21 Calculated conductance for a 600 nm long channel with a smooth potential (*curve A*) and potential fluctuations from a random distribution of ionized donors (*curves B to D*) (After Nixon et al. 1991)

[11]The degeneracy may already be lifted at zero magnetic field if there is a spontaneous spin polarization due to electron interactions (Chen et al. 2008), see the section on the "*0.7 conduction anomaly*" below.

with a potential along z like that illustrated in Fig. 2 with only the lowest state $E_{n=1}$ occupied. An additional parabolic potential along y like that shown in Fig. 17 creates the 1D quantum wire (QWR) along the x direction. An electron with momentum $\hbar k_{\|}$ moving along this x direction behaves like a free electron described by a plane wave with kinetic energy $\hbar^2 k_{\|}^2/(2m_n)$. Perpendicular to the QWR along y the energy is quantized into levels E_l of subbands. The single occupied 2D subband of the 2DEG with quantum number $n = 1$ hence splits into a number of higher lying 1D subbands specified by their quantum number l, see also Sect. 3.3 of ▶ Chap. 8, "Bands and Bandgaps in Solids"; since the confinement along y is usually quite shallow, the energy spacing of the 1D subbands is on the order of only a few meV, and several 1D subbands may easily be occupied. The total energy of the electron above the conduction-band edge E_c is

$$E_{n,l} = E_n + E_l + \hbar^2 k_{\|}^2/(2m_n), \quad n, l = 1,\ 2,\ 3,\ \ldots, \tag{18}$$

where E_l for a parabolic potential is given by the energies of a harmonic oscillator $E_l = (l + \frac{1}{2})\ \hbar\omega$. The total wavefunction of the electron is described by

$$\psi_{n,l}(\mathbf{r}) \propto \varphi_n(z)\ \phi_l(y)\ \exp\left(i k_{\|} x\right). \tag{19}$$

The electron wave in the QWR is analogous to an electromagnetic wave in a waveguide; the subbands are hence also referred to as *modes* and the QWR as a quantum-wire *waveguide*.

If a voltage V_{sd} is applied between source and drain, i.e., between the two ends of the QWR, with a higher potential at the source contact (see inset of Fig. 20), a net flow of electrons moves from source to drain. The velocity of an electron in a particular subband state is $v_x = \hbar k_{\|}/m_n$. The total current I is obtained by integrating ev_x over the k states in a small energy range of $\pm eV_{sd}/2$ around the Fermi energy E_F,

$$I = \int_{E_F - eV_{sd}/2}^{E_F + eV_{sd}/2} e\, v_x\, g_{1D}\ dE, \tag{20}$$

where the one-dimensional density of states g_{1D} in the quantum wire equals $1/(hv_x)$. The current is hence

$$I = \frac{e}{h} \int_{E_F - eV_{sd}/2}^{E_F + eV_{sd}/2} dE = \frac{e^2}{h} V_{sd}, \tag{21}$$

yielding a conductance $G = I/V_{sd} = e^2/h$ per spin direction. If we account for both spin directions and N occupied subbands, we obtain for the total conductance

$$G = N\frac{2e^2}{h}, \quad N = 0,\ 1,\ 2,\ .... \tag{22}$$

The 0.7 Conduction Anomaly A close inspection of the first conduction step shows a feature near $0.7 \times (2e^2/h)$, i.e., *below* the first plateau at $2e^2/h$ (Thomas et al. 1998). The structure is observed at zero magnetic field and may vary in appearance from a resonance to a plateau; it is called the 0.7 structure. If an increasing in-plane magnetic field is applied, the conductance value of the 0.7 structure smoothly approaches the value $0.5 \times (2e^2/h)$; e^2/h represents the normal spin-split value of the conductance when the spin degeneracy is lifted by a magnetic field. The continuous transition from zero field to this value indicates an origin from the electron spin. The 0.7 structure appears at a temperature below 5 K, but it disappears below 1 K by merging with the $2e^2/h$ plateau. The temperature behavior shows an Arrhenius dependence, indicating an origin at the spin-degenerate $2e^2/h$ state (Kristensen et al. 2000). Although the nature of this structure is not yet fully understood, an assignment to spin polarization induced by electron interactions is generally agreed. For a comprehensive review see Micolich (2011), for a brief survey Das and Green (2019).

3.3 Landauer-Büttiker Formalism

The classical picture of conduction based on electrons being localized in space was generalized by Landauer (1957) and Büttiker (1986), assuming a nonlocal character of the current flow by viewing conduction as a transmission problem. Such a description includes contacts on the same level as the mesoscopic active part of the device and is also appropriate for describing the conductance of multiple-terminal devices.

We consider the multiterminal structure depicted in Fig. 22 with an active region connected with leads to electron reservoirs, which are individually fixed at some chemical potential $E_{F,i}$ (here $i = 1 \ldots 5$). Analogous to Eq. 20 an electron fed from any reservoir i to the active region contributes to a current by

$$dI_i = e\, v_i\, g\ \left(E_{F,i} - E_{F,0}\right), \tag{23}$$

where $E_{F,0}$ is an arbitrarily chosen reference energy (e.g., the lowest potential of all reservoirs). For a one-dimensional wire the total current (per spin direction) in lead i is $(e/h) \times \left(E_{F,i} - E_{F,0}\right)$, see Eq. 21. The actual current injected from reservoir i is reduced by a fraction reflected back into reservoir i, expressed by a reflection coefficient

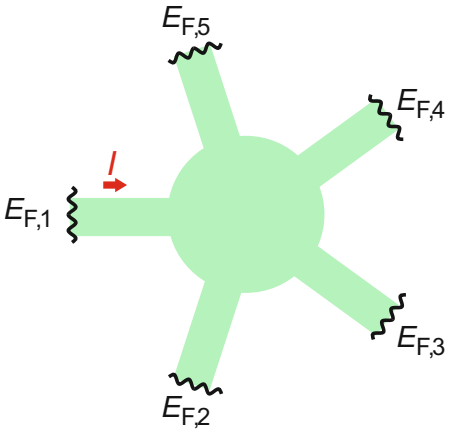


Fig. 22 Schematic of a five-terminal structure with all contacts kept at individually fixed Fermi energies

R_{ii}, and a part transmitted into reservoir i from all the other reservoirs j, analogously expressed by transmission coefficients T_{ij}. This yields a current per spin direction

$$I_i = \frac{e}{h}\left[(1 - R_{ii})E_{\mathrm{F},i} - \sum_{j \neq i} T_{ij} E_{\mathrm{F},j}\right]. \tag{24}$$

When the current in lead i is carried by N occupied subbands, the transmission in lead i is called multimoded, and the term $(1 - R_{ii})$ in Eq. 24 is replaced by $(N_i - R_{ii})$, yielding

$$I_i = \frac{e}{h}\left[(N_i - R_{ii})E_{\mathrm{F},i} - \sum_{j \neq i} T_{ij} E_{\mathrm{F},j}\right]. \tag{25}$$

The reflection and transmission coefficients then comprise all electrons in all modes; since reflection may also occur from an electron in mode n to a mode m and an electron may change the mode when transmitted from reservoir i to reservoir j, we obtain for the coefficients

$$R_{ii} = \sum_{m,n} R_{ii,mn} \quad \text{and} \quad T_{ij} = \sum_{m,n} T_{ij,mn}. \tag{26}$$

It is instructive to apply the general Landauer-Büttiker formalism to a 1D (two-terminal) conductor like those depicted in Fig. 19. The electrons in all modes N_1 fed from wire 1 into the conductor are either reflected back into wire 1 or transmitted into wire 2, yielding by conservation of the carrier flux $N_1 = R_{11} + T_{12}$

in terminal 1 and $N_2 = R_{22} + T_{21}$ in terminal 2; the transmission coefficients are $T_{21} = T_{12} \equiv T$. Inserting this into Eq. 24 we obtain a conductance (per spin direction)

$$G = I/V_{1,2} = I \; e/\left(E_{\mathrm{F},1} - E_{\mathrm{F},2}\right) = \frac{e^2}{h} \; T. \tag{27}$$

Equation 27 is the single-channel *Landauer formula*.[12] If we consider ballistic conduction, each electron injected from one reservoir into the conductor will reach the other reservoir without any backscattering. The transmission probability T in the conductor is then given by $T_{12} = \sum_m T_{12,m} = \sum_m \delta_{n,m} = N$, with the number N of occupied modes in the conductor. Considering the spin degeneracy (factor 2) yields the conductance of the ballistic conductor given in Eq. 22.

The general result of Eq. 25 can also be applied to multiterminal devices such as a four-terminal Hall probe, where a current flows between one pair of terminals and the voltage is measured across the other pair. For examples see Ferry et al. (2009).

4 Topological Insulators

A particular kind of low-dimensional carrier transport occurs in a new class of solids referred to as *topological insulator*. These solids have an energy gap like an ordinary insulator in the bulk, but they have states without such gap at the surface or, in case of 2D topological insulators, at the edges; the gapless states provide 1D or 2D conductivity for 2D or 3D topological insulators, respectively. There exists a large variety of topological insulators and topological superconductors distinct by space dimensionality and by the symmetry of their states with respect to particle/hole relation, chirality, and presence or absence of time reversal (Hasan and Kane 2010). In all these solids exists a *bulk-boundary correspondence*, which relates the topological structure of the bulk to the presence of gapless boundary states.

The first realization of an electronic phase, which is topologically distinct from the common phases (classified into insulated, conducting, superconducting, or magnetic), was the integer quantum Hall state described in Sect. 3.1 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients". This discovery led to a new classification scheme based on the notion of *topological order* (Thouless et al. 1982; Wen 1995), although the AlGaAs semiconductor applied in the discovery of the quantum Hall state is actually no topological insulator.

It is instructive to consider briefly some features of a topological classification and a respective view of the quantum Hall state. We then discuss its close relation to

[12] Instead of a factor T also a factor $T/(1-T) = T/R$ is found in literature. The difference arises from the location where $V_{1,2}$ is measured. In a two-terminal measurement with I and V measured through the same pair of leads a factor T results, while an ideal (noninvasive) four-terminal measurement yields a factor T/R; for a small T of the conductor both measurements coincide; for details see Engquist and Anderson (1981).

the spin quantum Hall effect, which was predicted and thereafter experimentally confirmed. Reviews are given in Yan and Zhang (2012), Shen (2012), Bernevig and Hughes (2013), and Ortmann et al. (2015); for a comprehensive tutorial review, see Tong (2016).

4.1 Topological Band Theory

Topology is a mathematical discipline considering properties of a geometric object that are preserved under continuous deformations; the object may be stretched, twisted, or bend, but not torn apart or glued. A property that is invariant under such deformations is a *topological* property; examples are the dimension (it allows distinguishing between line and surface) or the compactness (allows distinguishing between line and circle). Another example is the number of holes in uppercase letters of the alphabet; here {A, D, O, P, Q, R}, {B}, and {C, E, F, G, H, I, . . .} are three classes differing in the number of holes. In this respect, the elements in one class are equivalent, although they may still differ in other properties such as the number of "tails" (two for A, one for P, none for D).

In Fig. 3 of ▶ Chap. 8, "Bands and Bandgaps in Solids", we distinguished metal, semiconductor, and insulator; the first solid has no energy gap between occupied and unoccupied electron states, while the latter two have a gap separating the occupied valence band from an empty conduction band. According to this classical scheme, semiconductors and insulators are different (somewhat arbitrarily) due to the magnitude of the energy gap, which is usually considered with respect to the thermal energy at the temperature of device operation. In a simple topological classification, semiconductor and insulator are equivalent: an atomic interaction parameter can smoothly be tuned, so as to interpolate continuously between the two without closing the gap; in fact, at zero temperature, the semiconductor actually *is* an insulator.

Surface and Edge States

At the surface of solids, electronic surface states appear because the termination of a crystal leads to a deviation from perfect periodicity. The wavevector **k** then becomes complex, leading to the formation of states which differ from the bulk states. The wavefunction of a surface state is localized close to the crystal surface; it decays monotonically in the direction of vacuum and damps oscillatorily in the direction of the crystal. In the case of a two-dimensional solid, the edge states play the role of surface states in three-dimensional crystals.

Conventional Insulators Surface states of insulators are referred to as *Tamm states*. In metals, such states are generally called *Shockley states*, although their physical nature is similar. In the current topological nomenclature, these conventional surface states are called *topologically trivial*. They originate from one bulk band and return to the *same* bulk band as illustrated for the conduction band in Fig. 23a. Such trivial surface states may be metallic if the Fermi energy lies sufficiently high, but they can easily be removed by structural disorder at the surface – such as that induced by oxidation.

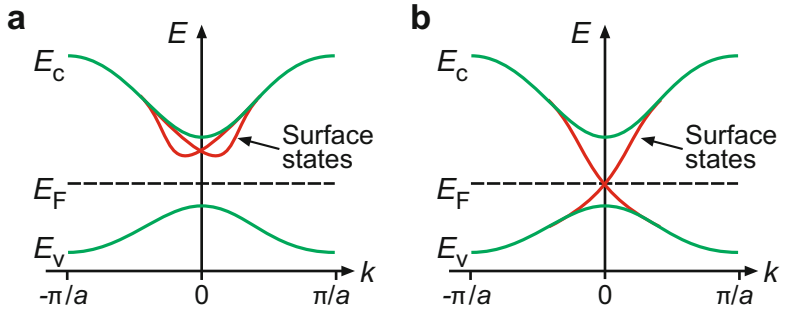


Fig. 23 Band structure $E(k)$ of (**a**) a conventional insulator and (**b**) a topological insulator. *Green curves* refer to bands of the bulk, *red curves* signify states existing at the surface. In the topological insulator, the surface states have no energy gap

Topological Insulators At the surface of a topological insulator exist states without an energy gap as indicated in Fig. 23b; these states are conductive, while the bulk states have a gap like those of any conventional insulator and thus cannot carry a current. According to the topological classification of solids introduced above, the topological invariant must change between bulk and surface. Unlike surface states of a conventional insulator, those of a topological insulator are robust and cannot easily be disrupted. The edge states occurring in the integer quantum Hall effect provide a descriptive example for such states.

The quantum Hall state

The quantum Hall state occurs when highly mobile two-dimensionally confined electrons are placed in a strong magnetic field B; see Sect. 3 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients". The circular orbits of the electrons shown in Fig. 7 of that chapter are then quantized, and their energy is given by the Landau levels $E_\nu = E_c + \left(\nu + \frac{1}{2}\right)\hbar\omega_c$, with $\nu = 1, 2, 3, \ldots$ and the cyclotron frequency $\omega_c = eB/m_n$. If N Landau levels are occupied and those above are empty, an energy gap of $\Delta E_c = \hbar\omega_c$ separates occupied and empty states. At low temperature with $kT \ll \hbar\omega_c$, the unoccupied states will remain empty. When a small electric field is applied, there are no unoccupied states in the occupied Landau levels where the electrons can move to. There is no scattering of electrons, and the longitudinal magnetoresistance is $\rho_{xx} = 0$ like in an insulator. The resistance ρ_{xx} only jumps to finite values when the highest Landau level is incompletely filled and states are accessible for the electrons to scatter into. Still the properties of electrons in a conventional insolator and those of electrons in a two-dimensional electron gas subjected to a strong field **B** (and Landau levels completely filled) differ. Electron states of a 3D insulator and of the topographically equivalent semiconductor are represented by a Fermi sphere in **k** space, while the states in the 2D electron gas are rather given by a toros in **k** space as illustrated in Fig. 24. The genus of these surfaces, which counts the number of holes, is a topological invariant which differs: it is zero for the sphere and one for the torus.

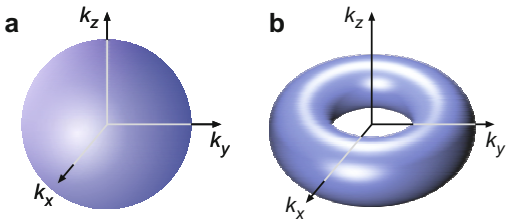


Fig. 24 Momentum space in a conventional 3D solid (**a**) and in a 2D electron gas subjected to a strong magnetic field (**b**)

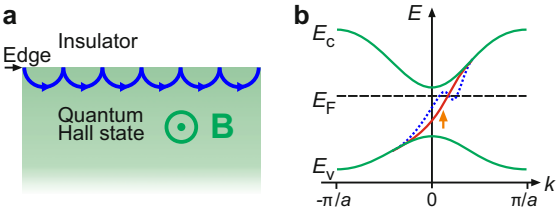


Fig. 25 (**a**) Chiral edge state at the interface between a two-dimensional Hall bar and an insulator. The magnetic field **B** points out of the drawing plane, the *blue curves* indicate skipping cyclotron orbits. (**b**) A single edge state (*red curve*) without energy gap. The *blue dotted curve* indicates a perturbed edge state

There exists a Hall current *perpendicular* to the electric field and a respective Hall resistivity ρ_{xy}; see Fig. 8 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients". The Hall conductivity $\sigma_{xy} = 1/\rho_{xy} = Ne^2/h$ is quantized in steps of the number of occupied Landau levels N. It was shown that N is actually identical to a quantized topological invariant, the *Cern number* n, with $n \in \mathbb{Z}$ (Thouless et al. 1982); it cannot change when the Hamiltonian varies smoothly. The geometrical meaning of n is comparable to the theorem of Gauß and Bonnet (Nakahara 1990): the integral of the Gaussian curvature over a closed surface of a body is given by the number of holes of the body; its value does not depend on the geometry. The number of holes expressed by the genus remains constant under continuous deformations. Analogously, the Cern number is given by the number of occupied Landau levels and does not change when the Hamiltonian changes, provided the number of filled Landau levels remains constant.

The edge states differ from the bulk states. At the edge of a two-dimensional Hall bar, the electrons cannot move in closed circles (or in trochoids in an electric field), since they collide with the boundary and bounce back as illustrated in Fig. 25. As a result, they move only in a single direction – a motion called *chiral*. On the other side of the sample, the electrons move in the other direction, i.e., with opposite chirality. Due to this chirality, the net current with no electric field is zero. According to the bulk-boundary correspondence, the number of chiral edge states is given by the Chern number (= the number of filled Landau levels). In the following, we consider only a single Landau level, i.e., a filling factor $\nu = 1$.

Electrons in edge states near the Fermi energy have unoccupied states, which are not separated by an energy gap; they hence have metallic conductivity. This is illustrated in Fig. 25b by a single band, which connects valence and conduction band with a positive group velocity. These chiral states are insensitive to disorder, since there are no states for backscattering available nearby (the corresponding edge state moving in the opposite direction is far away, confined to the opposite edge of the solid). Even if the edge state is disturbed and its dispersion develops a kink like that indicated by the dotted line in Fig. 25b, the Fermi energy is intersected by any other odd number of crossings than one. There hence remains one positive group velocity more than negative group velocities; the net motion remains chiral and is not disrupted.

4.2 Quantum Spin Hall Insulator

A two-dimensional topological insulator is called a quantum spin Hall (QSH) insulator. The name originates from a close relation to the integer quantum Hall effect. In a simple picture, the QSH effect may be viewed as a superposition of two integer quantum Hall effects with opposite magnetic field (see Fig. 26). Electrons on a given edge of the QSH insulator propagate in opposite directions; this differs from the one-way movement of electrons in the integer quantum Hall effect. Since electrons are fermions, the two counter-propagating edge currents of the QSH insulator must contain electrons with opposite spin: spin and momentum are locked. Such edge states are termed *helical* (Wu et al. 2006). In the QSH insulator, the effect occurs with zero external magnetic field; the spin-orbit interaction plays the role of an effective magnetic field, which acts in opposite direction for opposite spins.

The QSH insulator represents a topologically distinct new kind of matter. The helical edge states illustrated in Fig. 27a are invariant with respect to time reversal, in contrast to the integer quantum Hall states; they represent Kramers doublets. Edge

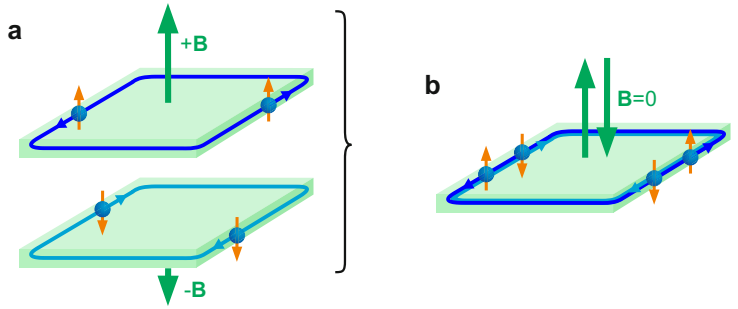


Fig. 26 (**a**) Integer quantum Hall effect with filling factor $\nu = 1$ for two antiparallel orientations of the magnetic field **B**; *orange arrows* signify electron spin. (**b**) Superposition of the two currents illustrated in (**a**); the resulting currents correspond to the quantum spin Hall effect at zero magnetic field

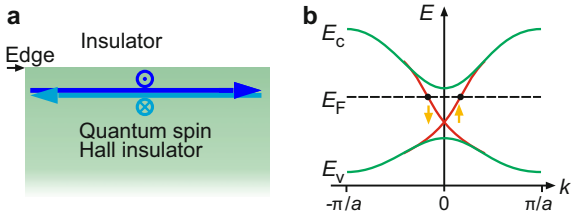


Fig. 27 (**a**) Counter-propagating helical edge states (*blue and light blue arrows*) at the interface between a quantum spin Hall insulator (2D topological insulator) and an insulator. There is no external magnetic field. The symbols ⊙ and ⊗ indicate up and down spins, respectively. (**b**) Dispersion of the edge states (*red curves*) with respective spin orientations (*orange arrows*) indicated

states with an odd number of Kramers' pairs are topologically non-trivial: Since a pair of counter-propagating modes have opposite spin, an electron cannot be scattered from a forward-mover to a backward-mover without reversing its spin. Without spin-flip, no backscattering occurs and transport along the edge proceeds without dissipation. This property makes the QSH states robust against disruption; the edge states are called topologically protected by time-reversal symmetry. The helical QSH edge states form a special one-dimensional conductor. Electronic states of a conventional 1D conductor are easily perturbed by disorder, yielding zero-dimensional Anderson localization (Lee and Ramakrishnan 1985). QSH states, however, cannot be localized even for strong disorder – basically due to the symmetry properties of states, which are reflected at a disordered region (Kane and Mele 2005).

It should be noted that such a protection exists only for an *odd* number of forward- and backward-moving Kramers pairs. In case of *two* forward and *two* backward edge channels, a forward-moving electron can be scattered to a backward channel without breaking the time-reversal symmetry; such a backscattering creates dissipation.

Experimental Confirmation of the QSH State

The first type of topological insulator was predicted (Bernevig et al. 2006) and experimentally confirmed (König et al. 2007) with a two-dimensional HgTe/$Hg_{0.3}Cd_{0.7}Te$ heterostructure. We consider this structure in more detail. CdTe (and also $Hg_{0.3}Cd_{0.7}Te$) has strong spin-orbit interaction and a band structure like that of other zincblende semiconductors: the conduction band has *s*-like symmetry and the valence band has *p*-like symmetry; HgTe has the same crystal structure, but the band structure is *inverted*: a *p*-like conduction band lies above the *s*-like valence band; see Fig. 28, and in ▶ Chap. 8, "Bands and Bandgaps in Solids", Fig 4.

When a HgTe quantum well (of type III) is sandwiched by $Hg_{0.3}Cd_{0.7}Te$ barriers, the *s*- and *p*-like bands have to cross at the interfaces; this can happen because they have different parity and do not show anticrossing. For thin HgTe layers (below d_c = 6.3 nm), the confined states have still the same order as in CdTe due to overlapping

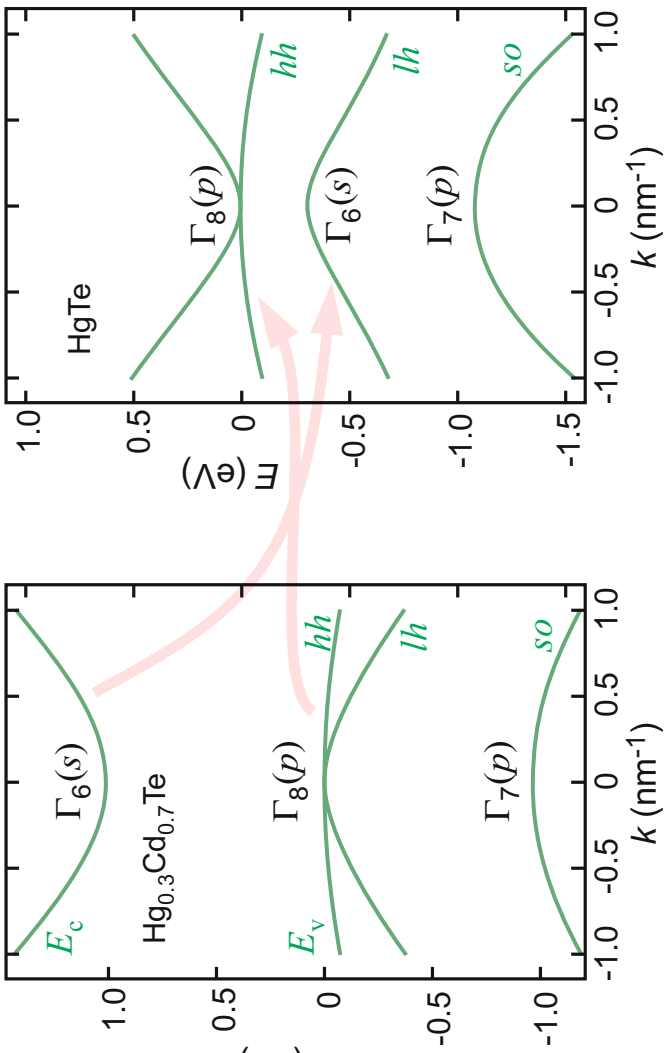


Fig. 28 Band structure of $Hg_{0.3}Cd_{0.7}Te$ and HgTe bulk crystals. At an interface between these solids, the order of bands must change as indicated by the *light red arrows*. HgTe is a semimetal with zero bandgap

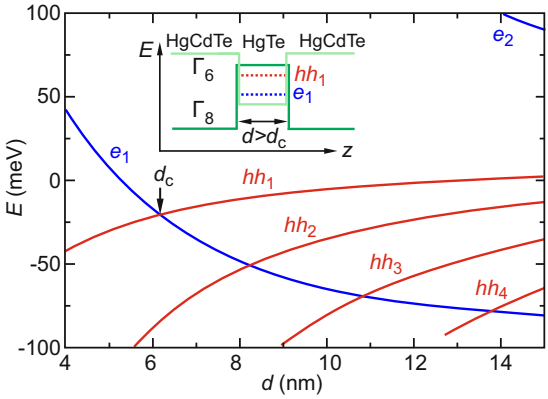


Fig. 29 Calculated subband energies of a HgTe type III quantum well clad by $Hg_{0.3}Cd_{0.7}Te$ barriers. Above the critical thickness d_c, the order of the subbands in the quantum well is inverted. (After König et al. 2008)

wavefunctions of the barriers, but for $d \geq d_c$, the bandgap inverts: the energy of the electron level e_1 decreases below the uppermost heavy-hole level hh_1; see Fig. 29.

In the $HgTe/Hg_{0.3}Cd_{0.7}Te$ quantum-well structure, the HgTe well is tensely strained due to a lattice mismatch of 0.3 %; thereby the Γ_8 degeneracy of the bulk crystal is lifted at the Γ point, and a small bandgap ($E_g = 10 \ldots 20$ meV) appears. At sufficiently low temperature, surface-dominated transport by helical edge states can then be detected. The measurement of helical states in such quantum wells is shown in Fig. 30. Samples with various thicknesses d were structured to Hall-bar shapes as indicated in the inset, and a top electrode was applied to control the concentration of carriers by shifting the Fermi energy with applied gate voltage V_g. Hall measurements proved that the conductivity could be continuously varied from the intrinsic n-type region at $V_g - V_{thr} > 0$ over a highly resistive range to p-type at $V_g - V_{thr} < 0$ (König et al. 2007). The quantum well in sample I ($d = 5.5$ nm $< d_c$) has the normal band ordering depicted in Fig. 28 (left). In the measurement shown in Fig. 30, a current I_{14} was injected from terminal 1 to terminal 4, and the voltage V_{23} was probed between terminals 2 and 4. At zero B-field, the four-terminal resistance $R_{14,23} \equiv V_{23}/I_{14}$ is exceedingly high (at detection limit) for $V_g - V_{thr}$ near 0: E_F lies in the bandgap, and at low sample temperature, no mobile carriers are available. Samples II to IV ($d = 7.3$ nm $> d_c$) have an inverted band structure at the edge and normal order in the center of the bar structure. There are hence zero-gap states near the edge and the resistance $R_{14,23}$ remains finite. For very small samples (III and IV) with sufficiently large mean free path (comparable to the device length of 1 μm), the resistance drops to a plateau at $R = h/(2e^2) \cong 12.9$ kΩ; this value corresponds to $G = 1/R = 2e^2/h$. In the larger device II (20 μm length), some residual scattering leads to a higher (still finite) resistance, i.e., lower G. The plateau at $G = 2e^2/h$ was found to persist also at higher temperature, provided $kT \ll E_g$ (König et al. 2008).

The currents in the device of Fig. 30 (inset) are $I = I_1 = -I_4$; all other currents are zero, since the other terminals serve as voltage probes. In the Landauer-Büttiker formalism (Sect. 3.3), the transmission coefficient is unity for neighboring contacts (zero otherwise), since the edge states do not interact. According to the single-

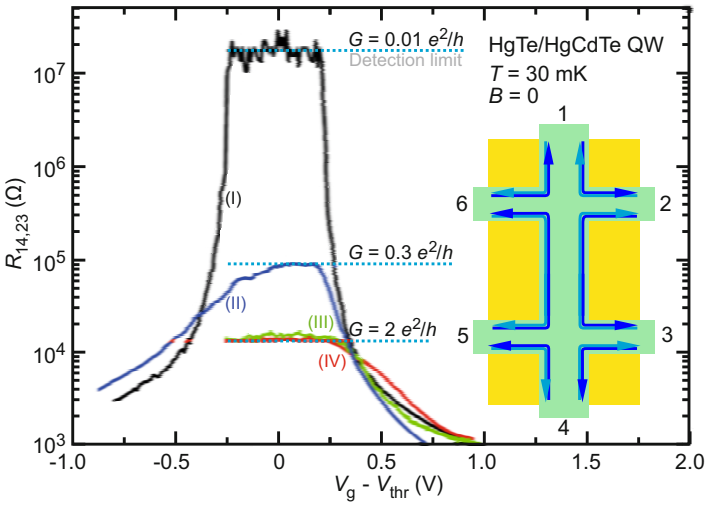


Fig. 30 Longitudinal resistance $R_{14,23} = V_{23}/I_{14}$ of various $Hg_{0.3}Cd_{0.7}Te$ quantum-well samples as a function of gate voltage V_g. *Sample I* is normal conductive, *samples II to IV* have inverted band structure. The *inset* illustrates the six-terminal device structure in Hall bar geometry (*green*) with ohmic contacts labeled *1 to 6*; the *yellow shading* indicates the top gate electrode. *Blue and light blue arrows* indicate couterpropagating spin-polarized edge channels. (After König et al. 2007)

channel Landauer formula (27), this yields $G = e^2/h$ per spin direction for neighboring voltage probes, i.e., $G = 2e^2/h$ for two spin directions.

The nonlocal property of the helical edge states was proved with various sample geometries (Roth et al. 2009). Using a H-bar geometry with 4 terminals (1 top left, 2 top right, 3 bottom right, 4 bottom left) and injecting a current I_{14}, a voltage V_{23} was measured. A nonlocal resistance of $R_{14,23} \cong 6.5$ kΩ was observed in the bulk-insolating regime of V_g, a value in agreement with theoretical results from the Landauer-Büttiker formalism for this geometry (Roth et al. 2009; Wiedmann 2015).

4.3 3D Topological Insulators

The three-dimensional topological insulator has metallic surface states which represent higher-dimensional analogues of the edge states of a 2D topological insulator, i.e., of the quantum spin Hall insulator. The surface states are confined to a very thin layer, which is estimated to be less than 5 nm, as concluded from the penetration depth of the wavefunction into the bulk (Liu et al. 2010; Zhang et al. 2010). The electronic surface structure of a typical 3D topological insulator resembles that of graphene, but with a single Dirac point instead of four (2 valley × 2 spin), and reasonably with a much larger spin-orbit coupling. Theoretical treatments initially focused on the strongly spin–orbit-coupled, band-inverted $Bi_{1-x}Sb_x$ compound for a 3D bulk realization of the quantum spin Hall phase (Fu et al. 2007), and soon after an

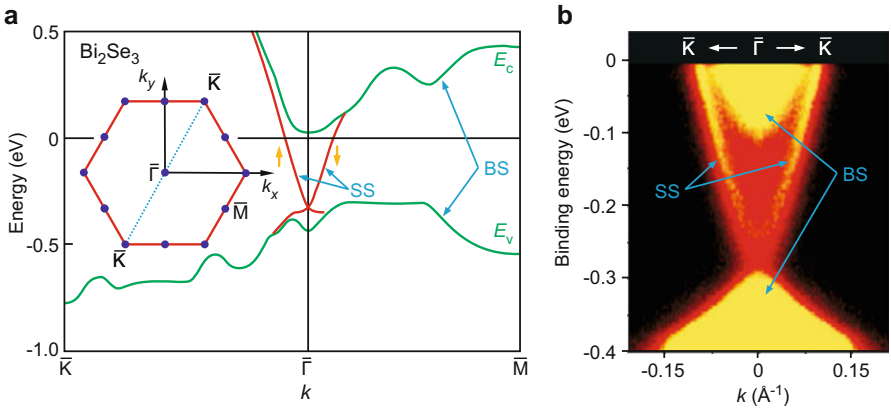


Fig. 31 (**a**) Calculated band structure of Bi_2Se_3 showing surface states (*SS*, *red*) and bulk states (*BS*, *green*). The *inset* shows the two-dimensional Brillouin zone of the (111) crystal surface with the direction $\overline{\Gamma}\rightarrow\overline{K}$ indicated by a *dotted line*. (**b**) Angle-resolved photoemission spectra of the electronic band dispersion on the (111) surface near the Dirac point $\overline{\Gamma}$ along $\overline{\Gamma}\rightarrow\overline{K}$. (After Xia et al. 2009)

experimental confirmation was given for such bulk crystals (Hsieh et al. 2008). Angle-resolved photoemission spectroscopy provided a mapping of the band dispersion, showing clearly the existence of Dirac particles with gapless surface electron bands and a direct energy gap in the bulk. The number of surface bands (three) in $Bi_{0.9}Sb_{0.1}$ is odd and thereby provides topologically protected states. However, the complexity of the states, the small bulk bandgap, and an intricate separation of bulk and surface contributions make transport measurements difficult in this compound.

A more simple band structure and larger bandgap in the bulk made the *second-generation* topological insulators attractive (Moore 2009). This V_2-VI_3 family comprises the narrow-gap rhombohedral chalcogenides Bi_2Se_3, Bi_2Te_3, and Sb_2Te_3 ($E_g = 0.15 \ldots 0.35$ eV) and their alloys, which have a single gapless surface band of nearly linear dispersion. The experimental confirmation of gapless surface states and bulk states with an energy gap is illustrated in Fig. 31. The Bi_2Se_3 crystal has a hexagonal Brillouin zone with a hexagonal projection[13] at its (111) surface shown in the inset of Fig. 31a. Angle-resolved photoemission spectra for the path along $\overline{\Gamma} \rightarrow \overline{K}$ given in panel *b* clearly show a V-shaped band pair of surface states (denoted SS) and a continuum-like U-shaped feature corresponding to the manifold of bulk states. The experiments agree well with results of LAPW calculations shown in panel *a* for the path along $\overline{\Gamma} \rightarrow \overline{K}$.

The time-reversal symmetry requires that the single-particle edge state of an electron must have a degenerate time-reversed partner (Kramers' theorem). The

[13]The Γ and Z points of the bulk Brillouin zone are projected on the $\overline{\Gamma}$ point of the hexagonal 2D surface Brillouin zone; L and F are projected on $\overline{M}$, and K is projected on $\overline{K}$.

two edge states of a Kramers pair then must cross each other at special points of the Brillouin zone, referred to as Kramers points or time-reversal-invariant momenta (TRIMs). TRIMs are points in **k** space which are connected to each other by a reciprocal lattice vector plus a time-reversal operation. The TRIMs of the Bi_2Se_3 (111) surface Brillouin zone depicted in Fig. 31a are the Dirac point $\overline{\Gamma}$ and points $\overline{M}$. The degeneracy at a TRIM cannot be lifted to open a gap by weak disorder or (nonmagnetic) impurities without violating the time-reversal symmetry. Because of the mandatory crossing of the edge-state levels at TRIMs, a topological insulator cannot be smoothly deformed into a topological trivial insulator with a gap but without helical edge states.

The measurement of Fig. 31b yields a cross section of the dispersion $E(k)$ in the $\overline{\Gamma} \rightarrow \overline{K}$ plane. Measurements of other directions show that the Fermi surface of the well-studied V_2-VI_3 chalcogenides is actually a nearly circular Dirac cone near the $\overline{\Gamma}$ point; at higher energies, the cone of Bi_2Se_3 remains nearly circular, while that of Bi_2Te_3 experiences a hexagonal deformation (Alpichshev et al. 2010). Due to the spin-momentum locking, the electron spin is always perpendicular to the momentum, i.e., to the direction of motion. The spin hence rotates on the constant-energy surface of the Dirac cone, being always perpendicular to the wavevector $\mathbf{k}_{\|}$ (parallel to the surface). The spin of a $\mathbf{k}_{\|}$ state is antiparallel to that of the corresponding $-\mathbf{k}_{\|}$ state, rendering these states orthogonal. The spin texture is illustrated in Fig. 32. This helical spin polarization was experimentally confirmed for Bi_2Se_3 and Bi_2Te_3 using spin- and angle-resolved photoelectron spectroscopy (Hsieh et al. 2009); the measurements yield nearly 100% spin polarization. More recently, such a spin polarization was measured up to room temperature in $BiSbTe_3$ by detecting the spin-voltage using a 4-tip STM technique (Leis et al. 2020).

A characteristic feature of the spin-textured edge states of topological insulators is their insensitivity to scattering. STM studies of the effect of disorder on the scattering of spin-polarized edge states in Bi_2Te_3 and $Bi_{1-x}Sb_x$ proved that despite strong disorder on an atomic scale, there is no backscattering between states of opposite momentum and opposite spin. Surface steps and Ag impurities adsorbed on a Bi_2Te_3 (111) surface lead to standing waves originating from the scattering of topological

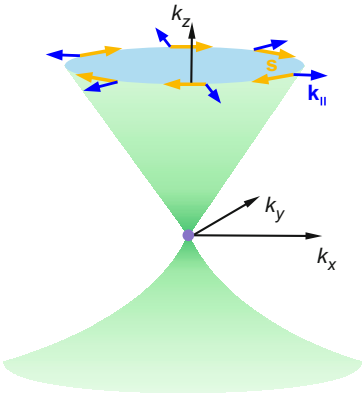


Fig. 32 Schematic of the spin-polarized dispersion of the surface states in a topological insulator such as Bi_2Se_3 near the Dirac point (*blue dot* at $\mathbf{k} = 0$); k_z is parallel to the surface normal. Orientations of the spins are indicated by *orange arrows*

states off these defects (Zhang et al. 2009). Measurement of the dispersion as a function of the scattering wavevector applying imaging by scanning tunneling spectroscopy (STS) confirmed calculated and experimental photoelectron data, and provided clear evidence for the complete suppression of backscattering. In $Bi_{1-x}Sb_x$, the gapless surface states were also measured using STS, and the Fourier-transformed data were compared to spin-dependent scattering probabilities evaluated from spin- and angle-resolved photoelectron data (Roushan et al. 2009). From the good agreement of these data, the absence of backscattering between states of opposite momentum and opposite spin could be concluded – despite the strong alloy disorder in this compound. The results demonstrate that the chiral nature of the surface states protects the spin of the carriers.

Although the existence of symmetry-protected helical surface states was clearly proved by surface-sensitive measurements, the experimental observation of surface transport with superior mobility remains challenging for all current 3D topological insulators. The surface conductivity is generally superimposed by a strong and usually dominating bulk contribution, originating from doping due to a high density of intrinsic and residual extrinsic defects (Analytis et al. 2010). Bi_2Se_3, e.g., has a tendency to form selenium vacancies, resulting in *n*-type doping of the bulk. Bi_2Te_3 forms antisite defects and has a Dirac point buried in the valence band, making it difficult to access. Furthermore, surface oxidation affects carrier transport. All these defects cannot destroy the protected surface states, but they reduce the mobility of the carriers and degrade the characteristics of the quantum effects.

The intrinsic transport behavior of topological surface states was demonstrated for Bi_2Te_2Se by applying a four-probe STM to allow for short distances s between probing tips as illustrated in Fig. 33 (Ko et al. 2018). While at room temperature always a dominant bulk conductance was obtained (only 7% surface contribution despite a small $s = 7.5$ μm), pure surface conductance with high mobility (6.1×10^4 $cm^2/(Vs)$, $s = 87$ μm) and 72% spin polarization[14] was reported, with a transition of

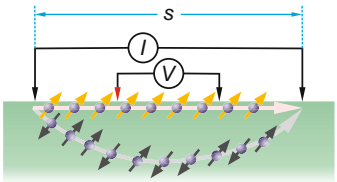


Fig. 33 Schematic (side view) of superimposed 2D surface and 3D bulk electron transport in a 3D topological insulator, measured by a four-probe scanning tunneling microscope. *Orange arrows* indicate spin orientation of the spin-polarized electrons, the *red tip* may either be nonmagnetic or ferromagnetic for spin detection

[14]The spin polarization was measured by using a ferromagnetic Ni tip instead of a nonmagentic W tip for the voltage probe (Hus et al. 2017).

the conductance type near 100 K; at 10 K, a mean free path of 1.4 μm was estimated, emphasizing the importance of micrometer-scale dimensions for device applications at ambient temperature.

Topological surface states were predicted also for many other solids, e.g., the large family of zincblende-related X_2YZ Heusler or XYZ half-Heusler compounds (Xiao et al. 2010; Yan and Zhang 2012), or the family of chalcopyrite I–III–IV_2, II–IV–V_2, or I_2–II–IV–VI_4 semiconductors (Feng 2011). Today applications of surface transport in current topological-insulator materials at ambient temperature are still hampered by significant bulk conduction.

Classical-Wave and Higher-Order Topological Insulators

In topological insulators, there is generally a bulk-boundary correspondence. We discussed above 2D and 3D systems; for a solid with d dimensions, gapless states exist correspondingly on the $(d-1)$-dimensional boundary. Higher-order topological insulators (HOTIs) have been predicted as a novel topological state (Benalcazar et al. 2017; Geier et al. 2018). Such HOTIs have corner and hinge states rather than edge states. A two-dimensional HOTI, e.g., has gapped 1D edge states and 0D corner states in the gap, described by a higher-order bulk-boundary correspondence. The lower-dimensional boundary states cannot be removed by any perturbations on the surface if the protective symmetry is preserved.

The concept of topological insulators has been applied to various fields of classical wave physics including photonics, phononics, acoustics, mechanics, and microwaves. The implementation for classical waves opens a new field of engineering-oriented applications, such as robust routing, lasing, signal processing, or switching, briefly outlined by Zangeneh-Nejad et al. (2020). Topological phases in photonic crystals are reviewed by Wang et al. (2020). Also higher-order topological insulators have been realized in various systems based on classical waves, such as photonic and phononic metamaterials as well as electric circuits; a comprehensive review is given by Xie et al. (2021).

5 Zero-Dimensional Transport

In a zero-dimensional semiconductor the electronic states are completely quantized, see Sect. 3.4 of ▶ Chap. 8, "Bands and Bandgaps in Solids". The electronic states in such a quantum dot are sensitive to the number of confined electrons due to their mutual Coulomb interaction. Transport through quantum dots is performed by connections to surrounding carrier reservoirs; generally transport occurs via tunneling through thick barriers with a tunnel resistance exceeding the quantum resistance[15] h/e^2. Adding a single electron to the dot may give rise to a Coulomb blockade

[15]This requirement for R_{tunnel} is related to Heisenberg's uncertainty relation: quantum fluctuations in the number of electrons on the dot due to tunneling through the barriers must be much less than one for the duration of measurement. In addition, the charging energy must exceed the thermal energy (see text). The characteristic discharging time of the dot is then $\Delta\tau = R_{\text{tunnel}}C_{\text{dot}}$, and the uncertainty relation yields $\Delta E\Delta\tau = (e^2/C_{\text{dot}})\, R_{\text{tunnel}}C_{\text{dot}} > h$, or $R_{\text{tunnel}} > h/e^2$.

of electron transfer, suppressing further transport (Kouwenhoven and McEuen 1999). Thereby single-electron turnstile devices can be realized with interesting applications in metrology.

The size of a quantum dot required for a sizeable Coulomb blockade can be estimated by a simplified consideration. We imagine a flat circular disk above a grounded conductive plane with parallel orientation. When the disk is charged by transferring a charge Q to it, it adopts an electric potential $V = Q/C$, where the capacitance of the disk depends on its radius r and the distance d above the plane; for $r \ll d$ the capacitance is $C = 8\varepsilon_0\varepsilon_r r$, with ε_r the dielectric constant of the insulator surrounding the disk. The work W performed to charge the disk[16] is stored in the electrostatic energy given by $W = Q^2/(2C)$. If the energy $e^2/(2C)$ is significantly larger than the thermal energy kT, the related voltage change gets relevant compared to thermal voltage fluctuations. This voltage represents a barrier for further electron transfer. The condition is met if $C \ll e^2/(2kT)$, or $r \ll e^2/(16\varepsilon kT)$. Applying a typical $\varepsilon_r = 13$, we obtain a range well below 10 nm at $T = 300$ K; although Coulomb blockade was observed with a very small (~2 nm) dot at room temperature (Shin et al. 2011), most structures are larger and observation requires very low temperatures.

Fabrication of such quantum dots for charge transport is often based on 1D structures discussed in Sect. 3 with additional constrictions introduced along the wire direction. Such barriers are readily realized when the split-gate technique defining depletion regions in a two-dimensional electron gas is applied. Often more sophisticated geometries with additional gates are applied to obtain an independent control of the barriers along the wire and the electron density. Besides these lateral quantum dots also vertical quantum dots were fabricated by etching small mesa structures from layers defining source, dot, and drain regions (e.g., Reed et al. 1988) or nanowires (e.g., Björk et al 2004); for a brief review of fabrication techniques see Ahmed and Nakazato (1996).

5.1 Single-Electron Tunneling

The effect of Coulomb blockade can be well observed with a single-electron transistor. This three-terminal device illustrated in Fig. 34a consists of source and gate electrodes, which are connected through tunnel junctions to the common quantum-dot island with a low capacitance. The quantum dot is decoupled from the environment by thick barriers, which allow to confine a defined number N of electrons. An additional gate contact provides a capacitive electrostatic coupling to the dot and continuously controls the electrochemical potential of the electrons in the dot.

The energy levels of the dot are shown in Fig. 34b; in a lateral quantum dot they are approximately evenly spaced with spacings ΔE. The (spin-degenerate)

[16] If the considered disk is replaced by a sphere, the factor 8 in the capacitance formula is replaced by a factor 4π. The simple consideration in the text does neither include single-particle energies of a dot with charge Ne nor external charges or capacities (e.g., from a gate electrode).

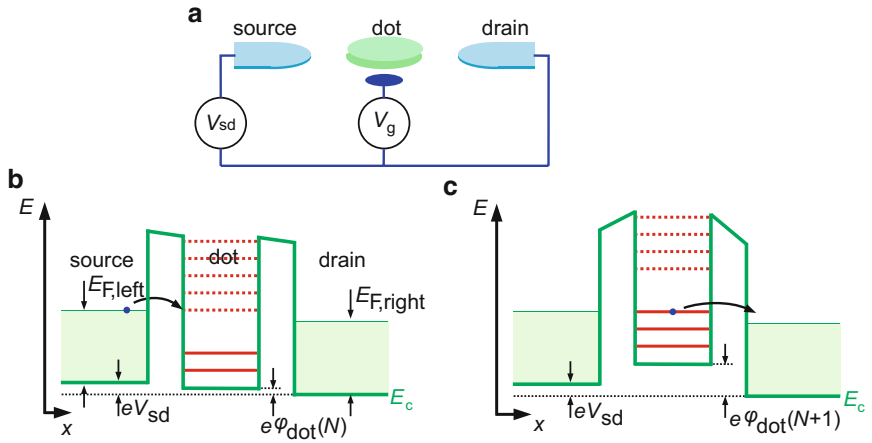


Fig. 34 (**a**) Schematic of a single-electron transistor with a lateral quantum dot coupled to source and drain by tunnel barriers. (**b**) Energy diagram of the device with bias V_{sd} adjusted for tunneling into an empty state of the dot. *Solid and dotted red lines* signify occupied and empty levels in the dot, respectively. (**c**) Energies with increased electrostatic dot energy $e\varphi_{dot}(N+1) = e\varphi_{dot}(N) + e^2/C$ due to the added electron

spacing can be estimated for a disk-shaped dot[17] by $\Delta E = \hbar^2/(m_n r^2)$; using the effective electron mass in GaAs ($0.067 \times m_0$) and a radius $r = 200$ nm, we obtain $\Delta E = 0.03$ meV. Observation of such a small spacing requires thermal energies well below this value. The *change* of the electrostatic potential for such a dot by adding an extra charge e is given by the charging energy $e^2/C = e^2/(8\pi\varepsilon_r\varepsilon_0 r)$, yielding with $\varepsilon_r = 13$ a value of 1 meV.

For simplicity we assume a dot capacitance being independent on the number of confined electrons, a coupling to the leads which does not affect the energies in the dot, and zero temperature (Kouwenhoven et al. 1997). We first consider a *fixed gate voltage* $V_g = \left(E_{F,\,source} - E_{F,\,drain}\right)/e$, with $E_{F,\,source} - E_{F,\,drain} < \Delta E$. The single-particle states E_N, measured from the bottom of the conduction band in the dot, represent excitations of the dot with constant electron number N; they are occupied up to a maximum electron number N, for which the condition $E_{F,dot}(N) < E_{F,source} \cong E_{F,drain}$ is met. Adding one electron to the dot requires an energy

$$E_{F,\,dot}(N+1) - E_{F,\,dot}(N) = \Delta E + e^2/C, \tag{28}$$

[17]The level spacing is given by the scaling $\hbar^2\pi^2/(m_n L^2)$ times a factor, which depends on the dimensionality of the "dot" (Kouwenhoven and McEuen 1999); L is the size of the confining box. In 1D, 2D, and 3D the factors are given by $N/4$, $1/\pi$, and $1/(3\,\pi^2 N)$, respectively. The spacing in lateral dots (2D) does not depend on the number of confined electrons N.

where the single-electron charging energy e^2/C creates an energy gap at the Fermi energy between the occupied and unoccupied states, see Fig. 34b, c. This Coulomb gap prevents tunneling through the dot, if $E_{F,dot}(N+1) > E_{F,source} \cong E_{F,drain}$. This effect is the previously introduced *Coulomb blockade*.

When V_g is *increased* to achieve $E_{F,source} > E_{F,dot}(N+1)$, an electron can tunnel to the dot as illustrated in Fig. 34b. The conduction-band edge E_c in the dot is then raised by the charging energy e^2/C as shown in Fig. 34c. Since $E_{F,dot}(N+1)$ exceeds $E_{F,drain}$, one electron can now tunnel from the dot to the drain; this causes the potential in the dot to return back to $E_{F,dot}(N)$. Since V_{sd} is unchanged, another electron can now tunnel to the dot and the cycle repeats. The device hence operates as an electron turnstile, producing a current by single-electron tunneling. Such a behavior is interesting for the realization of a quantum standard for current (Rossi et al. 2014) or capacity (Keller et al. 1999) by generating a current defined by $I = nef$ or a capacity $C = ne/V$, where n is an integer and f is the frequency of an externally applied ac pumping signal. The time constant $R_{tunnel}C_{dot}$ limits the rf frequency, but proper gate tuning or phase-shifted pumping (Blumenthal et al. 2007) yield high frequencies up to 3.4 GHz and currents of about 100 picoamperes (Rossi et al. 2014). A review on various approaches of such electron pumps is given by Giblin et al. (2019). Successive transfer of single electrons was also demonstrated across a linear array of coupled quantum dots (Mills et al. 2019); such a bucket brigade is interesting for quantum-state transfer within spin-based quantum processors.

When V_{sd} (smaller than the level spacing ΔE) is kept constant and the *gate voltage* V_g *is continuously increased*, the conductance G changes periodically from zero in Coulomb-blocking states to finite values in the conductive states described above (Meirav et al. 1990). Such an oscillation of G with sweeped gate voltage is shown in Fig. 35. At zero conductance the number of electrons on the dot is fixed and equals an integer N. In the subsequent valley the number of electrons is increased to $N+1$. In the conductive states between the valleys the number alternates between N and $N+1$ and leads to the current flow.

When V_g is kept constant and the *source-drain voltage* V_{sd} *is continuously increased*, the conductance G exhibits a nonlinear I-V_{sd} characteristic with a

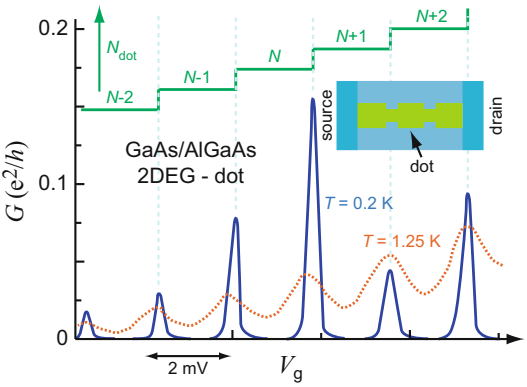


Fig. 35 Coulomb oscillations in the conductance as a function of the gate voltage, observed for a 1D channel with constrictions defining a dot region (*inset*) (After Meir et al. 1991). The stairs at the top of the measured curves indicates the number of electrons in the dot

Coulomb staircase: each time the number of electrons on the dot increases by one a new current step occurs; see, e.g., Kouwenhoven et al. (1991).

5.2 Coulomb Blockade in Few-Electron Dots

In dots carrying many electrons the simple approach of expressing the Coulomb interaction among the electrons in terms of a constant capacitance holds reasonably well; the Coulomb oscillations (Fig. 35) are then usually periodic. In a small dot with just a few electrons, the electron–electron interactions and quantum confinement effects become stronger, and the spacings between the peaks of Coulomb oscillations become irregular (see, e.g., Tarucha et al. 1996). This effect is well studied in vertical dot structures comprising heterostructure barriers, e.g., mesa structures or quantum wires. Tunnel barriers in such structures are abrupt and thin; these barriers are also less affected by the gate potential, which strongly increases the tunnel barriers in lateral dots when less than some 20 electrons are to be confined.

The conductance of a single-electron transistor fabricated from an InAs nanowire with an InAs dot section clad by InP barriers is shown in Fig. 36. The data show the conductance (in a gray scale) as a function of the source-drain voltage V_{sd}, which is required for current flow at a given gate voltage V_g; a horizontal cut, i.e., $G(V_g)$ at a fixed V_{sd}, yields Coulomb oscillations like those shown in Fig. 35. We observe the irregular spacings of G maxima at $V_{sd} = 0$ for the small electron numbers N noted in the figure. The yellow regions with zero conductance represent Coulomb blockades and are referred to as *Coulomb diamonds* due to their shape. The shape depends on the confining potential and hence on the particular device. The Coulomb diamond in the region $N = 0$ does not close even for more negative gate voltage since the dot contains no electron in this range.

There is a rich structure in stability diagrams like that of Fig. 36 comprising information on electron–electron interaction, spin, and the confining potential. For a more detailed discussion on zero-dimensional transport see Kouwenhoven et al. (1997, 2001), Reimann and Manninen (2002).

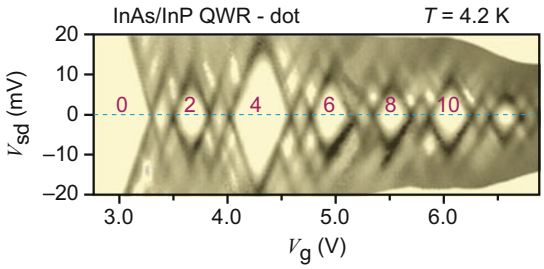


Fig. 36 Stability diagram of electron tunneling, given by the source-drain voltage V_{sd} at which current starts to flow for a given gate voltage V_g. *Dark shadings* indicate peak maxima in the conductance $G(V_g)$, *diamond-shaped regions* correspond to $V_g - V_{sd}$ combinations where Coulomb blockade occurs; *numbers* indicate electron occupations on the dot (After Björk et al. 2004)

6 Summary

In *two-dimensional semiconductor* structures the in-plane carrier mobility can reach very high values up to 10^7 $cm^2/(Vs)$ at low temperature, if the electron gas is spatially separated from ionized donor dopants by applying modulation doping. Occupation of only the lowest 2D subband is favorable to avoid intersubband scattering. At higher temperature, the mobility is mainly limited by scattering at optical phonons. Carrier transport perpendicular to the confining barriers yields lower mobility. For thin barriers tunneling occurs; the transmission coefficient increases for smaller barrier height and smaller effective mass. In a quantum well clad by thin barriers or a superlattice with thin barriers resonant tunneling occurs when the energy of incident electrons matches a quantized energy level; at increased bias a range of negative differential resistance is observed. Resonant tunneling and radiative intersubband relaxation is exploited in quantum-cascade lasers for designed emission wavelength in the mid-IR range.

Additional lateral confinement leads to a *one-dimensional quantum wire*, where each subband of the 2D semiconductor splits into 1D subbands with free carrier motion only along the wire direction. The split-gate technique provides a widely applied means to create such lateral confinement with tunable width and electron density of the quantum wire. For stronger lateral confinement etched mesa structures or nanowires are used. In quantum wires ballistic transport is observed, leading to quantization of the conductance in integer multiples of $2e^2/h$.

In a *zero-dimensional quantum dot* the electronic states are completely quantized and sensitive to the number of confined carriers. Transport through such a dot is controlled by the charging energy for single electrons; conductance proceeds by tunneling of individual electrons at certain bias values, while a Coulomb blockade yields zero conduction at other bias. Such turnstile behavior is interesting for applications in quantum metrology and quantum processing.

Topological insulators represent a new class of solids with low-dimensional carrier transport. Conductivity is provided by gapless states at the surface (edges in 2D structures), while the bulk is insulating. One-way carrier transport in edge states is inherently robust against scattering by disorder due to nonlocal symmetry reasons. Higher-order topological insulators have symmetry-protected lower-dimensional boundary states.

References

Ahmed H, Nakazato K (1996) Single-electron devices. Microelectron Engin 32:297

Ahmadi E, Keller S, Mishra UK (2016) Model to explain the behavior of 2DEG mobility with respect to charge density in N-polar and Ga-polar AlGaN-GaN heterostructures. J Appl Phys 120:115302

Alam MA, Si M, Ye PD (2019) A critical review of recent progress on negative capacitance field-effect transistors. Appl Phys Lett 114:090401

Alpichshev Z, Analytis JG, Chu JH, Fisher IR, Chen YL, Shen ZX, Fang A, Kapitulnik A (2010) STM imaging of electronic waves on the surface of Bi_2Te_3: topologically protected surface states and hexagonal warping effects. Phys Rev Lett 104:016401

Analytis JG, Chu J-H, Chen Y, Corredor F, McDonald RD, Shen ZX, Fisher IR (2010) Bulk Fermi surface coexistence with Dirac surface state in Bi_2Se_3: A comparison of photoemission and Shubnikov-de Haas measurements. Phys Rev B 81:205407

Ando T, Fowler AB, Stern F (1982) Electronic properties of two-dimensional systems. Rev Mod Phys 54:437

Beenakker CWJ, van Houten H (1991) Quantum transport in semiconductor nanostructures. Solid State Phys 44:1

Benalcazar WA, Bernevig BA, Hughes TL (2017) Quantized electric multipole insulators. Science 357:61

Bernevig BA, Hughes TL (2013) Topological insulators and topological superconductors. Princeton University Press, Princeton

Bernevig BA, Hughes TL, Zhang S-C (2006) Quantum spin Hall effect and topological phase transition in HgTe quantum wells. Science 314:1757

Björk MT, Thelander C, Hansen AE, Jensen LE, Larsson MW, Wallenberg LR, Samuelson L (2004) Few-electron quantum dots in nanowires. Nano Lett 4:1621

Blumenthal MD, Kaestner B, Li L, Giblin S, Janssen TJBM, Pepper M, Anderson D, Jones G, Ritchie DA (2007) Gigahertz quantized charge pumping. Nature Phys 3:343

Brennan KF, Summers CJ (1987a) Theory of resonant tunneling in a variably spaced multiquantum well structure: an Airy function approach. J Appl Phys 61:614

Brennan KF, Summers CJ (1987b) The variably spaced superlattice energy filter quantum well avalanche photodiode: a solid-state photomultiplier. IEEE J Quantum Electron 23:320

Büttiker M (1986) Negative resistance fluctuations at resistance minima in narrow quantum hall conductors. Phys Rev B 38:12724

Callebaut H, Hu Q (2005) Importance of coherence for electron transport in terahertz quantum cascade lasers. J Appl Phys 98:104505

Capasso F (1987) Band-gap engineering: from physics and materials to new semiconductor devices. Science 235:172

Capasso F (2010) High-performance midinfrared quantum cascade lasers. Opt Eng 49:111102

Capasso F, Mohammed K, Cho A (1986) Resonant tunneling through double barriers, perpendicular quantum transport phenomena in superlattices, and their device applications. IEEE J Quantum Electron 22:1853

Capasso F, Sen F, Beltram F, Cho AY (1990) Resonant tunneling and superlattices devices: physics and circuits. In: Capasso F (ed) Physics of quantum electron devices. Springer, Berlin, pp 181–252

Chen T-M, Graham AC, Pepper M, Farrer I, Ritchie DA (2008) Bias-controlled spin polarization in quantum wires. Appl Phys Lett 93:032102

Choi KK, Levine BF, Bethea CG, Walker J, Malik RJ (1987a) Photoexcited coherent tunneling in a double-barrier superlattice. Phys Rev Lett 59:2459

Choi KK, Levine BF, Malik RJ, Walker J, Bethea CG (1987b) Periodic negative conductance by sequential resonant tunneling through an expanding high-field superlattice domain. Phys Rev B 35:4172

Danneau R, Klochan O, Clarke WR, Ho LH, Micolich AP, Simmons MY, Hamilton AR, Pepper M, Ritchie DA, Zülicke U (2006) Zeeman splitting in ballistic hole quantum wires. Phys Rev Lett 97:026403

Das MP, Green F (2019) Anomalies in one-dimensional electron transport: quantum point contacts and wires. Adv Nat Sci Nanosci Nanotechnol 10:013001

Engquist H-L, Anderson PW (1981) Definition and measurement of the electrical and thermal resistances. Phys Rev B 24:1151

Esaki L, Chang LL (1974) New transport phenomenon in a semiconductor "superlattice". Phys Rev Lett 33:495

Esaki L, Tsu R (1970) Superlattice and negative differential conductivity in semiconductors. IBM J Res Dev 14:61

Faist J (2013) Quantum cascade lasers. Oxford University Press, Oxford

Faist J, Capasso F, Sivco DL, Sirtori C, Hutchinson AL, Cho AY (1994) Quantum cascade laser. Science 264:553

Faist J, Capasso F, Sirtori C, Sivco DL, Hutchinson AL, Hybertsen MS, Cho AY (1996) Quantum cascade lasers without intersubband population inversion. Phys Rev Lett 76:411

Feng W, Xiao D, Ding J, Yao Y (2011) Three-dimensional topological insulators in I−III−VI_2 and II−IV−V_2 chalcopyrite semiconductors. Phys Rev Lett 106:016402

Ferreira R, Bastard G (1989) Evaluation of some scattering times for electrons in unbiased and biased single- and multiple-quantum-well structures. Phys Rev B 40:1074

Ferry DK, Goodnick SM, Bird J (2009) Transport in nanostructures, 2nd edn. Cambridge University Press, Cambridge

Freeman W (2020) Longitudinal-optical phonon absorption and dephasing in three-level terahertz quantum cascade structures with different injector anticrossings. J Appl Phys 128:235702

Fu L, Kane CL, Mele EJ (2007) Topological insulators in three dimensions. Phys Rev Lett 98:106803

Geier M, Trifunovic L, Hoskam M, Brouwer PW (2018) Second-order topological insulators and superconductors with an order-two crystalline symmetry. Phys Rev B 97:205135

Giblin SP, Fujiwara A, Yamahata G, Bae MH, Kim N, Rossi A, Möttönen M, Kataoka M (2019) Evidence for universality of tunable-barrier electron pumps. Metrologia 56:044004

Giorgetta FR, Baumann E, Graf M, Yang Q, Manz C, Kohler K, Beere HE, Ritchie DA, Linfield E, Davies AG, Fedoryshyn Y, Jackel H, Fischer M, Faist J (2009) Quantum cascade detectors. IEEE J Quantum Electron 45:1039

Gurusinghe MN, Davidsson SK, Andersson TG (2005) Two-dimensional electron mobility limitation mechanisms in $Al_xGa_{1-x}N$/GaN heterostructures. Phys Rev B 72:045316

Harris JJ, Foxon CT, Lacklison DE, Barnham KWJ (1986) Scattering mechanisms in (Al, Ga)As/GaAs 2DEG structures. Superlattice Microstruc 2:563

Harris JJ, Pals JA, Woltjer R (1989) Electronic transport in low-dimensional structures. Rep Prog Phys 52:1217

Hartig M, Haacke S, Selbmann P, Deveaud B, Taylor R, Rota L (1998) Efficient intersubband scattering via carrier-carrier interaction in quantum wells. Phys Rev Lett 80:1940

Hasan MZ, Kane CL (2010) Topological insulators. Rev Mod Phys 82:3045

Hayden RK, Maude DK, Eaves L, Valadares EC, Henini M, Sheard FW, Hughes OH, Portal JC, Cury L (1991) Probing the hole dispersion curves of a quantum well using resonant magnetotunneling spectroscopy. Phys Rev Lett 66:1749

Hess K (1981) Lateral transport in superlattices. J de Physique 42:C7,3

Hillbrand J, Krüger LM, Dal Cin S, Knötig H, Heidrich J, Andrews AM, Strasser G, Keller U, Schwarz B (2021) High-speed quantum cascade detector characterized with a mid-infrared femtosecond oscillator. Opt Express 29:5774

Hirakawa K, Sakaki H (1986) Mobility of the two-dimensional electron gas at selectively doped *n*-type $Al_xGa_{1-x}As$/GaAs heterojunctions with controlled electron concentrations. Phys Rev B 33:8291

Hsieh D, Xia Y, Qian D, Wray L, Xia Y, Hor YS, Cava RJ, Hasan MZ (2008) A topological Dirac insulator in a quantum spin Hall phase. Nature 452:970

Hsieh D, Xia Y, Qian D, Wray L, Dil JH, Meier F, Osterwalder J, Patthey L, Checkelsky JG (2009) A tunable topological insulator in the spin helical Dirac transport regime. Nature 460:1101

Hus SM, Zhang X-G, Nguyen GD, Ko W, Baddorf AP, Chen YP, Li A-P (2017) Detection of the spin-chemical potential in topological insulators using spin-polarized four-probe STM. Phys Rev Lett 119:137202

Inoue M (1985) Hot electron transport in quantum wells. Superlattice Microstruc 1:433

Khandelwal S, Goyal N, Fjeldly TA (2013) A precise physics-based compact model for 2-DEG charge density in GaAs HEMTs applicable in all regions of device operation. Sol State Electron 79:22

Kane CL, Mele EJ (2005) Z_2 topological order and the quantum spin Hall effect. Phys Rev Lett 95:146802

Kastrup J, Grahn HT, Ploog K, Prengel F, Wacker A, Schöll E (1994) Multistability of the current–voltage characteristics in doped GaAs-AlAs superlattices. Appl Phys Lett 65:1808

Kastrup J, Hey R, Ploog KH, Grahn HT, Bonilla LL, Kindelan M, Moscoso M, Wacker A, Galán J (1997) Electrically tunable GHz oscillations in doped GaAs-AlAs superlattices. Phys Rev B 55:2476

Keever M, Kopp W, Drummond TJ, Morkoç H, Hess K (1982) Current transport in modulation-doped $Al_xGa_{1-x}As$/GaAs heterojunction structures at moderate field strengths. Jpn J Appl Phys 21:1489

Keller MW, Eichenberger AL, Martinis JM, Zimmerman NMA (1999) Capacitance standard based on counting electrons. Science 285:1706

Köhler R, Tredicucci A, Beltram F, Beere HE, Linfield EH, Davies AG, Ritchie DA, Iotti RC, Ross F (2002) Terahertz semiconductor-heterostructure laser. Nature 417:156

Ko W, Nguyen GD, Kim H, Kim JS, Zhang X-G, Li A-P (2018) Accessing the intrinsic spin transport in a topological insulator by controlling the crossover of bulk-to-surface conductance. Phys Rev Lett 121:176801

König M, Wiedmann S, Brüne C, Roth A, Buhmann H, Molenkamp LW, Qi X-L, Zhang S-C (2007) Quantum spin Hall insulator state in HgTe quantum wells. Science 318:766

König M, Buhmann H, Molenkamp LW, Hughes T, Liu C-X, Qi X-L, Zhang S-C (2008) The quantum spin Hall effect: theory and experiment. J Phys Soc Jpn 77:031007

Kouwenhoven LP, McEuen PL (1999) Single electron transport through a quantum dot. In: Timp GL (ed) Nanotechnology. Springer, New York, pp 471–535

Kouwenhoven LP, van der Vaart NC, Johnson AT, Kool W, Harmans CJPM, Williamson JG, Staring AAM, Foxon CT (1991) Single electron charging effects in semiconductor quantum dots. Z Phys B 85:367

Kouwenhoven LP, Marcus CM, McEuen PL, Tarucha S, Westervelt RM, Wingreen ND (1997) Electron transport in quantum dots. In: Sohn LL, Kouwenhoven LP, Schön G (eds) Mesoscopic electron transport. Springer, Dordrecht, pp 105–214

Kouwenhoven LP, Austing DG, Tarucha S (2001) Few-electron quantum dots. Rep Prog Phys 64:701

Kristensen A, Bruus H, Hansen AE, Jensen JB, Lindelof PE, Marckmann CJ, Nygård J, Sørensen CB, Beuscher F, Forchel A, Miche M (2000) Bias and temperature dependence of the 0.7 conductance anomaly in quantum point contacts. Phys Rev B 62:10950

Kwok SH, Grahn HT, Ramsteiner M, Ploog K, Prengel F, Wacker A, Schöll E, Murugkar S, Merlin R (1995) Nonresonant carrier transport through high-field domains in semiconductor superlattices. Phys Rev B 51:9943

Landauer R (1957) Spatial variation of currents and fields due to localized scatterers in metallic conduction. IBM J Res Dev 1:223

Laux SE, Frank DJ, Stern F (1988) Quasi-one-dimensional electron states in a split-gate GaAs/AlGaAs heterostructure. Surf Sci 196:101

Leisching P, Haring Bolivar P, Beck W, Dhaibi Y, Brüggemann F, Schwedler R, Kurz H, Leo K, Köhler K (1994) Bloch oscillations of excitonic wave packets in semiconductor superlattices. Phys Rev B 50:14389

Lee PA, Ramakrishnan TV (1985) Disordered electronic systems. Rev Mod Phys 57:287

Leis A, Schleenvoigt M, Jalil AR, Cherepanov V, Mussler G, Grützmacher D, Tautz FS, Voigtländer B (2020) Room temperature in-situ measurement of the spin voltage of a $BiSbTe_3$ thin film. Sci Rep 10:2816

Liu CX, Zhang HJ, Yan B, Qi XL, Frauenheim T, Dai X, Fang Z, Zhang SC (2010) Oscillatory crossover from two-dimensional to three-dimensional topological insulators. Phys Rev B 81:041307

Lu SL, Schrottke L, Teitsworth SW, Hey R, Grahn HT (2006) Formation of electric-field domains in $GaAs/Al_xGa_{1-x}As$ quantum cascade laser structures. Phys Rev B 73:033311

Lyssenko V, Leo K (2011) Bloch oscillations and ultrafast coherent optical phenomena. In: Bhattacharya P, Fornari R, Kamimura H (eds) Comprehensive semiconductor science and technology, vol 2: Physics and fundamental theory. Elsevier BV, Amsterdam, pp 343–399

Meir Y, Wingreen NS, Lee PA (1991) Transport through a strongly interacting electron system: theory of periodic conductance oscillations. Phys Rev Lett 66:3048

Meirav U, Kastner MA, Wind SJ (1990) Single-electron charging and periodic conductance resonances in GaAs nanostructures. Phys Rev Lett 65:771

Mendez EE, Agullo-Rueda F, Hong JM (1988) Stark localization in GaAs-GaAlAs superlattices under an electric field. Phys Rev Lett 60:2426

Micolich AP (2011) What lurks below the last plateau: experimental studies of the $0.7 \times 2e^2/h$ conductance anomaly in one-dimensional systems. J Phys Condens Matter 23:443201

Mills AR, Zajac DM, Gullans MJ, Schupp FJ, Hazard TM, Petta JR (2019) Shuttling a single charge across a one-dimensional array of silicon quantum dots. Nat Commun 10:1063

Moore J (2009) The next generation. Nat Phys 5:378

Morkoç H (1985) Modulation-doped $Al_xGa_{1-x}As/GaAs$ heterostructures. In: Parker EHC (ed) The technology and physics of molecular beam epitaxy. Plenum Press, New York, p 185

Morkoç H, Chen J, Reddy UK, Henderson T, Luryi S (1986) Observation of a negative differential resistance due to tunneling through a single barrier into a quantum well. Appl Phys Lett 49:70

Nakahara M (1990) Geometry, topology and physics. Hilger, Bristol

Nixon JA, Davies JH, Baranger HU (1991) Conductance of quantum point contacts calculated using realistic potentials. Superlattice Microstruc 9:187

Ortmann F, Roche S, Valenzuela SO (eds) (2015) Topological insulators. Wiley-VCH, Weinheim

Pfeiffer LN, West KW, Störmer HL, Baldwin KW (1989) Electron mobilities exceeding 10^7 cm^2/Vs in modulation-doped GaAs. Appl Phys Lett 55:1888

Piazza V, Casarini P, De Franceschi S, Lazzarino M, Beltram F, Jacoboni C, Bosacchi A, Franchi S (1998) Self-consistent electron-mobility calculation in a modulation-doped two-dimensional electron gas. Phys Rev B 57:10017

Reed MA, Randall JN, Aggarwal RJ, Matyi RJ, Moore TM, Wetsel AE (1988) Observation of discrete electronic states in a zero-dimensional semiconductor nanostructure. Phys Rev Lett 60:535

Reimann SM, Manninen M (2002) Electronic structure of quantum dots. Rev Mod Phys 74:1283

Rossi A, Tanttu T, Tan KY, Iisakka I, Zhao R, Chan KW, Tettamanzi GC, Rogge S, Dzurak AS, Möttönen M (2014) An accurate single-electron pump based on a highly tunable silicon quantum dot. Nano Lett 14:3405

Roth A, Brüne C, Buhmann H, Molenkamp LW, Maciejko J, Qi X-L, Zhang S-C (2009) Nonlocal transport in the quantum spin Hall state. Science 325:294

Roushan P, Seo J, Parker CV, Hor YS, Hsieh D, Qian D, Richardella A, Hasan MZ, Cava RJ, Yazdani A (2009) Topological surface states protected from backscattering by chiral spin texture. Nature 460:1106

Schomburg E, Henini M, Chamberlain JM, Steenson DP, Brandl S, Hofbeck K, Renk KF, Wegscheider W (1999) Self-sustained current oscillation above 100 GHz in a GaAs/AlAs superlattice. Appl Phys Lett 74:2179

Shen S-Q (2012) Topological insulators: Dirac equation in condensed matters. Springer, Heidelberg

Shin SJ, Lee JJ, Kang HJ, Choi JB, Yang S-RE, Takahashi Y, Hasko DG (2011) Room-temperature charge stability modulated by quantum effects in a nanoscale silicon island. Nano Lett 1:1591

Stern F (1972) Self-consistent results for *n*-type Si inversion layers. Phys Rev B 5:4891

Stern F, Das Sarma S (1984) Electron energy levels in $GaAs\text{-}Ga_{1-x}Al_xAs$ heterojunctions. Phys Rev B 30:840

Störmer HL, Dingle R, Gossard AC, Wiegmann W, Sturge MD (1979) Two-dimensional electron gas at a semiconductor-semiconductor interface. Solid State Commun 29:705
Störmer HL, Gossard AC, Wiegmann W (1982) Observation of intersubband scattering in a 2-dimensional electron system. Solid State Commun 41:707
Tarucha S, Austing DG, Honda T, van der Hage RJ, Kouwenhoven LP (1996) Shell filling and spin effects in a few electron quantum dot. Phys Rev Lett 77:3613
Thomas KJ, Nicholls JT, Appleyard NJ, Simmons MY, Pepper M, Mace DR, Tribe WR, Ritchie DA (1998) Interaction effects in a one-dimensional constriction. Phys Rev B 58:4846
Thornton TJ (1994) Mesoscopic devices. Rep Prog Phys 58:311
Thornton TJ, Pepper M, Ahmed H, Andrews D, Davies GJ (1986) One-dimensional conduction in the 2D electron gas of a GaAs-AlGaAs heterojunction. Phys Rev Lett 56:1198
Thornton TJ, Roukes ML, Scherer A, Van de Gaag BP (1989) Boundary scattering in quantum wires. Phys Rev Lett 63:2128
Thouless DJ, Kohmoto M, Nightingale MP, den Nijs M (1982) Quantized Hall conductance in a two-dimensional periodic potential. Phys Rev Lett 49:405
Tkachenko OA, Tkachenko VA, Baksheyev DG, Pyshkin KS, Harrell RH, Linfield EH, Ritchie DA, Ford CJB (2001) Electrostatic potential and quantum transport in a one-dimensional channel of an induced two-dimensional electron gas. J Appl Phys 89:4993
Tong D (2016) The quantum Hall effect. Tutorial accessible at http://www.damtp.cam.ac.uk/user/tong/qhe.html
van Wees BJ, van Houten H, Beenakker CWJ, Williamson JG, Kouwenhoven LP, van der Marel D, Foxon CT (1988) Quantized conductance of point contacts in a two-dimensional electron gas. Phys Rev Lett 60:848
Vitiello MS, Scalari G, Williams B, De Natale P (2015) Quantum cascade lasers: 20 years of challenges. Opt Express 23:5167
Vitiello MS, Tredicucci A (2021) Physics and technology of Terahertz quantum cascade lasers. Adv Phys X 6(1):1893809
Voisin P, Bleuse J, Bouche C, Gaillard S, Alibert C, Regreny A (1988) Observation of the Wannier-Stark quantization in a semiconductor superlattice. Phys Rev Lett 61:1639
Wacker A (2002) Semiconductor superlattices: a model system for nonlinear transport. Phys Rep 357:1
Wang H, Gupta SK, Xie B, Lu M (2020) Topological photonic crystals: a review. Front Optoelectron 13:50
Wen XG (1995) Topological orders and edge excitations in fractional quantum Hall states. Adv Phys 44:405
Wharam DA, Thornton TJ, Newbury R, Pepper M, Ahmed H, Frost JEF, Hasko DG, Peacockt DC, Ritchie DA, Jones GAC (1988) One-dimensional transport and the quantisation of the ballistic resistance. J Phys C Solid State Phys 21:L209
Wiedmann S, Molenkamp LW (2015) Topological insulators in two dimensions. In: Ortmann F, Roche S, Valenzuela SO (eds) Topological insulators. Wiley-VCH, Weinheim, pp 31–54
Wu C, Bernevig BA, Zhang S-C (2006) Helical liquid and the edge of quantum spin Hall systems. Phys Rev Lett 96:106401
Xia Y, Qian D, Hsieh D, Wray L, Pal A, Lin H, Bansil A, Grauer D, Hor YS, Cava RJ, Hasan MZ (2009) Observation of a large-gap topological-insulator class with a single Dirac cone on the surface. Nat Phys 5:398
Xiao D, Yao Y, Feng W, Wen J, Zhu W, Chen X-Q, Stocks GM, Zhang Z (2010) Half-Heusler compounds as a new class of three-dimensional topological insulators. Phys Rev Lett 105:096404
Xie B, Wang H-X, Zhang X, Zhan P, Jiang J-H, Lu M, Chen Y (2021) Classical higher-order topological insulators. Nat Rev Phys 3:520
Yan B, Zhang S-C (2012) Topological materials. Rep Prog Phys 75:096501
Zangeneh-Nejad F, Alù A, Fleury R (2020) Topological wave insulators: a review. CR Physique 21:467

Zhang T, Cheng P, Chen X, Jia J-F, Ma X, He K, Wang L, Zhang H, DaiX, Fang Z, Xie X, Xue Q-K (2009) Experimental demonstration of topological surface states protected by time-reversal symmetry. Phys Rev Lett 103:266803

Zhang Y, He K, Chang C-Z, Song C-L, Wang L-L, Xi Chen, Jia J-F, Fang Z, Dai X, Shan W-Y, Shen S-Q, Niu Q, Qi X-L, Zhang S-C, Ma X-C, Xue Q-K (2010) Crossover of the three-dimensional topological insulator Bi_2Se_3 to the two-dimensional limit. Nat Phys 6:584

Zhang Y, Klann R, Ploog K, Grahn HT (1997) Observation of bistability in GaAs/AlAs superlattices. Appl Phys Lett 70:2825

Zheng HZ, Wei HP, Tsui DC, Weimann G (1986) Gate-controlled transport in narrow GaAs/Al_xGa_{1-x}As heterostructures. Phys Rev B 34:5635

Carrier Transport Induced and Controlled by Defects

28

Karl W. Böer and Udo W. Pohl

Contents

1 Impurity-Band Conduction 1166
1.1 Concept of Impurity Bands 1166
1.2 The Impurity Band 1168
2 Phonon-Activated Conduction 1172
3 Heavily Doped Semiconductors 1175
3.1 Intermediate Doping Range 1175
3.2 Degenerate and Highly Compensated Heavily Doped Semiconductors 1177
4 Transport in Amorphous Semiconductors 1178
4.1 The Mobility Edge 1180
4.2 Diffusive Carrier Transport and Percolation 1181
4.3 Activated Mobility 1182
4.4 Temperature Dependence of the Conductivity 1185
5 Charge Transport in Organic Semiconductors 1187
5.1 Band Conductance in Organic Crystals 1188
5.2 Hopping Conductance in Disordered Organic Semiconductors 1192
6 Summary 1196
References 1198

Abstract

With a large density of impurities or other lattice defects, the carrier transport deviates substantially from the classical transport within the band. It is carried within energy ranges (within the bandgap), which are determined by the defect

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_28

structure. *Heavy doping* produces predominant defect levels split into two impurity bands. Below a density to permit sufficient tunneling, carrier transport requires excitation into the conduction band; at higher defect density, a diffusive transport within the upper impurity band becomes possible. At further increased defect density, metallic conductivity within the then unsplit impurity band occurs.

In *amorphous semiconductors*, tunneling-induced carrier transport can take place within the tail of states, which extend from the conduction or valence band into the bandgap. Major carrier transport starts at an energy referred to as the mobility edge. With statistically distributed defects, only some volume elements may become conductive. These volume elements widen at increasing temperature, eventually providing an uninterrupted *percolation path* through a highly doped or disordered semiconductor with a density-related threshold of conduction.

Conductance in *organic semiconductors* is governed by static and dynamic disorder. Band conductance in small-molecule crystals shows a decreasing carrier mobility at increased temperature with a power law similar to that of inorganic semiconductors. Small-molecule or polymer semiconductors with dominating static disorder show hopping conductance with a typically low but increasing mobility at higher temperatures.

Keywords

Amorphous semiconductor · Band conductance · Dispersive transport · Heavy doping · Hopping conduction · Hopping mobility · Impurity band · Mobility edge · Mott transition · Organic semiconductor · Percolation · Phonon-activated conduction · Tunneling-induced transport · Variable-range hopping

1 Impurity-Band Conduction

In addition to causing scattering, lattice defects can contribute directly to the carrier transport in two ways. They permit direct quantum-mechanical exchange of carriers from defect to defect (i.e., tunneling from one trap level to the next), or by thermal ionization of a carrier from a trap level into the band, intermediate transport within the band, and then a retrapping as shown in Fig. 1. The first type of carrier transport is called *tunneling* or *impurity-band conduction*; the second type is known as *hopping conduction* or *phonon-activated conduction.* These types of carrier transport are of major importance in highly disordered, highly doped, or amorphous semiconductors. For reviews, see Shklovskii and Efros (1984) and Mott (1993).

1.1 Concept of Impurity Bands

In a rather simple model, the overlapping hydrogen-like donor states can be used to form a *Hubbard band* (Hubbard 1963), which is centered about their ground-state energy. With partial compensation, there are $N_d - N_a$ free states in this Hubbard band, and conduction can occur (Adler 1980). The band width ΔE_{imp} is given by the overlap

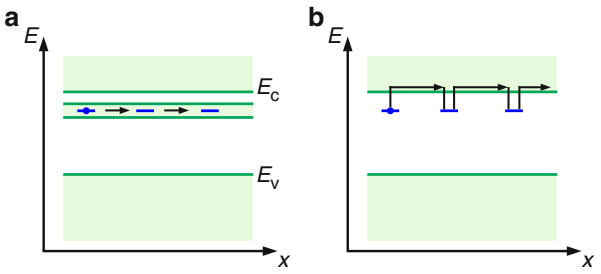


Fig. 1 (**a**) Impurity-band conduction and (**b**) hopping conduction in highly doped semiconductors

integral between equal centers at distance $1/\sqrt[3]{N}$, where N is the density of uncompensated donors ($= N_d - N_a$). This bandwidth is roughly equal to the interaction energy

$$\Delta E_{\text{imp}} \cong \frac{e^2 \sqrt[3]{N}}{4\pi\varepsilon_{\text{stat}}\varepsilon_0}. \tag{1}$$

The effective mass within this narrow impurity band is much larger than in the adjacent carrier band; hence, the *impurity-band mobility* is usually quite small ($< 10^{-2}$ cm^2/(Vs)). The mobility-effective mass in this impurity band should not be confused with the effective mass of a Bloch electron within the conduction band, which is hence related to each quasihydrogen state of the donor.

The carrier transport within such a narrow impurity band can no longer be described by the Boltzmann equation. The carrier transport must now be evaluated from the quantum-mechanical expectation value for the current, which is given by the *Kubo formula* (Kubo 1956, 1957). A somewhat simplified version was developed by Greenwood (1958)[1] in which the conductivity can be expressed as

$$\sigma = -\int \sigma_0(E) \frac{\partial f}{\partial E} dE, \tag{2}$$

with

$$\sigma_0(E) = \frac{2\pi e^2 \hbar^2 V}{m_n^2} g(E) \left| \int \psi^*(E') \frac{\partial}{\partial \mathbf{r}} \psi(E) d\mathbf{r} \right|^2, \tag{3}$$

being proportional to $g(E)$, the density of states, and the matrix element describing the electron transitions from E to E'; f is the Fermi distribution function and V is the sample volume. Equation 2 is referred to as the *Kubo-Greenwood formula*, which, when evaluated for $E = E'$, gives the tunneling current between equivalent defect centers.

[1] See also Zhang and Drabold (2011).

The distance between the impurities is not constant but fluctuates statistically; the impurity band is therefor substantially undulated. It is broader where impurities are closer together and narrower where they are more widely spaced.

Since there is no scattering during the tunneling process between adjacent defects, the tunneling is essentially temperature-independent. Except for thermal expansion, which has a small influence on the average distance between defects, and except for the broadening of the defect levels with increased lattice oscillation, the *trap conductivity* is almost temperature-independent when the Fermi level lies close to the extended states. Trap conductivity is important in highly doped semiconductors, semiconducting glasses, and inorganic semiconductors, that is, in all semiconductors with a high density of defects (see Sect. 4).

When carriers are created by optical excitation, trap conductivity persists to low temperatures and has a quasi-metallic behavior (Mott and Davis 1979). We will now describe in more detail the impurity band.

1.2 The Impurity Band

In semiconductors with high doping densities ($> 10^{18}$ cm^{-3}), shallow donors or acceptors can come close enough (<100 Å) to each other so that their eigenfunctions overlap significantly and therefore permit the exchange of carriers directly, without the involvement of the adjacent bands. Consequently, the defect level is split and develops into a narrow *impurity band* – see Fig. 1a. Such impurity-band formation is a basic effect that occurs whenever a defect level is present at sufficient density.[2] The formation of an impurity band is widely applied in blocked-impurity-band devices, where a heavily – but not degenerately – donor-doped layer creates an impurity band; this region is used as an infrared-active layer in IR photo detectors, where an incoming photon lifts an electron from the impurity band to the conduction band (Haegel et al. 2003, Wang et al. 2015).

The term "band" should, however, be used with caution, as it requires a more detailed density-of-states analysis and a distinction between localized and delocalized states; the latter are true band states. In principle, the Anderson Model (Sect. 1.1 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors") should be used to obtain some information about the localization aspect of the states. We will first discuss this behavior in a rather general fashion.

1.2.1 The Lifshitz-Ching-Huber Model

In the Lifshitz model, a statistical distribution of N identical potential wells is analyzed to obtain a density-of-states distribution of these defect levels and to identify a critical density at which the states within the center of the distribution become delocalized (Lifshitz 1965). This model is a forerunner of the Mott version,

[2] E.g., in *p*-type GaN:Mg the formation of an impurity band with a hole concentration in the mid 10^{18} cm^{-3} range was found (Gunning et al. 2012).

which is used to distinguish localized and nonlocalized states in band tails, see Sect. 1.2 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors" and Sect. 4.1.

When two identical defect centers are brought together, they show a split of eigenstates of the form

$$\psi_s = \frac{1}{\sqrt{2}}(\phi_1 + \phi_2) \quad \text{and} \quad \psi_a = \frac{1}{\sqrt{2}}(\phi_1 - \phi_2) \quad \text{with} \quad E_s - E_a = 2I, \tag{4}$$

where ϕ_1 and ϕ_2 are the wavefunctions of the two centers, E_s and E_a are the energies of the symmetric and antisymmetric states ψ_s and ψ_a, respectively, and I is the transfer integral (Eq. 2 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors"). When a third center is approaching at an arbitrary distance, it will not, however, participate in the resonance splitting. This is due to the fact that the doublet of the two centers closest to each other is far enough apart to be out of resonance with the third center. Hence, the Lifshitz model yields a band of *localized* states for statistically distributed traps. Only when these defects are close enough to fulfill

$$N^{1/3} r_0 \cong 0.3, \tag{5}$$

with r_0 as the fall-off radius of the wavefunction of an isolated defect ($= a_{qH}$ for a hydrogen-like defect) can delocalization of the states in the center of the band occur (Ching and Huber 1982). This result is close to the Mott-Anderson result for localization (see Eq. 17 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors". When interacting with each other, the splitting of the defect states also gives rise to a splitting of this defect band, causing the density of states to have a minimum near the center of the distribution.

1.2.2 Coulomb Gap and Mott Transition

The density minimum of the defect levels near the center of the density-of-states distribution may become complete in (partially) compensated semiconductors with a gap between filled and empty states. Such a splitting is caused by the long-range Coulomb interaction of *localized* electrons (Knotek and Pollak 1974; Efros and Shklovskii 1975) and occurs at the position of the Fermi level. For an inclusion of static screening, see Mazuruk et al. (1989); such a screening affects the density of states at the Fermi level and can replace the Coulomb gap by a dip in the density of states.

It can be shown that the Coulomb gap appears only for *localized* states. When the density of states becomes large enough, so that delocalization occurs (see below), then the Coulomb gap disappears (Aronov et al. 1979; Altshuler et al. 1980).

The impurity band with localized states cannot contribute to the conductivity at $T = 0$ K since it has an energy gap between filled and empty states. When the density of impurities is increased to an extent that delocalization occurs at the Fermi level, the gap disappears, and quasi-metallic conductivity is observed. This transition

within an impurity band is a *Mott-transition* and is related to a critical conductivity that was termed by Mott as *minimum metallic conductivity*

$$\sigma_{\min} \cong 0.05 \frac{e^2}{\hbar} N_{\text{Mott}}^{1/3}, \tag{6}$$

where N_{Mott} is the critical doping density–see below (Mott and Davis 1979). The Mott transition is observed to be smooth rather than abrupt, probably because of density fluctuation of impurities.

For Si:P, the Mott-transition occurs at a critical donor density of $N_{\text{Mott}} = 3.7 \times 10^{18}$ cm^{-3}; for this example, we obtain $\sigma_{\min} \cong 20\ \Omega^{-1}$ cm^{-1} (Rosenbaum et al. 1980) (see also Sect. 3.2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"). More recently the Anderson-Mott transition was studied from the isolated donor regime to the intermediate doping-density regime (characterized by impurity-band formation) by implanting single arsenic dopant atoms into the active channel of a Si transistor; at the critical density, experimental evidence for the formation of a lower and an upper Hubbard impurity band was found (Prati et al. 2012).

Scaling When, for example, hydrogen-like donors are close enough to each other, the donors' electrons no longer belong to a certain donor, but are able to move freely between donors even at $T = 0$ K, like electrons in a metal; that is, they belong to all of the donors. Critical to this transition are three units of length: the inter-donor distance $1/\sqrt[3]{N}$, the quasi-hydrogen radius a_{qH}, and the mean free path λ – for a diffusion-type of carrier migration. Their relative magnitude determines the type of conductivity, and its discussion is a subject of the theory of scaling. For a tight-binding analysis of localization in 3D to 1D systems with hopping matrix-elements that decay exponentially in the separation distance between neighboring sites, see Priour (2012); extended states in 3D are found to occur even for small decay lengths, but the interval of energies supporting extended states decreases exponentially for decreasing decay length.

Abrahams et al. (1979) suggested to use a dimensionless conductance of a cube of length L rather than the conductivity

$$G = \sigma L \times \frac{2\hbar}{e^2}, \tag{7}$$

measured in elementary units of $2\hbar/e^2$. They discussed the changes in G as a function of L; it should change when L approaches atomic dimensions. They argued that the scaling function

$$\beta(G) = \frac{\partial \ln G}{\partial \ln L} \tag{8}$$

is a universal function (Thouless 1974, 1980), which is ~1 for large conductances, becomes zero at a critical conductance G_c, and turns negative for $G < G_c$–see Fig. 2.

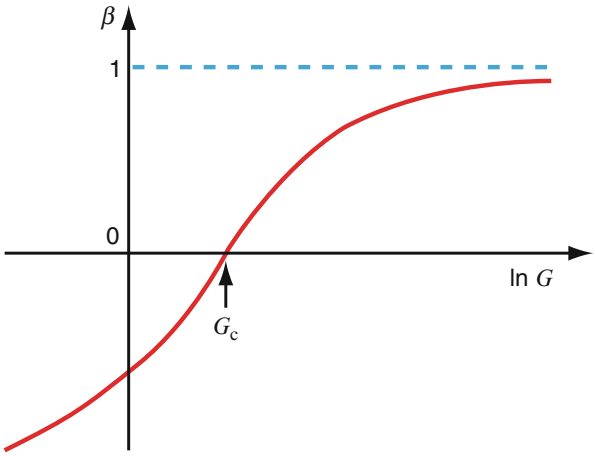


Fig. 2 Scaling function versus the dimensionless conductance (Eq. 7) for a 3D semiconductor

Within this theory, G_c is a universal constant[3] and indicates the transition between metal-like and semiconductor-type conductivity. Here, Mott obtains for the critical conductivity

$$\sigma_c \cong 0.03 \frac{e^2}{\hbar \ a_{qH}}, \tag{9}$$

a value close to σ_{min} given by Eq. 6, here for $\sqrt[3]{N_{Mott}} \cong 1/a_{qH}$.

Ioffe-Regel Rule For a further discussion of the concept of minimum metallic conductivity, we start from a metal and look for candidates of lower and lower mean free paths, that is, reduced conductivities. There are indications that with increased lattice disturbance, lower conducting metals (such as liquid metals) have a lower mean free path, but with a lower limit equal to the interatomic distance. The *Ioffe-Regel rule* argues that the electron wavefunction cannot lose phase memory faster than on the order of the interatomic distance a (Ioffe and Regel 1960). This means that the conductivity of a metal cannot be smaller than $\sigma = e\,\mu\,n$, with $n = a^{-1/3}$ and $\mu = (e/m_n)\,\tau = (e/m_n)\,(a/v_F)$. The Fermi-momentum is given by $\hbar\,k_F = m_n\,v_F = m_n(3\pi^2 n)^{1/3}$; hence, one obtains as minimum metallic (Ioffe-Regel) conductivity

$$\sigma_{I\text{-}R} = \frac{1}{\sqrt[3]{3\pi^2}} \frac{e^2}{\hbar\, a} = 0.32 \frac{e^2}{\hbar\, a}. \tag{10}$$

[3]It should, however, depend on the microscopic atomic arrangement and on the coordination (Mott and Kaveh 1985).

In doped semiconductors, two changes need to be introduced:

1. Instead of the interatomic distance, the quasi-hydrogen radius applies
2. Only a certain fraction of the impurity-band states are extended states, which, after Mott and Kaveh (1985), is on the order of 8.5%, yielding Eq. 9 as critical conductivity in a semiconductor. For $a_{qH} \cong 30$ Å, this critical conductivity is on the order of 20 Ω^{-1} cm^{-1}.

Carrier Localization in Strong Electric Fields When carriers are transported in narrow bands, independent of how such bands are produced, carrier localization can occur when the electric field is strong enough. Here, stationary electron states become localized in the direction of the electric field due to reflection at the boundaries of the Brillouin zone (Wannier 1960). This causes a *Stark ladder*, with the possibility of phonon-induced jumps between the levels of this ladder (Hacker and Obermair 1970). Resonance effects occur when the steps become equal to LO phonons (Maekawa 1970), causing current oscillations.

Another possibility of carrier localization occurs for small polarons in strong electric fields, where the mobility decreases with increasing field in the tunneling regime (Böttger and Bryksin 1979, 1980). For electric-field-induced carrier localization in one-dimensional semiconductors, see Pronin et al. (1994).

2 Phonon-Activated Conduction

For sufficiently high densities of impurities, the carrier transport within an impurity band occurs with a mean free path longer than the spacing of impurities. With less doping, the defect levels will become localized, and conduction can occur in one of two fashions:

1. By tunneling from one defect to the nearest neighboring defect of the same type
2. After thermal excitation into the adjacent band

Competition between these two processes is exponentially dependent on the temperature, due to a minor T dependence of the former and an Arrhenius-type dependence of the latter. At sufficiently high temperatures, the carrier transport via the conduction or valence band predominates.

If the mean free path of carriers is given by capture at impurities rather than by scattering, the conductivity can be described as a motion from one impurity center to another, but with electron transport *through the conduction band*[4] as illustrated in Fig. 1b (Fritzsche and Cuevas 1960; Butcher 1972). It can also be described as due to

[4]In highly disordered semiconductors the motion may occur through excited states with greater overlap of their eigenfunctions.

inelastic scattering at Coulomb-attractive centers, with phonon emission causing carrier capture. The corresponding carrier mobility is thermally activated:

$$\mu = \mu_0 \exp\left(-\frac{\Delta E_{\text{trap}}}{kT}\right), \tag{11}$$

where μ_0 is an effective mobility given by equivalent scattering mechanisms of carriers within the band, and ΔE_{trap} is the thermal activation energy of the trap.

Density Dependence of Hopping When the density of impurities increases, tunneling from center to center becomes more probable. The tunneling transport is accomplished by *hopping* from one to the adjacent center, resulting in a conductivity

$$\sigma_{\text{hop}} = \sigma_{0,\text{hop}} \exp\left(-\frac{\Delta E_{\text{hop}}}{kT}\right), \tag{12}$$

with ΔE_{hop} as the activation energy for hopping described below (Eq. 15) and with the preexponential factor given by

$$\sigma_{0,\text{hop}} = \sigma_{00} \exp\left(\frac{2r_c}{r_0}\right), \tag{13}$$

where r_0 is the fall-off radius of the impurity wavefunction ($= a_{\text{qH}}$ for quasi-hydrogen impurities). It is ~90 Å for Ga in Ge and ~47 Å for Cu in Ge; Cu has a larger ionization energy of ~40 meV. Also, r_c is the critical radius to establish a

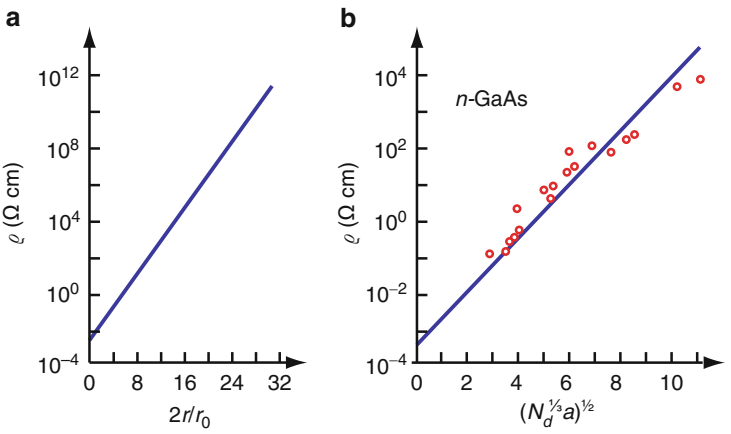


Fig. 3 (**a**) Calculated resistivity related to hopping conductivity as a function of the average separation between impurities. (**b**) Hopping resistivity as a function of the donor density in *n*-type GaAs (After Shklovskii and Efros 1984)

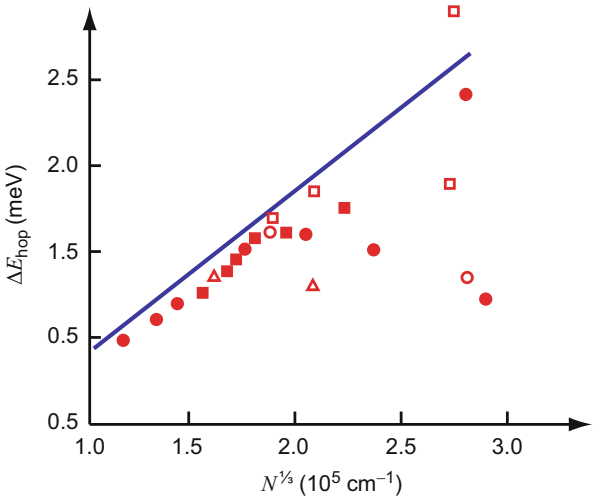


Fig. 4 Activation energy for hopping conduction in Ge doped with P, Ga, or Sb. The *solid line* is the theoretical dependence (Eq. 15) (After Shklovskii and Efros 1984)

percolation path (see Sect. 4.2) from one electrode to the other, and can be estimated (McInnes and Butcher 1979) as

$$r_c \cong (0.865 \pm 0.015) N^{-1/3}, \tag{14}$$

where N is the density of the specific impurities between which hopping occurs.

The computed relation of the resistivity $\rho = \sigma^{-1}$ versus the mean separation of the impurities is shown in Fig. 3a. The corresponding relation of the measured resistivity for different donor densities N_d in GaAs is given in Fig. 3b (Shklovskii and Efros 1984); the solid curves gives the theoretical estimate according to Eqs. 13 and 14.

Activation Energy for Hopping When impurities are spaced close enough to permit tunneling, the levels split to form a narrow band. Therefore, tunneling to arbitrary neighbors usually requires a slight thermal activation energy (see Sect. 1.2.1). The activation energy can be interpreted as the energy from the Fermi level to the energy of the maximum of the density of empty-state distribution. Typically, it is on the order of a few meV and can be approximated for low compensation (Efros et al. 1972) by ~60% of the Coulomb energy at the average separation between the impurities:

$$\Delta E_{hop} \cong 0.61 \frac{e^2}{4\pi \varepsilon_{stat} \varepsilon_0} \left(\frac{4\pi}{3} N\right)^{1/3}. \tag{15}$$

The experimental values for ΔE_{hop} for Ge doped with P, Ga, or Sb are shown together with the theoretical curves (Eq. 15) in Fig. 4.

With a distribution of defects in space and energy, the relation becomes more complex and is relevant for amorphous semiconductors, see Sect. 4.3. For reviews, see Mott and Davis (1979), Shklovskii and Efros (1984) and Mott (1993).
A special type of hopping conduction relates to the hopping of small polarons and is discussed by Holstein (1959) and Schnakenberg (1968) (see also Sect. 5).

3 Heavily Doped Semiconductors

The basic concepts discussed in the previous sections apply also for highly doped semiconductors, however, in a modified fashion relating to the specific level distribution. This permits a number of more transparent theoretical approximations.

A semiconductor is heavily doped when the condition

$$N\ a_{\mathrm{qH}}^3 \geq 1 \tag{16}$$

is fulfilled, which, dependent on the effective mass, is reached at vastly different doping densities in various semiconductors. For instance, $N\ a_{\mathrm{qH}}^3 = 1$ requires $N = 5 \times 10^{15}$ cm^{-3} in n-type InSb and $N{=}3{\times}10^{19}$ cm^{-3} in n-type Ge. In several semiconductors, the highly doped regime cannot be obtained by diffusion doping, since clusters will form with limited solubility. Here, ion implantation or radiation damage can be used.

We will give a short review of the phenomena related to heavy doping in the following sections.

3.1 Intermediate Doping Range

The distinction between light and heavy doping can be made in relation to the disappearance of the gap in the impurity band and the transition from an activated semiconductivity to a quasi-metallic conduction (see Sect. 1.2.2). There is, however, a large intermediate range between the Mott-transition at $Na_{\mathrm{qH}}^3 \cong 0.02$ (Eq. 65 of ▶ Chap. 21, "Equilibrium Statistics of Carriers") and the range of heavily doped semiconductors which starts at $Na_{\mathrm{qH}}^3 \cong 1$. In this intermediate range, some of the electrons are already delocalized.

The transition is related to the statistical distribution of the defects, which are frozen in and are located within the ensemble of free electrons, even at low temperatures. We will give some insight into this relation below. Other fluctuations are initiated at higher temperatures (*fluctuons*) and are reviewed by Krivoglaz (1974).

Near the Mott density transport may occur through two competing bands: the conduction band and the impurity band. In heavily doped n-type Ga_2O_3:Sn dominant conduction due to free electrons in the conduction band was observed at $T > 200$ K,

while conduction through the impurity band with a low mobility was found to prevail at $T < 200$ K; the latter takes place by variable-range hopping with a temperature-dependent resistivity $\ln(\rho) \propto T^{-1/4}$ (Kabilova et al. 2019). In heavily doped p-type 4H-SiC:Al with $c_{Al} = 1.8 \times 10^{20}$ cm^{-3} the transition from band conduction to variable-range hopping was found at $T \cong 50$ K, while at higher doping (3.5×10^{20} cm^{-3}) the latter becomes the dominant conduction mechanism over the entire range of measured temperatures (Matsuura et al. 2019).

Density of States in Heavily Doped Semiconductors In highly doped semiconductors, there are two major contributions to the density of states: (1) the states which are due to the extended eigenfunctions of the defects and (2) the states which are due to the perturbation in the surrounding host lattice. The latter may be described by analyzing the influence of heavy doping on free electrons. This influence can be expressed by band-edge perturbation, through the modulation of the band edges by the Coulomb potential of the defects (Kane 1963; Halperin and Lax 1966). In highly doped semiconductors, clusters of charged impurities often dominate. The charges of such clusters, however, are *not* Coulomb point charges.

In turn, the potential fluctuation near charged impurities results in an inhomogeneous distribution of electrons. When the potential fluctuation is smooth within the deBroglie wavelength of free electrons, the electron gas can be described classically. Its density varies according to the density of states, which is increased at positions near an attractive center where the conduction band is lowered. Near attractive centers there will be more carriers, while near repulsive centers there will be less of the corresponding type. With high doping densities, the potential fluctuations will have a higher amplitude. Complete state filling of the valleys occurs at sufficiently low temperatures, whereas higher parts of the potential mountains extend above this "electron lake".

The density of states now becomes space-dependent

$$g(E, \mathbf{r})\ dE = \frac{(2m_n)^{3/2}}{2\pi^2\hbar^3}\sqrt{E + eV(\mathbf{r})}, \tag{17}$$

with the fluctuating potential determined by a screened Coulomb potential

$$V(\mathbf{r}) = \frac{e}{4\pi\,\varepsilon_{stat}\varepsilon_0\,\mathbf{r}}\exp\left(-\frac{r}{r_0}\right). \tag{18}$$

When averaging over the space-dependent potential, we obtain from Eq. 17 the density of states tailing into the bandgaps (*Lifshitz tail*), as discussed in Sect. 3.1 of ▶ Chap. 17, "Optical Properties of Defects", and Sect. 1 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors".

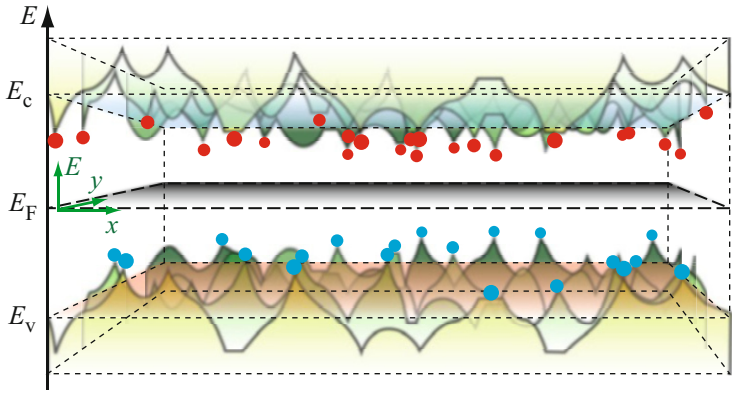


Fig. 5 Two-dimensional representation of the band-edge fluctuation in highly doped semiconductors. E_c and E_v refer to *average* values of conduction and valence band edges. *Red* and *blue circles* signify ionized donors and acceptors, respectively

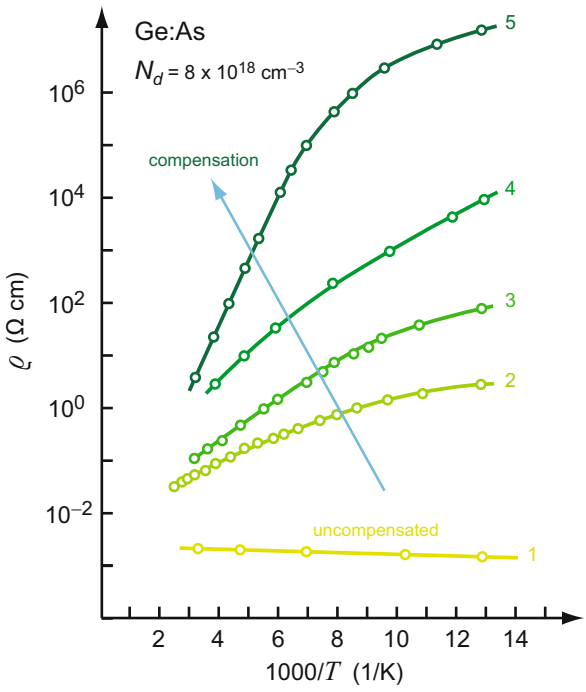


Fig. 6 Resistivity of an uncompensated (*lowest curve*) to more and more compensated samples of *n*-type Ge:As ($N_d = 8 \times 10^{18}$ cm^{-3}) as function of the inverse temperature (After Gadzhiev et al. 1972)

3.2 Degenerate and Highly Compensated Heavily Doped Semiconductors

The highly doped semiconductor with shallow impurities is usually degenerate, i.e., the Fermi-level is shifted to well within the band. Depending on its

position, the "lake" of electrons rises within a hilly terrain to fill only the lowest valleys as little lakes, or with a rising level connects more and more lakes until navigation from one electrode to the other becomes possible. This behavior is similar to that of a percolation conductivity, as described in Sect. 4.2.

When compensating a highly doped semiconductor, the level of the carrier lake within the modulated band drops, which causes a substantial decrease of the conductivity.

With sufficient compensation, the semiconductor reverts from metallic conduction to one with thermal activation over saddle points in the hilly terrain, as shown in Figs. 5 and 6. Here, carriers cannot contribute to percolation since they occupy only a small fraction of the volume, and tunneling is too expensive because of the high barriers between the remaining small puddles.

In completely compensated semiconductors, the potential fluctuation can increase further since the density of carriers is reduced below values, which are necessary for efficient screening. The maximum band fluctuation, however, is limited to $E_g/2$. Here, with penetration through the Fermi level, the deep valleys again become filled and screening reappears, thereby limiting any further increase in fluctuation of the band edges. For the effect of random defect density on the Fermi level in highly compensated semiconductors, see Donchev et al. (1997).

There is a large body of experimental and theoretical research on highly doped semiconductors, including the influence of light (*persistent photoconductivity* - Ryvkin and Shlimak 1973), of a magnetic field (*quantum screening* - Horring 1969, Aronzon and Chumakov 1994), and of low-temperature conductivity.

The carrier transport in such macroscopically fluctuating potentials is similar to that in semiconducting glasses (Ryvkin and Shlimak 1973, Overhof and Beyer 1981). In the following sections, we will analyze such a transport in more detail.

4 Transport in Amorphous Semiconductors

The carrier transport in semiconducting glasses deserves a separate discussion because of the lack of long-range order and the high density of defects specific to the amorphous material. This does not permit simple translation of the effective-mass picture and requires a reevaluation of carrier transport and scattering concepts.

There are two aspects with direct influence on the carrier transport: (1) the strong tailing of states into the bandgap, and (2) the absence of specific doping-induced defects in amorphous chalcogenides.

The *tailing of states* from the band into the bandgap is rather smooth and does not show a well-defined band edge. This necessitates a more careful analysis of the

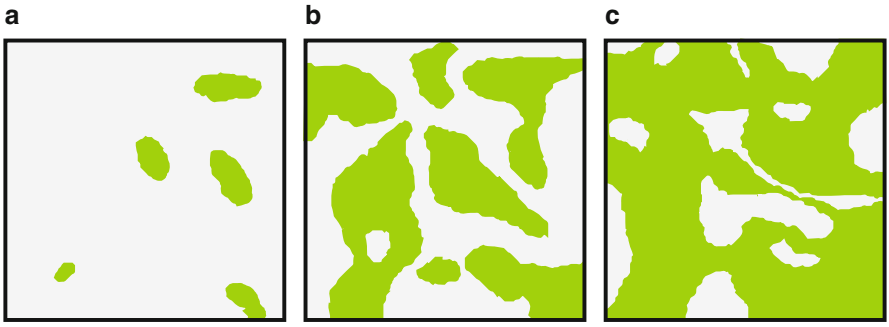


Fig. 7 Percolation regions (*green*), which become larger and interconnect with increasing energy from (**a**) to (**c**), connect as puddles in a hilly terrain to form small and larger lakes when the water level rises, finally leaving only small islands near the highest points of the terrain

mobility at different energies. At higher energies within the conduction band – i.e., closer to the surface of the Brillouin zone (see below for justification) – the electrons are quasi-free, except for scattering events, and have a mean free path λ larger than the interatomic spacing. Here, $\lambda k \gg 1$, and k may be used, albeit with some caution (Mott and Davis 1979).

With decreasing electron energy, the scattering probability increases, with scattering on potential fluctuations due to noncrystalline structures. Hence, the mean free path becomes comparable to the interatomic distance, and $\lambda k \cong 1$. Here, k is no longer a good quantum number.[5] Substantial differences between the crystalline and the amorphous semiconductor become important. Therefore, the carrier transport must now be described in terms of a transport between *localized states*; the Mott-Anderson localization threshold is reached (see Sect. 1.2 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors").

In taking a slightly different point of view, we expect the band states near the *edge* to become *perturbed* with a concurrent widening or narrowing of the bandgap, depending on the local degree of disorder. With charging of these defects, the Coulomb potential creates band undulations, as shown schematically in Fig. 5 (Böer 1972). In amorphous semiconductors with a much lower density of charged centers, a similar mountainous profile of the near-edge band states results from the local stress and other defect-induced perturbations.

When the Fermi- or quasi-Fermi level is moved above the lowest valleys of this edge (discussed in more detail in Sect. 4.2), these valleys will fill up with carriers. Assuming that only near the surface of such "lakes" a carrier transport is possible, one recognizes that a continuous current can only flow when the Fermi-level rises

[5] In a crystalline structure, k, when closer to the center of the Brilloin zone, represents points in real space farther away from the unit cell; in this case long-range deviation from periodicity becomes important. In contrast, when $\lambda k \cong 1$, the wavenumber is closer to the boundaries of the Brillouin zone; wheras in real space, the corresponding points are closer to the unit cell and the structure of the amorphous material resembles more that of a crystal.

enough to permit a *percolation path* from one to the other electrode (Fig. 7a–c); much below the "edge" carriers are trapped. We will now refine this roughly stated model.

4.1 The Mobility Edge

Carriers are able to travel readily when the eigenfunctions of traps overlap. There are two arguments for a larger overlap of shallower traps: (1) they usually have a larger fall-off radius of their eigenfunctions and (2) they are more plentiful, diminishing the intertrap distance. With an exponentially decreasing trap distribution, and with an adequately large overlap of shallow traps, the carrier transport through such shallow centers is almost band-like.

At slightly deeper trap levels, the carrier transport proceeds from trap to neighboring trap and has a diffusive character, with a diffusion constant given by the exchange frequency ν_{trap}:

$$D = \frac{\nu_{\text{trap}}\, a_{\text{trap}}^2}{6}; \tag{19}$$

here a_{trap} is the distance between these traps.

Carriers in yet deeper traps will have to penetrate through increasingly thicker barriers via tunneling. Finally, such carrier transfer via tunneling becomes negligible and requires thermal excitation into higher states.

In summary, the type of carrier transport depends on the depth of the traps between which such transport takes place. Carriers are significantly more mobile in shallower traps. There is a major step in the mobility of carriers between "localized" deeper and "extended" shallow trap states. This step at an energy E_μ is referred to as the *mobility edge*; see also Sect. 1.3 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors".

A material in which the Fermi level at $T = 0$ K coincides with the mobility edge is called a *Fermi glass*. This material displays metallic conductivity.

The distance between two defect centers at the mobility edge is approximately that of the nearest neighbors (Mott and Davis 1979), yielding for approximately cubic atom configurations (see Sect. 3.1 and Eq. 80 of ▶ Chap. 15, "Crystal Defects"):

$$D_\mu \cong \frac{\nu_\mu\, a^2}{6}, \tag{20}$$

with the tunneling frequency ν_μ approximated by the atomic electron frequency

$$\nu_\mu \cong \frac{\hbar}{m_0\, a^2}. \tag{21}$$

We are *not* assuming hydrogen-like defects here. Therefore, tunneling is much reduced, and the distance of defects centers between which tunneling becomes significant can no longer be much larger than the normal interatomic distance *a*. Using the Einstein relation, this yields, as an order of magnitude estimate for the mobility at the mobility edge,

$$\mu_{\text{edge}} = \frac{e\ D_\mu}{kT} \cong \frac{e\ \hbar}{6 m_0 kT} \cong 7.5 \times \left(\frac{300\,\text{K}}{T}\right)\ \left(\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}\right). \tag{22}$$

Electrons that are excited much above the mobility edge contribute to the current, as they do in a crystal, by being scattered at defects, and usually have a mean free path that is much larger than the interatomic distance. Electrons closer to the mobility edge contribute via exchange interaction to neighboring traps, and electrons that have relaxed much below the mobility edge contribute through tunneling or after thermal activation.

Since the mobility decreases very steeply at the mobility edge, whereas the density of states does not, it is customary to identify the *bandgap* in amorphous semiconductors as a *mobility gap*, i.e., the distance between the mobility edges for electrons and holes.

4.2 Diffusive Carrier Transport and Percolation

We will now look a bit closer at the carrier transport around the mobility edge E_μ. With decreasing trap energy E_{trap}, the trap density is reduced and the average distance between these defects is increased. At any given energy, the distance will fluctuate about an average value, making carrier transfer preferred in directions in which the distance is shortest. With further decreasing E_{trap}, preferred paths become rarer. The carrier has to move in a diffusive path along preferred intertrap connections. This indicates that the carrier motion, which was randomly diffusive at higher energies, now becomes direction-selective toward the closest neighbor, thereby reducing the effective diffusion constant. Finally, the path connecting the two electrodes will be broken. From this point on, thermally activated conductivity becomes the sole possibility for carrier transport.

The selection of paths between neighboring sites at the mobility edge is significant in that it is a determinant of the *Hall mobility*. In amorphous semiconductors, the Hall effect cannot be calculated from Lorentz forces, but must be computed from quantum-mechanical jump probabilities between localized states (Grünewald et al. 1981). Paths following the Lorentz force become *slightly* preferred. Because of this structure-determined path selection, the Hall voltage becomes dependent on the average microscopic geometry of the atomic arrangement. An anomalous sign of the Hall effect may occur at weak fields for the mean free path below a critical magnitude (Okamoto et al. 1993). Mott (1991) has shown that the sign anomaly of the Hall coefficient in α-Si:H should be expected when the disorder causing the mobility edge originates from stretched bonds. With a scattering length of 20 Å in such amorphous materials at the mobility edge, only 1% of

the bonds need to be stretched to produce such an anomaly in the sign. Also preferred even- or odd-numbered rings (see Sect. 3.1 of ▶ Chap. 3, "The Structure of Semiconductors") cause a sign reversal of the Hall effect for *n*- or *p*-type material (see Dresner 1983).

Percolation We return to the carrier transport near the mobility edge. When filling traps by raising the Fermi-level, carrier diffusion is eased. This would appear homogeneously throughout the semiconductor if it were not for the mountainous profile of the potential, as illustrated in Fig. 5. Here, in a mountain, the mobility edge is pushed above the Fermi-level, whereas in a valley, the Fermi-level lies above the mobility edge. In these lakes, the mobile electrons show diffusive motion along the surface of the lakes, but have to tunnel through the mountains. This type of transport, which can be understood from classical arguments (Broadbent and Hammersley 1957), is commonly referred to as *percolation*. For a review, see Shante and Kirkpatrick (1971) or Böttger and Bryksin (1985).

The analysis of percolation was facilitated by the simple model of Miller and Abrahams (1960), using a network of random resistors and Kirchhoff's law to calculate the corresponding resistivity between the electrodes in a semiconductor with percolating conductivity. Many aspects of carrier percolation can be discussed in the frame work of *fractal networks*, i.e., a network of resistors in which a statistically increasing number of the interconnecting resistors are omitted. Multi-fractality in carrier transport at the mobility edge in amorphous semiconductors is discussed by Huckestein and Schweitzer (1993).

4.3 Activated Mobility

Further below the mobility edge, the defect centers are sufficiently separated so that tunneling between them can be neglected compared to the thermal excitation into levels near the mobility edge. From here, electrons can be retrapped, excited again, etc. This process can be described as *thermally activated hopping*[6] and requires a periodic interplay with phonons, i.e., carriers alternately absorb or emit phonons. Consequently, the hopping mobility depends exponentially on the temperature. For excitation from centers at an energy E_{trap}, we obtain

$$\mu_{\text{hop}} = \mu_0 \exp\left(-\frac{E_\mu - E_{\text{trap}}}{kT}\right). \tag{23}$$

The thermally activated hopping mobility can be described in the form of a diffusion relation (Butcher 1972):

[6]Hopping conduction can also involve small polarons which move by hopping from site to site, ions which hop from interstitial to interstitial site, electrons which hop between soliton-bound states in one-dimensional conductors (e.g., acetylene) (Kivelson 1982), or Frenkel excitons in molecular crystals (see Sect. 5 and references in Böttger and Bryksin 1985).

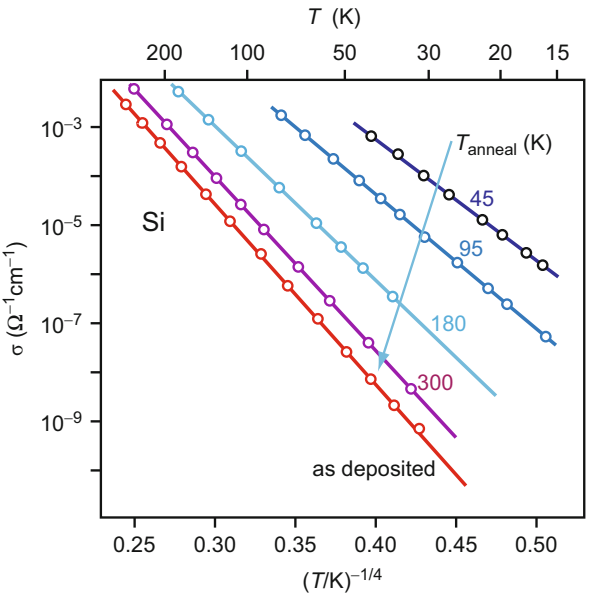


Fig. 8 Electrical conductivity of amorphous Si as a function of $T^{-1/4}$. Family parameter is the annealing temperature after damage with Si^+ bombardment (*uppermost curves*) (After Apsley et al. 1978)

$$\mu_{\text{hop}} = \frac{e\ D_{\text{hop}}}{kT} = \frac{e}{kT}\frac{\nu_{\text{hop}}\, r_{\text{hop}}^2}{6}. \tag{24}$$

The thermally activated effective hopping frequency is given by

$$\nu_{\text{hop}} = \frac{\omega_{\text{phon}}}{2\pi}\ \exp\left(-2\frac{r_{\text{hop}}}{r_0} - \frac{W_{\text{hop}}}{kT}\right), \tag{25}$$

where W_{hop} is the average energy difference between the two states for hopping, r_{hop} is the hopping distance, r_0 is the radius of the center, and ω_{phon} is an effective phonon frequency to match the energies of initial and scattered states. For hops of distance r_{hop}, the corresponding hopping energy is given by the band width of centers located at the Fermi energy $\Delta E_{\text{imp}}(E_{\text{F}})$, which in turn is given by

$$W_{\text{hop}} = \Delta E_{\text{imp}}(E_{\text{F}}) = \frac{3}{4\pi\ r_{\text{hop}}^3\, g_N(E_{\text{F}})}, \tag{26}$$

with $g_N(E_{\text{F}})$ (dimension $\text{eV}^{-1}\ \text{cm}^{-3}$) the density of states for defects with an energy at the Fermi level, from which such activation makes the largest contribution to the mobility (see Mott 1969, Pollak 1972).

Variable Range Hopping In amorphous semiconductors with the Fermi level below the mobility edge, thermal activation becomes essential to carrier transport. With reduced temperature, the width of the energy band decreases, thereby involving less centers, i.e., the distance between the active centers increases

(Mott 1968, 1969). The average hopping distance, which maximizes the hopping rate, is given by

$$\overline{r}_{\text{hop}} = \left(\frac{3\ r_0}{2\pi\ g_N(E_\text{F})\ kT} \right)^{1/4}. \tag{27}$$

This results in a hopping frequency of

$$\nu_{\text{hop}} \propto \exp\left(-\frac{C}{(kT)^{1/4}} \right), \text{with} C = 2\left(\frac{3}{2\pi} \right)^{1/4} \left(\frac{1}{r_0^3 g_N(E_\text{F})} \right)^{1/4}. \tag{28}$$

Introducing this relation into Eq. 23, we obtain a hopping mobility

$$\mu_{\text{hop}} = \mu_0 \exp\left(-\frac{T_0}{T} \right)^{1/4}, \tag{29}$$

which is experimentally observed in some of the amorphous semiconductors (see Fig. 8). In thin layers, the $T^{-1/4}$ relation changes to a $T^{-1/3}$ relation (Knotek et al. 1973). Here, percolation paths are cut open by the layer surfaces normal to the current flow (Hauser 1975). For a review of hopping conduction, see Böttger and Bryksin (1985).The frequency dependence of the low-temperature optical (AC) conductivity by variable-range hopping in highly doped Si:P – on the insulating side of the metal-insulator transition – is discussed by Hering et al. (2005) and Mahmood et al. (2021); the studies indicate a crossover from interacting to non-interacting (Fermi glass) behavior and measure relaxation times of the optical excitations in the THz range. For variable-range hopping in *n*-channel α-SiGe quantum well structures, see Shin et al. (1999).

Hopping Mobility of Polarons The strong interaction of trapped carriers with phonons suggests the involvement of polarons in the carrier transport of amorphous semiconductors (Emin 1975, Mott and Davis 1979). In certain amorphous semiconductors, the carrier transport may also be caused by hopping of bipolarons (Schlenker and Marezio 1980, Elliott 1977, and Elliott 1978).

Dispersive Carrier Transport One of the most convincing arguments about the carrier transport involving a quasi-exponential trap distribution stems from experiments with excess carriers, e.g., injected or photo-excited carriers. These are trapped, reemitted from shallow traps, retrapped, and so on; during the period between trapping, they are mobile and drift in an electrical field. The first carriers that traverse the device have not been trapped, followed by carriers that have been trapped once, twice, etc. Consecutive trapping causes further slow-down of carrier traversal. When

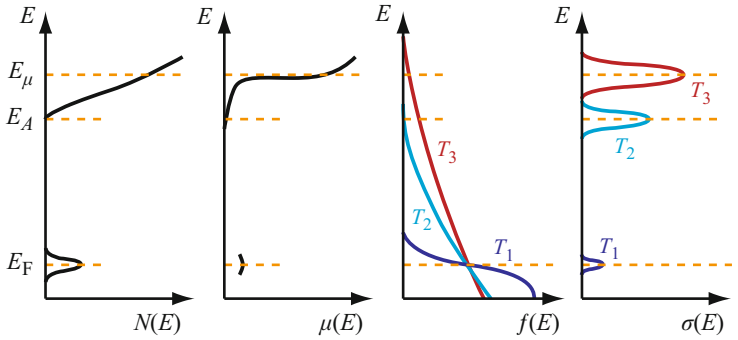


Fig. 9 Typical carrier distributions resulting in three different modes of conductivity at temperatures T_1, T_2, and T_3

being retrapped, energy is dissipated by emitting phonons, and successively deeper traps are filled; from here escape is much slower. This behavior results in a typical distribution of these excess carriers as a function of the time while in transit, in agreement with the experiment. This confirms the intimate involvement of a trap distribution in carrier transport.

Transport with a nonconstant hopping rate due to multiple trapping of carriers into progressively deeper states is referred to as *dispersive transport*. The change in hopping rates effects the broadening of a generated carrier ensemble. We will return to this subject when we discuss carrier kinetics in chapter Sect. 1.1.3 of ▶ Chap. 32, "Dynamic Processes". For a review, see Tiedje (1984).

4.4 Temperature Dependence of the Conductivity

The relative magnitude of the different contributions to the conductivity is a function of the temperature. This is depicted in Fig. 9, in which the Fermi level remains pinned.

(a) At low temperatures (T_1), only carriers near E_F can contribute to the conductivity. Since pinning of the Fermi level requires a high density of defect states at E_F, such a conductivity is similar to the impurity conduction in crystalline semiconductors. We can distinguish two cases of this *impurity conductivity*:
   (a1) The impurity density near E_F is large enough to permit sufficient tunneling within a band of width ΔE_1; then impurity conduction similar to a crystalline semiconductor dominates, with

$$\sigma = \sigma_{a1} \exp\left(-\frac{\Delta E_1}{2\ kT}\right). \tag{30}$$

(a2) The impurity density is smaller and its bandwidth is larger than kT; then *variable-range hopping* occurs with the characteristic $(kT)^{-1/4}$ dependence:

$$\sigma = \sigma_{a2} \exp\left(-\frac{C}{\sqrt[4]{kT}}\right), \tag{31}$$

with C as given by Eq. 28. Such a dependence is shown in Fig. 8.

(b) At medium temperatures (T_2), sufficient carriers are excited into tailing states near E_A. They show sufficient overlap for tunneling, so that hopping is activated. With ΔE_2 as the activation energy for hopping, we obtain

$$\sigma = \sigma_b \exp\left(-\frac{E_A - E_F + \Delta E_2}{kT}\right). \tag{32}$$

However, since $E_A - E_F$ is usually much larger than ΔE_2, one observes a constant slope in the $\ln(\sigma)$ versus $1/T$ diagram.

(c) At higher temperatures (T_3), when sufficient carriers are excited into non-localized, i.e., band states with an energy above the mobility-edge energy E_μ, the conductivity is given by

$$\sigma = \sigma_c \exp\left(-\frac{E_\mu - E_F}{kT}\right). \tag{33}$$

(d) With further increasing temperatures, the mobility may increase sufficiently above the saddle point between the undulating band edges to provide yet one more significant contribution to the conductivity:

$$\sigma = \sigma_d \exp\left(-\frac{E'_g}{2kT}\right), \tag{34}$$

where E'_g is a shifted effective bandgap energy: the mobility gap related to the mobility edges of electrons and holes.

The pre-exponential factors are

$$\sigma_{a2} = e^2 \omega \; g(E_F) \; \bar{r}^2, \tag{35}$$

for variable range hopping (see Eq. 27),

$$\sigma_b = 0.03 \frac{e^2}{\hbar \lambda_i}, \tag{36}$$

for hopping from tailing states, and

$$\left.\begin{matrix}\sigma_{\mathrm{al}}\\ \sigma_{\mathrm{c}}\end{matrix}\right\} = \sigma_{\min} \cong \frac{e^2}{2\pi^2 \hbar\, a_{\mathrm{qH}}} \cong \frac{610}{a_{\mathrm{qH}}/\text{Å}} \quad \left(\Omega^{-1}\,\mathrm{cm}^{-1}\right) \tag{37}$$

for band conductivity. Finally,

$$\sigma_{\mathrm{d}} = \frac{e^2\ g_N\left(E'_{\mathrm{g}}\right)\ kT\ \tau_e}{m_n} \tag{38}$$

for conduction above the saddle points of the band edges. Here, $\sigma_{\min}$ is the *minimum metallic conductivity* (see Mott and Davis 1979), τ_e is the energy-relaxation time, and $g_N\left(E'_{\mathrm{g}}\right)$ is the joint density of states at E'_{g}, the shifted effective bandgap energy.

The type of predominant conductivity depends on material preparation (Connell and Street 1980). This becomes rather sensitive in tetrahedrally bound amorphous semiconductors, such as α-Ge:H or α-Si:H, where a wide range of $\sigma(T)$ behavior is observed, depending on deposition parameters, doping, hydrogenation, and annealing treatments (LeComber et al. 1972; Bullot and Schmidt 1987).

5 Charge Transport in Organic Semiconductors

Organic semiconductors comprise small-molecule crystals and polymers. Quite a few of them have been obtained as single crystals and highly purified to obtain their intrinsic semiconductor properties (Sect. 1.5 of ▶ Chap. 3, "The Structure of Semiconductors"). Most organic solids are excellent insulators and become semiconductive only after doping (Pope and Swenberg 1982, Scaccabarozzi et al. 2022). Also organic polymers show semiconducting properties (Goodings 1976). They typically consist of polymer chains with a semiconducting backbone. Organic semiconductors are employed today in a wide field of applications, e.g., in organic LEDs (OLEDs) and displays, radio-frequency tags, solar cells, and integrated devices (Sirringhaus et al. 1998); for reviews see Hung and Chen (2002), Gather et al. (2011), Arias et al. (2010), Peumans et al. (2003), and Hains et al. (2010).

Conductance in organic semiconductors is governed by disorder. Even in perfect small-molecule crystals, the weak intermolecular van der Waals bonds give rise to a dynamical disorder which affects the mobility of carriers. Highly pure single crystals hence show generally an increase of mobility at decreased temperature. In contrast, less ordered organic solids exhibit a *decrease* of the mobility at lower temperatures due to localization of carriers and a required thermal activation for transport. We consequently observe both band conductance and hopping conductance in organic semiconductors.

Carriers in organic semiconductors couple strongly to molecular oscillations, suggesting a polaron-state contribution (Spear 1974). An early review on the theory

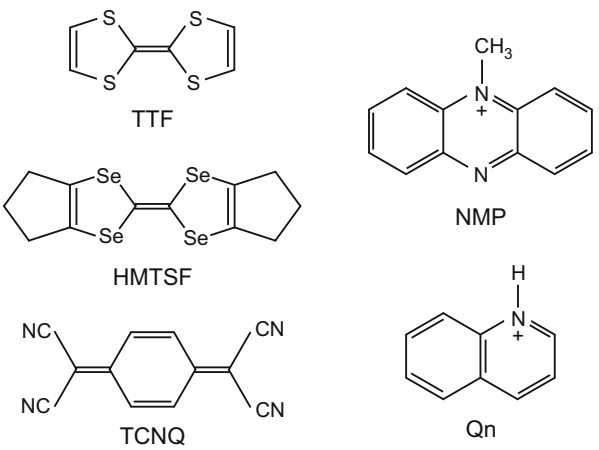


Fig. 10 Structure of some typical organic donor and acceptor molecules in charge-transfer crystals: Tetrathiofulvalene (*TTF*), Hexamethylenetetra-selenofulvalene (*HMTSF*), Tetracyanoquinodimethan (*TCNQ*), *N*-methylphenazinium (*NMP*), Quinolinium (*Qn*)

of carrier mobility in organic semiconductors is given by Druger (1975). Karl (1984) gives a critical review of mobility and other physical data for both one-component and mixed organic semiconductors. A comprehensive review is given by Bässler and Köhler (2012).

5.1 Band Conductance in Organic Crystals

Band conductance is found particularly in small-molecule crystals. These are predominantly van der Waals bonded (Sect. 3.3 of ▶ Chap. 2, "Crystal Bonding"); some of them have other bonding superimposed, such as ionic, hydrogen, and charge-transfer bonding. We distinguish single-component and two-component (charge-transfer) semiconductors. The first group contains the classical organic semiconductors, such as the acenes (Sect. 1.5 of ▶ Chap. 3, "The Structure of Semiconductors"), which show comparatively high mobilities; the second group includes highly conductive compounds, some of which show semiconductor-metal transitions and even superconductivity (Sects. 1.1 and 2 of ▶ Chap. 26, "Superconductivity").

5.1.1 Single-Component and Two-Component Semiconductors

Single-Component Semiconductors Single-component organic crystals are usually good insulators, but may become photoconductive with sufficient optical excitation. The class of aromatic hydrocarbons like the acenes has been more thoroughly investigated. They have bandgap energies between 2 and 5 eV (see Sect. 3 of ▶ Chap. 6, "The Origin of Band Structure") and are considered a class of one-dimensional

conductors (Kivelson and Chapman 1983), albeit with a comparably small anisotropy ratio. The bandgap energy of organic semiconductors can rarely be determined by optical absorption, since valence-to-conduction band transitions are masked by transitions to excited molecular states referred to as excitons (Davidov 1962), which lie within the bandgap and do not support charge transport. Transport bandgap and optical bandgap are hence distinguished, see Sect. 4.1 of ▶ Chap. 8, "Bands and Bandgaps in Solids". The transport bandgap can be measured directly from the threshold of intrinsic photoconductivity (Marchetti and Kearns 1970) or from photoelectron spectroscopy.

The mobility is usually low compared to inorganic semiconductors, for electrons and holes typically in the 10^{-2} to 10 $cm^2/(Vs)$ range at 300 K, and falls with increasing temperature.

Two-Component Semiconductors Two-component semiconductors consist of pairs of complementary molecules with large differences in their redox properties: the organic molecules with a low ionization energy act as electron donors D and the other molecules with a high electron affinity act as acceptors A. Such a combination produces organic crystals that can show very low or vanishing activation energies and comparatively high conductivities. The crystals are formed by a sandwich-like stacking of planar molecules, where donors and acceptors form $D^{\delta+}A^{\delta-}D^{\delta+}A^{\delta-}$ structures, or they are located in separated stacks, i.e., the stacking contains $D^{\delta+}D^{\delta+}\ldots$ and $A^{\delta-}A^{\delta-}\ldots$ complexes for segregated stacking with face-to-face stacks. δ denotes the transferred charge per molecule in units of elementary charges. The charge transfer (CT) may be incomplete, yielding two limiting cases: weak CT complexes and strong CT complexes, also referred to as radical ion salts (Soos 1974).

In *radical ion salts*, an organic radical cation (such as perylene$^+$) is combined with a counter anion (such as a halogen or PF_6^-), or an organic radical anion (such as $TCNQ^-$) is combined with a counter cation. The solids have a pronounced ionic character, i.e., δ is often close to 1; usually $\delta < 1$ for conductive radical ion salts.

Charge-transfer complexes with δ significantly smaller than unity can show high conductivities, caused by the incomplete charge transfer between D and A, which results for the ground state in partially filled bands. Typical donor and acceptor molecules that form such charge-transfer crystals are given in Fig. 10. Many such semiconductors have low-lying electronically excited states in which an electron is transferred from D to A. The charge-transfer transition in the excited state[7] may be written as

$$D\ A \rightarrow D^+A^- \text{ with } E_{CT} = I_D - A_A - C, \tag{39}$$

[7]The excited state has essentially ionic charge character.

Table 1 Bandwidth W of acene crystals for holes in the valence band (HOMO) and electrons in the conduction band (LUMO) (After Cheng et al. 2003)

Crystal	Bandwidth W	
	Valence band	Conduction band
Naphthaline	409	372
Anthracene	509	508
Tetracene	626	503
Pentacene	738	728

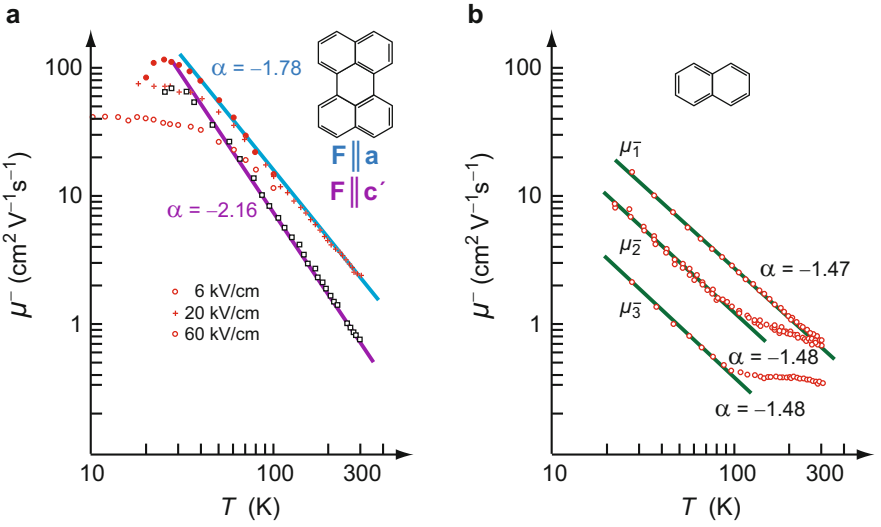


Fig. 11 Temperature dependence of the electron mobility (**a**) in a perylene crystal, and (**b**) in a naphthalene crystal (After Warta et al. 1985 and Karl 2001)

where I_D is the ionization energy of the donor and A_A is the electron affinity of the acceptor (both in the gas phase), and C is a Coulomb binding-energy of the excited state. E_{CT} is the "energy gap" between the ground state and the excited charge-transfer state (Mulliken 1952).

The resulting structures are termed *neutral charge-transfer crystals*, typically with stacks of alternating D and A molecules: $DADADA\ldots$. The lowest excited state is $DADAD^{+}A^{-}DADA\ldots$; the respective activation energy for semiconductivity is typically (Kuroda et al. 1962)

$$E_s \cong \frac{1}{2}E_{\mathrm{CT}}. \tag{40}$$

The incomplete electron exchange and consequently partially filled bands results in a rather large semiconductivity – or even metallic conductivity. The resistivity of these semiconductors lies between 10^2 and 10^6 Ω cm at room temperature with a transfer-energy gap in the 0.1–0.4 eV range (Braun 1980). The conductivity is

Table 2 Mobility μ in units of $cm^2\ V^{-1}\ s^{-1}$ for organic crystals at 300 K and exponent α of the temperature dependence according to Eq. 42

Crystal	Direction	Electrons		Holes	
		μ	α	μ	α
Naphthalene	*a*	0.62	−1.4	0.94	−2.8
	b	0.64	−0.55	1.84	−2.5
	c'	0.44	+0.04	0.32	−2.8
Anthracene	*a*	1.73	−1.45	1.13	−1.46
	b	1.05	−0.84	1.84	−1.26
	c'	0.39	+0.16	0.32	−1.43
Perylene	*a*	2.37	−1.78		
	b	5.53	−1.72		
	c'	0.78	−2.15		

usually highly anisotropic, with the electron transfer-integral in the stacking direction typically a factor of 10 larger than in the direction perpendicular to the stacks (Keller 1977). Trapping is of minor importance in these semiconductors with a high carrier density (Karl 1984). The charge-transfer crystals provide an opportunity for fine tuning of the semiconductive properties by replacing TTF-type donors and TCNQ-type acceptors with other similar molecules (Bloch et al. 1977), which could render such materials attractive for some technical applications.

5.1.2 Carrier Mobility in Pure Organic Crystals

Organic molecules show a strong structural relaxation on introducing a charge. The band structure calculated for a crystal composed of weakly bonded neutral molecules may hence not be preserved in the presence of carriers. As a rule of thumb, band conductance occurs despite lattice relaxation if the transfer integral between molecules is sufficiently large: a large transfer integral delocalizes the carrier wavefunction over several molecules. A more quantitative estimate follows from the width W of the energy band for carrier transport. If the mean scattering time τ is in the range or smaller than $\hbar/W$, no wavevector $\mathbf{k}$ can be assigned to the carrier. The description in terms of conduction in a band with dispersion $E(\mathbf{k})$ hence requires

$$\tau \gg \frac{\hbar}{W}. \tag{41}$$

The bands in organic crystals are rather narrow due to a small amount of wavefunction overlap of π electrons, see Sect. 4.1 of ▶ Chap. 8, "Bands and Bandgaps in Solids". Typical bandwidths W are in the range of some hundred meV (see Table 1), yielding $\tau > 10^{-15}$ s. If scattering results from molecular relaxation, τ is given by the characteristic time of molecular vibration; the carrier must leave the molecule before a significant relaxation and consequential trapping can occur. Equation 41 then reads $W \gg \hbar/\tau$. This yields bandwidths of 100–200 meV, a condition reasonable well fulfilled for crystals of acenes and comparable aromatic compounds.

A clear indication for band conduction is provided by the temperature dependence of the mobility. In inorganic semiconductors scattering at acoustic phonons leads to a $T^{-3/2}$ dependence of the carrier mobility (Eq. 15 of ▶ Chap. 23, "Carrier Scattering at Low Electric Fields"). A comparable decrease of mobility at higher temperature is observed for the carrier mobility in organic crystals with band conduction.

The mobility data in Fig. 11 show typical features of transport in organic crystals. The temperature dependence at low electric field F is described by a power law

$$\mu(T) = \mu_{300\mathrm{K}}\, T^{\alpha}, \tag{42}$$

where $\mu_{300\mathrm{K}}$ is the mobility at $T = 300$ K. The exponent α deviates somewhat from the ideal value of $-3/2$, see Table 2. The mobilities do not depend on the value of the electric field and show a pronounced anisotropy; the principal axes of the mobility tensor deviate slightly from the crystallograpic crystal axes. Typical mobilities are on the order of $10^{-2}\ldots 10\ \mathrm{cm}^2/(\mathrm{Vs})$. Data for a large number of organic semiconductors are tabulated by Schein (1977); they decrease at higher temperature according to Eq. 42 with $0 > \alpha > -3$.

At high electric field F, the mobility shows a sublinear velocity-field relation, see Fig. 11a. The drift velocity of the carrier saturates at high fields, similar to the transport in inorganic semiconductors discussed in Sect. 1.1 of ▶ Chap. 24, "Carrier Scattering at High Electric Fields".

5.2 Hopping Conductance in Disordered Organic Semiconductors

Many organic compounds – small molecules or polymers – cannot be prepared as single crystals. They are usually prepared as thin films by evaporating or spin coating. The carrier mobility of these semiconductors is by orders of magnitudes lower than that of crystalline semiconductors considered above. In these semiconductors, the *static* disorder dominates at most temperatures, and the mobility *increases* for increasing temperature. The low coupling between the molecules in the solid state leads to a strong localization of the carriers on a molecule; transport occurs via a sequence of charge-transfer steps from one molecule to another, similar to the hopping between defect states in inorganic semiconductors. The transport properties are thus described by the formalism of hopping conductance developed for amorphous inorganic semiconductors (Sect. 4): the charge carriers are assumed to hop in a time-independent disordered energy landscape as illustrated in Fig. 5.

The basic difference between amorphous inorganic semiconductors and disordered organic semiconductors is the shape of the density of states (DOS). In an amorphous solid, the DOS is found to have a mobility edge and a tail of localized states with an *exponentially* decreasing distribution extending into the bandgap, see

Figs. 2 and 3 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors", and Sect. 4. In contrast, the DOS in organic materials has a *Gaussian* shape:

$$g(E) = \frac{G_{\text{tot}}}{\sqrt{2\pi}\sigma} \exp\left[\frac{(E - E_{\text{center}})^2}{2\sigma^2}\right], \tag{43}$$

where G_{tot} is the total DOS, E_{center} is the center of the energy distribution, and σ is the variance of the distribution (Bässler 1993). Hopping of carriers is determined by both the energy difference ΔE and the spatial separation Δr of initial and final states; in addition, hopping is affected by an electric field F.

The hopping rate ν_{ij} between two localized states i and j depends on whether a hop-up ($\uparrow$) occurs with $\Delta E = E_j - E_i - eF(x_j - x_i)$ or a hop-down ($\downarrow$) with $\Delta E < 0$, where E_j and E_i are the energies within $g(E)$ at $F = 0$. Adopting the model of Miller and Abrahams (1960) we obtain

$$\nu_{ij}(\uparrow) = \nu_0 \exp\left(-2\gamma a \frac{\Delta r_{ij}}{a}\right) \exp\left(-\frac{E_j - E_i - eF(x_j - x_i)}{kT}\right), \tag{44}$$

$$\nu_{ij}(\downarrow) = \nu_0 \exp\left(-2\gamma a \frac{\Delta r_{ij}}{a}\right). \tag{45}$$

Here the constant γ is the spatial decrease of the wavefunction (invers Bohr radius for hydrogen-like wavefunctions), and a is the mean separation of localization sites. We note that a hop-down does not require a thermal activation.

Analogous to transport in amorphous inorganic semiconductors different transport regimes can be distinguished:

(a) Very low temperatures – *nearest neighbor hopping*: since $kT \ll \Delta E$, the spatial separation Δr controls the transfer, favoring next neighbors
(b) Low temperatures – *variable-range hopping*: the thermal energy kT allows for hopping within a narrow energy band around E_{center}, thereby relaxing the next-neighbor constraint
(c) Medium temperatures –*hopping in a wider energy range*: similar to (b) within a wider energy band, possibly opening percolation paths which are not restricted to next neighbors
(d) High temperatures – *multiple trapping and release*: carriers are excited from localized to extended states above the mobility edge, where band transport takes place until trapping at other localized states occurs. This regime requires a material where extended states exist.

The model of Bässler (1993) results in a thermally activated mobility

$$\mu = \mu_0 \exp\left[-\left(\frac{2\sigma}{3kT}\right)^2 + C\left(\left(\frac{\sigma}{kT}\right)^2 - \Sigma^2\right)\sqrt{F}\right] \tag{46}$$

determined by the spread σ of the energy distribution in the conducting band, the structural disorder parameter Σ, the applied electric field F, and the parameter μ_0 representing the mobility of the hypothetic not-disordered semiconductor at high temperature. The field dependence of Eq. 46 is comparable to the Poole-Frenkel effect (Frenkel 1938). For a more detailed review of various transport models, see Noriega and Salleo (2012) and Simonetti and Giraudet (2019). A review on experimental techniques for measuring transport properties is given by Coropceanu et al. (2007).

The model of Bässler (1993) – based on Monte-Carlo simulation – was developed further, yielding closed analytic models for hopping transport in organic semiconductors (Fishchuk et al. 2012, Baranovskii 2014). Fishchuk et al. construct an effective-medium model using the concept of an effective transport energy, while the approach of Baranovskii is based on the Gaussian disorder model of Bässler (1981) and also the multiple-trap model, which replaces the mobility edge by a slightly temperature-dependent energy level called *transport energy* E_{trans}. This energy is located near the maximum of the Gaussian density of states and is up to high carrier concentrations ($n \lesssim 0.05N$) not depending on n. In addition to expressions for the dependence of the mobility μ on defect sites N, temperature T, and carrier density n, the dependence on an electric field F is described by replacing the ambient temperature T by a field-dependent effective temperature $T_{\text{eff}}(F,T)$. While the results largely agree with the parameterized zero-field mobility given by Bässler, the field dependence of the analytic form is considered more realistic. For details see the comprehensive review by Baranovskii (2014). It was also pointed out that the dependence of $\mu(F)$ in the case of a spatially disordered material may drastically differ from that in regular grids, indicating a flaw of approaches based on lattices (Oelerich et al. 2017).

Polymers Typical examples of disordered organic semiconductors are the one-dimensional organic polymers, such as polyacetylene, as the simplest member (for *cis*-isomer see Fig. 15a of ▶ Chap. 2, "Crystal Bonding"), or the aromatic linear polymer poly(para-phenylenevinylene) (Fig. 16g of ▶ Chap. 3, "The Structure of Semiconductors"). Polyacetylene has been investigated most extensively – see the review by Heeger and MacDairmid (1980). It can easily be doped with donors or acceptors, which are incorporated between the polymer chains of the *trans* isomer, resulting in controlled changes of the conductivity over 13 orders of magnitude up to $3 \times 10^3\ \Omega^{-1}\ \text{cm}^{-1}$. At lower doping densities, devices with a *pn* junction can be formed from polyacetylene. At high doping densities, i.e., above 1%, a

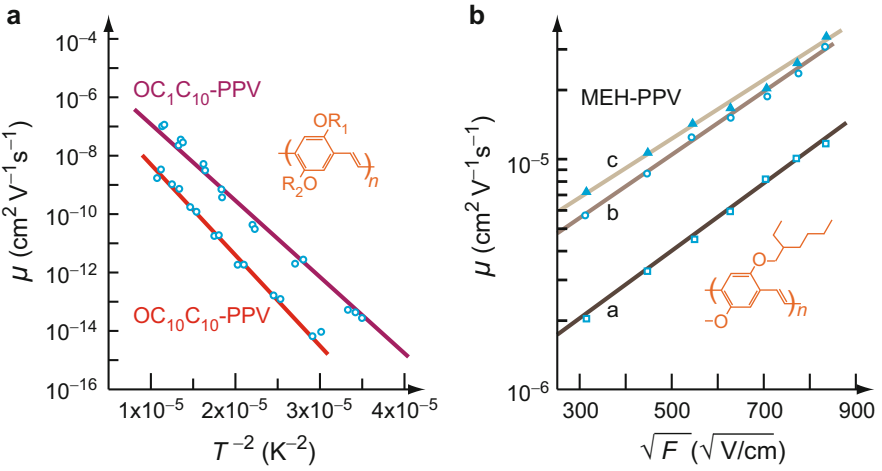


Fig. 12 Hole mobility in derivatives of PPV. (**a**) Mobility in derivatives with different side groups at zero electric field; OC_1C_{10}-PPV: $R_1 = CH_3$, $R_2 = C_{10}H_{21}$; $OC_{10}C_{10}$-PPV: $R_1 = R_2 = C_{10}H_{21}$ (After Blom and Vissenberg 2000). (**b**) Dependence of the hole mobility on the electric field F for MEH-PPV ($R_1 = C_8H_{16}$, $R_2 = CH_3$) prepared without (*a*) and with application of an electric field of 3 and 6 kV/cm for *b* and *c*; after Shi et al. (2006)

semiconductor-metal transition occurs. In the metallic state, polyacetylene has the optical appearance of a highly reflecting metal.

The typical structure of organic polymers has a bond alternation between conjugated single and double bonds within the backbone of the chain, see Sect. 3.3 of ▶ Chap. 2, "Crystal Bonding"; in polyacetylene there are two CH groups per unit cell, with one π electron for each CH group. At the boundary of two bond-alternation sequences with different phase, an unpaired electron is created; this bond-alternation defect, illustrated in Fig. 14 of ▶ Chap. 20, "Defects in Amorphous and Organic Semiconductors", has attracted substantial interest as a manifestation of a *soliton*. Such a soliton can be described as a kink in the electron-lattice symmetry, rather than a spread-out transition; according to minimum-energy calculations some spreading occurs, typically over about 10 lattice constants, remaining unchanged during the kink motion. Highly mobile, the soliton has a room-temperature hopping rate in excess of 10^{13} s^{-1}. The soliton seems to be responsible for a wide variety of unusual electrical, optical, and magnetic properties of these polymers; e.g., fast soliton transport of pulses injected into insulating polymers was reported by Dissado et al. (2011). A short review is given by Heeger (1981).

Mobility in Polymers The mobility is widely measured using time-of-flight experiments. Such studies of disordered organic semiconductors show both dispersive and nondispersive components, depending on temperature, electric field, and chemical purity. Nondispersive carrier mobilities show typical features (Bässler 1993): a field-independent activated mobility at low fields, characterized by an activation energy, and a field-dependent contribution in the form of a

stretched exponential at high fields F. These findings can in many cases be described by the empirical dependence

$$\mu \cong \mu_0 \exp\left[-\frac{E_A}{kT} + \beta\sqrt{F}\right], \tag{47}$$

with the zero-field mobility μ_0, an activation energy E_A, and the field-activation factor

$$\beta = B\left(\frac{1}{kT} - \frac{1}{kT_0}\right); \tag{48}$$

B and T_0 are empirical parameters. Since polymers are stable only in a very limited temperature range, the temperature dependence can often be described in this interval by different relations, such as Eqs. 46 or 47. For a discussion of the field dependence in terms of an analytical model see Baranovskii (2014).

The hole mobility of the poly(paraphenylene vinylene) (PPV) is shown in Fig. 12. Holes dominate the current in many polymers. The polymer PPV is a prominent compound due to its electroluminescence properties; PPV derivatives are soluble and can be spin coated for, e.g., fabrication of organic LEDs. The mobility is reasonably described by a $\log(\mu) \propto T^{-2}$ dependence, although also an Arrhenius dependence $\log(\mu) \propto T^{-1}$ fits well (Blom et al. 1997); in the limited temperature range a clear distinction is not possible. The activation energies E_A for PPV derivatives (Eq. 47) range between 0.3 and 0.5 eV, and widths σ of the Gaussian DOS according to Eq. 46 are near 100 meV, with mean separations a of localization sites in the range 1.1 ... 1.7 nm (Blom and Vissenberg 2000). We note in Fig. 12 the comparatively low mobility of disordered organic semiconductors and the characteristic increase at higher temperature.

The enhancement of the mobility in an electric field is shown in Fig. 12b. The dependence on $F^{1/2}$ is well fulfilled. This behavior has also been observed from time-of-flight measurements in many molecularly doped polymers and amorphous glasses. An electrically induced polarization of MEH-PPV during the preparation of films significantly enhances the mobility.

6 Summary

Carrier transport is generally influenced by defects; however, in highly doped or disordered semiconductors, the carrier transport becomes *induced* by defects. If doping produces a well-defined predominant defect level with increasing density, it will split into two (bonding and antibonding) bands separated by an energy gap. Below a density to permit sufficient tunneling, excitation from the filled, lower impurity band into the conduction band is required for carrier transport. When the impurity band is wider than $2kT$ and the Fermi level lies in the middle of this band, *variable range hopping* occurs: with decreasing T, the predominant excitation occurs

from a narrower range of width kT of these centers, and causes a semilogarithmic slope $\propto T^{-1/4}$. With increased defect density, a diffusive transport within the upper impurity band becomes possible; the conductivity in this band requires a small activation energy to bridge the gap. With further increase of the defect density, the carriers become delocalized, the gap disappears, and the conductivity within the impurity band becomes metallic.

In amorphous semiconductors, a similar tunneling-induced carrier transport can take place within the tail of states that extends from the conduction or valence band into the bandgap, when the states are close enough to each other to permit significant tunneling. Here carriers become delocalized. The edge at which delocalizing occurs is referred to as the mobility edge. At this edge, major carrier transport starts; below the mobility edge, carriers are trapped rather than being mobile.

With a statistical distribution of defects within semiconductors, at a given threshold, only some volume elements become conductive. With increasing temperature these volume elements will widen, will start to interconnect, and finally will provide an uninterrupted path from electrode to electrode. Such a percolation character is typical for most of the conduction phenomena in highly doped or disordered semiconductors, which have a density-related threshold of conduction. The mobility of carriers in such semiconductors is typically on the order of 10 $cm^2/(Vs)$ or lower. At low temperatures it is determined by tunneling (hopping) from neighbor to neighbor and is very sensitive to the density of defects and their distribution in space and energy. At sufficiently high temperatures, carrier transport higher within the conduction or valence band may compete significantly with the conduction mechanisms described above. This band conduction may have a mean free path compatible to the one in crystalline semiconductors.

Organic semiconductors can be *single-component* materials, which can be doped, such as aromatic hydrocarbons (e.g., anthracene). These materials are good insulators and show photoconductivity. Another group of single-component semiconductors is comprised of certain linear polymers, such as polyacetylene. With doping this group can change its conductivity up to 13 orders of magnitude and can become metallic in electrical behavior and optical appearance. *Two-component* semiconductors contain molecules or molecular layers, which act as donors and others which act as acceptors. Variation of their donor/acceptor ratio can change the behavior from highly compensated to *n*- or *p*-type, with a wide range of conductivities, depending on the deviation from a donor to acceptor ratio of 1: 1. There is a great variety of such crystals that exhibit a broad range of properties, including metallic conductivity and, at low temperatures, superconductivity.

The carrier mobility in organic semiconductors is substantially lower than in good semiconducting inorganic compounds and typically is in the 10^{-3} to 10 $cm^2/(Vs)$ range. It is controlled by static and dynamic disorder. Small-molecule crystals show band conductance with a decreasing carrier mobility at increased temperature; it is affected by dynamic disorder and described by a power law similar to that of inorganic semiconductors. Films of small-molecules or polymer semiconductors have dominating static disorder; they show hopping conductance with a typically very low mobility, which increases at higher temperatures.

References

Abrahams E, Anderson PW, Licciardello DC, Ramakrishnan TV (1979) Scaling theory of localization: absence of quantum diffusion in two dimensions. Phys Rev Lett 42:673

Adler D (1980) Electronic correlations, polarons, and hopping transport. In: Moss TS, Paul W (eds) Handbook on semiconductors, Theory and transport properties, vol 1. North Holland Publishing Company, Amsterdam, pp 805–841

Altshuler BL, Aronov AG, Lee PA (1980) Interaction effects in disordered Fermi systems in two dimensions. Phys Rev Lett 44:1288

Apsley N, Davis EA, Troup AP, Yoffee AD (1978) Electronic properties of ion-bombarded evaporated germanium and silicon. J Phys C 11:4983

Arias AC, MacKenzie JD, McCulloch I, Rivnay J, Salleo A (2010) Materials and applications for large area electronics: solution-based approaches. Chem Rev 110:3

Aronov AG, Mirlin DN, Nikitin LP, Reshina II, Sapega VF (1979) Determination of the inter-band transition time in the valence band of gallium arsenide. Sov Phys JETP 29:62

Aronzon BA, Chumakov NK (1994) Quantum effects and the phase diagram of the electronic system of highly compensated semiconductors in magnetic field. Physica B 194–196:1167

Baranovskii SD (2014) Theoretical description of charge transport in disordered organic semiconductors. Phys Stat Sol B 251:487

Bässler H (1981) Localized states and electronic transport in single component organic solids with diagonal disorder. Phys Stat Sol B 107:9

Bässler H (1993) Charge transport in disordered organic photoconductors. Phys Status Solidi B 175:15

Bässler H, Köhler A (2012) Charge transport in organic semiconductors. Top Curr Chem 312:1

Bloch AN, Carruthers TF, Poehler TO, Cowan DO (1977) The organic metallic state: some physical aspects and chemical trends. In: Keller HJ (ed) Chemistry and physics in one-dimensional metals. Plenum Press, New York, pp 47–85

Blom PWM, Vissenberg MCJM (2000) Charge transport in poly(p-phenylene vinylene) light-emitting diodes. Mater Sci Eng 27:53

Blom PWM, de Jong MJM, van Munster MG (1997) Electric-field and temperature dependence of the hole mobility in poly(p-phenylene vinylene). Phys Rev B 55:R656

Böer KW (1972) Perturbed bands in real semiconducting glasses. J Non-Cryst Sol 8–10:586

Böttger H, Bryksin VV (1979) Effective medium theory for the hopping conductivity in high electrical fields. Phys Status Solidi B 96:219

Böttger H, Bryksin VV (1980) Investigation of non-Ohmic hopping conduction by methods of percolation theory. Philos Mag B 42:297

Böttger H, Bryksin VV (1985) Hopping conduction in solids. VCH-Verlagsgesellschaft, Weinheim

Braun CL (1980) Organic semiconductors. In: Moss TS, Keller SP (eds) Handbook on semiconductors, Materials properties and preparation, vol 3. North Holland Publications, Amsterdam, pp 857–873

Broadbent SR, Hammersley JM (1957) Percolation processes I. Crystals and mazes. Proc Camb Philos Soc 53:629

Bullot J, Schmidt MP (1987) Physics of amorphous silicon-carbon alloys. Phys Status Solidi B 143:345

Butcher PN (1972) On the rate equation formulation of the hopping conductivity problem. J Phys C 5:1817

Cheng YC, Silbey RJ, da Silva Filho DA, Calbert JP, Cornil J, Brédas JL (2003) Three-dimensional band structure and bandlike mobility in oligoacene single crystals: a theoretical investigation. J Chem Phys 118:3764

Ching WY, Huber DL (1982) Numerical studies of energy levels and eigenfunction localization in dilute three-dimensional systems with exponential interactions. Phys Rev B 25:1096

Coropceanu V, Cornil J, da Silva Filho DA, Olivier Y, Silbey R, Bredas JL (2007) Charge transport in organic semiconductors. Chem Rev 107:926

Davidov AS (1962) Theory of molecular excitons. McGraw-Hill, New York

Dissado LA, Montanari GC, Fabiani D (2011) Fast soliton-like charge pulses in insulating polymers. J Appl Phys 109:064104
Donchev V, Shtinkov N, Germanova K (1997) Effect of random defect density fluctuations on the Fermi level in highly compensated semiconductors. Mater Sci Eng B 47:131
Dresner J (1983) Hall effect and hole transport in B-doped a-Si:H. J Non-Cryst Sol 58:353
Druger SD (1975) Theory of charge transport processes. In: Birks JB (ed) Organic molecular photophysics, vol 2. Wiley, London, pp 313–394
Efros AL, Shklovskii BI (1975) Coulomb gap and low temperature conductivity of disordered systems. J Phys C 8:L49
Efros AL, Shklovskii BI, Yanchev IY (1972) Impurity conductivity in low compensated semiconductors. Phys Status Solidi B 50:45
Elliott SR (1977) A theory of a.c. conduction in chalcogenide glasses. Philos Mag 36:1291
Elliott SR (1978) Defect states in amorphous silicon. Philos Mag B 38:325
Emin D (1975) Phonon-assisted transition rates I. Optical-phonon-assisted hopping in solids. Adv Phys 24:305
Fishchuk II, Kadashchuk A, Ullah M, Sitter H, Pivrikas A, Genoe J, Bässler H (2012) Electric field dependence of charge carrier hopping transport within the random energy landscape in an organic field effect transistor. Phys Rev B 86:045207
Frenkel J (1938) On pre-breakdown phenomena in insulators and electronic semi-conductors. Phys Rev 54:647
Fritzsche H, Cuevas M (1960) Impurity conduction in transmutation-doped *p*-type germanium. Phys Rev 119:1238
Gadzhiev AR, Ryvkin SM, Shlimak IS (1972) Fast-neutron-compensated *n*-germanium as a model of amorphous semiconductors. Sov Phys JETP 15:428
Gather MC, Köhnen A, Meerholz K (2011) White organic light-emitting diodes. Adv Mater 23:233
Goodings EP (1976) Conductivity and superconductivity in polymers. Chem Soc Rev 5:95
Greenwood DA (1958) The Boltzmann equation in the theory of electrical conduction in metals. Proc Phys Soc Lond 71:585
Grünewald M, Thomas P, Würtz D (1981) The sign anomaly of the Hall effect in amorphous tetrahedrally bonded semiconductors: a chemical-bond orbital approach. J Phys C 14:4083
Gunning B, Lowder J, Moseley M, Doolittle AW (2012) Negligible carrier freeze-out facilitated by impurity band conduction in highly *p*-type GaN. Appl Phys Lett 101:082106
Hacker K, Obermair G (1970) Stark ladders in a one-band-model. Z Phys 234:1
Haegel NM, Samperi SA, White AM (2003) Electric field and responsivity modeling for far-infrared blocked impurity band detectors. J Appl Phys 93:1305
Hains AW, Liang Z, Woodhouse MA, Gregg BA (2010) Molecular semiconductors in organic photovoltaic cells. Chem Rev 110:6689
Halperin BI, Lax M (1966) Impurity-band tails in the high-density limit. I. Minimum counting methods. Phys Rev 148:722
Hauser JJ (1975) Hopping conductivity in amorphous carbon films. Solid State Commun 17:1577
Heeger AJ (1981) Semiconducting and metallic polymers: new science with potential for new technology. Comments Sol State Phys 10:53
Heeger AJ, MacDairmid AG (1980) Conducting organic polymers: doped polyacetylene. In: Alcàcer L (ed) The physics and chemistry of low-dimensional solids. Reidel, Dordrecht, pp 353–391
Hering M, Scheffler M, Dressel M, v Löhneysen H (2005) Crossover from Coulomb glass to Fermi glass in Si:P. Physica B 359:1469
Holstein T (1959) Polaron band narrowing. Ann Phys 8:325. ibid 8:343
Horring NJ (1969) Quantum theory of static shielding of an impurity charge by an electron gas plasma in a magnetic field. Ann Phys 54:405
Hubbard J (1963) Electron correlations in narrow energy bands. Proc R Soc Lond A 276:238

Huckestein B, Schweitzer L (1993) Multifractality and anomalous diffusion at the mobility edge in disordered systems. J Non-Cryst Sol 164:461
Hung LS, Chen CH (2002) Recent progress of molecular organic electroluminescent materials and devices. Mater Sci Eng 39:143
Ioffe AF, Regel AR (1960) Non-crystalline, amorphous and liquid electronic semiconductors. Prog Semicond 4:237
Kabilova Z, Kurdak C, Peterson RL (2019) Observation of impurity band conduction and variable range hopping in heavily doped (010) β-Ga_2O_3. Semicond Sci Technol 34:03LT02
Kane EO (1963) Thomas-Fermi approach to impure semiconductor band structure. Phys Rev 131:79
Karl N (1984) Organic semiconductors. Landoldt-Börnstein/Springer, Heidelberg
Karl N (2001) Charge-carrier mobility in organic crystals. In: Farchioni R, Grosso G (eds) Organic electronic materials: conjugated polymers and low molecular weight organic solids. Springer, Berlin, pp 283–326
Keller HJ (ed) (1977) Chemistry and physics in one-dimensional metals. Plenum Press, New York
Kivelson S (1982) Electron hopping in a soliton band: conduction in lightly doped $(CH)_x$. Phys Rev B 25:3798
Kivelson S, Chapman OL (1983) Polyacene and a new class of quasi-one-dimensional conductors. Phys Rev B 28:7236
Knotek ML, Pollak M, Donovan TM, Kurtzman H (1973) Thickness dependence of hopping transport in amorphous-Ge films. Phys Rev Lett 30:853
Knotek ML, Pollak M (1974) Correlation effects in hopping conduction: a treatment in terms of multielectron transitions. Phys Rev B 9:664
Krivoglaz MA (1974) Fluctuon states of electrons. Sov Phys Usp (Engl Transl) 16:856
Kubo R (1956) A general expression for the conductivity tensor. Can J Phys 34:1274
Kubo R (1957) Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. J Phys Soc Jpn 12:570
Kuroda H, Kobayashi M, Kinoshita M, Takemoto S (1962) Semiconductive properties of tetracyanoethylene complexes and their absorption spectra. J Chem Phys 36:457
LeComber PG, Madan A, Spear WE (1972) Electronic transport and state distribution in amorphous Si films. J Non-Cryst Sol 11:219
Lifshitz IM (1965) Energy spectrum structure and quantum states of disordered condensed systems. Sov Phys Usp (Engl Transl) 7:549
Maekawa S (1970) Nonlinear conduction of ZnS in strong electric fields. Phys Rev Lett 24:1175
Mahmood F, Chaudhuri D, Gopalakrishnan S, Nandkishore R, Armitage NP (2021) Observation of a marginal Fermi glass. Nat Phys 17:627
Marchetti AP, Kearns DR (1970) Photoelectron emission from aromatic and metalloorganic hydrocarbons. Mol Cryst Liq Cryst 6:299
Matsuura H, Takeshita A, Imamura T, Takano K, Okuda K, Hidaka A, Ji S, Eto K, Kojima K, Kato T, Yoshida S, Okumura H (2019) Transition of conduction mechanism from band to variable-range hopping conduction due to Al doping in heavily Al-doped 4H-SiC epilayers. Jpn J Appl Phys 58:098004
Mazuruk K, Benzaquen M, Walsh D (1989) Coulomb gap and screening in impurity bands. Solid State Commun 69:337
McInnes JA, Butcher PN (1979) Numerical calculations of dc hopping conductivity. Philos Mag B 39:1
Miller A, Abrahams E (1960) Impurity conduction at low concentrations. Phys Rev 120:745
Mott NF (1968) Conduction in glasses containing transition metal ions. J Non-Cryst Sol 1:1
Mott NF (1969) Conduction in non-crystalline materials. Philos Mag 19:835
Mott NF (1991) The sign of the Hall effect in amorphous silicon. Philos Mag B 63:3
Mott NF (1993) Conduction in non-crystalline materials, 2nd edn. Clarendon Press, Oxford
Mott NF, Davis EA (1979) Electronic processes in noncrystalline materials. Claredon Press, Oxford, UK

Mott NF, Kaveh M (1985) Metal-insulator transitions in non-crystalline systems. Adv Phys 34:329
Mulliken RS (1952) A comparative survey of approximate ground state wave functions of helium atom and hydrogen molecule. Proc Nat Akad Sci 38:160
Noriega R, Salleo A (2012) Charge transport theories in organic semiconductors. In: Klauk H (ed) Organic electronics II. Wiley-VCH, Weinheim, pp 67–104
Oelerich JO, Nenashev AV, Dvurechenskii AV, Gebhard F, Baranovskii SD (2017) Field dependence of hopping mobility: lattice models against spatial disorder. Phys Rev B 96:195208
Okamoto H, Hattori K, Hamakawa Y (1993) Hall effect near the mobility edge. J Non-Cryst Sol 164–166:445
Overhof H, Beyer W (1981) A model for the electronic transport in hydrogenated amorphous silicon. Philos Mag B 43:433
Peumans P, Yakimov A, Forrest SR (2003) Small molecular weight organic thin-film photodetectors and solar cells. J Appl Phys 93:3693
Pollak M (1972) A percolation treatment of dc hopping conduction. J Non-Cryst Solids 11:1
Pope M, Swenberg CE (1982) Electronic processes in organic crystals. Oxford University Press, Oxford
Prati E, Hori M, Guagliardo F, Ferrari G, Shinada T (2012) Anderson-Mott transition in arrays of a few dopant atoms in a silicon transistor. Nat Nanotech 7:443
Priour DJ Jr (2012) Electronic states in one-, two-, and three-dimensional highly amorphous materials: a tight-binding treatment. Phys Rev B 85:014209
Pronin KA, Bandrauk AD, Ovchinnikov AA (1994) Electric-field-induced localization and non-perturbative response of a one-dimensional conductor. J Phys Condens Matter 6:4721
Rosenbaum TF, Andres K, Thomas GA, Bhatt RN (1980) Sharp metal-insulator transition in a random solid. Phys Rev Lett 45:1723
Ryvkin SM, Shlimak IS (1973) A doped highly compensated crystal semiconductor as a model of amorphous semiconductors. Phys Status Solidi A 16:515
Scaccabarozzi AD, Basu A, Aniés F, Liu J, Zapata-Arteaga O, Warren R, Firdaus Y, Nugraha MI, Lin Y, Campoy-Quiles M, Koch N, Müller C, Tsetseris L, Heeney M, Anthopoulos TD (2022) Doping approaches for organic semiconductors. Chem Rev 122:4420
Schein LB (1977) Temperature independent drift mobility along the molecular direction of As_2S_3. Phys Rev B 15:1024
Schlenker C, Marezio M (1980) The order–disorder transition of Ti^{3+}–Ti^{3+} pairs in Ti_4O_7 and $(Ti_{1-x}V_x)_4O_7$. Philos Mag B 42:453
Schnakenberg J (1968) Polaronic impurity hopping conduction. Phys Status Solidi B 28:623
Simonetti O, Giraudet L (2019) Transport models in disordered organic semiconductors and their application to the simulation of thin-film transistors. Polym Int 68:620
Shante VKS, Kirkpatrick S (1971) An introduction to percolation theory. Adv Phys 20:325
Shi Q, Hou Y, Lu J, Jin H, Yunbai L, Yan L, Sun X, Liu J (2006) Enhancement of carrier mobility in MEH-PPV film prepared under presence of electric field. Chem Phys Lett 425:353
Shin D-H, Becker CE, Harris JJ, Fernández JM, Woods NJ, Thornton TJ, Maude DK, Portal J-C (1999) Variable-range hopping transport in modulation-doped n-channel $Si/Si_{1-x}Ge_x$ quantum well structures. Semicond Sci Technol 14:762
Shklovskii BI, Efros AL (1984) Electronic properties of doped semiconductors. Springer, Berlin
Sirringhaus H, Tessler N, Friend RH (1998) Integrated optoelectronic devices based on conjugated polymers. Science 280:1741
Soos ZG (1974) Theory of π-molecular charge-transfer crystals. Annu Rev Phys Chem 25:121
Spear WE (1974) Electronic transport and localization in low mobility solids and liquids. Adv Phys 23:523
Thouless DJ (1974) Electrons in disordered systems and the theory of localization. Phys Rep 13:93
Thouless DJ (1980) Resistance and localization in thin films and wires. J Non-Cryst Sol 35:3
Tiedje T (1984) Information about band-tail states from time-of-flight experiments. In: Willardson RK, Beer AC, Ji P (eds) Semiconductors and semimetals, vol 21C. Academic, New York, pp 207–238

Wang X, Wang B, Hou L, Xie W, Chen X, Pan M (2015) Design consideration of GaAs-based blocked-impurity-band detector with the absorbing layer formed by ion implantation. Opt Quant Electron 47:1347

Wannier GH (1960) Wave functions and effective hamiltonian for Bloch electrons in an electric field. Phys Rev 117:432

Warta W, Stehle R, Karl N (1985) Ultrapure, high mobility organic photoconductors. Appl Phys A Mater Sci Process 36:163

Zhang M, Drabold DA (2011) Comparison of the Kubo formula, the microscopic response method, and the Greenwood formula. Phys Rev E 83:012103

Part VII

Generation - Recombination

Carrier Generation

29

Karl W. Böer and Udo W. Pohl

Contents

1 Thermal and Optical Carrier Generation 1206
1.1 Optical Carrier Generation 1207
1.2 Thermal Ionization 1209
2 Field Ionization 1213
2.1 Frenkel-Poole Ionization 1213
2.2 Impact Ionization 1215
2.3 Electron Tunneling 1225
3 Summary 1235
References 1237

Abstract

When the semiconductor is exposed to an external electromagnetic field, a phonon field, or an electric field, free carriers can be generated, resulting in semiconductivity or photoconductivity. Carriers can also be generated by high-energy particles, such as fast electrons or ions. Optical carrier generation proceeds as band-to-band direct or indirect generation or from defect levels with photons of sufficient energy. Thermal generation of free carriers is substantially enhanced by defect centers. Shallow centers may absorb a phonon of sufficient energy or a few phonons involving intermediate steps into excited states; generation from deep centers requires multiphonon-induced giant oscillations.

Generation of carriers by an electric field can at low fields be caused by the Frenkel-Poole effect: a field-enhanced thermal generation from Coulomb-attractive defect centers. At high fields, impact ionization from deep centers or

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_29

band-to-band impact ionization is observed. At still higher fields in the 10^6 V/cm range, tunneling from deep defect centers or from the valence band occurs. Besides thickness and height of the barrier, the tunneling probability depends on the shape of the barrier potential.

Keywords

Avalanche multiplication · Band-to-band tunneling · Defect-assisted carrier generation · Dielectric breakdown · Field ionization · Franz-Keldysh effect · Frenkel-Poole effect · Impact ionization · Multiplication factor · Optical carrier generation · Thermal ionization · Transition probability · Transition rate · Tunneling · Tunneling current · Tunneling spectroscopy

1 Thermal and Optical Carrier Generation

Typical Electron Transitions The density of carriers[1] in semiconductors can be determined from the difference between generation rate g_n, recombination rate r_n, and the net influx of carriers from surrounding regions described by the divergence of the current density:

$$\frac{dn}{dt} = g_n - r_n + \frac{1}{e}\frac{\partial}{\partial \mathbf{r}}\mathbf{j}_n \tag{1}$$

In homogeneous semiconductors, this net influx vanishes and the electron density changes, just as a change in population occurs when there is a difference between birth and death rates.

Three types of *generation processes* can be distinguished. Each one of the generation processes needs energy, which can be supplied as one of the following forms:

- Thermally via a phonon field
- Optically via an electromagnetic (photon) field
- Electrically via an electric field

Two types of *recombinations* are distinguished. Each of the recombination processes releases energy in one of the following forms:

- Thermal energy via nonradiative recombination
- Luminescence via radiative recombination

[1]We consider *n*-type carriers (electrons, indicated by the index *n*) unless stated otherwise.

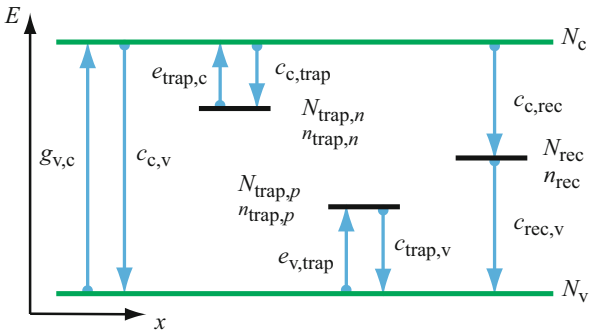


Fig. 1 Electron transitions between localized (in the bandgap) and nonlocalized states (bands). Labels e and c denote rates of excitation and capture, respectively

The type of generation selected is usually the choice of the experimentalist, whereas the type of predominant recombination is mostly a function of the defect structure of the selected material. It is also influenced by the temperature and sometimes by the electric field and other parameters. A number of examples will be discussed for each type of generation in this chapter and for the recombination in ▶ Chap. 30, "Carrier Recombination and Noise".

1.1 Optical Carrier Generation

Typical electronic transitions among a variety of states are shown in Fig. 1. For consistency in the following descriptions, only *electron* transitions will be identified; *hole* transitions proceed in the opposite direction. The *transition coefficients* for capture c_{ik} and excitation e_{ik} can then be unambiguously defined by the first and second subscripts, indicating the initial and final states, respectively.

A *transition rate* R_{ik} is defined as the product of the electron density in the original state n_i, the hole density in the final state p_k, and the transition coefficient c_{ik}:

$$R_{ik} = c_{ik}\, n_i\, p_k. \tag{2}$$

For example, the capture rate of an electron from the conduction band into an electron trap is given by

$$R_{c,\mathrm{trap}} = c_{c,\mathrm{trap}}\, n\ \left(N_{\mathrm{trap},n} - n_{\mathrm{trap},n}\right), \tag{3}$$

where $N_{\mathrm{trap},n}$ and $n_{\mathrm{trap},n}$ are the densities of electron traps and of captured electrons in these traps, respectively. In deviation from the above-given rule, the index c is left off from the electron density in the conduction band ($n_c \rightarrow n$) to conform with common notation. This implies that an approximation is used here by describing the conduction and valence bands as levels at the respective band edges and assuming that transition coefficients are independent of the actual carrier distribution within the bands.

The transition coefficients c_{ik} or e_{ik} have the units cm^3s^{-1}. The product of these coefficients with the hole density in the final state, $c_{ik}\, p_k$, is the *transition probability* with the unit s^{-1}. For further detail, see Sect. 1.2.2.

The following sections present a brief review of the different excitation mechanisms into higher-energy states.

Optical Generation Rate In Sect. 1.1 of ▶ Chap. 13, "Band-to-Band Transitions", the optical excitation of electrons across the bandgap was discussed as the product of the joint density of the initial and final states and the probability for each of these transitions. This probability is given by the product of the matrix element and the probability of finding the initial state occupied and the final state empty before the transition. The optical excitation is the part of the optical absorption that generates free carriers – electrons and holes.

We distinguish bound-to-free and free-to-free excitation, depending on whether the excitation started from a center with an occupied state, i.e., below the Fermi level, or from the valence band. In both cases, the optical absorption has an edge-like character – see Sect. 2.4 of ▶ Chap. 17, "Optical Properties of Defects", and 2.2 of ▶ Chap. 13, "Band-to-Band Transitions". When the excitation proceeds in energy beyond the exciton spectrum, free carriers are produced (▶ Chap. 31, "Photoconductivity").

From the flux of impinging photons ϕ_o traveling in the direction z normal to the semiconductor surface, a certain fraction, depending on the wavelength, will be absorbed. The photon flux within the semiconductor is given by

$$\phi_\lambda(\lambda, z) = \phi_0(\lambda)\exp[-\alpha_o(\lambda)\, z]. \tag{4}$$

Here, $\phi_o(\lambda)$ is the *photon flux per unit wavelength* $\Delta\lambda$, which penetrates, after reflection is subtracted, through the top layer of the solid, measured in $cm^{-2}s^{-1}\Delta\lambda^{-1}$; α_o is the *optical absorption coefficient*, measured in cm^{-1}. When polychromatic light is used, the total carrier-generating photon flux as a function of the penetration depth z is obtained by integration

$$\phi(z) = \int_{\lambda_1}^{\lambda_2} \phi_\lambda(\lambda, z)\, d\lambda, \tag{5}$$

where $\phi(z)$ is measured in $cm^{-2}s^{-1}$. Although $\phi_\lambda(\lambda,z)$ depends exponentially on the penetration depth, $\phi(z)$ usually does not, since, with polychromatic light of various absorption coefficients $\alpha_o(\lambda)$, the superposition of a wide variety of such exponential functions causes a dependence of ϕ on z of much lesser steepness than an exponential function.

The optical generation rate $g(z)$ is given by the absorbed photoelectrically active light in each slab of infinitesimal thickness; hence

$$g(z) = -\frac{d\phi(z)}{dz}. \tag{6}$$

This rate depends *exponentially* on *z for monochromatic* light:

$$g(z,\lambda_0) = \alpha_o(\lambda_0)\,\phi_0(\lambda_0)\,\Delta\lambda \exp[-\alpha_o(\lambda_0)\,z], \tag{7}$$

where $\Delta\lambda$ is a small wavelength range in which $\alpha_o(\lambda)$ is constant. For polychromatic excitation, the decline of g with increasing depth is much more gradual; often a constant space-independent generation rate $g = g_0$ can be used as a better approximation (Böer 1977). Optical excitation and carrier generation were discussed in ▶ Chaps. 13, "Band-to-Band Transitions", and ▶ 21, "Equilibrium Statistics of Carriers", and consequent photoconductivity is described in ▶ Chap. 31, "Photoconductivity", where more detail about the various types of optical excitation and related references can be found.

1.2 Thermal Ionization

Thermally induced *ionization* can occur from any type of lattice defect as *extrinsic ionization* or from lattice atoms as *intrinsic ionization.* Such ionization was discussed in Sects. 1 and 2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers", using an equilibrium approach without requiring an understanding of the microscopic mechanism involved in the actual process of ionization. The previous analysis required only the magnitude of the ionization energy and the density of levels in equilibrium. In this section, the discussion is extended to include information relating to the excitation process, which is necessary for a kinetic evaluation.

1.2.1 Thermal Ionization Mechanism

Thermal ionization is a process requiring statistical consideration, since it usually needs the presence of several phonons simultaneously – or in a short time interval – to supply sufficient energy. Such a "simultaneous" supply of phonons is necessary for centers that do not have eigenstates spaced closely enough, into which an intermediate thermal excitation can take place. The simultaneous phonon excitation can be interpreted as a process involving a transient giant oscillation of a lattice atom.

For *deep defect centers* that possess a phonon ladder, there exist two possibilities: a multiphonon absorption or a sequential absorption of phonons, a cascade process. For the latter case, the absorption of the next phonon must occur within the lifetime of an excited state to accomplish the subsequent step of excitation, and so on, until the total ionization energy is supplied. In both cases, a transient giant oscillation of the defect center results, as indicated by the large double-headed arrow in Fig. 2a. This results in the total energy of the center exceeding that of the bottom of the conduction band, from which the electron can tunnel into the adjacent states of the conduction band.

Such a deep defect center can induce transitions from the valence to the conduction band. In this case, a giant oscillation of a lattice atom, i.e., a multiphonon process, can result in a transfer of an electron from the valence band into the defect center. If the lifetime of the electron within the center is long enough, a consequent giant oscillation of the center, as previously discussed, would bring this electron into the conduction band. The sum of these two processes is equivalent to a thermally

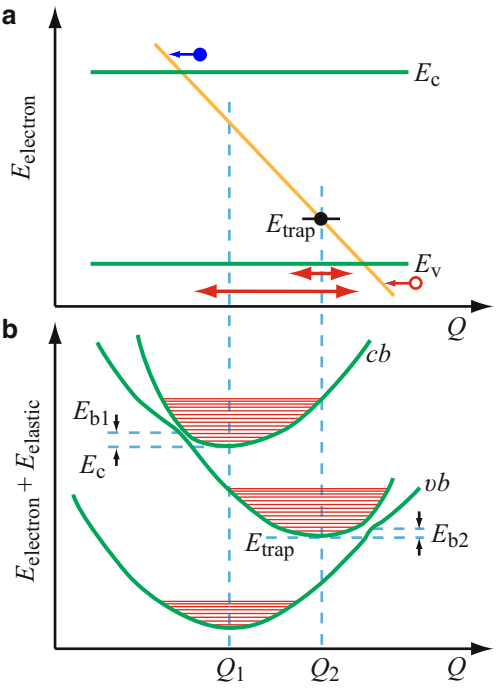


Fig. 2 Excitation of a trapped electron and possible retrapping. **(a)** Band model with amplitude of normal and giant oscillations indicated by the *red double-headed arrows*. **(b)** Corresponding configuration coordinate diagram with total energy $E_{total} = E_{electron} + E_{elastic}$; *horizontal red lines* are spaced by the breathing-mode phonon energy

induced band-to-band transition, which is more difficult to understand without this intermediate step.

For every excitation process, there exists an inverse de-excitation process. The lifetime of an electron in any intermediate state and the supply of phonons with the proper energy for accomplishing the next excitation step decide with what probability any higher excitation is accomplished. For further discussion of the actual microscopic phonon stepping, or "simultaneous" multiphonon excitation, see the review of Stoneham (1981). A semiclassical model of such ionization, presented below, will provide some insight.

1.2.2 Thermal Excitation Probability

In a simple configuration-coordinate diagram, we can represent the eigenstate of an electron as that of an oscillator as illustrated in Fig. 2. With its surrounding lattice, it will oscillate between different electronic eigenstates. With each of the oscillations of variable amplitude, the electron has a varying probability of escaping from the center into the adjacent band.

Determining when enough phonons have been supplied for such an escape is a matter of statistics. This probability is given by the Boltzmann factor $\exp[-E_i/(k\ T)]$. Hence, the total ionization probability is

$$P_i = \nu_i \exp[-E_i/(kT)], \tag{8}$$

where ν_i is the *attempt-to-escape frequency* $\nu_i = \omega_i/2\pi$, which may be approximated by the breathing mode or a specific vector-mode phonon frequency ω_i of the defect center; see Sect. 1.2.1 of ▶ Chap. 17, "Optical Properties of Defects". For Coulomb-attractive shallow centers, there is a different, rather crude way of estimating an electronic attempt-to-escape frequency from the ionization radius r_i of such a center, as well as the thermal velocity v_{rms} of the electron, assuming that it behaves like an effective-mass particle:

$$\nu_i = \frac{v_{\mathrm{rms}}}{2\pi r_i}. \tag{9}$$

This radius also determines the *capture cross-section* $s_n = \pi r_i^2$; we hence obtain

$$\nu_i = \frac{v_{\mathrm{rms}}}{2\sqrt{\pi s_n}}. \tag{10}$$

This approximation, however, is only useful for a Coulomb-attractive center, wherein the effective-mass picture can be applied, and therefore the appropriate capture cross-section can be determined. With an rms velocity of $\sim 10^7$ cm/s, and a capture cross-section of $\sim 10^{-13}$ cm^2, we estimate the frequency factor of such centers to be $\sim 10^{13}$ s^{-1}. This simple model, however, should *not* be applied to other types of centers, which would yield values that are substantially too high.

Thermal excitation probabilities are readily obtained from thermodynamic arguments. This will be shown for a defect center that contains only the ground level and interacts preferably with the conduction band: an electron trap. In equilibrium, all transitions into the level must equal all transitions out of this level between each group of two states, since its population remains constant within each volume element. This fundamental *detailed-balance principle* yields (see Fig. 1)

$$e_{\mathrm{trap,c}}\, n_{\mathrm{trap}}\, p_{\mathrm{c}} = c_{\mathrm{c,trap}}\, n \ \left(N_{\mathrm{trap}} - n_{\mathrm{trap}}\right). \tag{11}$$

Equation 11 holds independently of the position of the Fermi level. For ease of the following computations, let us assume that the Fermi energy coincides with the energy of the trap level ($E_{\mathrm{F}} = E_{\mathrm{trap}}$); then, in thermal equilibrium, the population of these traps is $N_{\mathrm{trap}}/2$, and $(N_{\mathrm{trap}} - n_{\mathrm{trap}})/n_{\mathrm{trap}} = 1$ applies. Therefore, with $p_{\mathrm{c}} \cong N_{\mathrm{c}}$ (the density-of-state distribution in the conductance band), we obtain

$$e_{\mathrm{trap,c}} N_{\mathrm{c}} \cong c_{\mathrm{c,trap}}\, n = c_{\mathrm{c,trap}} N_{\mathrm{c}} \exp\left[-\frac{E_{\mathrm{c}} - E_{\mathrm{trap}}}{kT}\right], \tag{12}$$

yielding for the ratio of emission-to-capture coefficients

$$\frac{e_{\mathrm{trap,c}}}{c_{\mathrm{c,trap}}} = \exp\left[-\frac{E_{\mathrm{c}} - E_{\mathrm{trap}}}{kT}\right]. \tag{13}$$

Although this condition was obtained for a specific case, namely, in thermal equilibrium with $E_F = E_{trap}$, the ratio holds true in general, since both coefficients are constant and do not change with trap population.

The *emission probability* $e_{trap,c}N_c$ can also be expressed in a microscopic model as the product of the attempt-to-escape frequency ν_i and the Boltzmann factor:

$$e_{trap,c}N_c = \nu_i \exp\left[-\frac{E_c - E_{trap}}{kT}\right]. \tag{14}$$

From gas-kinetic arguments, we can describe the capture coefficient as the product of the capture cross-section s_n of the center and the rms velocity of the mobile carrier:

$$c_{c,trap} = s_n v_{rms}. \tag{15}$$

Along its path through the lattice, a carrier can be thought of as sweeping out the cylinder of a cross section of the capturing defect, as it recombines when it touches the defect at any point on its cross section. In combining Eqs. 11, 12, 13, 14, and 15, we obtain for Coulomb-attractive centers a useful relation between the capture cross-section s_n and an effective attempt-to-escape frequency:

$$\nu_{trap} = N_c s_n v_{rms}. \tag{16}$$

The typical parameters characterizing an electron trap are given in Table 1. The use of Eq. 16 becomes problematic for centers with a capture cross-section less than 10^{-14} cm^2; thus, the attempt-to-escape frequency for these centers is left open in Table 1. In centers with a capture cross-section less than the geometric cross section of the defect center, other arguments need to be considered, such as resonance transitions or tunneling.

The population of a defect center in thermal equilibrium is determined by its energy alone, while the attainment of this equilibrium, i.e., the time it takes to follow changes in excitation, is determined by the center's kinetic parameters $c_{c,trap}$ and $e_{trap,c}$ – i.e., by two parameters: the energy of the level E_{trap} and either its capture cross-section or its attempt-to-escape frequency ν_{trap}.

Table 1 Typical capture and ionization parameters for electron traps

Center type	Coulomb-attractive	Neutral	Coulomb-repulsive	Tight binding	Dimension
s_n	10^{-13}	10^{-16}	10^{-21}	10^{-18}	cm^2
$c_{c,trap}$	10^{-6}	10^{-9}	10^{-14}	10^{-11}	cm^3/s
ν_{trap}	10^{13}				s^{-1}

2 Field Ionization

There are three field-ionization mechanisms:

- Frenkel-Poole ionization
- Impact ionization
- Tunnel ionization

All of these mechanisms produce free carriers predominantly by inducing bound-to-free transitions. At sufficiently high fields, band-to-band transitions can also be initiated.

2.1 Frenkel-Poole Ionization

The Frenkel-Poole effect requires the lowest field[2] for the ionization of Coulomb-attractive centers (Frenkel 1938; Poole 1921). This ionization is achieved by tilting the bands and thus lowering the thermal ionization energy of such a center as illustrated in Fig. 3. The ionization probability is thereby increased.

The lowering of the potential barrier is obtained by superimposing an external electric field F upon the Coulomb potential:

$$V(x) = \frac{e\ Z}{4\pi\,\varepsilon_{\text{stat}}\,\varepsilon_0\,x} - F\,x, \tag{17}$$

which shows a maximum where dV/dx vanishes:

$$x(V_{\max}) = \sqrt{\frac{e\ Z}{4\pi\,\varepsilon_{\text{stat}}\,\varepsilon_0\,F}}\,; \tag{18}$$

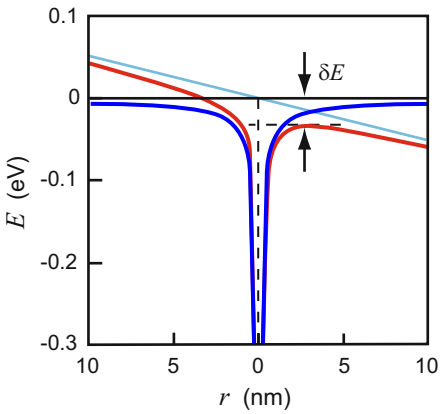


Fig. 3 Lowering of the electron-binding energy by δE for a Coulomb-attractive center with an external electric field (Frenkel-Poole effect), assuming $\varepsilon_{\text{stat}} = 10$ and $F = 50$ kV/cm

[2]Except for high-mobility semiconductors for excitation from shallow centers at low temperatures, where impact ionization favorably competes – see Sect. 2.2.

at $x(V_{max})$ the barrier is lowered by $\delta E = eV_{max}$:

$$\delta E = e\sqrt{\frac{e\ F\ Z}{\pi\,\varepsilon_{stat}\,\varepsilon_0}}\,. \tag{19}$$

The field-enhanced thermal ionization probability of such a center can now be approximated by

$$e_{trap,c} = \nu^{(0)}_{trap,c}\,\exp\left[-\frac{E_c - E_{trap} - \delta E}{kT}\right], \tag{20}$$

where $\nu^{(0)}_{trap,c}$ is the frequency factor for thermal ionization.

By setting $\delta E = kT$, a simple estimate yields a critical electric field for marked Frenkel-Poole ionization at

$$F_{FP} = \frac{\pi\,\varepsilon_{stat}\,\varepsilon_0}{e\ Z}\left(\frac{kT}{e}\right)^2 \cong 1.08\times 10^4\left(\frac{\varepsilon_{stat}}{10}\right)\left(\frac{T}{300\,\mathrm{K}}\right)^2 \quad (\mathrm{V/cm}). \tag{21}$$

Effect of Local Fields The classical Frenkel-Poole model yields an ionization value which is often too large when compared to experimental data. At sufficient density, Coulomb-attractive centers will interact with each other due to their local field F_{loc}, which needs to be added to the external field F (Dallacasa and Paracchini 1986, Schroeder 2015). The ionization probability consequently is given as

$$e_{trap,c}\,g = e^{(0)}_{trap,c}\,\exp\left[\sqrt{\frac{e^3 Z}{\pi\,\varepsilon_{stat}\,\varepsilon_0}}\,\frac{\sqrt{F+F_{loc}}}{kT}\right]; \tag{22}$$

the local field F_{loc} is on the order of 10^4 V/cm for typical densities of charged defects of 10^{16} cm^{-3}. In Fig. 4, the relative ionization probability is given as a function of the normalized electric field for a number of different approximations.

Increased leakage currents in metal/insulator structures is often attributed to the Frenkel-Poole effect in the bulk of the insulator,[3] as discussed, e.g., by Miron et al. (2019) and studied for AlGaN by Reddy et al. (2021). The facilitated release of carriers from traps by the Frenkel-Poole effect leads to a current density according to $j \propto F\exp\left(\sqrt{F}\right)$. A detailed simulation of metal/insulator/metal structures indicates that a respective interpretation deduced from plots ln(j/F) versus $\sqrt{F}$ may be misleading (Schroeder 2015), since various parameters and conditions enter the evaluation. Instead, multiphonon trap ionization was recently assigned to the origin of leakage

[3]Other assignments refer to interface effects such as the Schottky effect or thermionic emission.

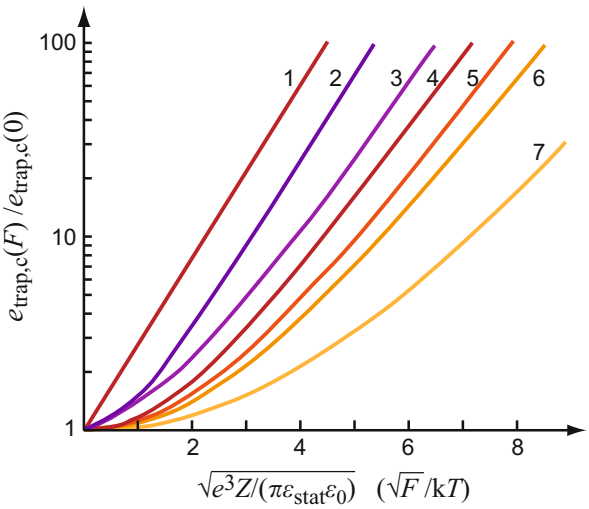


Fig. 4 Relative ionization probability as a function of the normalized external field for: (1) the original Frenkel-Poole model, (2) Dallacasa and Paracchini (1986) for $\gamma = 1$ (=abscissa value with F replaced by F_{loc}), (3) the Hartke (1968) model, (4) the Hill (1971) and Connell et al. (1972) model for emission in field direction, (5) the Pai (1975) model; and (6) the Hill (1971) model for isotropic emission, (7) the Dallacasa and Paracchini (1986) model plotted for $\gamma = 10$. After Dallacasa and Paracchini (1986)

current in BN and SiN_x insulators (Novikov and Gritsenko 2021, 2022). For Frenkel-Poole emission due to phonon-assisted tunneling, see Ganichev et al. (1999). The intricate interplay between acceptors at threading dislocations in GaN and deep electronic states inducing Frenkel-Poole emission is reported in Wang et al. (2020).

2.2 Impact Ionization

An electron that gains enough energy between scattering events to exceed the ionization energy can set free an additional electron on impact. This ionization energy can be that of a defect center, e.g., a donor or, at sufficient fields, the bandgap energy. The two resulting electrons will in turn gain energy, each one setting free another electron. Sequentially, two, four, eight, etc., free electrons will result in an *avalanche formation*. In homogeneous semiconductors, it takes place over the length of the high-field region, i.e., throughout the entire distance between electrodes. With avalanche formation, the current increases dramatically and can lead to a *dielectric breakdown*. In very thin semiconductors, however, there is not enough space from electrode to electrode to develop a breakdown avalanche. Consequently, breakdown fields at which the avalanche reaches critical values (see the following sections) are higher in such thin layers.

Impact Ionization of Shallow Donors Shallow donors can easily be ionized by an impact with free carriers at relatively low fields, i.e., a few V/cm in the 4–10 K range in Ge, until all donors are depleted at ~50 V/cm. In Fig. 5, we see two ohmic branches with moderate slope dI/dV and a steeply increasing branch in between when impact ionization occurs. The magnitude of the shift between the low- and high-field branches is determined by the initial population of the shallow donors. This

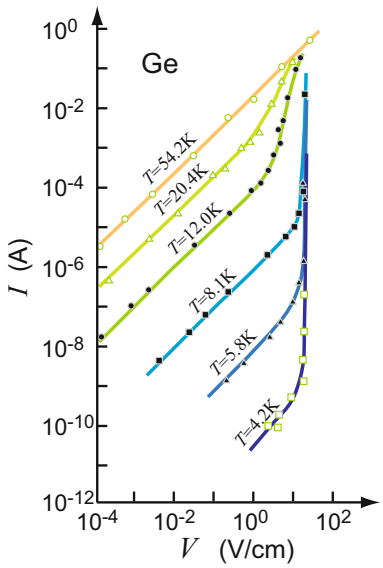


Fig. 5 Current-voltage characteristics of Ge with the temperature as the family parameter. Impact ionization is indicated near 10 V/cm, depleting electrons from shallow donors. Thermal depletion is completed at 54 K. (After Lautz 1961)

population is reduced at higher temperatures because of thermal ionization. For more detail, see Sclar and Burstein (1957) and Bratt (1977). For the measurement of the impact ionization cross-section, see Schappe et al. (1996).

2.2.1 Impact Ionization Across the Bandgap

Impact ionization can proceed by the impact of a conduction-band electron with an electron from the valence band, thereby creating an additional electron and a hole. The critical energy for such ionization can be estimated from the conservation of momentum and energy. We will examine first a rather simple model that explains the main principles.

The momentum conservation between initial ($k_{n,\mathrm{init}}$) and final states ($2k_n + k_p$), while conserving the group velocity after impact (Anderson and Crowell 1972) $\mathbf{k}_n/m_n = \mathbf{k}_p/m_p$, yields

$$k_{n,\mathrm{init}} = k_n\left(2 + \frac{m_p}{m_n}\right), \tag{23}$$

and the energy conservation with $E_{n,\mathrm{init}} = 2E_n + E_p$ reads

$$E_{n,\mathrm{init}} = \frac{\hbar^2 k_n^2}{2m_n}\left(2 + \frac{m_p}{m_n}\right) + E_\mathrm{g}. \tag{24}$$

Also, as long as a parabolic band approximation holds, we have

$$E_{n,\text{init}} = \frac{\hbar^2 k_{n,\text{init}}^2}{2m_n} = \frac{\hbar^2 k_n^2}{2m_n}\left(2 + \frac{m_p}{m_n}\right)^2; \tag{25}$$

for narrow bandgap material this may be a reasonable approximation. When combining Eqs. 23, 24, and 25, we obtain for the threshold energy for impact ionization $E_{\text{ii}} = E_{n,\text{init}}$:

$$E_{\text{ii}} = E_g \frac{2m_n + m_p}{m_n + m_p}, \tag{26}$$

which becomes the often-cited

$$E_{\text{ii}} \cong \begin{cases} 1.5E_g & \text{for } m_n = m_p \\ E_g & \text{for } m_n \ll m_p \end{cases} \tag{27}$$

as the condition for the electron-energy threshold. In typical semiconductors, the observed threshold energy is larger, since the bands are nonparabolic and the effective mass increases with increasing energy; for example, for GaAs, we find $E_{\text{ii}} \cong 2.0\ E_{\text{g}}$, and for Si, we have $E_{\text{ii}} \cong 2.3\ E_{\text{g}}$. In indirect-bandgap semiconductors, momentum conservation requires that intervalley phonons are involved in the ionization process: holes are located at the Γ point, while conduction electrons are accelerated from a side valley – see Anderson and Crowell (1972).

We must now evaluate how conduction electrons reach this threshold. This was originally done by Wolff (1954), who estimated the balance between energy gain from the field and losses due to scattering and ionization, in analogy to a gas discharge analyzed by Townsend. We first will briefly review the main concepts of Wolff's theory.

The characteristic parameter of the impact ionization is the ionization rate $\tilde{\alpha}_{\text{i}}$ per unit path length, measured in cm^{-1}. Electrons gain energy from the field and lose part of the incremental gain during scattering events. Since the scattering is a statistical process, there is a finite probability that an electron can accumulate enough energy for impact ionization over a sufficiently long path length [a "lucky" electron, as it is referred to later (Shockley 1961)].

The ionization rate increases rapidly with increasing field. The incremental speed Δv of an electron gained from the electric field F is given by

$$m_n \frac{dv_x}{dt} = m_n \frac{\Delta v_x}{\tau_{\text{m}}} = eF_x, \tag{28}$$

with τ_{m} as the *momentum*-relaxation time. The average energy gain is therefore

$$\overline{\Delta E} = \frac{m_n}{2}\Delta v_x^2 = \frac{e^2 F_x^2 \tau_{\text{m}}^2}{2m_n}. \tag{29}$$

The energy gain can be dissipated most efficiently by scattering with longitudinal optical phonons. Wolff estimates that the average gain during a free path λ is approximately equal to the optical phonon energy $\hbar\omega_{\mathrm{LO}}$ near threshold fields:

$$\overline{\Delta E_{\mathrm{LO}}} = \frac{e^2 F_x^2 \tau_{\mathrm{e}}^2}{2m_n} = \frac{e^2 F_x^2 \lambda^2}{2m_n v_e^2} \cong \hbar\omega_{\mathrm{LO}}, \tag{30}$$

where v_e is the average electron velocity in the given electric field F_x, i.e., a velocity *above* the rms velocity of electrons, not to be confused with the drift velocity. We have also used the energy-relaxation time τ_{e} here, rather than the momentum-relaxation time at lower energies, in order to consider the strong interaction with LO phonons. With the kinetic energy of the accelerated electron $E = m_n v_e^2/2$, we obtain from Eq. 30

$$\frac{\overline{\Delta E_{\mathrm{LO}}}}{E} = \frac{e^2 F_x^2 \lambda^2}{4\hbar\omega_{\mathrm{LO}}} \frac{1}{(m_n/2)\, v_e^2} \cong 1 \tag{31}$$

in the hot-electron field range. The field-modified electron distribution-function can then be written as

$$f(E) \cong A\exp\left(-\frac{E}{\overline{\Delta E_{\mathrm{LO}}}}\right) = A\exp\left(-\frac{B\ E}{F^2}\right), \tag{32}$$

where A is a proportionality constant on the order of 1, and $B \cong 4\ \hbar\omega_{\mathrm{LO}}/(e^2\ \lambda^2)$. The calculation of the ionization rate results in a similar functional dependence on the electric field:

$$\tilde{\alpha}_{\mathrm{i}} = C\exp\left(-\frac{B\,E}{F^2}\right). \tag{33}$$

Shockley (1961), however, argued that impact ionization is caused by a few "lucky" carriers that escape scattering altogether and are accelerated within one free path to the ionization energy E_{ii}. He obtains for the ionization rate

$$\tilde{\alpha}_{\mathrm{i}} = C^*\exp\left(-\frac{B^*\,E}{F}\right). \tag{34}$$

Baraff (1962) assumed a more general distribution function and obtained results that contain Shockley's $\tilde{\alpha}_{\mathrm{i}}$ as the low-field limit and Wolff's $\tilde{\alpha}_{\mathrm{i}}$ as the high-field limit.

All of these theories have a shortcoming in as much as they do not recognize the actual band structure, i.e., major deviation from the parabolic (effective mass) approximation of the conduction band. Except for very narrow bandgap semiconductors, the impact ionization energy lies high within the conduction band and

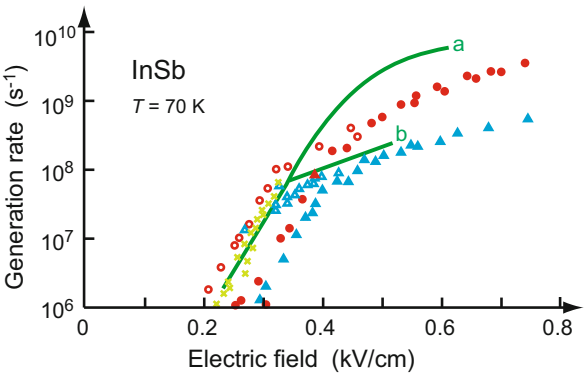


Fig. 6 Computed electron generation rate due to impact ionization in InSb at 70 K for $n = 10^{14}$ and 2×10^{16} cm^{-3} for curves (*a*) and (*b*), respectively. *Data points* are measured values for various electron densities. After Curby and Ferry (1973)

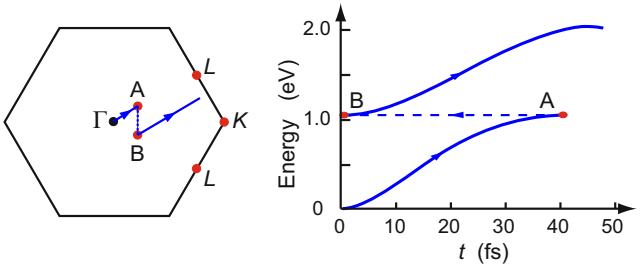


Fig. 7 **(a)** Momentum vector trajectory in the Γ *K L* plane of the Brillouin zone with an electric field in the ⟨1 1 1⟩ direction and one elastic scattering event ($A \rightarrow B$). **(b)** Variation of the electron energy in time for the process shown in (*a*). After Shichijo and Hess (1981)

requires a more accurate accounting of $E(k)$. See Curby and Ferry (1973) for a Monte Carlo analysis of impact ionization in InAs and InSb; see also Fig. 6.

Shichijo and Hess (1981) have used a Monte Carlo method to account for the competing field acceleration and polar optical or intervalley scattering within the Brillouin zone of GaAs. They observed characteristic anisotropies of scattering and ionization thresholds and recognized that electrons can never gain sufficient energy within one ballistic path (Shockley's "lucky" electrons) while confined within the Brillouin zone. However, after an *elastic* scattering event, such electrons could start over again and attain sufficient energy for impact ionization – see Fig. 7. At an energy sufficient for ionization, the electron in GaAs lies in the second conduction band.

The total phonon-scattering rate is shown in Fig. 8 as a function of the electron energy; the data agree well with more recent calculations (Grupen 2011). Above the impact-ionization energy, the scattering rate increases dramatically, as given by Keldysh (1965):

$$e_{\mathrm{ii}} = \frac{1}{\tau(E_{\mathrm{ii}})} P \left(\frac{E - E_{\mathrm{ii}}}{E_{\mathrm{ii}}} \right)^2, \tag{35}$$

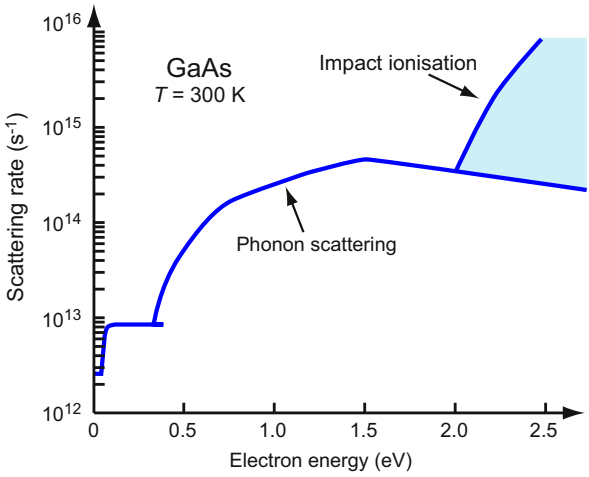


Fig. 8 Phonon and impact-ionization scattering rate in GaAs at 300 K as function of the electron energy. After Shichijo and Hess (1981)

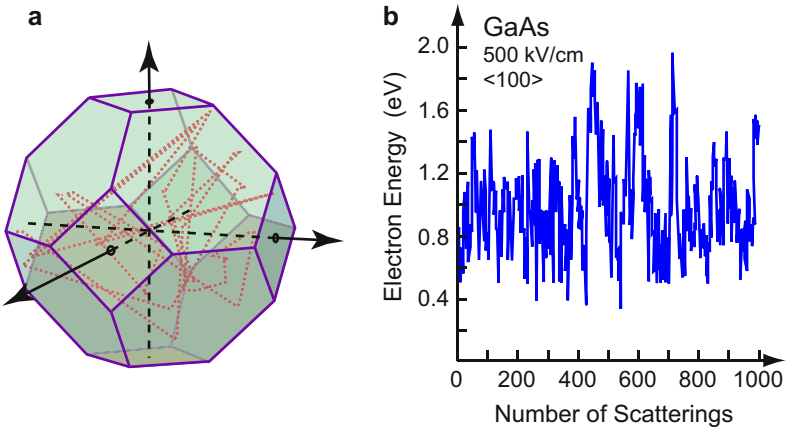


Fig. 9 **(a)** Typical electron trajectory within the Brillouin zone of GaAs at an electric field of 500 kV/cm in the ⟨1 0 0⟩ direction. *Dashed lines* represent the scatterings, the short drift-related sections of the trajectory are hardly visible in this scale. **(b)** Variation of the electron energy in GaAs between scattering events for an applied field of 500 kV/cm in the ⟨1 0 0⟩ direction (impact ionization energy threshold is at 2 eV), obtained by Monte Carlo simulation. After Shichijo and Hess (1981)

where $e_{ii} = 1/\tau(E_{ii})$ represents the scattering (impact ionization) rate at the threshold energy, $P \cong 50 \ldots 500$ is a dimensionless factor, and E is the electron energy.

Typical trajectories in the Brillouin zone near threshold fields are shown in Fig. 9, with corresponding variations of the electron energy computed by Shichijo and Hess during a sample of 1,000 scattering events. This figure shows the typical behavior in which only a few events bring the electron to favorable starting positions from which

they can be accelerated to reach the threshold energy. At lower fields, the amplitudes of the energy variation are much more quiet and electrons stay well below the threshold energy.

The generation rate of electrons due to impact ionization can be obtained from the deformed distribution function f_F, density of states $g(E)$, and energy-relaxation time according to

$$g_\mathrm{i} = \int \frac{f_F(E)\, g(E)}{\tau_\mathrm{e}(E)}\ dE, \tag{36}$$

and is shown for InSb in Fig. 6. The estimate includes the electron-electron interaction, which tends to shorten the relaxation time at higher electron energies. The ionization rate and the generation rate are related by

$$\tilde{\alpha}_\mathrm{i} = \frac{g_\mathrm{i}}{v_\mathrm{d}}, \tag{37}$$

with v_d as the drift velocity. For reviews, see Stillman and Wolfe (1977), Ridley (1983), Dmitriev et al. (1987), Higman et al. (1991), and Fischetti and Vandenberghe (2016).

Impact-ionization is often also conveniently defined in terms of impact-ionization coefficients. These quantities are defined as the number of electron-hole pairs created per unit length and charge carrier, and are efficiently evaluated using the semiempirical model reported by Thornber (1981). The analytical expressions of ionization coefficients for electrons and holes contain the impact-ionization threshold energy [conveniently assumed 1.5 E_g according to (27)], the mean free path, the optical phonon energy, and the thermal energy. Using literature data cited in Nouketcha et al. (2020), this study calculates the critical field and breakdown voltage within *p-i-n* structures for various semiconductors (Si, AlGaN, 4H-SiC, β-Ga_2O_3, C_diamond). Results indicate that solids with larger bandgap do not necessarily have larger critical fields and breakdown voltages – due to a low dielectric constant and high impact-ionization coefficients.

2.2.2 Avalanche Current and Multiplication Factor

With impact ionization, the electron current increases with increasing distance from the cathode. The increment of electrons in a distance dx is given by

$$dn = \tilde{\alpha}_n n\ \ dx, \tag{38}$$

where $\tilde{\alpha}_n$ is the ionization rate of electrons. For band-to-band ionization, this increment is equal to the increment of holes ($dn = dp$) because of the mutual creation of electrons and holes during each ionizing impact.

In a similar fashion, the hole current increases with increasing distance from the anode, with an increment given by

$$dp = \tilde{\alpha}_p p\ \ dx. \tag{39}$$

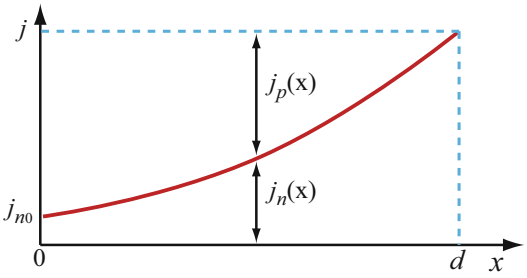


Fig. 10 Current distribution as a function of the distance from the cathode (*at left*) with impact ionization. j_{n0} is the original current at the cathode; $\Delta j_n = j - j_{n0}$ is the current multiplied by impact ionization. Holes are minority carriers; thus $j_p(x = d) = 0$. Pair production by impact ionization makes the incremental currents for electrons and holes equal to each other: $\Delta j_n = \Delta j_p = j_p(x = 0) - j_p(x = d)$

In an arbitrary slab of a semiconductor with planar electrodes at a distance x from the cathode (shown at the left in Fig. 10), the change in electron density is given by the change due to electrons coming from the left, $\tilde{\alpha}_n (n_0 + n_1)\ dx$, plus the change due to the holes coming from the right, $\tilde{\alpha}_p p_2\ dx = \tilde{\alpha}_n n_2\ dx$, yielding

$$dn_1 = \tilde{\alpha}_n (n_0 + n_1)\ dx + \tilde{\alpha}_p n_2\, dx, \tag{40}$$

where n_0 is the electron density at the cathode. Shifting x to the anode, we obtain as the electron density at the anode

$$n_\mathrm{a} = n_0 + n_1 + n_2, \tag{41}$$

and as a *multiplication factor* for electrons

$$M_n = \frac{n_\mathrm{a}}{n_0}. \tag{42}$$

A similar multiplication factor for holes can be defined as $M_p = p_c/p_0$, where p_c is the hole density at the cathode.

After eliminating n_2 from Eqs. 40 and 41, we obtain the linear differential equation:

$$\frac{dn_1}{dx} = \left(\tilde{\alpha}_n - \tilde{\alpha}_p\right)(n_0 + n_1)\ \ + \tilde{\alpha}_p n_\mathrm{a}. \tag{43}$$

This yields for the boundary conditions $n(x = 0) = n_0$ and $n(x = d) = n_0 + n_1 = n_\mathrm{a}$ the electron-multiplication factor:

$$M_n = \frac{1}{1 - \int_0^d \tilde{\alpha}_n \left[\exp\left\{-\int_0^x \left(\tilde{\alpha}_n - \tilde{\alpha}_p\right) dx'\right\}\right] dx}. \tag{44}$$

The electron current increases from cathode to anode by

$$dj_n = j_n \tilde{\alpha}_n \, dx + j_p \tilde{\alpha}_p \, dx, \tag{45}$$

and the hole current increases similarly from anode to cathode, so that the sum of both remains independent of x, as shown in Fig. 10. The total current increases with increasing M. When M becomes infinite, *dielectric breakdown* occurs. This occurs when the integral in Eq. 44 is equal to 1. In materials for which $\tilde{\alpha}_n \cong \tilde{\alpha}_p$, e.g., in GaP, this breakdown condition reduces to

$$\int_0^d \tilde{\alpha} \; dx = 1. \tag{46}$$

This means that every electron or hole passing through the semiconductor creates another electron-hole pair, with a probability equal to 1. As this process goes on with more and more new carriers generated, breakdown occurs.

On the other hand, as long as the integral in Eq. 46 remains below 1, a stationary, increased carrier density is obtained and no breakdown occurs. For a review, see Dmitriev et al. (1987). For impact-ionization characteristics in group III–V semiconductors, see Yuan et al. (2000); for Si, see Maes et al. (1999); for GaN, see Maeda et al. (2021). In noncubic semiconductors, impact-ionization coefficients depend on the orientation of the electric field, as measured, e.g., for 4H-SiC by Stefanakis et al. (2020). Different experimental methods of estimating the impact-ionization coefficients are compared by Ji et al. (2022).

Avalanche multiplication is applied particularly in sensitive photodetectors used for single-photon detection (Ceccarelli et al. 2021; Liu et al. 2022; Vines et al. 2019). Multiplication factors for InAs obtained from Monte Carlo simulation are shown in Fig. 11. The simulation assumed photogeneration of electron-hole pairs at random locations in the bulk. Due to the random creation of the initial carriers within the device and an inherent dead space of the impact-ionization process, the breakdown threshold-field, which corresponds to the sharp vertical increases in the

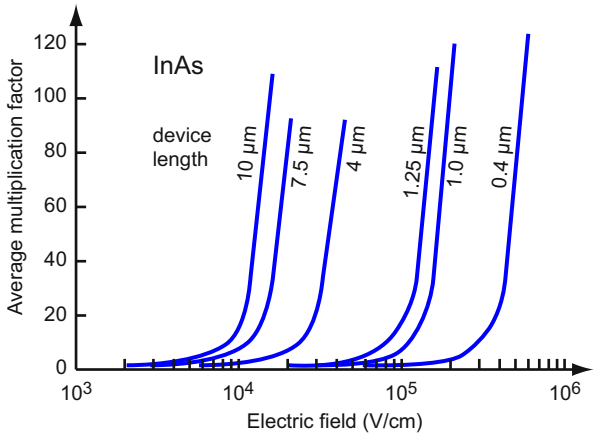


Fig. 11 Multiplication factor for InAs as a function of the electric field for randomly photogenerated initial electron-hole pairs in devices of different lengths. After Satyanadh et al. (2002)

curves, is higher for shorter devices (Satyanadh et al. 2002); in addition, the multiplication factors depend on the length of the device and are smaller than those obtained for electrical injection. The threshold fields are roughly invers to the device length, yielding comparable avalanche voltages for devices of varying lengths due to a comparable total energy gain for carrier multiplication.

2.2.3 Ionization via Energetic Particles

Carriers can be generated by a wide variety of high-energy particles, such as fast electrons, x-ray photons, and various nuclear particles. Such particles are used, e.g., for cathodoluminescence in cathode-ray tubes, imaging, or dosimetry. In all of these excitation processes with high-energy particles, a large number of carriers are produced per incident particle. In addition, however, lattice damage is also produced depending on the energy and the mass of the particle. Such a damage varies in severity from photochemical reactions with x-rays to severe impact damage with protons and α-particles. The excitation process depends on the type of particle. It may occur in the bulk via x-rays or γ-rays or near the surface by fast electrons or α-particles; it may also be of a cascade character, i.e., the initially generated fast carriers may in turn create secondary carriers of sequentially lower energy.

Radiation with energetic particles, such as electrons, nuclei, or ions, results in the creation of lattice defects, if the energy and momentum of the impacting particle are sufficiently large. The energy transfer depends on the scattering angle θ in the center-of-mass reference frame. The actual collision (scattering) event is particularly angle-dependent when far-reaching Coulomb forces provide the interaction, e.g., for protons or α-particles. This type of collision is described as *Rutherford scattering*. On the other hand, the event does *not* depend on θ for so-called *hard-sphere collisions*, i.e., for neutrons or fast ions; here θ is the angle between impacting and scattered particle trajectories.

The result of a particle impact can be divided into metastable atomic displacements (i.e., radiation damage) and electronic ionization. The minimum energy necessary for displacing a lattice atom after an impact is typically on the order of 5 . . . 50 eV and is shown in Table 2 for some semiconductors. From energy and momentum conservation, we estimate that the minimum energy for *impacting electrons* is on the order of 100 keV to cause radiation damage. Atoms displaced in such a manner have insufficient energy to cause secondary damage; however, displacement occurs preferably in the neighborhood of lattice defects.

Bombardment with *protons or neutrons* transfers substantially more energy, and results into numerous secondary damage events, which are wider-spaced for protons and in a more compact region for neutrons. A similar, compact region of high disorder is observed after bombardment with *high-energy ions*. Ion bombardment from a gas discharge creates disorder in a region close to the surface. Ion implantation leaves even larger defect clusters. Careful posttreatment annealing is required to restore sufficient order to reobtain attractive semiconductive properties (i.e., high mobility and low recombination rates), since these defect clusters usually act as major scattering or recombination centers. For a review, see Hayes and Stoneham (1984).

Table 2 Measured atomic displacement energies E_{dis} and calculated displacement energy thresholds $E_{dis,thr,Frenkel}$ for A or B Frenkel-pair formation by radiation damage in AB compound semiconductors, after Hayes and Stoneham (1984) and van Vechten (1980), respectively. Values of $E_{dis,thr,Frenkel}$ in brackets are experimental data

	E_{dis} (eV)		$E_{dis,thr,Frenkel}$ (eV)	
Crystal	Cation A	Anion B	Cation A	Anion B
GaAs	9	9.4	12.4 (15–17)	17.6
InP	6.7	8.7	12.2	16.5
InAs	6.7	8.3	10.8	14.9
InSb	5.7	6.6	9.1	12.9
ZnO	30–60	60–120	18.5	41.4 (57)
ZnS	7–9	15–20	12.1	27.5
ZnSe	7–10	6–8	10.8 (10)	24.8
CdS	2–7	8–25	11.0 (8)	24.3
CdSe	6–8	8–12	10.1	22.4
CdTe	5.6–9	5–8	8.7	19.5

2.3 Electron Tunneling

Electrons can penetrate potential barriers if these barriers are thin and low enough, as pointed out by Oppenheimer (1928), Fowler and Nordheim (1928), and others. We addressed this effect briefly in the discussion of perpendicular carrier transport in two-dimensional structures in Sect. 2.1 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors". Here, we consider this quantum mechanical phenomenon in more detail; it can easily be understood by recognizing that the wavefunction of an electron cannot immediately stop at a barrier but rather decreases exponentially into the barrier with a slope determined by the barrier height. If the barrier is thin, there is a nonzero amplitude of the wavefunction remaining at the end of the barrier, i.e., a nonzero probability for the electron to penetrate, as illustrated in Fig. 12. For a comprehensive review, see Duke (1969) or Wolf (1975). This *electron tunneling* phenomenon will first be recalled from ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors" for a simple one-dimensional rectangular barrier and then be discussed for differently shaped barriers.

Tunneling Through a Rectangular Barrier When an electron wave with an energy E impinges on a rectangular barrier with height eV_0 and width d, we distinguish three wavefunctions in the three regions: before, in, and after the barrier, see Fig. 12:

$$\begin{aligned} \varphi_1(x) &= A_1 e^{ik_0 x} + B_1 e^{-ik_0 x} \text{ for } x < -d/2 \\ \varphi_2(x) &= A_2 e^{ik_1 x} + B_2 e^{-ik_1 x} \text{ for } -d/2 < x < d/2 \\ \varphi_3(z) &= A_3 e^{ik_0 x} \text{ for } x > d/2 \end{aligned} \tag{47}$$

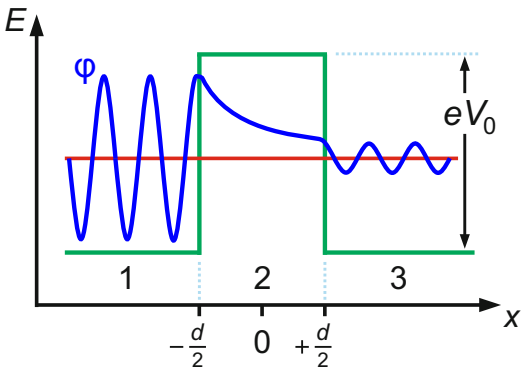


Fig. 12 One-dimensional rectangular barrier (*green lines*) with incoming and attenuated transmitted waves (*blue curves*)

with amplitude coefficients A_i for the incoming wave and B_i for the reflected wave at $-d/2$ and $+d/2$, respectively. The wavenumbers k_0 and k_1 outside and inside the barrier are given by

$$k_0^2 = 2m_n E/\hbar^2 \text{ and } k_1^2 = 2m_n (eV_0 - E)/\hbar^2. \tag{48}$$

Continuity of φ and $d\varphi/dx$ at $x = -d/2$ and at $x = +d/2$ provides four conditions for the coefficients A_i and B_i, from which we derive an expression for the transmission and reflection probabilities T_e and R:

$$T_e = \left(\frac{A_3}{A_0}\right)^2 \text{ and } R = B_1^2 = 1 - \left(\frac{A_3}{A_1}\right)^2. \tag{49}$$

Solving the set of condition equations for A_i and B_i yields

$$T_e = \left[1 + \left(\frac{k_0^2 + k_1^2}{4k_0 k_1}\right)^2 \sinh^2\left(\frac{k_1 d}{2}\right)\right]^{-1} \cong \left(\frac{4k_0 k_1}{k_0^2 + k_1^2}\right)^2 \exp(-k_1 d); \tag{50}$$

this approximation is for $k_1 d \gg 1$. After introducing the expressions for k_0 and k_1, we obtain for the transmission probability:

$$T_e \cong 16 \left(\frac{E}{eV_0}\right)^2 \left(\frac{eV_0}{E} - 1\right) \exp\left(-d\sqrt{2m_n(eV_0 - E)/\hbar^2}\right). \tag{51}$$

For a barrier height eV_0 much larger than the kinetic energy of the tunneling electrons, the -1 in Eq. 51 can be neglected yielding a more simple approximation. For example, thermal electrons [$E = kT$ (300 K)] are attenuated by a factor of 2.2×10^{-3} when impinging on a 10 Å thick barrier of 1 V height.

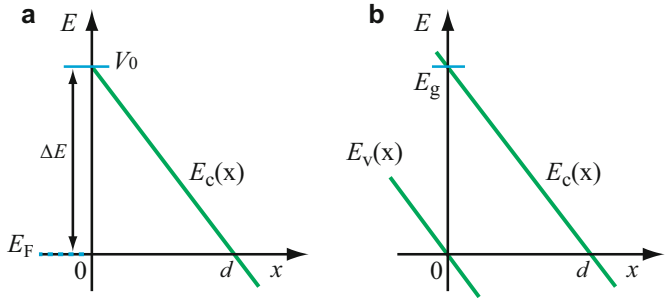


Fig. 13 Triangular potential barrier at **(a)** a metal-semiconductor contact neglecting image forces, **(b)** for a band-to-band transition in tilted band edges

2.3.1 Tunneling Through Triangular or Parabolic Barriers

Tunneling Through a Triangular Barrier The transmission probability also depends on the shape of the barrier. In semiconducting devices, the potential barrier often can be approximated by a *triangular shape*, for instance, at a simplified metal-semiconductor (Fig. 13a) or heterojunction interface or for a band-to-band transition (Fig. 13b). At sufficiently high fields, the band edges are substantially tilted and tunneling through such a barrier can become important.

The transmission probability T_e for a wide class of barriers is given in the WKB approximation[4] by

$$T_e \cong \exp\left(-\int_0^d |k(x)|\ dx\right), \tag{52}$$

neglecting the pre-exponential factor which is on the order of 1 and with the shape of the barrier contained in the wavevector:

$$k(x) = \sqrt{2m_n[eV_0 - E(x)]/\hbar^2}. \tag{53}$$

For the triangular barrier, $k(x)$ is given by

$$k(x) = \sqrt{\frac{2m_n}{\hbar^2}(\Delta E - eFx)}. \tag{54}$$

Integration of Eq. 52 with Eq. 54 yields

[4]This semiclassical approximation of the one-dimensional, stationary Schrödinger equation is named after G. Wentzel, H.A. Kramers, and L. Brillouin.

$$\int k(x)dx = \sqrt{\frac{2m_n}{\hbar^2}}\ \frac{2}{3}\ \frac{(\Delta E - eFx)^{3/2}}{eF/2}\Bigg|_0^d, \tag{55}$$

and with $\Delta E - eF{\cdot}0 = \Delta E$, and $\Delta E - eF{\cdot}d = 0$, we obtain

$$T_{e,\text{triangle}} = \exp\left(-\frac{4}{3}\ \sqrt{\frac{2m_n}{\hbar^2}}\frac{(\Delta E)^{3/2}}{eF}\right). \tag{56}$$

The pre-exponential factor is similar in form to that given in Eq. 51; compared with the exponential, its ΔE- and F-dependences are usually neglected. The barrier height to field relation is superlinear; thus, doubling the barrier height requires $2^{3/2} = 2.83$ times the field to result in the same tunneling probability.

For **band-to-band tunneling**, ΔE in Eq. 56 is replaced by the bandgap energy E_g:

$$T_{e,\text{band-band}} \cong \exp\left(-\frac{4}{3}\ \sqrt{\frac{2m_n}{\hbar^2}}\frac{E_g^{3/2}}{eF}\right). \tag{57}$$

Although this transmission probability is multiplied by a large density of electrons acting as candidates for tunneling from band to band, very high fields ($>10^6$ V/cm) are needed to produce significant tunneling currents, except for very-narrow-gap semiconductors.

A **parabolic barrier** is better suited as an approximation for barriers in which two fields overlap, e.g., the Coulomb-attractive field of a center or the image force of a metal-semiconductor barrier and the external field. This barrier type is shown in Fig. 14. The corresponding wavevector is given by

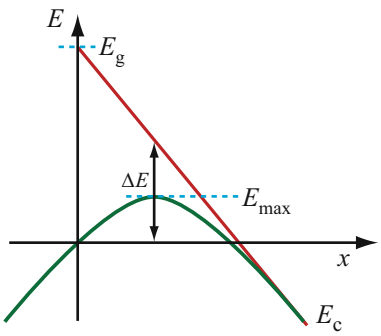


Fig. 14 Parabolic barrier showing the lowering of a triangular barrier from E_g to E_{max}

$$k(x) = \sqrt{\frac{2m_n}{\hbar^2}\frac{(\Delta E)^2 - (eFx)^2}{\Delta E}}, \tag{58}$$

and yields after integration

$$\int_0^d k(x)dx = \sqrt{\frac{2m_n}{\hbar^2}}\ \frac{\pi}{8}\ \frac{(\Delta E)^{3/2}}{eF/2}. \tag{59}$$

Therefore, the transmission probability for a parabolic barrier is given by

$$T_{e,\text{parabola}} = \exp\left(-\frac{\pi}{4}\ \sqrt{\frac{2m_n}{\hbar^2}}\frac{(\Delta E)^{3/2}}{eF}\right), \tag{60}$$

with an exponent that is reduced by a numerical factor of $3\pi/16 = 0.589$ from the expression for the triangular barrier.

2.3.2 Tunneling in a Three-Dimensional Crystal

In the one-dimensional model, we are concerned only with the momentum in the direction of the normal on the barrier plane; it decreases exponentially during the tunneling transition. In a *three*-dimensional crystal, however, there is a three-dimensional distribution of momenta. Assuming a planar barrier, only the component in the direction of tunneling is influenced by the tunneling process; the two components perpendicular to the tunneling are not: *these components are conserved.* The total transition probability is therefore reduced by a factor given by the fraction of electrons having a favorable momentum component to those that do not. This fraction η was calculated by Moll (1964):

$$\eta = \exp\left(-E_\perp/\overline{E}\right), \tag{61}$$

with $E_\perp$ the energy associated with the momentum perpendicular to the direction of tunneling, equal to $\hbar^2 k_\perp^2/(2m_n)$, and with $\overline{E}$ given by

$$\overline{E} = \sqrt{\frac{\hbar^2}{2m_n}}\frac{eF}{\pi\sqrt{\Delta E}}. \tag{62}$$

Hence, for the tunneling probability through a flat plate barrier with a potential of parabolic shape, we obtain

$$T_{e,\text{parabola, 3D}} = \eta\ T_{e,\text{parabola, 1D}} = \exp\left(-\frac{E_g + 4E_\perp}{4\overline{E}}\right). \tag{63}$$

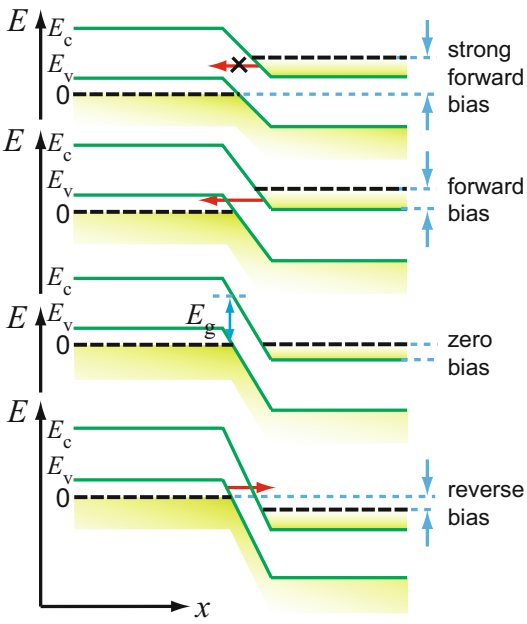


Fig. 15 Band model with idealized tunnel junction for four different bias conditions. The current flow is indicated by *red arrows*

Tunneling Currents The current is proportional to the product of the tunneling transmission-probability and the incident carrier flux. As an example, an idealized n^+ p^+-junction is shown in Fig. 15 for various bias conditions.

Depending on the bias, the current can flow across this junction from filled states in the valence to empty states in the conduction band with reverse bias or from filled states in the conduction band to the empty states in the valence band with forward bias. The net current is the difference of both (Moll 1964): $j_n = \vec{j}_n - \overleftarrow{j}_n$. For any given bias, we have

$$\vec{j}_n = j_{n,\text{v-c}} = A \int_{E_{c,r}}^{E_{v,l}} \{N_v(E)\ f_n(E)\}_l\ \left\{N_c(E)\ f_p(E)\right\}_r T_{e,\text{triangle,3D}}\ dE, \tag{64}$$

$$\overleftarrow{j}_n = j_{n,\text{c-v}} = A \int_{E_{c,r}}^{E_{v,l}} \{N_c(E)\ f_n(E)\}_r\ \left\{N_v(E)\ f_p(E)\right\}_l T_{e,\text{triangle,3D}}\ dE, \tag{65}$$

with $T_{e,\text{triangle,3D}} = \eta T_{e,\text{triangle,1D}}$ from Eqs. 56 and 61. The subscripts l and r stand for the left and right sides in Fig. 15, $N_v(E)$ and $N_c(E)$ are the respective densities of states in the valence and conduction bands, and f_n and $f_p = 1 - f_n$ are the Fermi distributions for electrons and holes. The first factor in parentheses under the integral identifies the density of available electrons; the second factor gives the density of holes into which tunneling can proceed. We therefore obtain

$$j_n = A \int_{E_{c,r}}^{E_{v,l}} N_v(E) N_c(E) \{f_{n,r}(E) - f_{n,l}(E)\} \; T_{e,\text{triangle,3D}} \; dE. \tag{66}$$

The proportionality constant A can be obtained by accounting for the charge and the velocity of electrons in k space; with these, the current density is

$$j_n = \frac{e \; m_n}{2\pi^2\hbar^3} \exp\left(\frac{\pi \; m_n^{1/2} E_g^{3/2}}{2\sqrt{2} e\hbar \; F}\right) \int\int \{f_{n,r}(E) - f_{n,l}(E)\} \; \exp\left(-\frac{2E_\perp}{\overline{E}}\right) \; dE \; dE_\perp, \tag{67}$$

assuming the same isotropic effective mass at the left and right sides of the junction. After integrating over $E_\perp$, which yields $\overline{E}/2$, and approximating the integral over E, we obtain

$$j_n = \frac{e^3 \; m_n^{1/2} d}{2\sqrt{2}\pi^3\hbar^2 E_g^{1/2}} \; F^2 \; \exp\left(-\frac{\pi \; m_n^{1/2} E_g^{3/2}}{2\sqrt{2} e\hbar \; F}\right) \tag{68}$$

for $V_a \gg kT/e$ and $V_a \gg \overline{E}/e$; V_a is the applied voltage across the barrier. Equation 68 has the field dependence given by the Fowler-Nordheim formula (Fowler and Nordheim 1928): $j \propto F^2 \exp(-F_0/F)$. Equation 68 yields a substantial tunneling ($>10^{-3}$ A/cm^2) for fields in excess of 1.5×10^6 V/cm for the parameters $d = 100$ Å and $m_n \cong m_0$. A reduced effective mass lowers the critical field by a factor of 3 for $m_n = 0.1 \; m_0$.

2.3.3 Tunneling Spectroscopy

The dependence of the tunneling current on the level-density distribution (Eq. 66) permits the use of this current for obtaining information about the distribution. There are several methods available; they involve a similar principle in shifting the Fermi level or quasi-Fermi level on one side of a barrier with respect to the Fermi level on the other side (see Fig. 15), thereby permitting the tunneling of carriers through the barrier at variable energies. Thus, the level distribution is profiled near the top of the valence band or in the conduction band, depending on forward or reverse bias, respectively: the increment in tunneling current becomes larger when more levels become available. The probing side of the barrier can be either a highly doped semiconductor, a metal electrode, or a superconductor.

Examples for tunneling spectroscopy were given for a double-barrier QW and superlattices in Figs. 10, 11, and 16 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors". An example for an asymmetric GaAs/AlGaAs/GaAs double heterostructure is shown in Fig. 16. Two 2D electron gases exist on either sides of the doped AlGaAs barrier; contacts to GaAs on both sides allow for shifting the 2D subbands on each side with respect to each other. The relative energy shift of the 2D systems equals the bias eV to a very good approximation, since all of

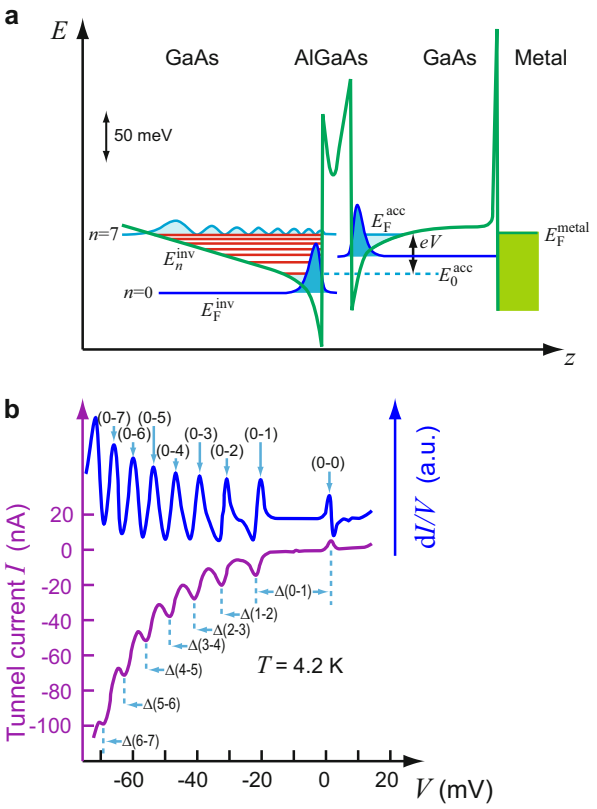


Fig. 16 **(a)** Calculated profile of the conduction-band edge (*green curves*) of a GaAs/AlGaAs/GaAs double heterostructure for a given bias *V*. E_n and E_F denote the subband and Fermi energies; indices *inv* and *acc* indicate inversion and accumulation. **(b)** *I-V* and *dI/dV* characteristics of the structure with a 200 Å wide AlGaAs barrier, recorded at low temperature. After Smoliner (1996)

the applied voltage drops across the barrier (Smoliner 1996). The self-consistently calculated structure of the conduction-band edge is shown in Fig. 16a. Bound states E_n are clearly seen in the current-voltage (*I-V*) characteristics and even more pronounced in the *dI/dV* derivative.

Other examples of tunnel spectroscopy are listed in the following sections (see also Tsui 1982; Hayes et al. 1986).

Phonon Assistance Indirect bandgap materials show a much lower band-to-band tunneling probability, since an indirect transition requires an additional phonon to accomplish the change in momentum. An estimate of the tunneling probability by Keldysh (1958) shows a similar functional behavior as given by Eq. 57, but reduced by a factor on the order of 10^{-3}.

Other changes in the tunneling probability are seen when the phonon energy is used in addition to the electron energy to bridge the barrier:

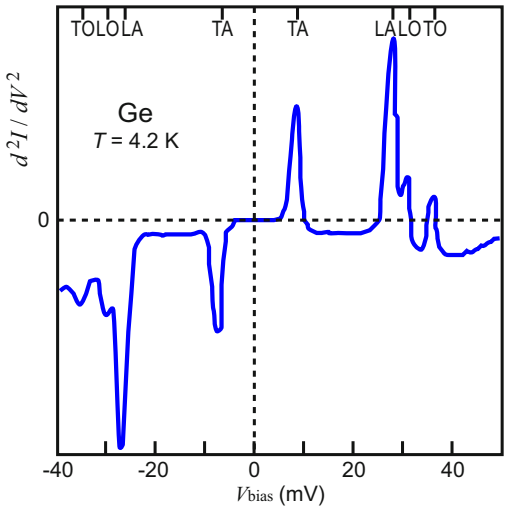


Fig. 17 Second derivative of tunnel current to emphasize the structure for a Ge tunnel diode, indicating the phonon spectrum. After Payne (1965)

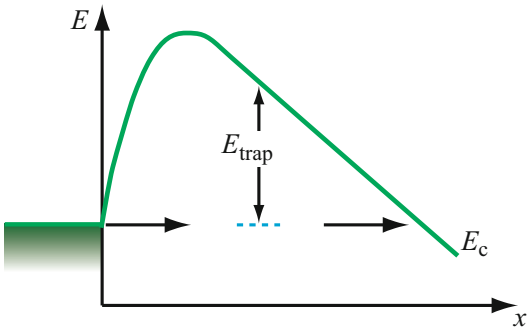


Fig. 18 Tunneling from a contact metal into a semiconductor, which becomes marked when the trap level is lowered by the electric field to coincide with the Fermi level of the metal

$$T_e = \exp\left(-\frac{4}{3}\sqrt{\frac{2m_n}{\hbar^2}}\frac{(E_g - \hbar\omega_0)^{3/2}}{eF}\right). \tag{69}$$

When the bias V reaches the phonon energy, $V \geq \hbar\omega_0/e$, the tunneling current increases measurably. This change can be used to identify the corresponding phonon energies as shown in Fig. 17 for Ge.

Phonon-assisted tunneling from a localized level to the conduction band also occurs in direct-bandgap materials. The effect was first demonstrated and modeled for the EL2 donor level in GaAs (Makram-Ebeid and Lannoo 1982) and recently assigned to (non-resonant) trap-assisted charge transport in insulating SiN_x (Novikov and Gritsenko 2022).

Trap Assistance Tunneling through a barrier into the conduction band can be assisted by *resonant* tunneling first into a trap and then from the trap into the conduction band – a *two-step tunneling* process, as indicated in Fig. 18. When

defect centers are spaced close enough to an interface, and are present at sufficient densities, the two-step tunneling can substantially increase the overall tunneling probability, which is calculated from the sum of the reciprocal individual probabilities:

$$T_{e,\text{two-step}} = \left(\frac{1}{T_{e1}} + \frac{1}{T_{e2}}\right)^{-1}. \tag{70}$$

Photon Assistance (Franz-Keldysh Effect) The Franz-Keldysh effect is an important tunneling phenomenon in which only a small fraction of the energy is supplied from the electric field. Most of the energy comes from an optical excitation from a state in the gap near the valence-band edge to a symmetrical state close to the conduction-band edge; see Fig. 19a.

This three-step process was suggested by Franz (1958) and Keldysh (1958) and was observed first by Böer et al. (1959). It results in a shift of the absorption edge toward lower energies. The amount of the shift can be estimated from the photon-assisted tunneling – compared with Eq. 69:

$$T_e = \exp\left(-\frac{4}{3}\sqrt{\frac{2m_n}{\hbar^2}}\frac{(E_g - h\nu)^{3/2}}{eF}\right), \tag{71}$$

where m_n is the effective mass in the direction of the electric field. As a result, the optical absorption edge is shifted (Fig. 19b) by the same amount:

$$\alpha_o = \alpha_{o0} \exp\left(-\frac{4}{3}\sqrt{\frac{2m_n}{\hbar^2}}\frac{(E_g - h\nu)^{3/2}}{eF}\right). \tag{72}$$

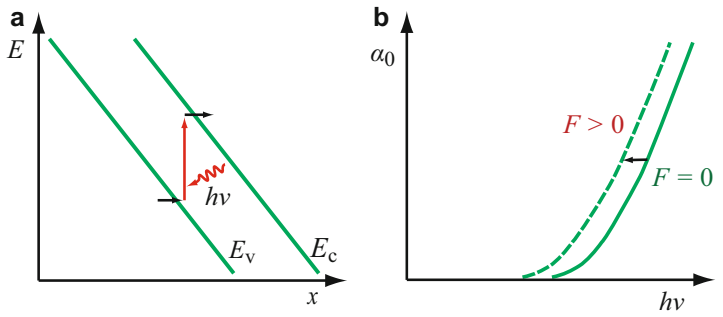


Fig. 19 Franz-Keldysh effect: **(a)** Photon-assisted tunneling from band to band; **(b)** resulting shift of the band-edge absorption

The shift $\Delta E_{g,opt}$ can be obtained from the condition that the exponent remains constant and $\cong 1$, resulting in

$$\Delta E_{g,opt} = E_g - h\nu = \left(\frac{4}{3}\ eF\sqrt{\frac{\hbar^2}{2m_n}}\right)^{2/3}, \tag{73}$$

or

$$\Delta E_{g,opt} = 7.25 \times 10^{-6}(F/(\mathrm{V/m}))^{2/3} \quad (\mathrm{eV}). \tag{74}$$

This means, for a band-edge shift of ~10 meV, an electric field of ~50 kV/cm is required. A theoretical analysis of the optical absorption tails near the absorption edge is given by Kuwamura and Yamada (1996).

The photon-assisted tunneling is not restricted to the transition from the valence to conduction band near its principal edge; an electron can optically be excited to a state close to any higher band and complete the transition via tunneling. The Franz-Keldysh effect thereby provides a relatively simple method for measuring the energy of characteristic points in the $E(k)$ behavior of any band (Seraphin 1964; Aspnes 1967). At first view, this may suggest no advantage over a purely optical transition. When applying an ac electric field, the maximum modulation signal of the optical response is observed where the absorption edge has the highest slope, i.e., at the inflection point of $\kappa(\lambda)$. With overlapping higher bands, the optical absorption itself is not very structured. In contrast, however, the Franz-Keldysh modulation is highly structured and shows unusually sharp features with characteristic oscillations; see Fig. 20. For reviews, see Frova and Handler (1965), Seraphin and Bottka (1965), and Aspnes (1980).

3 Summary

Free carriers are obtained by thermal, optical, or electrical field generation or by a combination of these processes. In addition, carriers can be generated by high-energy particles, such as fast electrons, ions, or x-ray photons; these excitation processes produce a large number of carriers, but also lattice damage depending on the energy and mass of the particle. *Thermal carrier generation* from shallow centers requires the absorption of a phonon of sufficient energy or at most a few phonons involving intermediate steps into excited states. Thermal generation from deep centers requires multiphonon-induced giant oscillations. Such oscillations cause a crossover of the defect level and levels of the conduction band and provides sufficient probability for the defect-level electron to be ejected into band states. Similarly, a giant oscillation of an adjacent lattice atom could result in a replenishing of the emitted electron from the valence band. Thereby, a defect-assisted, band-to-band thermal excitation can take place.

Optical carrier generation proceeds from defect levels or as band-to-band direct or indirect generation with photons of sufficient energy. *Electrical field generation of carriers* can be caused by a field-enhanced thermal generation from Coulomb-

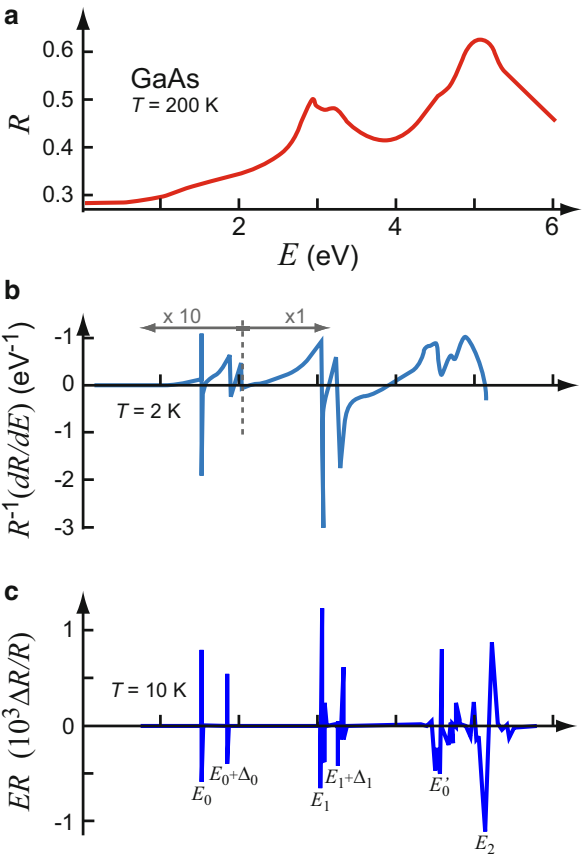


Fig. 20 Comparison of reflectance **(a)**, energy-derivative reflectance **(b)**, and low-field Franz-Keldysh reflectance **(c)** of GaAs. After Aspnes (1980)

attractive defect centers (Frenkel-Poole effect) at fields typically in the low 10 kV/cm range. Impact ionization from shallow impurity levels can be induced at very low fields, typically 10 V/cm at low temperatures. Impact ionization from deep centers is observed at much higher fields in materials of sufficient thickness and carrier density. Band-to-band impact ionization in typical semiconductors requires fields in the 10^5 V/cm range. At still higher fields, typically 10^6 V/cm, tunneling from deep defect centers or from the valence band takes place. These field-excitation mechanisms increase rapidly with increased bias and, when unchecked by a sufficiently large series resistance, cause dielectric breakdown.

Interaction between two or more of the generation processes provides a means of sensitive detection. For instance, optically stimulated thermal or field generation of carriers yields information about defect levels or band spectra which are difficult to obtain otherwise. Field-enhanced thermal generation provides an ionization mechanism that acts selectively on Coulomb-attractive centers and can be controlled at relatively low-field strengths. The generation of carriers, controlled by external parameters such as temperature, optical excitation, or bias (electric field), is the primary process for many semiconducting devices.

References

Anderson CL, Crowell CR (1972) Threshold energies for electron-hole pair production by impact ionization in semiconductors. Phys Rev B 5:2267

Aspnes DE (1966) Electric-field effects on optical absorption near thresholds in solids. Phys Rev 147:554

Aspnes DE (1967) Electric field effects on the dielectric constant of solids. Phys Rev 153:972

Aspnes DE (1980) Modulation spectroscopy/electric field effects on the dielectric function of semiconductors. In: Moss TS, Balkanski M (eds) Handbook on semiconductors, Optical properties of solids, vol 2. North-Holland Publ, Amsterdam, pp 109–154

Baraff GA (1962) Distribution functions and ionization rates for hot electrons in semiconductors. Phys Rev 128:2507

Böer KW (1977) The CdS/Cu_2S solar cell I. Minority carrier generation and transport in the Cu_2S emitter. Phys Status Solidi A 40:355

Böer KW, Hänsch HJ, Kümmel U (1959) Anwendung elektro-optischer Effekte zur Analyse des elektrischen Leitungsvorgangs in CdS Einkristallen. Z Phys 155:170. (Application of electro-optical effects fort the analysis of electrical conductance in CdS single crystals, in German)

Bratt PR (1977) Impurity germanium and silicon infrared detectors. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals vol 12. Academic Press, New York, pp 39–142

Ceccarelli F, Acconcia G, Gulinatti A, Ghioni M, Rech I, Osellame R (2021) Recent advances and future perspectives of single-photon avalanche diodes for quantum photonics applications. Adv Quantum Technol 4:2000102

Connell GAN, Champhausen DL, Paul W (1972) Theory of Poole-Frenkel conduction in low-mobility semiconductors. Philos Mag 26:541

Curby RC, Ferry DK (1973) Impact ionization in narrow gap semiconductors. Phys Status Solidi A 15:319

Dallacasa V, Paracchini C (1986) Field-enhanced electronic transport in solids. Phys Rev B 34:8967

Dmitriev AP, Mikhailova MP, Yassievich IN (1987) Impact ionization in $A^{III}B^{V}$ semiconductors in high electric fields. Phys Status Solidi B 140:9

Duke CB (1969) Tunneling in solids. Academic Press, New York

Fischetti MV, Vandenberghe WG (2016) Advanced physics of electron transport in semiconductors and nanostructures. Springer, Switzerland

Fowler RH, Nordheim L (1928) Electron emission in intense electric fields. Proc R Soc Lond A 119:173

Franz W (1958) Einfluß eines elektrischen Feldes auf eine optische Absorptionskante. Z Naturforsch A 13:484 (Effect of an electric field on the optical absorption edge, in German)

Frenkel JI (1938) On pre-breakdown phenomena in insulators and electronic semi-conductors. Phys Rev 54:647

Frova A, Handler P (1965) Franz-Keldysh effect in the space-charge region of a germanium $p-n$ junction. Phys Rev 137:A1857

Ganichev SD, Ziemann E, Prettl W, Istratov AA, Weber ER (1999) High field limitation of Poole-Frenkel emission caused by tunneling. Mater Res Soc Symp Proc 560:239

Grupen M (2011) Energy transport model with full band structure for GaAs electronic devices. J Comput Electron 10:271

Hartke JL (1968) The three-dimensional Poole-Frenkel effect. J Appl Phys 39:4871

Hayes W, Stoneham AM (1984) Defects and defect processes in nonmetallic solids. John Wiley & Sons, New York

Hayes JR, Levi AFJ, Gossard AC, English JH (1986) Base transport dynamics in a heterojunction bipolar transistor. Appl Phys Lett 49:1481

Higman JM, Bude J, Hess K (1991) Electronic transport in semiconductors at high energy. Comput Phys Commun 67:93

Hill RM (1971) Poole-Frenkel conduction in amorphous solids. Philos Mag 23:59

Ji D, Zeng K, Bian Z, Shankar B, Gunning BP, Binder A, Dickerson JR, Aktas O, Anderson TJ, Kaplar RJ, Chowdhury S (2022) A discussion on various experimental methods of impact ionization coefficient measurement in GaN. AIP Adv 12:030703
Keldysh LV (1958) Influence of the lattice vibrations of a crystal on the production of electron-hole pairs in a strong electrical field. Sov Phys JETP 7:665
Keldysh LV (1965) Concerning the theory of impact ionization in semiconductors. Sov Phys JETP 21:1135
Kuwamura Y, Yamada M (1996) Analysis of modified Franz-Keldysh effect under influence of electronic intraband relaxation phenomena. Jpn J Appl Phys 35:6117
Lautz G (1961) Elektrische Eigenschaften der Halbleiter bei tiefen Temperaturen. In: Sauter F (ed) Halbleiterprobleme, vol VI. Vieweg, Braunschweig, pp 21–117. (Electrical properties of semiconductors at low temperatures, in German)
Liu C, Ye H-F, Shi Y-L (2022) Advances in near-infrared avalanche diode single-photon detectors. Chip 1:5
Maeda T, Narita T, Yamada S, Kachi T, Kimoto T, Horita M, Suda J (2021) Impact ionization coefficients and critical electric field in GaN. J Appl Phys 129:185702
Maes W, De Meyer K, Van Overstraeten R (1999) Impact ionization in silicon: a review and update. Solid State Electron 33:705
Makram-Ebeid S, Lannoo M (1982) Quantum model for phonon-assisted tunnel ionization of deep levels in a semiconductor. Phys Rev B 25:6406
Miron D, Krylov I, Baskin M, Yalon E, Kornblum L (2019) Understanding leakage currents through Al_2O_3 on $SrTiO_3$. J Appl Phys 126:185301
Moll JL (1964) Physics of semiconductors. McGraw-Hill, New York
Nouketcha FLL, Cui Y, Lelis A, Green R, Darmody C, Schuster J, Goldsman N (2020) Investigation of wide- and ultrawide-bandgap semiconductors from impact-ionization coefficients. IEEE Transact Electron Devices 67:3999
Novikov YuN, Gritsenko VA (2021) Multiphonon ionization of deep centers in amorphous boron nitride. JETP Lett 114:433
Novikov YuN, Gritsenko VA (2022) Multiphonon trap ionization mechanism in amorphous SiN_x. J Non-Crystalline Sol 582:121442
Oppenheimer JR (1928) Three notes on the quantum theory of aperiodic effects. Phys Rev 31:66
Pai DM (1975) Electric-field-enhanced conductivity in solids. J Appl Phys 46:5122
Payne RT (1965) Phonon energies in germanium from phonon-assisted tunneling. Phys Rev 139: A570
Poole HH (1921) On the electrical conductivity of some dielectrics. Philos Mag 42:488
Reddy P, Khachariya D, Mecouch W, Breckenridge MH, Bagheri P, Guan Y, Kim JH, PavlidisS, Kirste R, Mita S, Kohn E, Collazo R, Sitar Z (2021) Study on avalanche breakdown and Poole-Frenkel emission in Al-rich AlGaN grown on single crystal AlN. Appl Phys Lett 119:182104
Ridley BK (1983) A model for impact ionisation in wide-gap semiconductors. J Phys C 16:4733
Satyanadh G, Joshi RP, Abedin N, Singh U (2002) Monte Carlo calculation of electron drift characteristics and avalanche noise in bulk InAs. J Appl Phys 91:1331
Schappe RS, Walker T, Anderson LW, Lin CC (1996) Absolute electron-impact ionization cross section measurements using a magneto-optical trap. Phys Rev Lett 76:4328
Schroeder H (2015) Poole-Frenkel-effect as dominating current mechanism in thin oxide films – an illusion?! J Appl Phys 117:21510
Sclar N, Burstein E (1957) Impact ionization of impurities in germanium. J Phys Chem Solids 2:1
Seraphin BO (1964) The effect of an electric field on the reflectivity of germanium. In: Hulin M (ed) Proc 7th Int Conf Phys Semicond. Dunod, Paris, pp 165–170
Seraphin BO, Bottka N (1965) Franz-Keldysh effect of the refractive index in semiconductors. Phys Rev 139:A560
Shichijo H, Hess K (1981) Band-structure-dependent transport and impact ionization in GaAs. Phys Rev B 23:4197
Shockley W (1961) Problems related to *p-n* junctions in silicon. Solid State Electron 2:35

Smoliner J (1996) Tunnelling spectroscopy of low-dimensional states. Semicond Sci Technol 11:1
Stefanakis D, Chi X, Maeda T, Kaneko M, Kimoto T (2020) Experimental determination of impact ionization coefficients along $\langle 11\bar{2}0 \rangle$ in 4H-SiC. IEEE Transact Electron Devices 67:3740
Stillman GE, Wolfe CM (1977) Avalanche photodiodes. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals vol 12. Academic Press, New York, pp 291–393
Stoneham AM (1981) Non-radiative transitions in semiconductors. Rep Prog Phys 44:1251
Thornber KK (1981) Applications of scaling to problems in high-field electronic transport. J Appl Phys 52:279
Tsui DC (1982) Semiconductor tunneling. In: Moss TS, Paul W (eds) Handbook on semiconductors, Band theory and transport properties, vol 1. North-Holland Publ, Amsterdam, pp 661–688
van Vechten JA (1980) A simple man's view of the thermochemistry of semiconductors. In: Moss TS, Keller SP (eds) Handbook of Semiconductors. Vol 3: Materials properties and preparation. North Holland Publication, Amsterdam, pp 1–111
Vines P, Kuzmenko K, Kirdoda J, Dumas D, Mirza MM, Millar RW, Paul DJ, Buller GS (2019) High performance planar germanium-on-silicon single-photon avalanche diode detectors. Nat Commun 10:1086
Wang R, Tong X, Xu J, Dong C, Cheng Z, Zhang L, Zhang S, Zheng P, Chen F-X, Zhang Y, Tan W (2020) Acceptor decoration of threading dislocations in (Al,Ga)N/GaN heterostructures. Phys Rev Appl 14:024039
Wolf EL (1975) Nonsuperconducting electron tunneling spectroscopy. In: Seitz F, Turnbull D, Ehrenreich H (eds) Solid state physics vol 30. Academic Press, New York, pp 1–91
Wolff PA (1954) Theory of electron multiplication in silicon and germanium. Phys Rev 95:1415
Yuan P, Hansing CC, Anselm KA, Lenox CV, Nie H, Holmes AL, Streetman BG, Campbell JC (2000) Impact ionization characteristics of III-V semiconductors for a wide range of multiplication region thicknesses. IEEE J Quantum Electron 36:198

Carrier Recombination and Noise

30

Karl W. Böer and Udo W. Pohl

Contents

1 Nonradiative Recombination 1242
1.1 Energy Transfer to Phonons 1243
1.2 Energy Transfer to Electrons 1251
2 Statistics of Recombination 1256
2.1 Trapping or Recombination 1256
2.2 Thermal Equilibrium, Steady State, and Shockley-Read-Hall Recombination 1259
3 Radiative Recombination 1262
3.1 Thermal Radiation 1263
3.2 Intrinsic Luminescence 1265
3.3 Extrinsic Luminescence 1274
4 Noise 1281
4.1 Description of Fluctuation 1281
4.2 Noise in Equilibrium Conditions 1283
4.3 Nonequilibrium Noise 1284
4.4 The 1/*f* Noise 1287
5 Summary 1291
References 1292

Abstract

In steady state, for each act of carrier generation or excitation there must be one inverse process of recombination or relaxation. Carriers can return immediately or after scattering to their original state, or they can recombine radiatively or nonradiatively with another state.

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_30

Nonradiative recombination is almost always defect-center controlled; it releases energy in the form of phonons, or in Auger recombination, by accelerating another carrier. Phonon emission occurs as a single-phonon process when trapping a carrier at a shallow defect center, or as a multiphonon emission when recombination occurs at a deep center. In *carrier traps*, which are located close to one band, excitation into the adjacent band and trapping at the center dominate, while in *recombination centers*, which are located closer to the center of the bandgap, carriers recombine from one band to the other. The capture cross-section of defect centers spread over more than 12 orders of magnitude.

Radiative recombination proceeds as an emission delayed by the lifetime of an excited state and changes in energy after relaxation of the excited state. The spectral distribution of the luminescence is related to the electronic structure of the semiconductor and its defects. The sharp low-temperature spectra of shallow-level defects in pure crystals are well understood, while the assignment of the broad emissions of deep defects with strong lattice coupling is usually difficult.

The random fluctuation of individual carrier motion and carrier generation-recombination creates *noise*. Equilibrium noise is caused by the Brownian motion of carriers and independent of the frequency. Nonequilibrium noise is generated upon optical excitation or current injection and has usually a typical $1/f$ frequency dependence. It is composed of various contributions; a fundamental part originates from energy loss by low-frequency bremsstrahlung in basically elastic scattering processes. Noise creates a lower limit for signal detection.

Keywords

$1/f$ noise · Auger recombination · Band-to-band recombination · Bound-exciton luminescence · Capture cross-section · Electron-lattice coupling · Equilibrium noise · Extrinsic luminescence · Geminate recombination · Generation-recombination noise · Luminescence · Luminescence centers · Near-band-edge emission · Noise · Nonequilibrium noise · Nonradiative recombination · Phosphors · Radiative recombination · Recombination · Recombination centers · Recombination coefficient · Recombination cross-section · Shockley-Read-Hall center · Shot noise · Spontaneous emission · Stimulated emission · Thermal noise · Thermal radiation · Trap

1 Nonradiative Recombination

Carrier recombination is the opposite process to carrier generation, which was discussed in the previous chapter. For reasons of detailed balance, a process of recombination must occur for every process of generation in equilibrium. Although there is a wide variety of such recombinations, depending on the type of energy released during recombination, we can distinguish two principle types:

- Nonradiative recombination
- Radiative recombination

The technically important radiative recombination is discussed in Sect. 3. In the following we consider nonradiative recombination.

Nonradiative recombination is usually an undesired effect, since it converts high electronic energy into heat, i.e., it increases the entropy of the system. It thereby decreases the performance of all but a few devices – such as bolometers, which are designed to measure the incident total energy; this is best accomplished by converting it into heat. Nonradiative recombination can occur by:

- Single-phonon emission when recombining with a shallow level
- A cascade emission of phonons
- Simultaneous multiphonon emission
- Auger generation of an accelerated carrier

Recombination can occur from a band edge at $k = 0$ or at $k \neq 0$ or from an excited state of a defect level. Recombination can proceed to another band, from a higher to a lower excited state or to the ground state of a defect level. Finally, recombination can proceed either directly back, or after some scattering back into the same state (geminate recombination), or after migration to another site.

In the following we first describe recombination mechanisms with an initial energy transfer to phonons and then describe mechanisms with an initial energy transfer to electrons.

1.1 Energy Transfer to Phonons

When a carrier collides with a defect center, it can be captured at this defect center and release its energy in a cascade emission of phonons (a phonon ladder), if the dissipated energy is at least kT. Its probability of reemission would be equal to its probability of further emitting a phonon, but multiplied with the probability of finding a phonon for the reemission. This probability is <1.

The capture cross-section of such a center is

$$s_{\text{nonrad}} = \pi \; r_0^2, \tag{1}$$

where r_0 is the radius of the electron eigenstate at an energy $\geq kT$ below the band edge. In the following section, we will present an estimate of r_0 for a Coulomb-attractive center.

Capture occurs with a high probability when, during its Brownian path within a band, a carrier approaches the center (a) within an energy kT of the band edge *and* (b) within a distance of less than r_0 of the recombination center. The first condition is usually fulfilled (it occurs for carriers in thermal equilibrium), and only the second condition needs to be considered. Therefore, using a gas-kinetic model, carrier capture, like ordinary scattering, takes place after the electron has traveled a distance

$$\lambda_{\text{r}} = v_{\text{rms}} \, \tau_{\text{r}}, \tag{2}$$

Table 1 Capture cross-sections s_n for electrons and s_p for holes of deep-level impurities at room temperature in cm^2 and level depth from the edge of the conduction band E_c or valence band E_v

	Host crystal					
	Si		GaAs		CdS	
Dopant	Depth (eV)	s (cm^2)	Depth (eV)	s (cm^2)	Depth (eV)	s (cm^2)
Ag		$s_n = 10^{-12}$[a]	$E_v + 0.24$	$s_p = 10^{-17}$	$E_c - 0.23$	$s_n = 10^{-13}$
Au	$E_v + 0.35$	$s_n = 10^{-14}$	$E_v + 0.40$	$s_p = 10^{-17}$	$E_c - 0.07$	$s_n = 10^{-19}$
	$E_c - 0.55$	$s_p = 10^{-15}$			$E_c - 0.15$	$s_n = 10^{-17}$
Ni	$E_c - 0.35$	$s_n = 10^{-15}$	$E_v + 0.2$	$s_p = 10^{-16}$		$s_n = 10^{-17}$
	$E_v + 0.23$	$s_n = 10^{-15}$			$E_v + 0.14$	$s_p = 10^{-15}$

[a]at $T = 77$ K

where τ_r is the lifetime of a carrier between generation and capture and is given by

$$\tau_r = \frac{1}{v_{rms}\, s_{nonrad}} (N_{rec} - n_{rec}), \tag{3}$$

with $(N_{rec} - n_{rec})$ as the density of unoccupied recombination centers and the carrier velocity v_{rms}. A few capture cross-sections for electrons or holes of some defect centers are listed in Table 1. When nonequilibrium conditions are involved at higher electric fields or with optical excitation, an energy-dependent capture cross-section must be considered that involves relaxation processes within the band – see Sect. 1.1.2; see also Sect. 2.1 of ▶ Chap. 24, "Carrier Scattering at High Electric Fields".

1.1.1 Recombination at Coulomb-Attractive Centers

The energy spectrum of a Coulomb-attractive quasi-hydrogen center is given by

$$E_{qH} = \frac{m^*/m_0}{\varepsilon_{stat}^2\, n^2} R_\infty, \tag{4}$$

where n is the quantum number and R_∞ is the hydrogen Rydberg energy, see Sect. 1 of ▶ Chap. 18, "Shallow-Level Centers". For capture to take place, we identify a quantum number n_{kT} as the closest integer of n, for which $E_{n=\infty} - E_n \geq kT$. Then

$$n_{kT}^2 = \frac{m^*/m_0}{\varepsilon_{stat}^2\, kT} R_\infty; \tag{5}$$

for the radius of the corresponding eigenstate, we obtain

$$r_{kT} = a_H \frac{\varepsilon_{stat}\, m_0}{m^*} n_{kT}^2, \tag{6}$$

where a_H is the Bohr radius of the hydrogen atom. Usually the ground state is only slightly below kT at room temperature. Thus, $n = 1$ is often used for the quasi-hydrogen recombination radius, yielding the well-known approximation

$$r_{kT} = a_H \frac{\varepsilon_{stat} m_0}{m^*}. \tag{7}$$

From Eqs. 5 and 6, we obtain

$$r_{kT} = a_H \frac{R_\infty}{\varepsilon_{stat} kT} = \frac{e^2}{8\pi \varepsilon_{stat} \varepsilon_0 kT} = 27.8 \times \left(\frac{10}{\varepsilon_{stat}} \frac{300\,\mathrm{K}}{T} \right) \quad (\text{Å}). \tag{8}$$

Except for a factor of 3/2, this is the same result as obtained by setting a random-walk velocity away from the center equal to a drift velocity due to the Coulomb potential (Bube 1974, p. 488). The derivation presented here is not limited to lower-mobility semiconductors, which are implicitly required ($\lambda \ll r_{kT}$) in the velocity criterion.

The resulting recombination cross-section is independent of m^*,

$$s_{nonrad} = \frac{e^4}{64\pi (\varepsilon_{stat} \varepsilon_0 kT)^2} = 2.43 \times 10^{-13} \times \left(\frac{10}{\varepsilon_{stat}} \frac{300\,\mathrm{K}}{T} \right)^2 \quad (\mathrm{cm}^2), \tag{9}$$

and decreases with increasing temperatures $\propto 1/T^2$. It becomes very large at low temperatures, e.g., ~$4 \cdot 10^{-12}$ cm^2 at $T = 70$ K or $6 \cdot 10^{-10}$ cm^2 at 5 K. Measured capture coefficients are given in Fig. 1, exhibiting giant cross sections at low temperatures.

From Eq. 8 we obtain for the Coulomb potential

$$2r_{kT} = \frac{e^2}{4\pi \varepsilon_{stat} \varepsilon_0 kT}; \tag{10}$$

the radius of the Coulomb well for $E_n \cong kT$ is just twice the corresponding quasi-Bohr radius.

Geminate Recombination In semiconductors with a relatively short mean free path, the excitation from a Coulomb-attractive center higher up into the band is followed by a number of scattering processes until the carrier is within kT of the band edge. If the carrier reaches the band edge within the same Coulomb funnel, it will recombine with the same center from which it was generated, see Fig. 2a. Such a recombination is referred to as *geminate recombination.* It is observed in organic crystals and provides a prominent loss mechanism in organic solar cells (see, e.g., Wilken 2020). It also plays an important role in some amorphous semiconductors. A similar process also holds for geminate band-to-band recombination.

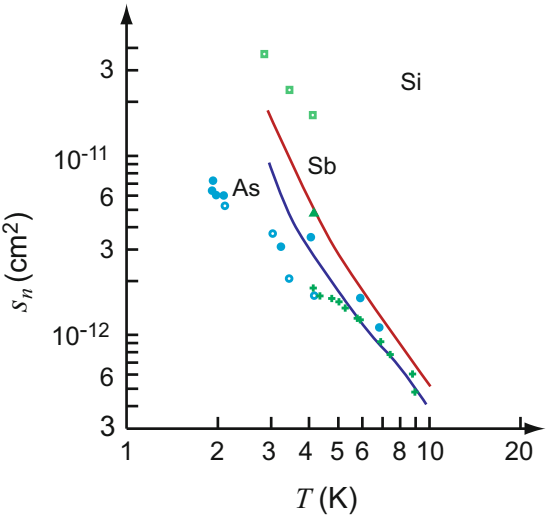


Fig. 1 Giant capture cross-section for electrons by positively charged donors in As- or Sb-doped Si as a function of temperature (After Ascarelli and Rodriguez 1961)

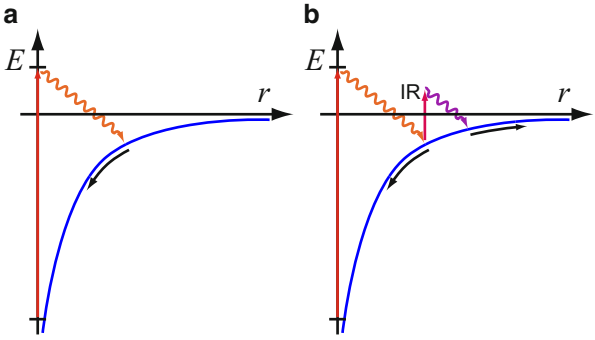


Fig. 2 (**a**) Geminate recombination after excitation from a Coulomb-attractive center. (**b**) Out-diffusion after excitation with an additional IR excitation. The *wavy arrows* pointing down symbolize scattering with phonon emission

With an additional IR excitation, it is possible to increase the probability for out-diffusion and thereby reduce geminate recombination as illustrated in Fig. 2.

Recombination in Amorphous Semiconductors After optical excitation higher into the bands of amorphous semiconductors, inelastic scattering will relax these carriers into the tailing states, which extend from the band edges. Near the mobility edge (Sect. 4.1 of ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects"), the carriers continue to diffuse, thereby losing energy and consequently being trapped at deeper and deeper centers until thermal reemission and tunneling from center to center is no longer possible, as the deeper centers are more widely spaced.

1.1.2 Nonradiative Recombination at Deep Centers

Electron eigenfunctions in deep centers are strongly coupled with lattice oscillations – see Sect. 1.2 of ▶ Chap. 17, "Optical Properties of Defects". The energy of such a

defect level depends on the relative position of the defect atom with respect to its surrounding lattice atoms. With vibrations of these atoms, the defect level moves up and down in the bandgap about its equilibrium position.

One mechanism of nonradiative capture is related to a sufficiently large lattice vibration, which moves the defect level into the conduction band. Here it can accept an electron from the conduction band. After the capture, the lattice is far from the equilibrium position, considering the recharging of the defect. This causes a violent vibration, which quickly relaxes with the emission of many phonons.

If the relaxation is very large, the defect level can move from the upper to the lower half of the bandgap. Here, it can act now as a similar trap for holes, thereby completing the process of band-to-band recombination (Henry and Lang 1977). This mechanism was alluded to in Sect. 1.2 of ▶ Chap. 17, "Optical Properties of Defects", and is pictured schematically in Fig. 3a. The equilibrium position of the defect after electron capture is given by the configuration coordinate Q_2. The extent of average oscillations before and after capture (short arrows), as well as the extent of the giant oscillation directly after electron capture but before relaxation (large arrow), are indicated at the bottom of Fig. 3a. The corresponding configuration-coordinate diagram, which was given in Sect. 1.2.1 of ▶ Chap. 17, "Optical Properties of Defects", is redrawn here for convenience – see Fig. 3b. The diagram identifies the activation energy E_{b1} for the electron trapping-transition and E_{b2} for hole trapping. The thermal binding energy of the electron in the center is $E_c - E_{trap}$.

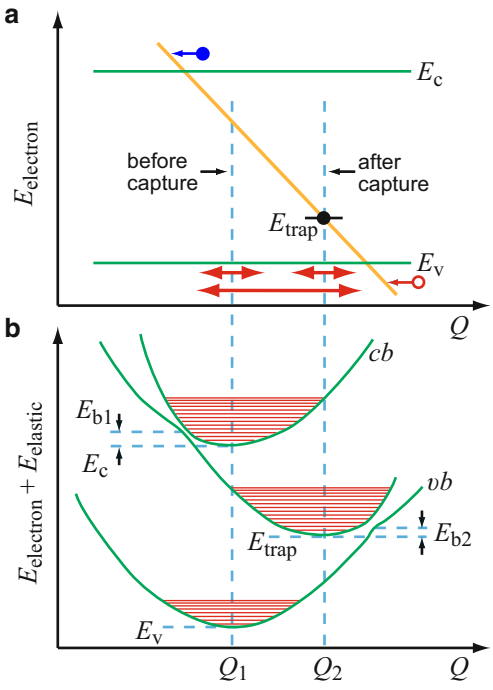


Fig. 3 (**a**) Electron energy versus configuration coordinate of a deep center before and after relaxation. (**b**) Electron plus elastic energy versus coordination coordinate, shown for the deep level between valence and conduction bands

As an important consequence of this model, we obtain a temperature-dependent capture cross-section:

$$s_n = s_\infty \exp\left(\frac{E_c - E_{b1}}{kT}\right). \tag{11}$$

The pre-exponential factor s_∞, obtained from detailed-balance arguments, relates capture and emission, as shown in Eqs. 11 and 12 of ▶ Chap. 29, "Carrier Generation", for an unrelaxed trap. With lattice relaxation, care must be taken to account for the different activation energies; we then obtain (Sumi 1983)

$$e_{\text{trap},c} = \frac{\nu_c}{\nu_{\text{trap}}} c_{c,\text{trap}} \exp\left(-\frac{S\ \hbar\omega_r}{kT}\right), \tag{12}$$

where ν_c is the number of equivalent valleys in the conduction band, ν_{trap} is the degeneracy of the deep trap level, and S is the number of phonons emitted (Huang-Rhys factor) during the relaxation process after electron capture. Here, ω_r is the relevant defect eigenfrequency of a breathing mode (see Fig. 2 of ▶ Chap. 17, "Optical Properties of Defects").

With $e_{\text{trap},c}$ also given by (see Fig. 3b)

$$e_{\text{trap},c} = \frac{\omega_r}{2\pi N_c} \exp\left(-\frac{E_c - E_{\text{trap}}}{kT}\right), \tag{13}$$

we obtain (with $c_{c,\text{trap}} = v_{\text{rms}}\ s_n$ and Eq. 13, setting $E_c - E_{\text{trap}} = S\ \hbar\omega_r$) for the pre-exponential factor of the capture cross-section in Eq. 11:

$$s_\infty = \frac{\nu_{\text{trap}}}{\nu_c} \frac{\omega_r}{2\pi} \frac{1}{N_c v_{\text{rms}}}; \tag{14}$$

N_c is the density of states in the conduction band. A more careful consideration of the approximations used to compute the capture requires the introduction of a factor η into Eq. 14, yielding a modified $s'_\infty = \eta\, s_\infty$. The correction factor η can be approximated as (Sumi 1983)

$$\eta \cong \begin{cases} (3\pi/4)\gamma & \text{for } \gamma \ll 1 \\ 1 - \left(5\pi/\left(9\sqrt{3},\right)\right)\gamma^{-2/3} & \text{for } \gamma \gg 1 \end{cases} \tag{15}$$

with the material parameter

$$\gamma = \frac{4\alpha_c\sqrt{\Delta E_c}}{3\,\hbar\omega_r} \left(E_c - E_{\text{trap}}\right)^{3/2} \tag{16}$$

and ΔE_c as the width of the conduction band (see also Kayanuma and Fukuchi 1984). Peuker et al. (1982) have given a short review of the different approaches to obtain transition probabilities between the states in the adiabatic

approximation. Such probabilities in turn are proportional to the capture cross-section discussed here.

The pre-exponential factor s_∞ in Eqs. 11 and 14 is for deep centers on the order of 10^{-15} cm^2 and depends on the effective mass and degeneracies in agreement with measurements in typical III–V compounds (Henry and Lang 1977). The thermal activation energies vary between a few meV and 0.6 eV for different defects; the larger energies represent a significant fraction of the level depth and require substantial thermal activation for recombination.

Competition Between Radiative and Nonradiative Recombination For transitions *within* a deep center, there is a simple rule as to whether radiative recombination from the upper minima to the lower curves in Fig. 4 or radiationless recombination from the upper curve via crossover to the lower curve is preferred. The Dexter-Klick-Russell rule (Dexter et al. 1956) states that radiative recombination occurs if the optical excitation E_n ends above the crossover energy E_b, i.e., for relatively weak coupling. Otherwise, the electron will cross over to the lower curve and reach the ground state in a nonradiative process via multiphonon emission. Depending on the strength of the coupling, such a crossover for nonradiative recombination may or may not require thermal activation; the nonthermal part is accomplished by tunneling to the lower curve.

This rule can be translated into the ratio of measurable energies, $E_n - E_0$ for optical excitation and $E_0' - E_0$ for thermal excitation:

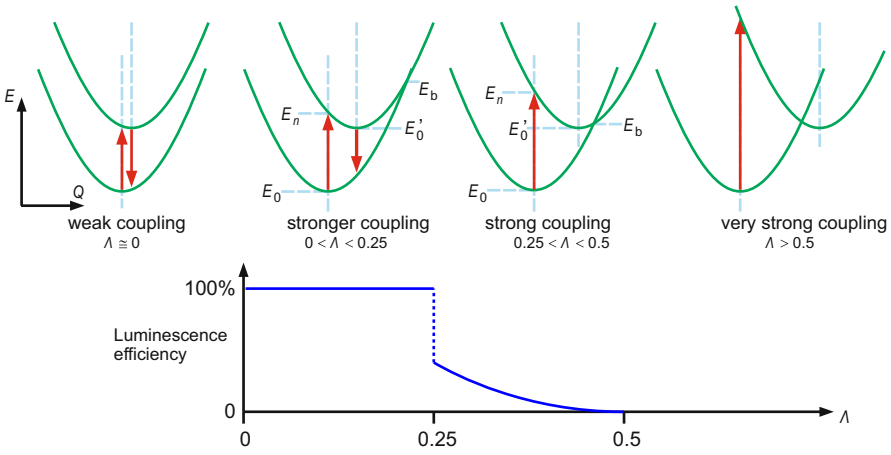


Fig. 4 *Bottom* Luminescence efficiency as a function of the electron-lattice coupling expressed as parameter Λ (Eq. 17). *Top* Four typical configuration-coordinate diagrams for ground and excited states of a deep-level center for various degrees of coupling (After Hayes and Stoneham 1984)

$$\Lambda = \frac{E_n - E_0'}{E_n - E_0} \quad \begin{cases} 0 < \Lambda < 0.25 & \text{luminescence} \\ 0.25 < \Lambda < 0.5 & \text{weak luminescence} \\ 0.5 < \Lambda & \text{no luminescence} \end{cases} \tag{17}$$

with E_n, E_0, and E_0' as defined in Fig. 4 and as discussed by Bartram and Stoneham (1975). These authors have shown that, for F centers in a variety of wide-bandgap materials, this rule is fulfilled reasonably well. The energy E_b can be estimated as

$$E_b = \frac{\left(E_0' - E_n\right)^2}{4(E_n - E_0)}. \tag{18}$$

Nonradiative Multiphonon Recombination Nonradiative recombination into tightly bound centers, or from band to band with *simultaneous emission* of multiple phonons (i.e., typically ~30 phonons for a 1 eV bandgap transition), are comparatively rare transitions. Haug (1972) estimates the transition probability for band-to-band recombination as

$$P_{cv} \cong A \exp\left(-\frac{(h\nu_e - S\ \hbar\omega_0)^2}{C\ kT}\right), \tag{19}$$

where $h\nu_e$ is the electron energy to be dissipated; $\hbar\omega_0$ is the relevant phonon energy, and S, the Huang-Rhys factor, gives the average number of phonons emitted in the recombination process. From Eq. 19 we see that P_{cv} increases exponentially with temperature and decreases with increasing energy dissipation. A comparison with a similar equation for the radiative transition, which shows only minor temperature dependence, indicates that nonradiative transitions will predominate at higher temperatures. There is a large body of investigation dealing with nonradiative transitions, which is based on nonadiabatic approximations, rather than the static approach used by Haug. For a detailed discussion, see Gutsche (1982).

Other types of radiationless transitions are required to explain the observed large rate of nonradiative recombination at elevated temperatures. These include the involvement of deep centers with large lattice relaxation (see begin of Sect. 1.1.2) or the acceleration of free carriers in the proximity of a recombination center, which takes up part of the energy set free during a recombination transition. This process is described in Sect. 1.2.

Calorimetric Absorption Spectroscopy The generation of phonons upon optical excitation raises the lattice temperature; this can directly be measured using calorimetry. A particularly sensitive method is the calorimetric absorption spectroscopy introduced by Bubenzer et al. (1980) with a high sensitivity at lowest temperatures (Gruhl et al. 1985; Juhl and Bimberg 1988). The absorbed power of the integral production of phonons is detected by a temperature-dependent resistor; absolute

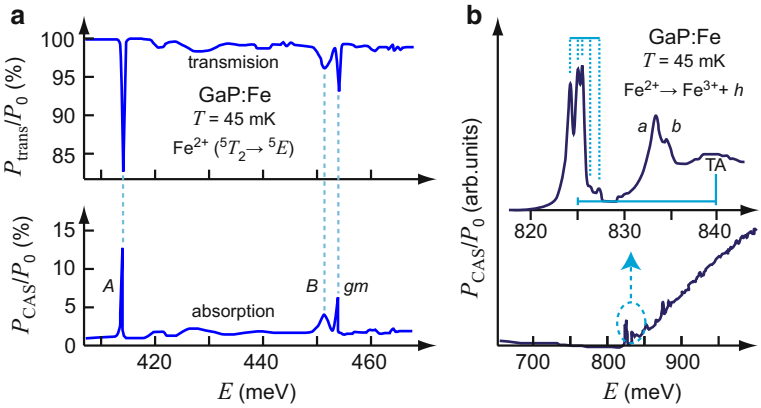


Fig. 5 *Bottom* Spectrally resolved power absorbed by an optically excited deep impurity, measured using calorimetric absorption spectroscopy. (**a**) Intra-ionic ${}^5E(D) \rightarrow {}^5T_2(D)$ transition of Fe^{2+} in GaP with the transmitted (*top*) and absorbed part of the spectrum (*bottom*). (**b**) Charge-transfer transition $Fe^{3+} \rightarrow Fe^{2+} + h$ of the deep iron acceptor in GaP (After Heitz et al. 1995)

values for quantum efficiencies are obtained by calibration via an attached heater. Figure 5 shows as an example the nonradiative part of transitions related to Fe, which is introduced as a deep impurity in III–V semiconductors to produce semi-insulating material. At lower photon energy, intra-ionic crystal-field transitions of iron in the 2+ state (see Sect. 3 of ▶ Chap. 19, "Deep-Level Centers") are optically excited, which relax both radiatively and nonradiatively. In Fig. 5a the transmitted part of the optical excitation is compared to the power absorbed in the sample; lines *A* and *B* are dipole-allowed transitions, line *gm* is assigned to a local-mode phonon replica of line *A*; from this measurement an external quantum efficiency of $(26 \pm 4)\%$ is determined for this transition (Heitz et al. 1995). The broad absorption band observed at higher photon energy originates from the $Fe^{3+} \rightarrow Fe^{2+} + h$ photoionization of the iron acting as an acceptor. The fivefold fine structure at the onset of this band (Fig. 5b) is due to the splitting of the ${}^6A_1(S)$ ground state of the Fe^{3+} ion; the structures *a*, *b* are attributed to an excited state ${}^2P_{3/2}$ of the hole in a (Fe^{2+},h) complex. For the charge-transfer transition, a similar quantum efficiency of 26% was evaluated.

1.2 Energy Transfer to Electrons

1.2.1 Auger Recombination

An electron in the conduction band will lose a large amount of energy while recombining with a defect center or a hole in the valence band if it can transmit this energy to another nearby electron. The second electron is thereby excited high into the band and can easily return to the bottom of the band by sequential LO phonon scattering. The process is called *Auger recombination* and is the inverse

process to the impact ionization. It was originally proposed by Beattie and Landsberg (1959). For a review, see Landsberg (1987) and Haug (1988).

Energy and momentum need to fit the excited state higher in the band. When a phonon is provided to facilitate the momentum match, we speak of *phonon-assisted Auger transitions* (Lochmann and Haug 1980). Several types of Auger recombination are possible, depending on whether the recombination occurs into an ionized defect center or into the valence band. The energy set free during the Auger recombination can be used to accelerate a second electron or a hole as illustrated in Fig. 6. The smaller electron mass is the reason for a preference for the acceleration of a second electron (Landsberg and Willoughby 1978; Landsberg 1987).

A typical electron lifetime as a function of the electron density is given in Fig. 7; it shows three ranges. At low densities, τ is independent of n, then it decreases $\propto 1/n$, and finally $\propto 1/n^2$. We will concentrate first on the final range, which is determined by band-to-band Auger recombination.

The recombination rate and the electron lifetime, which are limited by Auger recombination, are given by:

$$C_{cv}^{(A)} = B\ n^2 p \text{ and } \tau_A = \frac{1}{B\ n^2}; \tag{20}$$

B is typically on the order of $10^{-30} \ldots 10^{-22}$ cm^6 s^{-1}.

A quantum-mechanical derivation of the Auger-recombination rate for band-to-band transitions from thermally produced intrinsic carriers is given by

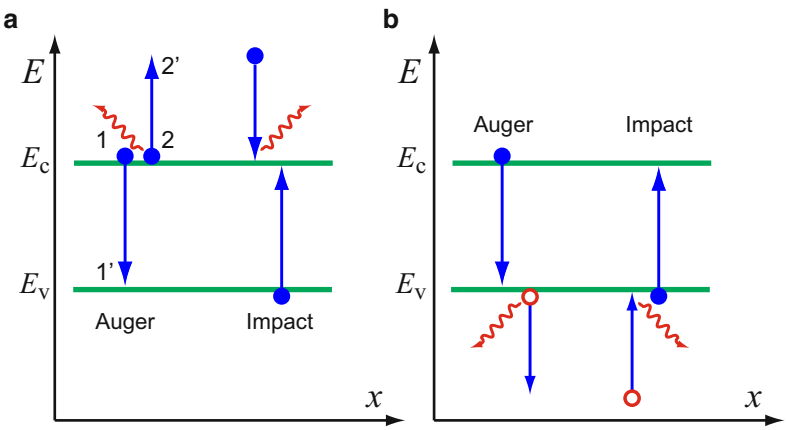


Fig. 6 Comparison between Auger recombination and impact ionization involving (**a**) a second electron or (**b**) a second hole (*open circles*), with an additional phonon indicated for momentum matching

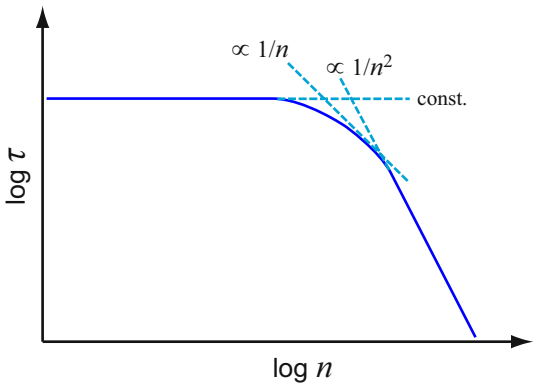


Fig. 7 Schematic of the dependence of carrier lifetime on the carrier density

$$C_{\mathrm{cv}}^{(\mathrm{A})} = 2\frac{2\pi}{\hbar}\frac{V^3}{(2\pi)^9}\int\int\int\int |M|^2 f(E_1) f(E_2)\ \left|1 - f\left(E_1'\right)\right|\ \left|1 - f\left(E_2'\right)\right| \times\ \delta\left(E_1 + E_2 + E_1' - E_2'\right) d^3k_1\, d^3k_2\, d^3k_1'\, d^3k_2' \tag{21}$$

with the Auger matrix

$$M = \int\int \phi_{k_1}^*(\mathbf{r}_1)\, \phi_{k_2}^*(\mathbf{r}_2)\, V(|\mathbf{r}_1 - \mathbf{r}_2|)\, \phi_{k_1'}(\mathbf{r}_1)\, \phi_{k_2'}(\mathbf{r}_2)\ d^3r_1\, d^3r_2, \tag{22}$$

where ϕ is the Bloch function, and $V(\mathbf{r})$ is the screened Coulomb potential

$$V(\mathbf{r}) = \int \frac{d^3q}{(2\pi)^3}\frac{4\pi\ e^2}{\varepsilon(q)\ \left(q^2 + \lambda^2\right)}\exp(i\mathbf{q}\cdot\mathbf{r}). \tag{23}$$

Here, λ is the electron screening factor, $\mathbf{q} = |\mathbf{k}' - \mathbf{k}|$ is the momentum transfer, and $\varepsilon(q)$ is the dielectric constant of the material. Assuming parabolical isotropic bands and $m_n \ll m_p$, Haug (1972) obtains for the electron lifetime

$$\tau_{\mathrm{A}} \cong \frac{\pi\hbar\left(4\pi\varepsilon_{\mathrm{opt}}\varepsilon_0\hbar\right)^2}{24\, e^4\ m_n}\frac{\Delta E}{kT}\frac{\sqrt{2}}{0.01}\exp\left(\frac{\Delta E}{kT}\right). \tag{24}$$

Haug's formula also contains two overlap integrals, the values of which are estimated (Beattie and Landsberg 1959) as $I_1 \cong 1$ and $I_2 \cong 0.1$, resulting in $(I_1\ I_2)^2 \cong 0.01$, and as such are included in the denominator of Eq. 24; for comments, see Haug (1988) and Laks et al. (1988). $\Delta E = [(2m_n + m_p)/(m_n + m_p)]E_{\mathrm{g}}$ is the energy dissipated in the Auger process. With an increasing bandgap energy, the

Auger-determined electron lifetime increases according to Eq. 24 rapidly[1] and reaches values not attained in a semiconductor with a bandgap $E_g > 0.35$ eV; here, one estimates $\tau_A \cong 10^{-6}$ s.

1.2.2 Auger Recombination at High Carrier Densities

Intrinsic Auger recombination at room temperature for thermally excited carriers is important only for narrow-gap semiconductors. With high doping densities, however, sufficient carrier densities can be created to obtain Auger recombination rates also in wider-gap semiconductors. Band-to-band Auger recombination is often the dominating carrier-loss mechanism in InP-based long-wavelength laser diodes (Agrawal and Dutta 1993) due to the high density of electrons and holes ($>10^{18}$ cm^{-3}) in the active region. Auger recombination is also identified as a main cause of the decrease in the conversion efficiency of GaN-based LED lamps at high current densities, an effect referred to as efficiency droop (Iveland et al. 2013).

For semiconductors with larger bandgap energy, the approximations used are too coarse. Computations using the empirical pseudopotential method, with self-consistent calculation plus Thomas-Fermi screening and $\varepsilon = \varepsilon(q)$, resulted in much improved results, as shown in Fig. 8 for *n*-type Si (Laks et al. 1988); more recent

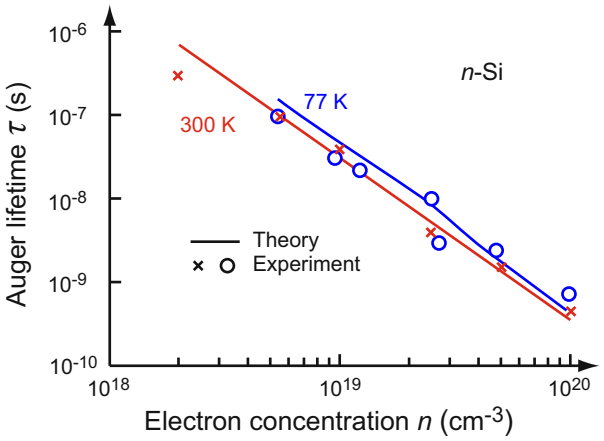


Fig. 8 Auger lifetime in *n*-type Si as function of the electron density at 77 and 300 K (After Dziewior and Schmid 1977) and computation for the corresponding temperatures (After Laks et al. 1988)

[1]It is interesting to see that Eq. 24 can be rewritten, using the quasi-hydrogen energy E_{qH}, as

$$\tau_A \cong \frac{h}{E_{qH}} \frac{E_g}{kT} \exp\left(\frac{E_g}{kT}\right). \tag{25}$$

The first part of Eq. 25 represents the Heisenberg uncertainty relation, indicating that τ_A cannot be smaller than the Heisenberg uncertainty time for an exciton, $\tau_A \geq 1.6 \cdot 10^{-13}$ s. This presents a lower limit for $E_g \cong kT$ for the approximation used.

DFT (G_0W_0) calculations for n- and p-type Si and GaAs with largely comparable results for n-Si are reported by Govoni et al. (2011). Similar results are obtained for Auger recombination with ionized defects, where again only shallow defects influence the observed lifetime.

With additional excitation, e.g., high-intensity optical carrier generation, sufficient carriers are available to render Auger recombination important. The Auger lifetime for recombination via recombination centers is given by (Haug 1981)

$$\tau_{\mathrm{A}} \cong \frac{1}{B N_{\mathrm{rec}}\, n} = \frac{v_{\mathrm{rms}}}{N_{\mathrm{rec}}\, n}\, \frac{m_n^2}{8\pi^2 e^4 \hbar^3}\, \frac{Q}{R} \sqrt{\frac{m_p}{m_n}} (E_{\mathrm{c}} - E_{\mathrm{rec}})^3, \tag{26}$$

with $Q = 0.5(m_0/m_n)^{3/2}(1 + m_n/m_0)^2$ and $R \cong 2.6$ being an enhancement factor; E_{rec} is the energy of the recombination center. This yields for typical values of $E_{\mathrm{c}} - E_{\mathrm{rec}} \cong 0.5$ eV and $m_n \cong 0.1\ m_0$ an Auger coefficient $B \cong 10^{-26}$ cm^6 s^{-1}. Somewhat lower values have also been suggested by Robbins and Landsberg (1980). For a density of recombination centers $N_{\mathrm{rec}} \cong 10^{17}$ cm^{-3} and $n \cong 10^{14}$ cm^{-3}, we obtain an Auger lifetime of 10^{-5} s. Computed Auger coefficients for halide perovskites are given in Shen et al. (2018). In nitride semiconductors, Auger processes are strongly assisted by an additional scattering due to phonons and alloying; respective *indirect* Auger processes are modeled by Kioupakes et al. (2015), and the related coefficients for GaN and InGaN are reported.

The Auger coefficient at recombination centers can be estimated from rough formulae with simple power-law dependences in $E_{\mathrm{c}} - E_{\mathrm{rec}}$ (Landsberg and Pimpale 1976). In particular, for GaAs it can be estimated as (Haug 1980)

$$B = \frac{2.5 \times 10^{-25}}{Q\ (E_{\mathrm{c}} - E_{\mathrm{rec}})^3}. \tag{27}$$

The capture cross-section for Auger recombination is given as

$$s_n = \frac{B\, n}{v_{\mathrm{rms}}}, \tag{28}$$

and is on the order of 10^{-11} cm^2 for shallow traps and 10^{-18} cm^2 for deep traps, when $n \cong 10^{14}$ cm^{-3}.

Evidence of intrinsic Auger recombination can be obtained from the dependence of the lifetime of minority carriers on the square of the density of majority carriers (Haynes and Hornbeck 1955; Dziewior and Schmid 1977).

Plasmon-Induced Recombination At very high excitation rates, the carrier density becomes high enough so that the plasmon energy $\hbar\omega_{\mathrm{p}}$ equals the bandgap energy; the recombination is much enhanced by such resonance transitions. Typical critical densities for a bandgap of 1 eV are on the order of 10^{20}–10^{21} cm^{-3} (see Eq. 4 of

► Chap. 12, "Photon–Free-Electron Interaction"). This *plasmon-induced recombination* can exceed the Auger recombination, which is also effective at high carrier densities (Malvezzi 1987).

2 Statistics of Recombination

All recombination transitions are spontaneous and follow the rules of statistics, except for *stimulated* transitions discussed in Sect. 3.3.6. The statistical description of the recombination does not consider how a carrier loses its energy, but with what state the carrier recombines and what changes in carrier densities occur with changing rates of generation and temperature. The probability of recombination is described by a capture rate, which is linked to the capture cross-section by Eq. 15 of ► Chap. 29, "Carrier Generation".

2.1 Trapping or Recombination

The return of a carrier to an ionized state is referred to as *recombination.* When a carrier is captured by a shallow level from which it can be thermally reemitted into the band before it finally recombines, the process is called *trapping.* A more precise way to distinguish between trapping and recombination is given in the following section.

2.1.1 Electron and Hole Traps

There are several transitions possible between the electronic state of any center and other states. All such transitions are described by their corresponding rates. These rates are additive and describe the change in the population of the center. For instance, the change of the electron density in an electron trap is determined by excitation from the trap into the conduction band, by electron capture from the band, by recombination with holes from the valence band, and by electron transfer to other localized states of nearby defects to which such transitions are sufficiently probable.

For reasons of detailed balance, there are always *pairs* of transitions between two states (Fig. 9), which must be equal to each other in thermal equilibrium. The magnitude of transition rates *varies from pair to pair over a wide range*. For example, thermal excitation of an electron from the more distant valence band into an electron trap is much less probable than thermal excitation of a trapped electron into the closer conduction band.

In *steady state* (see Sect. 2.2), the total rate of transitions from the center must equal the total rate into the center. Usually, we can *neglect all transitions compared to the one pair with the highest transition probability.* These pairs can now involve different states (see below). Thereby, we can identify different classes of centers according to the predominant types of transitions. Centers close to the conduction band are identified as *electron traps* and centers close to the

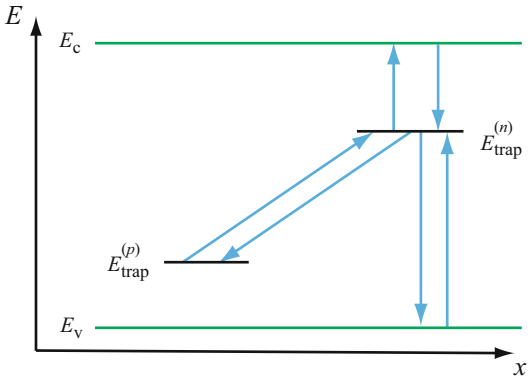


Fig. 9 Various possible transitions to and from a localized state of an electron trap with energy $E_{\text{trap}}^{(n)}$

valence band as *hole traps*, when these centers communicate predominantly with the adjacent bands.

2.1.2 Recombination Centers

Centers close to the middle of the bandgap readily communicate with both bands, since it is easier for a captured electron to recombine with a hole in the valence band than to be reemitted into the conduction band. These centers are called *recombination centers.* Relaxation of the center after trapping a carrier is not included here. Such relaxation can assist significantly in bridging the bandgap during recombination and is discussed in Sect. 1.1.2 and Sect. 2.3.2 of ▶ Chap. 17, "Optical Properties of Defects" – see also Henry and Lang (1977).

A *demarcation line* between electron traps and recombination centers is defined (Rose 1951) when the transition rates of electrons from these centers to the conduction band and to the valence band become equal:

$$n_{\text{trap}}\, e_{\text{trap,c}}\, N_c = n_{\text{trap}}\, c_{\text{trap,v}}\, p; \tag{29}$$

here we use the symbols introduced in Sect. 1.1 of ▶ Chap. 29, "Carrier Generation". Applying the expression for the excitation from the electron trap to the conduction band $e_{\text{trap,c}}$ and for the hole density p (see Sect. 2.4 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"),

$$e_{\text{trap,c}} = v_{\text{rms}}\, s_n \exp\left(-\frac{E_c - E_{\text{trap}}}{kT}\right) \text{ and } p = N_v \exp\left(\frac{E_v - E_{F_p}}{kT}\right), \tag{30}$$

we obtain an equation for this electron demarcation line E_{D_n} from Eq. 29, defining a specific E_{trap} (Eq. 30) $= E_{D_n}$:

$$E_c - E_{D_n} = E_{F_p} - E_v + \delta_i, \text{where } \delta_i = \frac{3}{2}\, kT \ln\left(\frac{m_n\, s_{ni}}{m_p\, s_{pi}}\right). \tag{31}$$

Although possibly confusing at first glance, the reference to the *hole* quasi-Fermi level E_{F_p} for determining the *electron* demarcation line is understandable, since the recombination path, competing with thermal ionization, depends on the availability of free holes. Figure 10 may help to clarify this dependence: the distance of the demarcation line for electrons from the conduction band, identified as Δ_p, is the same as the distance of the quasi-Fermi level for holes from the valence band plus a corrective δ_i or δ_j – see below. These correction terms are logarithmically related to the ratio of the capture cross-sections for electrons and for holes at this center. The electron demarcation line E_{D_n} defines the energy border between electron traps above and recombination centers below this energy.

A similar relationship holds for the *hole demarcation line* E_{D_p}:

$$E_{D_p} - E_v = E_c - E_{F_n} + \delta_j, \text{ where } \delta_j = kT \ln\left(\frac{m_p s_{pi}}{m_n s_{ni}}\right); \tag{32}$$

this line is also shown in Fig. 10. For *n*-type material with narrow $E_c - E_{F_n}$, there is a wide range of electron traps and a narrow range of hole traps and vice versa.

The correction terms δ_i and δ_j depend on the ratio of capture cross-sections for electrons and holes that change with the occupancy of the center. As an illustration, let us assume a simple example of a center that is neutral without an electron in it, having a cross section for an electron on the order of 10^{-16} cm^2. After it has captured the electron, it is negatively charged; its capture cross-section for a hole has thus increased, say to ~10^{-14} cm^2. For this example $s_{n_i}/s_{p_i} \cong 10^{-2}$ and $\delta_i \cong -0.12$ eV. For a similar type of hole trap, the charge character changes from neutral to positive after hole capture, making $s_{n_j}/s_{p_j} \cong 100$ and $\delta_j \cong +0.12$ eV. The shifts δ_i and δ_j in Fig. 10 have been chosen accordingly.

Since the capture cross-section varies from center to center, typically from ~10^{-12} to ~10^{-22} cm^2, δ_i varies for these different centers by as much as ~0.6 eV at room

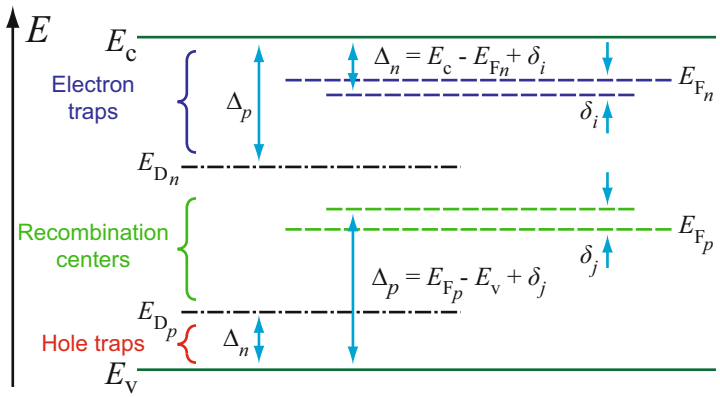


Fig. 10 Band model with quasi-Fermi potentials E_{F_n} and E_{F_p} and demarcation lines E_{D_n} and E_{D_p} separating electron traps, recombination centers, and hole traps

temperature. Hence, the demarcation lines of these centers are spread over a wide range within the bandgap. Therefore, it is not customary to plot demarcation lines of all the possible centers, while it is still instructive to discuss those that provide the most important transitions in the given model.

Neglect of the other centers is often justified, because all transitions enter additively, spanning many orders of magnitude; therefore, usually only one kind of transition is important to each type of trap or recombination center for a given situation.

2.2 Thermal Equilibrium, Steady State, and Shockley-Read-Hall Recombination

Thermodynamic (*thermal*) *equilibrium* is reached when a semiconductor without any external excitation is kept long enough at a constant temperature to reach such an equilibrium. Deviations from thermal equilibrium can occur because of nonthermal, additional excitation by light or an electric field. When such deviations occur but have become stationary, a nonequilibrium *steady state* is reached.

2.2.1 Thermal Equilibrium

In a semiconductor at constant temperature without optical or electrical excitation, thermal equilibrium becomes established. Electrons and holes are then generated by thermal excitation alone. The same number of carriers which are generated in any volume element must recombine in the same volume element except for statistical fluctuations considered in Sect. 4.2. There is no net transport of carriers. This also holds for space-charge regions, e.g., in a *pn* junction, in which n and p are rapidly changing functions of the spatial coordinate, whereas j_n and j_p vanish independently in every volume element.

For thermal equilibrium the carrier distribution is uniquely described by the Fermi level E_F. Consequently, when formally using quasi-Fermi levels, they must collapse to $E_{\mathrm{F}_n} = E_{\mathrm{F}_p} = E_\mathrm{F}$. We see from Eqs. 31 and 32 that the resulting demarcation lines then coincide: $E_{\mathrm{D}_n} = E_{\mathrm{D}_p}$, i.e., electron and hole traps join borders with no recombination-center range in between. Here, thermal ionization (generation) and recombination on the average attain a balance within each volume element of the semiconductor.

Steady State When nonthermal carrier generation is introduced, the equilibrium density-distribution is altered. As a consequence, the Fermi energy splits into two quasi-Fermi levels, and sets of two demarcation lines for each type of defect center appear. Thus, some levels, which previously acted as traps, will now act as recombination centers. Much of the content of the following sections deals with this steady-state condition.

2.2.2 The Shockley-Read-Hall Center

Changing the external excitation will alter the demarcation lines between traps and recombination centers. Therefore, we need to include all four transitions between the level and the two bands for such centers when variations of carrier distributions are considered (Shockley and Read 1952).

For a center in the bandgap that interacts only with the bands, the four transition rates are $e_{\mathrm{trap,c}}\, n_{\mathrm{trap}}\, p_{\mathrm{c}}$, $c_{\mathrm{c,trap}}\, n\, (N_{\mathrm{trap}} - n_{\mathrm{trap}})$, $e_{\mathrm{v,trap}}\, n_{\mathrm{v}}\, (N_{\mathrm{trap}} - n_{\mathrm{trap}})$, and $c_{\mathrm{trap,v}}\, n_{\mathrm{trap}}\, p$ – see Fig. 11. In *equilibrium*, the sum of each pair of transitions to each band must vanish. In *steady state*, this is no longer necessary; there may be a net flow of carriers from one band through such a center to the other band, which is balanced with another transition, e.g., an optical band-to-band generation of carriers. The sum of all four transition rates, however, must vanish to maintain a time-independent electron population in the center:

$$c_{\mathrm{c,trap}}\, n\, \left(N_{\mathrm{trap}} - n_{\mathrm{trap}}\right) - e_{\mathrm{trap,c}}\, n_{\mathrm{trap}}\, p_{\mathrm{c}} = c_{\mathrm{trap,v}}\, n_{\mathrm{trap}}\, p - e_{\mathrm{v,trap}}\, n_{\mathrm{v}}\, \left(N_{\mathrm{trap}} - n_{\mathrm{trap}}\right). \tag{33}$$

Equation 33 can be used to determine this population. After using Eq. 13 of ▶ Chap. 29, "Carrier Generation", and an analogous condition for holes to convert $e_{\mathrm{trap,c}}$ and $e_{\mathrm{v,trap}}$ into the respective capture coefficients, and with the parameters $n_1 = N_{\mathrm{c}} \exp\left((E_{\mathrm{trap}} - E_{\mathrm{c}})/(kT)\right)$ and $p_1 = N_{\mathrm{v}} \exp\left((E_{\mathrm{v}} - E_{\mathrm{trap}})/(kT)\right)$, we obtain

$$n_{\mathrm{trap}} = \frac{N_{\mathrm{trap}}\left(c_{\mathrm{c,trap}}\, n + c_{\mathrm{trap,v}}\, p_1\right)}{c_{\mathrm{c,trap}}\, (n + n_1) + c_{\mathrm{trap,v}}\, (p + p_1)}. \tag{34}$$

We now introduce this steady-state density of the trapped electrons into the net rate equation, permitting a net flow U of electrons through such a center:

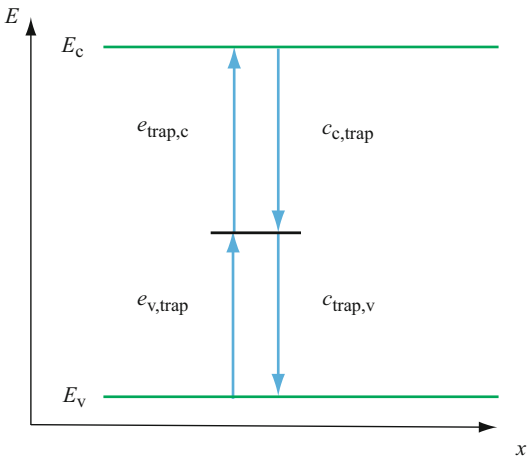


Fig. 11 Shockley-Read-Hall center with transitions to both, conduction band and valence band

$$\begin{aligned} U &= c_{\mathrm{c,trap}}\, n \ \left(N_{\mathrm{trap}} - n_{\mathrm{trap}}\right) - e_{\mathrm{trap,c}}\, n_{\mathrm{trap}} N_{\mathrm{c}} \\ &= c_{\mathrm{trap,v}}\, n_{\mathrm{trap}}\, p - e_{\mathrm{v,trap}}\, n_{\mathrm{v}} \ \left(N_{\mathrm{trap}} - n_{\mathrm{trap}}\right). \end{aligned} \tag{35}$$

By eliminating n_{trap} from the net carrier flow rate, we obtain

$$U = \frac{c_{\mathrm{c,trap}}\, c_{\mathrm{trap,v}} N_{\mathrm{trap}} (np - n_1 p_1)}{c_{\mathrm{c,trap}} \ (n + n_1) + c_{\mathrm{trap,v}} \ (p + p_1)}. \tag{36}$$

Using the condition

$$n_1 p_1 = n_{\mathrm{i}}^2 \tag{37}$$

and introducing the *intrinsic level* E_{i}

$$E_{\mathrm{i}} = \frac{E_{\mathrm{c}} - E_{\mathrm{v}}}{2} + \frac{kT}{2} \ln\left(\frac{N_{\mathrm{v}}}{N_{\mathrm{c}}}\right), \tag{38}$$

we obtain the expression for the net carrier flow through a Shockley-Read-Hall center:

$$U = \frac{c_{\mathrm{c,trap}}\, c_{\mathrm{trap,v}} N_{\mathrm{trap}} \left(np - n_{\mathrm{i}}^2\right)}{c_{\mathrm{c,trap}} \ \left[n + n_{\mathrm{i}} \exp\left(\frac{E_{\mathrm{trap}} - E_{\mathrm{i}}}{kT}\right)\right] + c_{\mathrm{trap,v}} \ \left[p + n_{\mathrm{i}} \exp\left(\frac{E_{\mathrm{i}} - E_{\mathrm{trap}}}{kT}\right)\right]} \tag{39}$$

From Eq. 39 we see immediately that U vanishes for thermal equilibrium, i.e., for $np = n_{\mathrm{i}}^2$. In steady state, however, with optical generation and the Shockley-Read-Hall center acting as a dominant recombination center, U must equal the generation rate g_0 in a homogeneous semiconductor.

Equation 39 becomes very valuable when deviations from the thermal equilibrium are analyzed. A separation into thermal generation rates

$$g_n = g_p = \frac{c_{\mathrm{c,trap}}\, c_{\mathrm{trap,v}} N_{\mathrm{trap}}\, n_{\mathrm{i}}^2}{c_{\mathrm{c,trap}} \ \left(n + n_{\mathrm{i}}^{+}\right) + c_{\mathrm{trap,v}} \ \left(p + n_{\mathrm{i}}^{-}\right)} \tag{40}$$

and recombination rates

$$r_n = r_p = \frac{c_{\mathrm{c,trap}}\, c_{\mathrm{trap,v}} N_{\mathrm{trap}}\, np}{c_{\mathrm{c,trap}} \ \left(n + n_{\mathrm{i}}^{+}\right) + c_{\mathrm{trap,v}} \ \left(p + n_{\mathrm{i}}^{-}\right)}, \tag{41}$$

is helpful. For brevity, the expression

$$n_{\mathrm{i}}^{\pm} = n_{\mathrm{i}} \exp\left(\pm \frac{E_{\mathrm{trap}} - E_{\mathrm{i}}}{kT}\right) \tag{42}$$

is used above.

A deviation from thermal equilibrium may be caused by optical excitation. With band-to-band excitation, the optical generation term g_{o} is simply added to Eq. 40, yielding

$$g_{n,\mathrm{o}} = g_{p,\mathrm{o}} = \frac{c_{\mathrm{c,trap}}\, c_{\mathrm{trap,v}}\, N_{\mathrm{trap}}\, n_{\mathrm{i}}^2}{c_{\mathrm{c,trap}}\ \left(n + n_{\mathrm{i}}^{+}\right) + c_{\mathrm{trap,v}}\ \left(p + n_{\mathrm{i}}^{-}\right)} + g_{\mathrm{o}}; \tag{43}$$

the necessarily increased recombination in steady state is automatically included in Eq. 41 through the increase in n and p.

In homogeneous semiconductors and for steady-state conditions, the generation rate will always be equal to the recombination rate; therefore, U will vanish. Deviations from $U = 0$ will occur during kinetics (see ▶ Chap. 32, "Dynamic Processes") and when spatial inhomogeneities are considered. In space-charge regions, such as in Schottky barriers or *pn* junctions, this formalism becomes most valuable – e.g., for the evaluation of nonradiative recombination; this is a major issue in the performance of solar cells and light-emitting diodes. For a discussion of recombination by Shockley-Read-Hall centers in terms of a configuration-coordinate model, see Das et al. (2020). In wide-bandgap semiconductors, intradefect excited states may play an important role in Shockley-Read-Hall recombination (Alkauskas et al. 2016). A general theory of inelastic scattering by defects with a distinction between capture under thermal equilibrium and under nonequilibrium conditions is reported by Barmparis et al. (2015).

3 Radiative Recombination

In the previous sections, carrier recombination was discussed in general terms. Recombination requires the dissipation of energy, which can proceed via the creation of phonons and can involve Auger collision of electrons, which results in *nonradiative recombination* or by the emission of photons as *radiative recombination.*

We distinguish between several types of radiative processes, depending on their relationship to the excitation process and on intermittent processes. One process occurring in *optical* excitation is the reflection of radiation impinging on the surface of the solid, i.e., the coherent reradiation. Reflection of electromagnetic radiation involves the polarization of the semiconductor lattice, which is characterized by its index of refraction and extinction coefficient. The polarization appears in phase with the electric vector of the incoming radiation and produces coherent reradiation, essentially without delay. Such reradiation occurs from the undamped fraction of the excitation processes. It is observed from any medium with a refraction index larger than 1, as seen from the basic reflectance equation, presented here for normal

incidence and interface to vacuum. For details see Sects. 1 of ▶ Chap. 10, "Interaction of Light with Solids" and 2 of ▶ Chap. 11, "Photon–Phonon Interaction".

Besides reflection the following radiative recombination processes occur and are discussed in this chapter:

- Thermal radiation (blackbody radiation)
- Luminescence
- Stimulated emission (lasing)
- Phosphorescence (delayed luminescence)

3.1 Thermal Radiation

The intrinsic band-to-band radiative recombination transition can be observed as thermal blackbody radiation. Electrons, thermally excited into the conduction band, can recombine with holes in the valence band with the emission of light. If the temperature is high enough, the emission becomes visible – the material will *glow*.

In the wavelength range of the intrinsic absorption, the semiconductor can be described as a blackbody. Here, the (thermally excited) emitted photon flux in the frequency range $\Delta\nu$ is given by Planck's formula

$$\Phi_{\mathrm{bb}} = \frac{8\pi h\nu^3}{\left(\frac{c}{n_{\mathrm{r}}}\right)^2} \frac{\Delta\nu}{\exp\left(\frac{h\nu}{kT}\right) - 1} \tag{44}$$

with the index of refraction n_{r} introduced in Eq. 11 of ▶ Chap. 10, "Interaction of Light with Solids". This emission, originating within a thickness of $d \cong 1/\alpha_{\mathrm{o}}$ of the semiconductor, requires a volume-generation rate of electron-hole pairs

$$g_{\mathrm{o}} = \int_{E_{\mathrm{g}}}^{\infty} \frac{\Phi_{\mathrm{bb}}(\nu)}{d}\, d\nu, \tag{45}$$

which must be equal to the radiative recombination rate of these electrons and holes

$$r_{\mathrm{c,v}}^{(\mathrm{rad})} = c_{\mathrm{c,v}}^{(\mathrm{rad})}\; n\; p. \tag{46}$$

In equilibrium, we have

$$n\; p = n_{\mathrm{i}}^2 = N_{\mathrm{c}} N_{\mathrm{v}} \exp\left(-E_{\mathrm{g}}/(kT)\right); \tag{47}$$

thus, we obtain from $g_{\mathrm{o}} = r_{\mathrm{c,v}}$ for the radiative recombination coefficient

$$c_{c,v}^{(rad)} \cong \frac{\pi \hbar^3 E_g^2 \alpha_o}{c^2 \left(m_n m_p\right)^{2/3} (kT)^2}. \tag{48}$$

Equation 48 yields for the cross section of radiative band-to-band recombination $s_{np} = c_{c,v}^{(rad)}/v_{rms}$. For typical values of $E_g = 1$ eV, $T = 1000$ K, and assuming $m_n = m_p = 0.1\, m_0$ and $\alpha_o = 10^5$ cm^{-1}, we obtain $s_{np} \cong 10^{-19}$ cm^2, which is a much smaller value than we would expect from a semiclassical model, i.e., an electron colliding with an ionized lattice atom with $s \cong 10^{-16}$ cm^2. It should, however, be recognized that Eq. 48 is based on a rather crude estimate and provides only an order-of-magnitude guidance for any specific transition.

Blackbody Radiation of Semiconductors Compared to the Planck's emission distribution of a blackbody, the emission spectrum of the semiconductor is modified, since in the bandgap range there is essentially no absorption and thus little emission (*Kirchhoff*'s *law*). Therefore, in a semiconductor, the long-wave emission is suppressed as illustrated in Fig. 12. If the band edge lies in the visible range, the thermal glow of the semiconductor is visibly different from the usual red or orange glow of a blackbody. For instance, ZnS with a bandgap of ~2.5 eV at 1000 K shows a green thermal glow: it is transparent for red and yellow light and thus shows no blackbody radiation there. This glow-light emission is closely related to the spontaneous band-to-band luminescence described in Sect. 3.2.1.

Thermal Radiation from Free Electrons In addition to the band-to-band emission, a bandgap emission can be observed when sufficient free electrons are created by doping, which in turn can recombine with the ionized donors and emit long-wavelength radiation. An increase of thermal radiation has also been observed after the injection of excess electrons (Ulmer and Frankl 1968).

Thermal Radiation from Lattice Vibration The thermal emission spectrum shows strong emission maxima in the *Reststrahl* range (Stierwalt and Potter 1967). Such an emission is indicative of the direct coupling of the different types of phonons to the

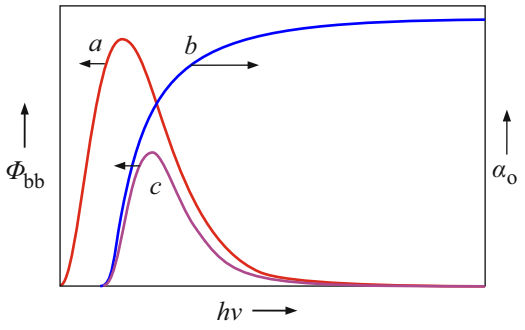


Fig. 12 Emission of a blackbody (curve *a*) and absorption of a semiconductor (*b*). Curve *c* is the emission spectrum of a semiconductor obtained by multiplication of curves *a* and *b*

electromagnetic radiation and can be used to obtain the corresponding spectrum of the optically active phonons.

3.2 Intrinsic Luminescence

Luminescence is defined as any radiative recombination from an excited electronic state, with the exception of reflection and blackbody radiation, and lasing.[2] There are various possibilities for the excitation of luminescence:

- Optical excitation (photoluminescence)
- Excitation by carrier injection (electroluminescence)
- Electron-beam excitation (cathodoluminescence)
- Excitation by high-energy nuclear radiation (α-, β -, γ -, or X-rays)
- High-field excitation (e.g., by impact ionization)

In photoluminescence, it is instructive to measure the intensity of a specific luminescence line as a function of the wavelength of the *exciting light*, called *photoluminescence excitation spectroscopy*. In this mode of excitation, the connection between the excitation process and the luminescent center becomes evident. We observe selective enhancement of the luminescence when resonance transitions take place. This is important for the analysis of bound excitons and the processes involved in the generation of such excitations (see Dean 1984).

In all of the excitation processes listed above, excited states or free carriers are produced, a certain fraction of which in turn recombine with emission of a photon, yielding the luminescence efficiency

$$\eta_{\mathrm{L}} = \frac{r_{\mathrm{rad}}}{r_{\mathrm{rad}} + r_{\mathrm{nonrad}}}, \tag{49}$$

where r_{rad} is the radiative, and r_{nonrad} is the nonradiative transition rate. Radiative transitions can occur along a variety of paths, which are indicated in Fig. 13a for intrinsic and Fig. 13b for extrinsic luminescence.

The *intrinsic transitions* are:

(a) Direct band to band
(b) Indirect band to band
(c) Free exciton
(d) Free-exciton molecule
(e) Electron-hole liquid
(f) Phonon-assisted band edge

[2]Other radiation due to nonlinear optical processes, such as photon (Raman or Brillouin) scattering and higher harmonic generation, are also excluded in this discussion. These are discussed in Sect. 3.3 of ▶ Chap. 11, "Photon–Phonon Interaction".

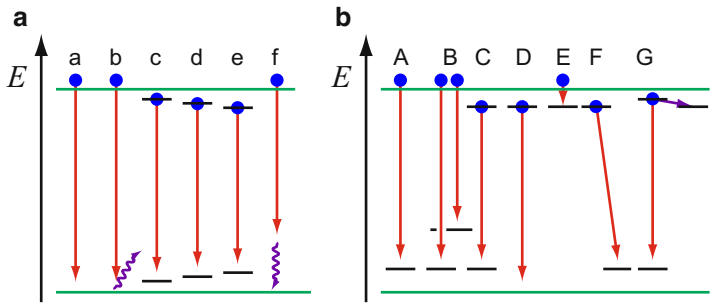


Fig. 13 Typical transitions for (**a**) intrinsic and (**b**) extrinsic luminescence (see text)

The *extrinsic transitions* (with or without assistance of phonons) are:

(A) Band to impurity
(B) Band-to-multilevel impurity
(C) Excited state to ground state of a localized defect
(D) Donor-to-valence band, similar to conduction-band-to-acceptor
(E) Conduction-band-to-donor, similar to acceptor-to-valence band
(F) Donor-acceptor pair
(G) Bound exciton

During the luminescent transition, various types of photons or phonons can be released or absorbed (at higher temperatures), in addition to the emitted luminescence photon, which provides additional structure or broadening in the luminescence spectrum. A luminescent line spectrum can also be broadened due to a kinetic energy fraction of free carriers or excitons participating in the transition. Other effects, too numerous to list here, influence the luminescence. They are described in many reviews of various fields relating to luminescence (see Voos et al. 1980). A few examples of some of the more important luminescence effects are given in the following sections.

3.2.1 Band-to-Band Luminescence

Direct band-to-band radiative recombination proceeds whenever an electron finds a hole with the same momentum vector within its recombination cross-section. The highest probability, therefore, is realized when both the electron and hole are close to the band edge near $\mathbf{k} = 0$.

Band-to-band luminescence is called *spontaneous emission*. It is time reversed to the optical excitation process, and thus the same matrix elements and selection rules apply – see Sect. 1 of ▶ Chap. 13, "Band-to-Band Transitions". The connection between optical absorption and luminescence emission can be seen by thermodynamic arguments. When a semiconductor is in equilibrium with a radiation field of brightness B (in photons/cm^2 per unit bandwidth, unit time and solid angle 4π), this

field loses entropy, which is gained by the optical excitation of the semiconductor and is expressed in the spread of quasi-Fermi levels:

$$\Delta S = -k\ln\left\{1 + \frac{8\pi\nu^2}{B}\left(\frac{n_\mathrm{r}}{\mathrm{c}}\right)^2\right\} = \frac{1}{T}\left\{h\nu - \left(E_{\mathrm{F}_n} - E_{\mathrm{F}_p}\right)\right\}. \tag{50}$$

This equation yields for the brightness of a radiation field in equilibrium with the semiconductor

$$B = 8\pi\nu^2\left(\frac{n_\mathrm{r}}{\mathrm{c}}\right)^2 \frac{1}{\exp\left\{h\nu - \left(E_{\mathrm{F}_n} - E_{\mathrm{F}_p}\right)/(kT)\right\} - 1}. \tag{51}$$

For reasons of detailed balance, we can now express the radiative transition rate r_rad for the spontaneous luminescence as the product of absorption and brightness of the impinging radiation field

$$r_\mathrm{rad} = \alpha_\mathrm{o} B, \tag{52}$$

which yields for weak optical excitation $\left(h\nu \ll E_{\mathrm{F}_n} - E_{\mathrm{F}_p}\right)$ the van Roosbroeck-Shockley (1954) relation

$$r_\mathrm{rad} = \alpha_\mathrm{o} 8\pi\nu^2\left(\frac{n_\mathrm{r}}{\mathrm{c}}\right)^2 \exp\left(\frac{E_{\mathrm{F}_n} - E_{\mathrm{F}_p} - h\nu}{kT}\right). \tag{53}$$

Equation 53 connects spontaneous luminescence emission with the optical absorption spectrum under the assumption that the electron distributions in conduction and valence bands are thermalized. In a microscopic description, we obtain the intensity of the luminescence by multiplying the matrix element with the joint density of states and the product of the distributions of electrons and holes in the conduction and valence bands, respectively (see Yariv 1975).

The recombination rate given above can be expressed as the product of electron and hole densities with a recombination coefficient. The *recombination coefficient* is given as the product of the capture cross-section (of electrons finding the hole) and the thermal velocity of these carriers. In Table 2, the recombination coefficients are listed for a number of typical semiconductors at two temperatures, indicating recombination cross-sections of 10^{-22}–10^{-15} cm^2 that decrease with increasing temperature. These cross sections are low for indirect-bandgap semiconductors; they require a matching phonon for momentum conservation. In direct-bandgap materials, e.g., in CdS (Reynolds 1960) and GaAs (Shah and Leite 1969), the recombination cross-sections are more than three orders of magnitude higher and are on the order of the atomic cross section.

A band-to-band luminescence spectrum is shown in Fig. 14 for Ge at 77 K. It shows two peaks: one due to an indirect transition near 0.71 eV and one with six orders of magnitude smaller intensity near 0.88 eV due to direct transitions. The intensity relation can be easily understood from the smaller occupation for the higher

Table 2 Recombination coefficients and cross sections for radiative band-to-band recombination at various temperatures (After Varshni 1967)

Semiconductor	T (K)	$c_{cv}^{(rad)}$ (cm^3 s^{-1})	s_{rad}
Si	90	$1.3\cdot10^{-15}$	$2.4\cdot10^{-22}$
	290	$1.8\cdot10^{-15}$	$1.9\cdot10^{-22}$
Ge	77	$4.1\cdot10^{-13}$	$8.5\cdot10^{-20}$
	300	$5.3\cdot10^{-14}$	$5.5\cdot10^{-21}$
GaAs	90	$1.8\cdot10^{-8}$	$3.3\cdot10^{-15}$
	294	$7.2\cdot10^{-9}$	$7.6\cdot10^{-17}$
GaSb	80	$2.8\cdot10^{-8}$	$5.6\cdot10^{-15}$
	300	$2.4\cdot10^{-10}$	$2.5\cdot10^{-17}$

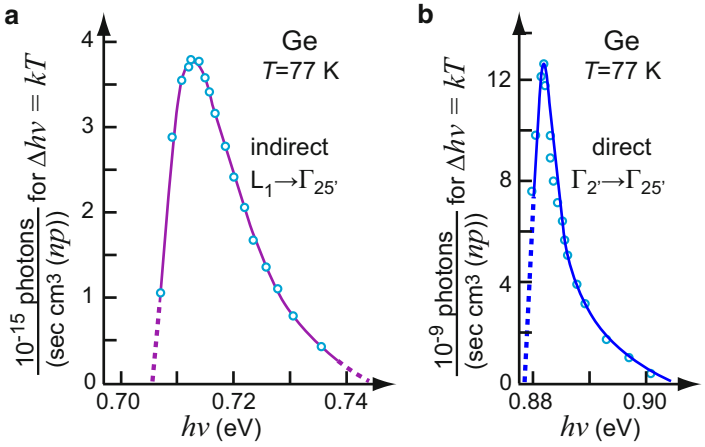


Fig. 14 Band-to-band recombination luminescence spectrum for Ge at 77 K for (**a**) indirect and (**b**) direct recombination (After Haynes and Nilsson 1965)

valley: the Boltzmann factor indicates a ratio of ~10^{-10}; however, the fact that we compare the direct with the indirect transition makes the direct emission ~10^4 times more probable. Therefore, the observed ratio of 10^6 has a reasonable explanation.

The luminescence photoflux is given by

$$I_L = \vartheta \; r_{rad}\, d, \tag{54}$$

where ϑ is a geometry factor, taking into consideration that only a fraction of the light can exit the front surface where it can be observed; it is assumed here that reabsorption of the emitted light can be neglected. r_{rad} is the radiative recombination rate, and d is the thickness of the sample. The spectral distribution of the spontaneous luminescence intensity shows a steep rise at the band edge and an exponential decay

toward higher energies, which reflects an exponential decrease of the distribution of thermalized electrons and holes with further distance from the band edges.

Spontaneous Luminescence from Hot Electrons The high-energy tail of the spontaneous luminescence, after sufficient relaxation (see Sect. 1.2 of ▶ Chap. 32, "Dynamic Processes"), is a direct measure of the electron temperature. When electrons are excited with light substantially above the bandgap energy, electron heating occurs and results in $T_e > T$. The intensity distribution of the spontaneous luminescence is given by (see Eq. 53):

$$I_{L,\,\mathrm{spont}}(h\nu) \propto \alpha_o \exp\left(-\frac{h\nu - E_g}{kT_e}\right). \tag{55}$$

The electron temperature T_e can be obtained directly from the high-energy slope of the luminescence peak, as shown in Fig. 15 for GaAs at a lattice temperature of $T = 2$ K. The hole distribution is little influenced by the optical excitation because of the larger hole mass.

The temperature of the electron gas increases logarithmically with the generation rate between 10^2 and 10^4 W/cm^2 and with a slope of $-k/\hbar\omega_{LO}$ due to a shift of the quasi-Fermi level in agreement with the experiment. At higher rates, the electron-hole interaction modifies the results (Shah 1974); at lower rates, additional piezoelectric deformation-potential scattering dominates, which keeps the electron temperature low and explains the deviation from the straight line in Fig. 15b.

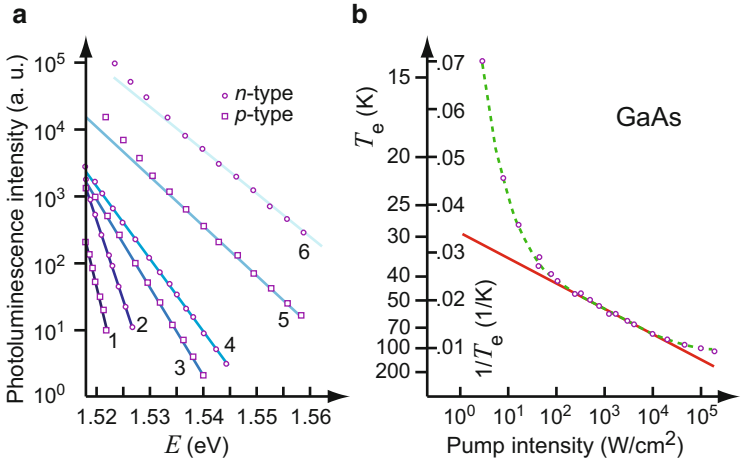


Fig. 15 (**a**) High-energy tails of the spontaneous luminescence of GaAs at 2 K with the optical generation rate as the family parameter: $g_o = 0.0011$, 0.0016, 0.0034, 0.0049, 0.64, and 0.76 g_o for curves 1–6, respectively. Corresponding electron temperatures are $T_e = 14$, 21, 36, 45, 64, and 76 K (After Shah and Leite 1969). (**b**) Inverse electron temperatures as a function of the optical excitation rate in GaAs (After Shah 1978)

Phonon-Assisted Oscillatory Luminescence When optical excitation is used to an energy level high within the band, rapid relaxation occurs with the preferred emission of LO phonons. Such relaxation is observed before the recombination of an electron with a hole occurs for spontaneous luminescence. Alternatively, a sufficiently slowed down electron can capture a hole to form an exciton with consequent exciton luminescence – see Sect. 3.2.3. The luminescence probability increases with decreasing kinetic energy of the carriers, hence causing oscillatory behavior with maxima when the LO phonon fits to bring the carrier close to the band edge. This is shown in Fig. 16. This condition yields luminescence maxima at energies given by

$$h\nu = E_g + n\ \hbar\omega_{LO}\left(1 + \frac{m_n}{m_p}\right), \tag{56}$$

with n an integer number.

3.2.2 Recombination of Free Excitons

At low temperatures, free electrons and holes first recombine to form excitons. Consequent recombination radiation occurs from free excitons or, after trapping at lattice defects (impurities or dislocations), from bound excitons. The intensity of this exciton recombination increases relative to the free electron-hole luminescence with decreasing temperature. We distinguish free excitons in direct- and indirect-bandgap materials. Their recombination luminescence is different.

Exciton Luminescence in Direct-Gap Semiconductors Direct-gap semiconductors have excitons close to $\mathbf{K} = 0$, where mixing with photons is important. This creates *exciton polaritons*, which have been discussed in Sect. 1.3 of ▶ Chap. 14, "Excitons". The shape of the dispersion curves for $|\mathbf{K}| > 0$ near $\mathbf{K} = 0$ approaches a parabola (see Fig. 14b of ▶ Chap. 14, "Excitons") for the different branches, since the exciton can acquire a kinetic energy that renders its total energy $\propto K^2$:

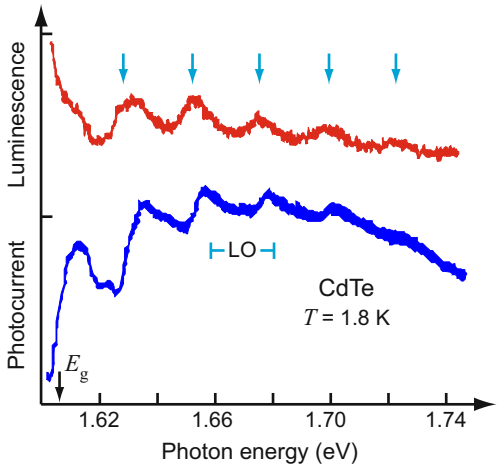


Fig. 16 Photoexcitation spectrum of bound-exciton luminescence (*upper curve*) and photoconductivity (*lower curve*) in CdTe at 1.8 K. *Arrows* according to Eq. 56 (After Nakamura and Weisbuch 1978)

$$E_{\mathrm{exc},(t,l)} \cong E_{\mathrm{g}} - E^{(\mathrm{exc})}_{\mathrm{qH}(t,l)} + \frac{\hbar^2 K^2}{2(m_n + m_p)}, \tag{57}$$

where $E^{(\mathrm{exc})}_{\mathrm{qH}}$ is the ground state of the quasi-hydrogen exciton; the subscripts t and l are for transverse and longitudinal excitons. The third term represents the kinetic energy of the exciton, which can be a substantial contribution. Therefore, the free-exciton line is relatively broad.

An example of a line spectrum for GaAs is given in Fig. 17, superimposed on the $E(k)$ dispersion diagram for identification of the upper and lower polariton branches (observe the stretched energy scale). The structure below the lower-branch exciton peak is due to impurities, which dominate the luminescence spectrum in bulk GaAs discussed in Sects. 3.3.2 and 3.3.3. For exciton-absorption spectra, which according to Eq. 53 are similar to exciton-luminescence spectra, see Figs. 7 and 8 of ▶ Chap. 14, "Excitons". For a review of spectra in II–VI compound semiconductors see Shionoya (1991).

The splitting of the exciton into longitudinal and transverse excitons is caused by the difference in polarizability; their energy difference is proportional to $\sqrt{\varepsilon_{\mathrm{stat}}/\varepsilon_{\mathrm{opt}}}$. Only the transverse exciton couples with the photon (Knox 1963). For a more extensive discussion, see Sect. 1.3 of ▶ Chap. 14, "Excitons". For phonon assistance or phonon scattering, see the review by Voos et al. (1980).

Excitons in Indirect-Gap Semiconductors In indirect-bandgap semiconductors, the luminescence emission-lines are usually phonon assisted:

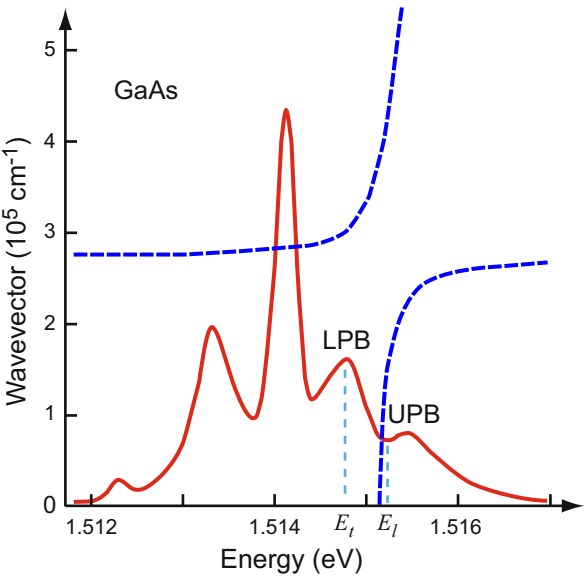


Fig. 17 Exciton-luminescence spectrum of the free-exciton polariton (*upper two maxima*); excitons bound at lattice defects (*lower three maxima*). *UPB* and *LPB* stand for upper and lower polariton branches, respectively. The *dashed curve* shows the $E(k)$ diagram (After Weisbuch and Ulbrich 1978)

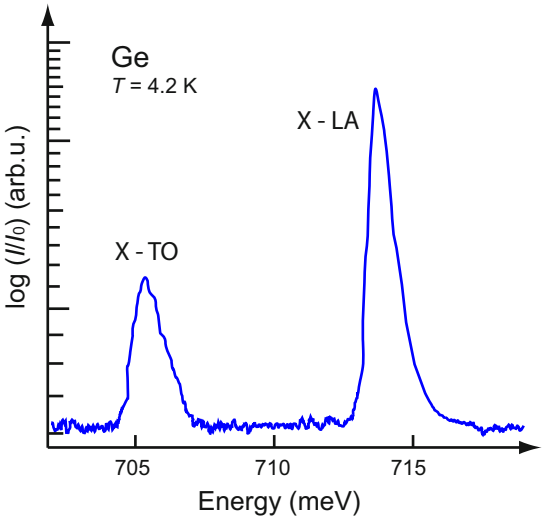


Fig. 18 LA and TO phonon-assisted free-exciton (X) luminescence of Ge at 4.2 K (After Etienne 1975)

$$E_{\text{exc}} \cong E_{\text{g}} - E_{\text{qH}}^{(\text{exc})} - \hbar\omega. \tag{58}$$

Selection rules determine which of the phonons is dominant, e.g., TA and LO, TO phonons for Si or LA and TO phonons for Ge as shown in Fig. 18.

3.2.3 Recombination of Exciton Molecules and of Electron-Hole Liquids

Exciton Molecules Trions and biexcitons are formed at higher excitation density with binding energies $E_{\text{B,trion}}$ or $E_{\text{B,XX}}$, which are on the order of meV or below (see Fig. 19 of ▶ Chap. 14, "Excitons"). The resulting luminescence energy is, e.g., for biexcitons,

$$E_{\text{B, XX}}(k) \cong E_{\text{g}} - E_{\text{qH}}^{(\text{exc})} - E_{\text{B, XX}} + \frac{\hbar^2 K^2}{4\left(m_n + m_p\right)}. \tag{59}$$

The existence of such excitonic molecules can be seen by observing the growth of this emission peak with increasing generation rate, see Figs. 20 and 25 of ▶ Chap. 14, "Excitons". Excitonic molecules have been observed in direct- and indirect-bandgap semiconductors. For more details see Sect. 1.4 of ▶ Chap. 14, "Excitons". Due to the low binding energy, further condensation into an electron-hole liquid is favored at high excitation density, as also observed in Fig. 19.

Electron-Hole Liquid Luminescence Electron-hole condensation can be observed from a bandgap reduction to the renormalized bandgap energy E_{g}' and from a broad line that appears at a well-defined temperature and at an intensity-related threshold –

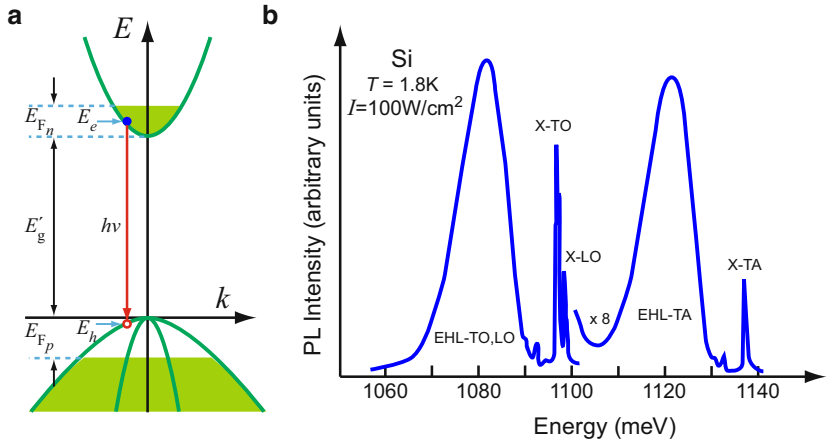


Fig. 19 Electron-hole liquid luminescence transitions; (**a**) band diagram $E(k)$, (**b**) photoluminescence spectrum of intrinsic Si at high excitation intensity, showing radiative recombination from an electron-hole liquid (EHL) (After Rostworowski and Bergersen 1978)

see Sect. 3.2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers". The emission related to the electron-hole liquid occurs at

$$E_{\mathrm{EHL}} = E_{\mathrm{g}}' + \left(E_{\mathrm{F}_n} + E_{\mathrm{F}_p}\right), \tag{60}$$

where E_{F_n} and E_{F_p} represent the kinetic energies of free electrons and holes, see Fig. 19. The luminescence intensity is given by

$$I(h\nu) \propto \int g_n(E_{\mathrm{F}_n})\ g_p\left(E_{\mathrm{F}_p}\right)\ f_{\mathrm{FD}}(E_{\mathrm{F}_n})\ \left\{1 - f_{\mathrm{FD}}\left(E_{\mathrm{F}_p}\right)\right\}\ \delta\left(E_{\mathrm{F}_n} + E_{\mathrm{F}_p} + E_{\mathrm{g}}' - h\nu\right), \tag{61}$$

with g as the density of states and f_{FD} as the Fermi-Dirac distribution. The luminescence spectrum of a typical electron-hole liquid is shown in Fig. 19, which also contains free-exciton-related peaks. The characteristic energies can be obtained from such spectra. From amplitude and shape of the EHL peak, we can obtain the pair density and carrier temperature. With decreasing temperature and pair density, the EHL peak increases at the expense of the free-exciton-related peak. For further discussion, see the reviews by Reinecke (1982), Rice (1977), and Hensel et al. (1978). For confined semiconductor structures, see Kalt (1994).

At temperatures *above* the critical temperature T_{c} (Sect. 3.2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers") and at exciton densities approaching the *Mott transition density* n_{Mott} (Eq. 63 of ▶ Chap. 21, "Equilibrium Statistics of Carriers"), the exciton gas becomes unstable and dissociates into a dense electron-hole plasma,

indicating the semiconductor-metal transition (Mott 1974). Consequently, a broadening of the recombination emission lines occurs, as observed by Shah et al. (1977).

3.3 Extrinsic Luminescence

Impurities in semiconductors cause a wealth of optical transitions in addition to the intrinsic transitions discussed in the previous sections. They usually appear at energies below the band-to-band and free-exciton luminescence described in Sects. 3.2.1 and 3.2.2. At very high doping level, however, the band-to-band transition is found at increased energy.

3.3.1 Band-Edge Luminescence in Heavily Doped Semiconductors

Heavy doping moves the Fermi level into the conduction band – see Sect. 3.2 of ▶ Chap. 17, "Optical Properties of Defects". Optical excitation then requires larger energies, labeled E_{g1} in Fig. 20a; band-to-band luminescence, however, has a lower energy threshold, shown as E_{g2} in the figure. The width of the emission band is shown as ΔE_{g12}. In elemental indirect-gap semiconductors, this luminescence band involves momentum-conserving TA or TO phonons with their corresponding replica, as shown for Si in Fig. 20b. In Ge, the LA and TO replica are most prominent. In direct-gap compound semiconductors, LO phonons are employed in this process.

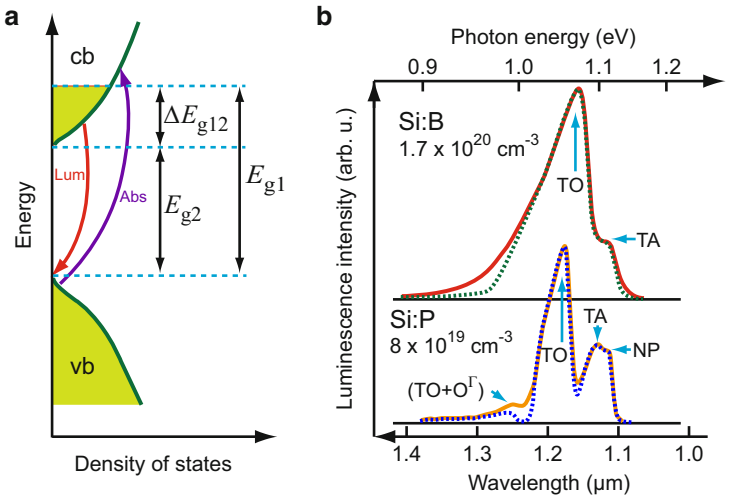


Fig. 20 (**a**) Density-of-state distribution of a heavily doped *n*-type semiconductor with partially filled conduction band (schematic). (**b**) Band-to-band luminescence of *p*-type Si:B ($p = 1.7 \times 10^{20}$ cm^{-3} – *upper curve*) and *n*-type Si:P ($n = 8 \times 10^{19}$ cm^{-3} – *lower curve*); *NP* denotes the no-phonon transition. *Dotted curve* theoretical fit (After Wagner 1985)

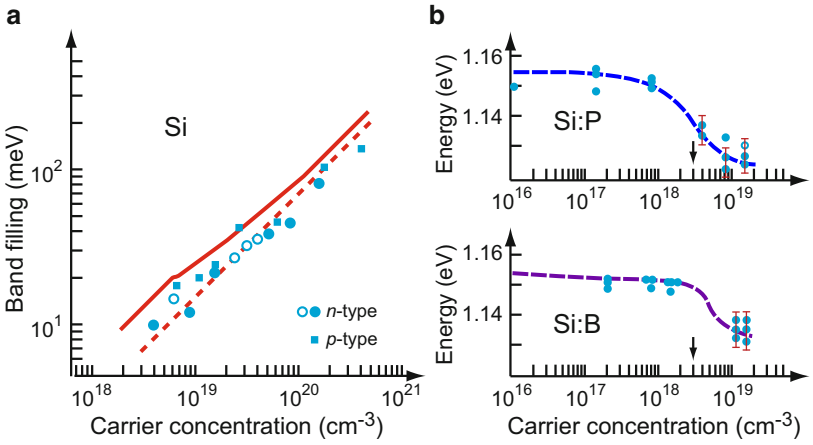


Fig. 21 (**a**) Band filling of Si as a function of carrier density: calculated curves for *n*-type (*dashed*) and for *p*-type (*solid curve*) Si at $T = 20$ K. (**b**) Bandgap reduction as a function of carrier density for Si obtained from optical absorption and from the photoluminescence cutoff at $T = 5$ K. The critical Mott density indicated by the *arrows* (After Wagner 1985)

The *band filling*, obtained by heavy doping and observed by the Burstein-Moss effect (Sect. 3.2 of ▶ Chap. 17, "Optical Properties of Defects"), is obtained from the luminescent line shape. The band filling is shown in Fig. 21a. The corresponding line broadening for Si is significant only for a carrier density in excess of 10^{18} cm^{-3}; it amounts to ~60 meV at 10^{20} cm^{-3} and is slightly larger for holes than for electrons, since it relates to the density-of-states mass, with $m_{nds} \cong 1.062\, m_0$ and $m_{pds} \cong 0.55\, m_0$ (see Barber 1967).

The *decrease* of *the bandgap* with band tailing at higher doping levels (see also Sect. 3.2 of ▶ Chap. 17, "Optical Properties of Defects") is obtained by a square-root fitting of the low-energy tail of the luminescence. It shows for Si an essentially constant bandgap energy up to carrier densities of ~10^{18} cm^{-3} and a significant decrease of E_g starting near the critical Mott density, as shown in Fig. 21b. For further detail, see Wagner (1985).

3.3.2 Free-to-Bound and Donor-Acceptor-Pair Transitions

A large variety of luminescent transitions occurs between states of free carriers and those of defect centers or within a defect center from a higher to a lower excited state or to the ground state. Shallow, hydrogen-like donors or acceptors are known to have high luminescence efficiencies; in GaAs, for instance, these are Be_{Ga}, Mg_{Ga}, Zn_{Ga}, Cd_{Ga}, C_{As}, Si_{As}, Ge_{As}, and Sn_{As} as acceptors and S_{As}, Se_{As}, Te_{As}, Si_{Ga}, Ge_{Ga}, and Sn_{Ga} as donors.

The hydrogen-like impurities produce spectra below the energy of the bandgap referred to as *near-band-edge emission*. Luminescence transitions, involving these defects, occur from the band to an acceptor or a donor (*free-to-bound* transitions), within such centers (*two-electron* or *two-hole* transitions), or between a donor and an

acceptor (*donor-acceptor-pair* transitions). With increasing temperature, recombination transitions involving at least one of the bands are favored, since the lifetime of carriers in shallow donors is decreased. This is indicated by a relation of the type given in Eq. 64 – see Sect. 3.3.5 (William and Hall 1978).

Luminescence transitions from a band into a hydrogen-like donor or acceptor reflect the optical ionization energy, as discussed for the example of the transition of an electron from the conduction band into an acceptor in GaAs shown in Fig. 15a of ▶ Chap. 18, "Shallow-Level Centers".

When donors are located close to acceptors, their relative positions must have certain values dictated by their occupied lattice sites. The Coulomb interaction between donor and acceptor is hence "quantized" according to the "shell" number, giving rise to a sequence of closely spaced discrete lines in the recombination-luminescence spectrum – see in ▶ Chap. 18, "Shallow-Level Centers" Sect. 1.6 and Fig. 13. Distant donor-acceptor pairs are numerous and have a broad distribution of donor-acceptor distances and a related spread in Coulomb energy. This yields broad luminescence lines as shown in Fig. 20 of ▶ Chap. 18, "Shallow-Level Centers".

Virtually all emission lines are accompanied by phonon replica. Pronounced phonon-assisted emissions appear at energies $n \times E_{LO}$ ($n = 1, 2 \ldots$) below the zero-phonon line, where E_{LO} is the energy of an LO phonon (e.g., 35 meV for GasAs and 92 meV for GaN). The intensity of phonon replica depends on the strength of phonon coupling expressed by the *Huang-Rhys factor* (Sect. 1.2.2 of ▶ Chap. 17, "Optical Properties of Defects"). Since the Huang-Rhys factor of donor-acceptor pairs is usually substantially larger than that of excitons, donor-acceptor pair transitions have pronounced replica of LO phonons at the low-energy side of the no-phonon emission.

3.3.3 Bound-Exciton Luminescence

Excitons can interact with point defects via an attractive dipole-monopole coupling when the defects are charged or via a somewhat weaker dipole-dipole coupling to neutral defects. These excitons (X) are identified as (D^+,X), (D^0,X), or (A^0,X) complex when bound to a charged donor, a neutral donor, or an acceptor, respectively (Lampert 1958). Excitons, bound to neutral or ionized donors or acceptors may be considered akin to an H atom, H^- ion, or H_2 molecule (Hopfield 1964). In many materials, the line spectrum of bound excitons is abundant because of a variety of involved defects, various phonon replica, and the narrow line shape that permits the resolution of rather complex spectra at low temperatures. The transition energies are decreased with respect to the free-exciton line by the binding energy to the defect. Acceptors provide a stronger binding potential to excitons than donors: due to the larger effective mass of the hole, the central-cell potential (i.e., chemical individuality) has a more pronounced effect[3]; recombinations of (A^0,X) complexes hence appear at lower energy than those of (D^0,X) complexes.

[3]The binding energy of excitons bound to donors or acceptors is proportional to the ionization energy of these defects (*Hayne's rule*, Sect. 2.2 of ▶ Chap. 18, "Shallow-Level Centers").

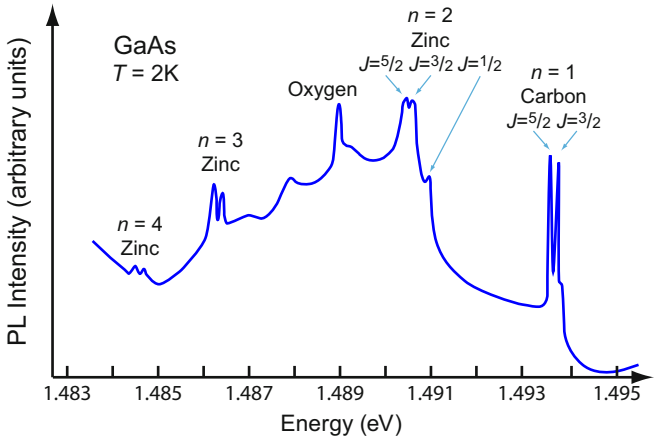


Fig. 22 Luminescence of excitons bound to neutral C and Zn acceptors in GaAs at 2 K. Termination of the luminescent transition occurs in the 2*s*, 3*s*, or 4*s* excited states of the acceptor. The fine structure is due to the initial-state splitting as indicated by $J = 1/2$, 3/2, or 5/2 as a result of combining two $J = 3/2$ holes and one $J = 1/2$ electron (After Reynolds et al. 1985)

In order to separate and identify the closely spaced lines, a magnetic field can be applied, which compresses the wavefunction; it also splits states with different orbital angular momentum. A splitting of the free-exciton state is also found in crystals of reduced symmetry, such as wurtzite crystals (like GaN or CdS), crystals under uniaxial stress (Benoit a la Guillaume and Lavallard 1972), or *strained* cubic crystals. Emission lines of free excitons, usually labelled X or FE, are then labelled X_A to X_C for split *A* to *C* valence bands or X_{hh} and X_{lh} for light- and heavy-hole excitons, respectively.

No-phonon transitions occur from the ground or excited state of the bound exciton leaving the state of the donor or acceptor unchanged or with a final state of these impurities at $n = 2, 3, \ldots$. When the radiative recombination of an exciton bound to a donor or an acceptor leaves the shallow impurity in an *excited* state, the transition energy is reduced accordingly. The related luminescence is referred to as two-electron transitions or two-hole transitions, respectively.

An example of the luminescent spectral distribution in GaAs with C and Zn acceptors is shown in Fig. 22. The transitions involve the annihilation of an exciton bound to these acceptors, with the acceptor at the end of the transition in the excited 2*s*, 3*s*, or 4*s* states as indicated by $n = 2$, 3, and 4, respectively. The high degree of resolution in the small range of the spectrum shown – here a total of 6 meV – is emphasized (see Reynolds et al. 1985).

Excitons can also be bound by neutral atoms of the same valency as the replaced lattice atoms. Such *isoelectronic centers* have a short-range potential and therefore extend to rather large *k* values. This permits electron transitions from $k \neq 0$ without phonon assistance and hence yield a high luminescence efficiency at room

temperature, an important fact for indirect-bandgap materials such as N, As, or Bi in GaP (Thomas et al. 1965; Trumbore et al. 1966; Dean et al. 1969).

For a detailed discussion of bound-exciton states see Sect. 2 of ▶ Chap. 18, "Shallow-Level Centers".

3.3.4 Luminescence Centers

Materials that show highly efficient luminescence are called *luminophores* or *phosphors.*[4] They consist of a host material such as aluminates, oxides, phosphates, sulfates, selenides, or silicates of Ba, Ca, Cd, Mg, or Zn and of a light-emitting dopant referred to as *luminescence center* or *activator.* They are of high technical interest for luminescent displays, such as CRT screens or luminescent lamps (LEDs). For reviews, see Curie (1963), Goldberg (1966), and Yen et al. (2007).

By introduction of an activator with preferred radiative transitions the luminescence of a semiconductor can greatly be enhanced. In contrast to the exciton and band-edge transitions, which were discussed in the previous sections, such activator-related luminescence is shifted from the band edge well into the bandgap, where the optical absorption is reduced. These luminescent transitions have a very high transition probability, approaching 100% at room temperature.

Activators are often transition-metal ions, which produce optical transitions within their uncompleted inner *d* or *f* shell, see Sect. 3 of ▶ Chap. 19, "Deep-Level Centers". White-emitting LEDs, e.g., superimpose blue light of an InGaN LED emitting at 450 nm with the broad yellow emission of a phosphor peaking near 550 nm; a common yellow phosphor is yttrium aluminum garnet ($Y_3Al_5O_{12}$ or YAG) with a bandgap energy of 4.7 eV at room temperature, doped with cerium ($YAG:Ce^{3+}$). Another example of a good luminophore is ZnS with a bandgap energy of 3.6 eV at room temperature; with Cu as an activator, it yields a green luminescence peak at 2.3 eV. Other typical activators of ZnS are Mn, Ag, Tl, Pb, Ce, Cr, Ti, Sb, Sn, and other transition metal ions except Fe, Ni, and Co – see Sect. 3.3.5. Some prominent phosphors are listed in Table 3. For more information on the numerous types of phosphors and their applications, see, e.g., Blasse and Grabmayer (1994), Ronda (2008).

The intensity of the luminescence can be obtained from reaction-kinetic considerations

$$\frac{dn}{dt} = g_o - n \left(c_{c,act} p_{act} + c_{c,rec} p_{rec}\right), \tag{62}$$

with the optical generation rate g_o and $n = p_{act} + p_{rec}$; $c_{c,act}$ is the coefficient for the transition from the conduction band to the activator, and p_{act} is the hole density in the activator, yielding the luminescent transition as the product of these two quantities.

[4]The term phosphor should not be confused with the chemical element phosphorus, which emits a faint glow when exposed to oxygen. This property is related to its Greek name *Φωσφόρος*, meaning *light bearer*, which also lead to the term *phosphorescence* to describe a glow after illumination, and the general term phosphor for fluorescent materials.

Table 3 Emission-peak wavelength, peak photon energy, and peak width of phosphors used in fluorescent lamps and cathode-ray/projection-TV tubes

Color	Phosphor	λ (nm)	$h\nu_{peak}$ (eV)	$h\Delta\nu$ (eV)
UV	$BaSi_2O_5:Pb^{2+}$	350	3.56	0.22
Blue	$ZnS:Ag^{+},Cl^{-}$	438	2.83	0.15
	$ZnS:Cu^{+},Cl^{-}$	453	2.74	0.19
	$ZnS:Ag^{+}$	457	2.71	0.36
Green	$Zn_2SiO_4:Mn^{2+}$	520	2.38	0.11
Yellow	$Y_3Al_5O_{12}:Ce^{3+}$	539	2.30	
Red	$Y_2O_3:Eu^{3+}$	620	2.0	

Correspondingly $c_{c,rec}\, p_{rec}$ is the nonradiative transition, see Sect. 1 of ▶ Chap. 29, "Carrier Generation". For steady-state conditions ($dn/dt \equiv 0$), we obtain for the luminescent intensity I_L (proportional to the radiative recombination transition)

$$I_L \propto c_{c,act}\, n\ p_{act} = g_o c_{c,rec}\, n p_{rec} \tag{63}$$

for reasons of detailed balance. Losses in luminescence efficiency relate to partial absorption, to internal reflection, and to competing nonradiative recombination transitions. The luminescence intensity ($\propto c_{c,act}\ n\ p_{act}$) increases also with increasing activator concentration. Doping with a higher density of activators is usually restricted by their solubility, which is related to neutrality considerations – see Sect. 2.5 of ▶ Chap. 15, "Crystal Defects". The neutrality is provided by intrinsic lattice defects, which are generated during the incorporation of the activator according to thermodynamic relations – see Sect. 2.4 of ▶ Chap. 15, "Crystal Defects". Further *increased doping* can be achieved by the incorporation of atoms with an opposite charge character. In the example of ZnS doped with Cu, an oppositely charged defect is Cl, which results in ZnS:Cu,Cl. This additional dopant is called a *co-activator*. Heavily doped luminophores often show additional emission maxima. ZnS:Cu,Cl first shows a green emission; increased densities of Cu and Cl result in a blue emission maximum. This is due to the fact that Cu can be incorporated at two different sites in the ZnS lattice, replacing one Zn atom as Cu_{Zn} and as an interstitial Cu_i. The blue emission is attributed to a $Cu_{Zn}Cu_i$ associate.

3.3.5 Quenching of Luminescence

When holes are released into the valence band from electron-activated luminescent centers, the luminescence is reduced (*quenched*) since these holes can become trapped at other, competing centers, the *quenchers* with nonradiative transitions. Typical luminescent quenchers are Fe, Ni, and Co.

The release of such holes can be caused by thermal ionization, resulting in thermal quenching of the blue luminescence of ZnS:Cu, which is thermally quenched already below room temperature or by optical (IR) irradiation resulting in quenching also of the green luminescence. These transitions are similar to those that will be discussed in the section on photoconductivity, see ▶ Chap. 31,

"Photoconductivity". The intensity of luminescent transitions, competing with thermal quenching, can be described by (see Eq. 49)

$$I = \frac{I_0}{1 + \frac{\tau_{\text{nonrad}}}{\tau_{\text{rad}}}} = \frac{I_0}{1 + \gamma \exp\left(\frac{\Delta E_{\text{th}}}{kT}\right)}, \tag{64}$$

where $\gamma = \tau_{\text{nonrad}} \exp[\Delta E_{\text{th}}/(kT)]/\tau_{\text{rad}}$, and ΔE_{th} is the thermal activation energy for the quenching transition (William and Hall 1978). The parameter γ contains the capture coefficients $c_{n,p}$ for electron and hole transitions, which more recently were computed from first principles using hybrid density functional theory (Alkauskas et al. 2014, Dreyer et al. 2020). The calculations considered electronic structure and lattice relaxations, and obtained, e.g., c_n results for GaAs:V_{Ga}-Te_{As} and GaN:C_N of 3.5 and 0.7 (in units of 10^{-13} cm^3s^{-1}), in good agreement with time-resolved photoluminescence studies of *n*-type GaAs (Glinchuk et al. 1977) and GaN (Reshchikov et al. 2016). This indicates the validity of the configuration-coordinate model outlined in Sect. 1.2 of ▶ Chap. 17, "Optical Properties of Defects", and Sect. 1.1.2, even for large lattice relaxations. A discussion of this model in harmonic approximation is given by Das et al. (2020). The importance of anharmonicity is emphasized for the bondbreaking relaxation of the DX center in GaAs, where a first-principles procedure with a Morse-like potential accounts for anharmonic carrier capture (Kim et al. 2019).

With increasing concentration of activators, first an increase and, at higher densities, a decrease of luminescence efficiency is observed. One reason for such a *concentration quenching* is a carrier transfer from activator to activator, facilitated by their close proximity, until a quencher is encountered before luminescent recombination occurs.

When deep traps are present, they may store optically excited or injected electrons for a substantial amount of time. These carriers are then slowly released by thermal excitation and consequently recombine under radiative transitions. This phenomenon is called *phosphorescence.* Release of electrons from such traps can be enhanced by increasing the temperature. The distinction between luminescence and phosphorescence is kinetic in nature.

3.3.6 Stimulated Emission

In the previous sections, *spontaneous* luminescent emission was discussed. This emission type dominates when the density of photogenerated carriers is small enough so that each act of photon emission is generally spontaneous; that is, the different acts of emission do not communicate significantly with each other, and the emission is *incoherent*.

When the density of photogenerated carriers becomes very large, the light generated by luminescent recombination can stimulate other electrons at nearby luminescent centers to also recombine and emit light *in phase* with the stimulating luminescent light. This *stimulated emission* increases with increasing excitation. When reflections create standing waves of the corresponding wavelength in an

optical resonator and thereby identify a well-defined phase relation, in-phase emission occurs, i.e., *coherent emission*. *Lasing* can finally be achieved with sufficient pumping to a critical intensity of the exciting light at this wavelength. In a somewhat simplified fashion, such a critical intensity can be deduced from the condition that the internal luminescent efficiency, given by Eq. 49: $\eta_L = 1/(1 + \tau_{spont}/\tau_{nonrad})$ must increase sufficiently to overcome the optical losses in the luminophore. The relaxation time for spontaneous luminescence τ_{spont} is usually much longer than that for nonradiative recombination τ_{nonrad}; hence, η_L is usually $\ll$ l. When an inverted population of the excited and ground states of the luminescence center is produced, the luminescent lifetime decreases sharply from τ_{spont} to a stimulated-emission lifetime τ_{stim}; thus, η_L rapidly rises and lasing occurs. Such an inverted population can be achieved when the excitation is strong enough so that there are more electrons in the excited than in the ground state.

The pumping of a laser, in order to achieve a sufficient population of excited states, can be performed by optical excitation. This was done first with a light flash exciting a ruby rod from its circumference (Maiman 1960). Carrier injection provides another efficient mode for pumping semiconductor lasers, originally carried out by Hall et al. (1962), Nathan et al. (1962), and Quist et al. (1962) in GaAs. Today, many kind of semiconducting lasers are available. For reviews of semiconductor lasers, see, e.g., Agrawal and Dutta (1993), Chow et al. (1997), Kapon (1999), and Coleman et al. (2012).

4 Noise

In previous sections and chapters, carrier generation, recombination, and transport were regarded as stationary with a time-independent current or luminescence. However, statistical fluctuations of carrier density, mobility, and electric field are the cause for *current-density fluctuations*, called *noise*, as in the acoustic impression one hears in a loudspeaker after sufficient amplification:

$$\delta j = e\mu n \delta F + e\mu \; F \delta n + e n F \delta \mu. \tag{65}$$

Fluctuations of the electric field F are of importance for inhomogeneous semiconductors and will not be discussed here. We will first discuss the basic elements of current fluctuation (see van der Ziel 1986) and then will give a brief review of the different noise mechanisms, with some emphasis on $1/f$ noise – also denoted here as $1/\nu$ noise for consistency.

4.1 Description of Fluctuation

The current in a semiconductor fluctuates about an average value. Such fluctuations are observable after amplification; band-pass filtering transforms the fluctuation $j(t)$ from the time (t) domain into the frequency (ν) domain $j(\nu)d\nu$, mathematically

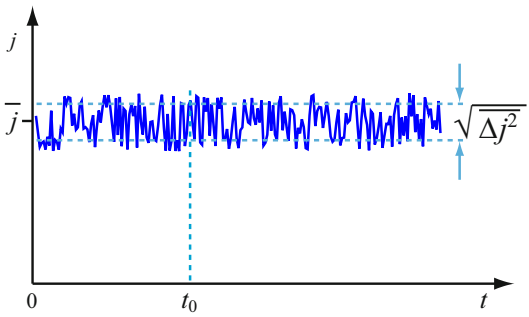


Fig. 23 Fluctuating current j with average value and mean square deviation identified. Probing occurs in specific time intervals of length t_0 after amplification through an amplifier of bandwidth $\Delta\nu = 1/t_0$

requiring a Fourier transformation. Usually, we sample the random fluctuation of $j(t)$ shown in Fig. 23, or of $V(t) = R\ I(t)$ in a small frequency interval $\Delta\nu$, with its variance given by the mean square

$$\overline{\left(j(t) - \bar{j}\right)^2} = \overline{\Delta j^2} = \int_{\nu_1}^{\nu_2} S_j(\nu)\ d\nu \cong S_j(\nu)\ \Delta\nu. \tag{66}$$

Here, $S_j(\nu)$ is the spectral intensity of the current fluctuation, which can be measured after amplification, and is the subject of most theoretical investigation. Alternatively, sometimes the spectral intensity $S_V(\nu)$ of the fluctuation of an emf across the device is analyzed.

When expressing a fluctuating current $j(t)$ by its Fourier components

$$j(t) = \sum_{n\ =\ -\infty}^{\infty} a_n \exp(i\,\omega_n t) \text{ with } \omega_n = n\frac{2\pi}{t_0}, \tag{67}$$

the amplitudes can be expressed as

$$a_n = \int_0^{t_0} j(t)\exp(-i\,\omega_n t)\ dt; \tag{68}$$

here $1/t_0 = \Delta\nu$, and $\Delta\nu$ is the sampling bandwidth. The spectral density is given by

$$S_j(\nu)\ = \lim_{t_0\to\infty} 2t_0\overline{a_n a_n^*}. \tag{69}$$

In the following sections, we will give a simplified picture of the general noise phenomenon. We then will present the results for the different types of noise for a homogeneous semiconductor, rather than deriving the spectral density in a vigorous manner. Such derivations can be found in several monographs (see, e.g., Ambrozy 1982; van der Ziel 1986; Kogan 1996).

Electronic Noise Electronic noise can be described as a train of pulses, which are generated by each movement of an electron between scattering events or between generation and recombination of carriers. The averaged square of the deviation from the average current, divided by $\bar{i}^2$, is inversely proportional to the number n of current pulses. Hence, with $i = n\,e/t$, and t as the length of each current pulse, we have, following the square-root law of Poisson statistics,

$$\frac{\overline{\Delta i^2}}{\bar{i}^2} = \frac{1}{n} = \frac{e}{\bar{i}\,t}, \text{or } \overline{\Delta i^2} = \frac{e\,\bar{i}}{t}. \tag{70}$$

After Fourier transformation of each current pulse, by taking the squared absolute value and integrating over all amplitudes in a small frequency interval $\Delta\nu$ of amplification, we obtain the total mean square current fluctuation within $\Delta\nu$.

Commonly we distinguish:

- Thermal noise
- Shot noise
- Generation-recombination noise
- $1/f$ noise

The first three classifications can be derived from basic classic principles, while the fourth is derived from quantum electrodynamics. For reviews, see van der Ziel (1970, 1986), Dutta and Horn (1981), Hooge et al. (1981), Ambrozy (1982), van Kampen (1992), and Handel (1994, 1996). We also distinguish noise under equilibrium and nonequilibrium conditions.

4.2 Noise in Equilibrium Conditions

The equilibrium noise is given by the thermodynamic behavior of carriers. Little can be done to reduce its amplitude, except by changing the temperature or the bandwidth of the amplifier.

Thermal Noise Thermal noise, also referred to as *Johnson-Nyquist* noise, is caused by the random motion of carriers in a semiconductor. It was originally discussed by Nyquist (1928) from thermodynamic arguments and was later analyzed by Bakker and Heller (1939). The respective mean square of the current is given by

$$\overline{(\Delta j)^2} = \frac{4\,\Delta\nu}{R}\;\frac{h\nu}{\exp[h\nu/(kT)] - 1} \cong \frac{4\,kT}{R}\Delta\nu, \tag{71}$$

where R is the resistance of the semiconductor and $\Delta\nu$ is the bandwidth of the amplifier. This noise is caused by the Brownian motion of the carriers. Equation 71 aptly describes the noise observed in metals and semiconductors in thermodynamic

equilibrium, i.e., for vanishing current or optical excitation. However, the noise can be larger by many orders of magnitude due to density and field fluctuation in nonequilibrium conditions (see below).

Equivalent Noise Resistor The actual noise in semiconducting devices is often described by an *equivalent noise resistor,* using Eq. 71, but describing the *potential* fluctuation

$$\overline{V^2} = \overline{j^2} R^2 = 4kTR_{\mathrm{n}}\Delta\nu \tag{72}$$

by an equivalent resistor R_{n}. When such a resistor becomes larger than given by the actual device resistivity, nonequilibrium noise is also included in this description.

Another way of assigning an elevated noise output to a device is by defining an *equivalent noise temperature* T_{n}, according to a similar formula,

$$\overline{V^2} = \overline{j^2} R^2 = 4kT_{\mathrm{n}}R\,\Delta\nu. \tag{73}$$

Temperature Fluctuations The phonon content in a small semiconductor in thermal equilibrium can fluctuate with respect to its internal energy distribution as well as its connection to the surrounding. We hence expect temperature fluctuations which in turn affect the resistance noise. This was first suggested by Voss and Clarke (1976) as a cause for $1/f$ noise, but was later rebutted when measurements of noise correlations disagreed with their predictions; see, e.g., Dutta and Horn (1981) and Kilmer et al. (1982). van Vliet and Mehta (1981) proved that thermal fluctuations do not yield $1/f$ noise in any geometry or number of dimensions.

4.3 Nonequilibrium Noise

Nonequilibrium noise is generated by an injected current or by extrinsic excitation. This noise can be changed by doping, variation of interfaces, or contacts.

4.3.1 Shot Noise or Injection Noise

The *shot noise* was derived for vacuum diodes. It is caused by the discreteness of elementary charges crossing the diode and is related to the statistical exit of electrons from the cathode. It may be applied to carrier injection from one (or two, in double injection) of the electrodes of a semiconductor (see Nicolet et al. 1975). The shot noise was first estimated by Schottky (1918) with a current fluctuation obtained for $1/2\ \Delta\nu$ from Eq.70

$$\overline{(\Delta I)^2} = 2eI\,\Delta\nu. \tag{74}$$

Except for a factor of 2, it is equal to the thermal noise[5] given in Eq. 71, when using thermal emission from the electrode $I = I_s \exp[-e(\psi_{MS} - V)/(k\ T)]$, where ψ_{MS} is the metal-semiconductor work function and I_s is the saturation current, linearizing, and setting $\partial I/\partial V = 1/R$. For a review of the measurement techniques of intrinsic noise, see the classic work of Bittel (1959) and van der Ziel (1970, 1986); see also Sikula and Levinshtein (2004).

Noise Due to Mobility Fluctuation The fluctuation of carrier mobilities due to a variety of scattering processes was evaluated by Kousik et al. (1985, 1989) and Tacano (1993); it yields a $1/f$ noise for ionized-impurity scattering as well as for electron-phonon, intervalley, and umklapp scattering. See also Pellegrini (1986).

Noise in Quantum-Ballistic Systems In point contacts and devices that are smaller than the free path of carriers, the carriers move in a ballistic fashion without scattering, see Sect. 3.1 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors". Still, the carrier motion in the entire device is dissipative because the heated carriers loose their energy in the adjacent parts of the device by scattering with phonons.

In thermal equilibrium, such a ballistic system follows the Johnson-Nyquist formula (Eq. 72). With an applied voltage the noise increases; however, it reaches the full level of the shot noise $2\ e\ I$ only when the current is low. With higher currents this noise is substantially reduced. The suppression of the shot noise is due to the Fermi statistics (i.e., the Pauli principle), as first shown by Levitov and Lesovik (1993): In "small" quantum devices the transfer of electrons becomes restricted as soon as their density exceeds a quantum limit (the Fermi limit) of electron correlation. To emphasize this correlation, we speak of a *quantum shot noise*. The magnitude of this noise suppression is seen from Fig. 24, which shows for different currents the spectral density of the noise in a quantum-point contact as a function of the gate voltage defining a two-dimensional electron gas (2DEG).

With larger contacts or devices, this suppression disappears. González et al. (1998) discuss this crossover between the diffusive and the ballistic regime. The joint effect of Coulomb and Fermi correlation in shot-noise suppression in ballistic conductors is emphasized by González et al. (1999); see also Liefrink et al. (1991). For the analysis of quantum shot noise in mesoscopic systems, see Reznikov et al. (1998). Contacts also produce $1/f$ noise, see Sect. 4.4.

Noise in Tunnel Junctions The noise in tunnel junctions is governed by two contributions: the statistical transit of carriers through a constant junction and the fluctuation of the barrier height and its resistance produced by the motion of defects at the boundary or inside the junction (Kogan 1996). The first effect yields the full shot noise, except for small-area tunnel junctions; here the capacitance is small enough to change its energy $e^2/2C$, when a single electron passes, by more than kT (see Sect. 4.1 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors").

[5] In contrast to the vacuum diode, the current in a semiconductor is bidirectional, hence the factor 2.

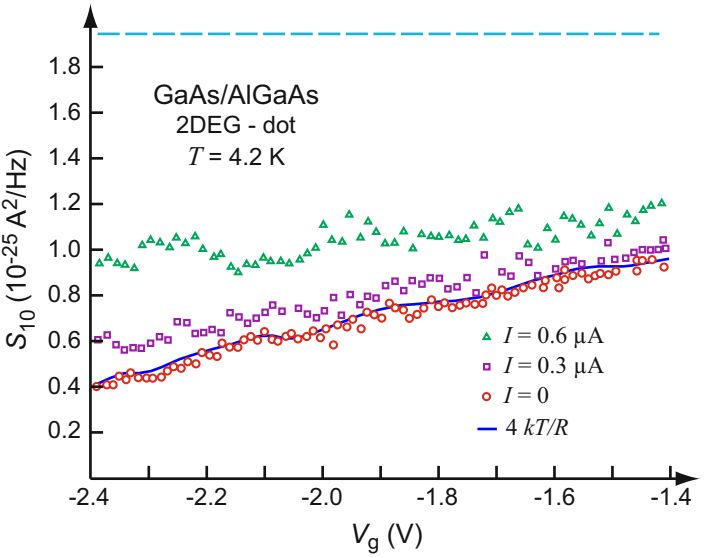


Fig. 24 Spectral density of the noise in a quantum-dot contact versus gate voltage *V*. The *solid line* represents the Johnson-Nyquist noise; the *dashed line* on *top* is the full shot noise for $I = 0.6$ μA (After Li et al. 1990)

This regime is referred to a Coulomb blockade and is reviewed by Grabert and Devoret (1992); see also Korotov (1994).

4.3.2 Generation-Recombination Noise

Carriers can be generated by injection from the electrode, as discussed before, or within the semiconductor (e.g., by thermal generation or by light – see Sect. 1 of ► Chap. 29, "Carrier Generation") and annihilated by recombination or temporarily immobilized by trapping. This causes a fluctuation in carrier density. This contribution was first calculated by Gisolf (1949) and was later modified by van Vliet (1958), yielding

$$\overline{\Delta I_n^2} = 4\left(\frac{I}{N}\right)^2 \frac{\tau_n}{1+4\pi^2\nu^2\tau_n^2} \quad \frac{N(N_A+N)(N_D-N_A-N)}{(N_A+N)(N_D-N_A-N)+NN_D}\ \Delta\nu, \quad (75)$$

where τ_n is the electron lifetime in the conduction band, and N, N_D, and N_A represent the numbers (not concentrations) of electrons, donors, and acceptors in the actual device (see also van Vliet and van der Ziel 1958). An example of the spectral distribution for a device with two pronounced centers is shown in Fig. 25; the figure also contains two other noise components. The observed frequency distributions of the *generation-recombination noise* component usually also shows a dispersive

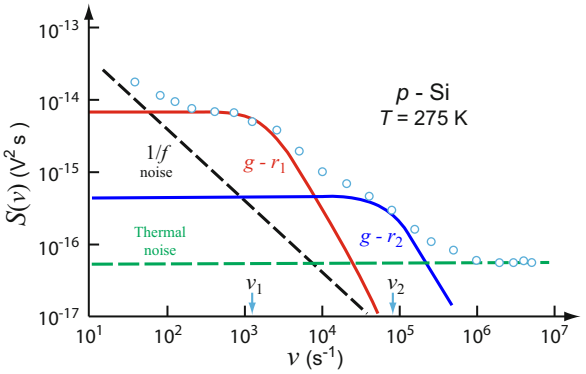


Fig. 25 Spectral intensity $S(\nu)$ of open-circuit voltage fluctuation of a *p*-type Si single-injection device at an electric field of 83 V/cm. *Solid curves* indicate two generation-recombination noise components $g - r_{1,2}$ with substantially different characteristic frequencies. Also *plotted* are the limiting $1/f$ noise and the thermal noise contribution (After Bosman 1981)

behavior, indicating a wide *distribution* of lifetimes (Böer and Junge 1953; McWorther 1955).

4.4 The 1/*f* Noise

The $1/f$ noise, in the early literature referred to as *flicker noise*, was first observed by Johnson (1925) and prevails in many semiconducting devices. This noise contains, in general, both fundamental and nonfundamental fluctuations (see van der Ziel 1950). Nonfundametal quasi-$1/f$-like contributions include superposition of Lorentzian generation-recombination noise spectra, with a distribution of electron-relaxation times τ_n (McWorther 1955), and mobility fluctuation. The fundamental $1/f$ noise is the *quantum* $1/f$ *noise* and is described by Handel (1980, 1994, 1996) and Sherif and Handel (1982).

In collision-free devices (diodes and ballistic devices), the $1/f$ noise can be explained by accelerated electrons emitting *bremsstrahlung*, which has a $1/\nu$ -spectrum. This is caused by the feedback of the emitted photon on the decelerating electron, yielding

$$\overline{I^2} = \frac{\alpha_{\mathrm{H}}}{\nu N} I^2 \, \Delta\nu, \tag{76}$$

where N is the number of carriers in the system, and α_{H} is the Hooge parameter (Hooge 1969)

$$\alpha_{\mathrm{H}} = \frac{4\alpha}{3\pi}\left(\frac{\Delta \mathbf{v}}{c}\right)^2. \tag{77}$$

Here, $\alpha = e^2/\hbar c = 1/137$ is the fine-structure constant, $\Delta\mathbf{v}$ is the change of carrier velocity along its path, and c is the light velocity (see Handel 1982; van der Ziel 1986). A similar formula can also be applied to semiconductors in which the carrier transport is *collision limited*, and α_{H} is calculated by averaging Eq. 77 over the various scattering processes. Often it is sufficient to focus on a single, dominant large-angle scattering, recombination, or tunneling process to obtain the average $\overline{(\Delta \mathbf{v}/c)^2}$. For larger samples or devices, Handel obtains for the coherent quantum-1/*f*-effect $\alpha_{\mathrm{H}} \cong 2\alpha/\pi$, independent of the scattering mechanism.

The coherent or conventional quantum 1/*f* effect is of basic nature and is generally referred to as *quantum* 1/*f noise* as discussed below; for reviews see van der Ziel (1986).

4.4.1 The Conventional Quantum 1/*f* Effect

The physical origin of quantum 1/*f* noise can be understood by considering the example of Coulomb scattering of electrons at a charged defect. The scattered electrons reach a detector at different angles with respect to the incident direction. When they lose energy in this process due to the emission of (low frequency) bremsstrahlung photons, a part of the outgoing de Broglie waves is slightly shifted to lower frequencies. This makes the scattering slightly inelastic. When calculating the probability density of the scattered electron ensemble, we obtain also cross terms, linear in both the part scattered elastically and inelastically, i.e., with and without emission of bremsstrahlung. These cross terms oscillate with the same frequency as the emitted bremsstrahlung photons. The emission of photons at all frequencies therefore results in probability-density fluctuations. The corresponding current-density fluctuations are obtained by multiplying the probability-density fluctuations by the velocity of the scattered carriers; they are detected as low-frequency current fluctuations and can be interpreted as fundamental cross-section fluctuations in the scattering cross-section of the scatterer. While incoming carriers may have been Poisson distributed, the scattered ensemble will exhibit super-Poissonian statistics with more probable bunching; due to this effect, we may call this a quantum 1/*f* effect.

The quantum 1/*f* effect is thus a collective many-body effect; it can be analyzed as a two-particle effect, described by a two-particle wavefunction and two-particle correlation function. The two-particle wavefunction is the product of two single-particle functions ψ with negligible interaction between them, so its square is $|\psi|^4$.

In a semiclassical approach, a particle of charge e and acceleration $\mathbf{a}$ radiates a power given by the classical (Larmor) formula $2e^2\mathbf{a}^2(\Delta\mathbf{v})^2/(3\hbar\, c^3)$. The acceleration can be approximated by a delta function $\mathbf{a}(t) = \Delta\mathbf{v}\delta(t)$, whose Fourier transform $\Delta\mathbf{v}$ is constant and is the change in the velocity vector of the particle during the almost instantaneous inelastic scattering process; the spectral density of the emitted bremsstrahlung power is therefore also constant. The number of emitted photons per unit frequency-interval is obtained by dividing the bremsstrahlung power by the energy of one photon; the probability amplitude of photon emission is then given by the

square root of this photon-number spectrum. If ψ is a representative single-particle Schrödinger wavefunction of the scattered charged particles, the beat term in the probability density $\rho = |\psi|^2$ is linear in both the bremsstrahlung amplitude and in the non-bremsstrahlung amplitude, i.e., in the square root of the probability for inelastic and elastic scattered electrons. Its spectral density will therefore be given by the product of the squared probability amplitude of photon emission ($\propto 1/\nu$) with the squared non-bremsstrahlung amplitude which is independent of ν. This yields the spectral density of fractional probability-density fluctuations

$$|\psi|^{-4} S\,|\psi|^2(\nu) = \frac{8\,e^2(\Delta \mathbf{v})^2}{3\,h\nu N\,c^3} = \frac{2\alpha A}{\nu N} = \frac{\alpha_0}{\nu N} = j^{-2} S_j(\nu), \tag{78}$$

where α is the fine-structure constant, α_0 is the quantum $1/f$ noise coefficient, and αA is known as the infrared exponent in quantum-field theory. From Eq. 78 we obtain the current fluctuation in the $1/f$ regime by insertion of the spectral density $S_j(\nu)$ into Eq. 76. The spectral density of the current-density fluctuations is obtained by multiplying the probability-density fluctuation spectrum with the squared velocity of the outgoing particles. When we calculate the spectral density of fractional fluctuations in the scattered current j, the outgoing velocity simplifies, and therefore Eq. 78 also gives the spectrum of current fluctuations S_j. The quantum $1/f$ noise contribution of each carrier is independent, and therefore the quantum $1/f$ noise from N carriers is N times larger; since also the current j is N times larger, a factor N was included in the denominator of Eq. 78.

The fundamental fluctuations of cross sections and process rates are reflected in various kinetic coefficients in condensed matter, such as the mobility μ, the diffusion constant D, and the thermal diffusivity in semiconductors, the surface and bulk recombination speeds v_{rec}, recombination times τ_{rec}, or the rate of tunneling j_{t}. Therefore, the spectral density of fractional fluctuations in all these coefficients is given also by Eq. 78. This conventional quantum $1/f$ theory was verified by calculations of van Vliet (1990). The application of the theoretical approach outlined above yields excellent agreement with the experiment in a large variety of samples (Tacano 1993), devices (Berntgen et al. 1999), and microstructures (Balandin et al. 1999).

4.4.2 The Coherent Quantum 1/*f* Effect

This effect arises for any electric current in matter (solids, liquids, gases) or in vacuum from the definition of the physical electron (or other charged particle) as a bare particle plus a coherent state of the electromagnetic field (Handel 1996). It is caused by the energy spread (uncertainty) in characterizing any coherent state of electromagnetic field oscillators. This energy spread causes nonstationarity, i.e., fluctuations that have a $1/f$ spectrum both in probability density and in the current.

The spectral density of these fundamental fluctuations was obtained through both derivation from the quantum-electrodynamical propagator of a physical charged particle and through an earlier elementary derivation similar to Eq. 78. The spectral

density for the particle concentration $|\psi|^2$, or for the current density $j = e(k/m)\ |\psi|^2$, is given by

$$S_{|\psi|^2}\overline{|\psi|^{-4}} = S_j\bar{j}^{-2} = \frac{2\alpha}{\pi\nu N} = 4.6 \times 10^{-3}\ \ \nu^{-1}N^{-1}. \tag{79}$$

Here, we have included the total number N of charged particles, which are observed simultaneously in order to define the current j, in the denominator, because the noise contributions from each particle are independent. This result is related to the conventional quantum $1/f$ effect introduced in the previous section.

Relation to the Conventional Quantum $1/f$ Effect The coherent state in a conductive sample is a state defined by the collective motion, i.e., by the drift current of the carriers. It is expressed in the Hamiltonian by the magnetic energy E_{mag} per unit length of the current. In very small samples or electronic devices, this coherent magnetic energy

$$E_{\text{mag}} = \int \frac{B^2}{8\pi}\ d^3x = \left(\frac{N e v S}{c^2}\right)^2 \ln(R/r) \tag{80}$$

is much smaller than the total kinetic energy E_{kin} of the drift motion of the carriers

$$E_{\text{kin}} = \sum \frac{m v^2}{2}. \tag{81}$$

Here B is the magnetic field, n the carrier concentration, S the cross-sectional area, r the radius of the cylindrical sample (e.g., a current-carrying wire), and R is the radius of the electric circuit. Introducing the "coherence ratio" $s = E_{\text{mag}}/E_{\text{kin}}$, the total observed spectral density of the mobility fluctuations can be written by a relation of the form (Handel 1994, 1996)

$$\left(1/\mu^2\right) S_\mu(\nu) = \frac{1}{1+s}\ \ \frac{2\alpha A}{\nu N} + \frac{s}{1+s}\ \ \frac{2\alpha}{\pi\nu N} = \frac{2\alpha}{(1+s)\,\nu N}(A + s/\pi), \tag{82}$$

which can be interpreted as an expression of the effective Hooge parameter a_{H} if ν and the number N of carriers in the (homogeneous) sample are brought to the numerator of the left-hand side. In this equation $\alpha\, A = 2\alpha(\Delta v/c)^2/3\pi$ is the usual nonrelativistic expression of the infrared exponent, present in the familiar form of the conventional quantum $1/f$ effect. This equation is limited to quantum $1/f$ mobility (or diffusion) fluctuations and does not include other contributions. Including all contributions finally yields the empirical Hooge parameter α_{H}.

Contact Noise A major contribution to noise can be generated at the contact interface or in the high-field region of a Schottky barrier (Çeik-Butler 1996). This *contact noise* also has $1/f$ behavior, is part of the flicker noise, and can be calculated as the sum of various types of $1/f$ noise. Two important contributions are known; with R the contact resistance, the first is given by $(1/R^2)S_R(\nu) = \pi\alpha R^3/(10n\nu\rho^3)$. It originates from a coherent quantum $1/f$ effect in metals or degenerated semiconductors of carrier concentration n and resistivity ρ, in agreement with data of Hooge (1969). The second contribution occurs from the conventional quantum $1/f$ effect in very small, less-doped semiconductors with a coherence ratio $s \ll 1$; in metals, a similar effect arises from electron-hole waves at the Fermi surface. In nanoscale devices with inhomogeneous contact areas, this contact noise can become the dominating noise.

5 Summary

After generation or excitation, carriers can return to their original state, or they can recombine radiatively or nonradiatively with another state. *Nonradiative recombination* of carriers releases energy in the form of phonons or by accelerating another electron, with Auger recombination, which consequently leads to LO phonon emission by the accelerated electron. Phonon emission occurs as a single-phonon process when trapping a carrier at a shallow defect center or as a multiphonon emission when recombination at a tightly bound deep center occurs; a sequential release of a phonon cascade is possible during the lattice-relaxation process of a deep center.

We distinguish between *carrier traps*, in which excitation into the adjacent band and trapping at the center dominate, and *recombination centers*, through which carriers recombine from one band to the other. The former are located close to one band, the latter closer to the center of the bandgap. They are separated from each other by demarcation lines, which lie near the quasi-Fermi levels and appear when deviating from thermal equilibrium. As a characteristic parameter for carrier recombination, the capture cross-section, which measures the "effectiveness" of a defect center in trapping or recombination, can be defined. The capture cross-sections spread over more than 12 orders of magnitude, ranging from 10^{-10} cm^2 for Coulomb-attractive centers at low temperatures to less than 10^{-22} cm^2 for some Coulomb-repulsion centers or for deep centers with large lattice coupling. At high carrier densities, both Auger- and plasmon-induced recombinations become effective and transfer the relaxed recombination energy to single carriers or the collective of carriers, respectively. Nonradiative recombination is one of the most important loss mechanisms in most semiconductor devices; it is almost always defect center controlled and can be reduced by avoidance of deep-level defects, low doping levels, or proper doping with luminescence centers.

Radiative recombination proceeds as an emission delayed by the lifetime of an excited state and changed in energy after relaxation of the excited state. Depending on the lifetime, luminescence and the slowlier phosphorescence are distinguished. In thermal blackbody radiation, excitation is accomplished by phonons alone. Such a

radiation has many similarities with the phonon-induced luminescence in the far-IR emission range. At higher frequencies, the intrinsic coupling to a phonon distribution causes quantitative changes in the behavior, which makes thermal radiation distinctly different from luminescence.

The spectral distribution of the luminescence offers a wide variety of methods to analyze the electronic structure of the semiconductor and its defects. The spectra of shallow-level defects in stress-free ultrapure crystals at low temperatures are extremely sharp and well understood. When lattice relaxation by phonon emission becomes involved, these spectra become broader and it is more difficult to unambiguously relate these to specific deep-defect level centers. Band-to-band excitation preceding spontaneous emission, the so-called *band-edge luminescence,* provides direct information concerning the carrier distribution near the edge. Clear luminescence signatures are obtained from excitons bound to defects and from the rich spectrum of defect pairs, which assist in unambiguous identification. Electron-hole drop and carrier plasma luminescence help at high carrier densities to identify carrier phase-transitions.

Random fluctuations of the carrier density, the local carrier mobility, and electric fields of local space charges in semiconductors cause electronic noise. When these sources are independent of each other, superposition holds, and each of the noise-frequency components is added to obtain the total noise. *Equilibrium* noise, also described as thermal noise, is independent of the frequency (*white noise*) and directly proportional to the bandwidth of the amplifier. *Nonequilibrium* noise requires optical excitation or a current and depends on doping or space-charge effects. The discreteness of elementary charges gives rise to *shot noise* in the carrier injection from electrodes to a semiconductor. Carrier generation by thermal generation or by light and annihilation by recombination or trapping cause a fluctuation in carrier density referred to as *generation-recombination noise*. Noise components usually increase with decreasing frequency with a typical $1/f$ behavior, composed of fundamental and nonfundamental contributions. The first can be considered as energy loss from low-frequency bremsstrahlung in a basically elastic scattering process, while the latter arises from the energy uncertainty in characterizing any coherent state of electromagnetic field oscillators.

Electronic noise presents a lower limit for signal detection in electronic devices and hence of reliable device operation. It is difficult to separate unambiguously the different components of nonequilibrium noise. However, avoidance of material and electrode defects has a major effect in reducing device noise.

References

Agrawal GP, Dutta NK (1993) Semiconductor lasers. Springer, Kluwer Academic Publishers, Boston

Alkauskas A, Yan Q, Van de Walle CG (2014) First-principles theory of nonradiative carriercapture via multiphonon emission. Phys Rev B 90:075202

Alkauskas A, Dreyer CE, Lyons JL, Van de Walle CG (2016) Role of excited states in Shockley-Read-Hall recombination in wide-band-gap semiconductors. Phys Rev B 93:201304

Ambrozy A (1982) Electronic noise. McGraw-Hill Book Co, New York
Ascarelli G, Rodriguez S (1961) Recombination of electrons and donors in *n*-type germanium. Phys Rev 124:1321
Bakker CJ, Heller G (1939) On the Brownian motion in electric resistances. Physica 6:262
Balandin A, Wang KL, Svizhenko A, Bandyopadhyay S (1999) The fundamental 1/*f* noise and the Hooge parameter in semiconductor quantum wires. IEEE Trans Electron Dev 46:1240
Barber HD (1967) Effective mass and intrinsic concentration in silicon. Solid State Electron 10:1039
Barmparis GD, Puzyrev YS, Zhang X-G, Pantelides ST (2015) Theory of inelastic multiphonon scattering and carrier capture by defects in semiconductors: Application to capture cross sections. Phys Rev B 92:214111
Bartram RH, Stoneham AM (1975) On the luminescence and absence of luminescence of *F* centers. Sol State Commun 17:1593
Beattie AR, Landsberg PT (1959) Auger effect in semiconductors. Proc Roy Soc A 249:16
Benoit a la Guillaume C, Lavallard P (1972) Piezoemission of GaSb: impurities and bound excitons. Phys Rev B 5:4900
Berntgen J, Heime K, Daumann W, Auer U, Tegude F-J, Matulionis A (1999) The 1/*f* noise of InP based 2DEG devices and its dependence on mobility. IEEE Trans Electron Dev 46:194
Bittel H (1959) Schwankungserscheinungen bei der Elektrizitätsleitung in Festkörpern. Erg exakt Naturwiss 31:84 (Fluctuation phenomena in the electrical conduction of solids, in German)
Blasse G, Grabmayer BC (1994) Luminescent materials. Springer, Berlin
Böer KW, Junge K (1953) Zur Frequenzabhängigkeit von Elektronenschwankungserscheinungen in Halbleitern. Z Naturforsch 8a:753 (On the frequency dependence of electron fluctuation phenomena in semiconductors, in German)
Bosman G (1981). PhD Thesis, University of Utrecht, Netherlands
Bube RH (1974) Electronic properties of crystalline solids. Academic Press, New York
Bubenzer A, Hunklinger S, Dransfeld K (1980) New method for measuring extremely low optical absorptions. J Non-Cryst Sol 40:605
Çeik-Butler Z (1996) 1/*f* noise as an electromigration characterization tool for W-plug vias between TiN/Al-Cu/TiN metallizations. Solid State Electron 39:999
Chow WW, Koch SW, Sargent M (1997) Semiconductor laser physics. Springer, Berlin/New York
Coleman JJ, Bryce C, Jagadish C (eds) (2012) Advances in semiconductor lasers. Semiconductors and semimetals, vol 86. Academic Press, San Diego
Curie D (1963) Luminescence in crystals. Wiley, New York
Das B, Aguilera I, Rau U, Kirchartz T (2020) What is a deep defect? Combining Shockley-Read-Hall statistics with multiphonon recombination theory. Phys Rev Materials 4:024602
Dean PJ (1984) Collective excitation in solids. Plenum Press, New York
Dean PJ, Cuthbert JD, Lynch RT (1969) Interimpurity recombinations involving the isoelectronic trap bismuth in gallium phosphide. Phys Rev 179:754
Dexter DL, Klick CC, Russell GA (1956) Criterion for the occurrence of luminescence. Phys Rev 100:603
Dutta P, Horn PM (1981) Low-frequency fluctuations in solids: 1/*f* noise. Rev Mod Phys 53:497
Dreyer CE, Alkauskas A, Lyons JL, Van de Walle CG (2020) Radiative capture rates at deep defects from electronic structure calculations. Phys Rev B 102:085305
Dziewior J, Schmid W (1977) Auger coefficients for highly doped and highly excited silicon. Appl Phys Lett 31:346
Etienne B (1975) Thesis. L'Ecole Normale Superieure, Paris
Gisolf JH (1949) On the spontaneous current fluctuations in semiconductors. Physica 15:825
Glinchuk KD, Prokhorovich AV, Rodionov VE, Vovnenko VI (1977) The scheme of electronic transitions via 0.94, 1.0, 1.2, and 1.3 eV radiative centres in *n*-GaAs. Phys Stat Sol A 41:659
Goldberg PG (ed) (1966) Luminescence of inorganic solids. Academic Press, New York
González T, Mateos J, Pardo D, Bulashenko OM, Reggiani L (1998) Microscopic analysis of shot-noise suppression in nondegenerate ballistic transport. Semicond Sci Technol 13:714

González T, Mateos J, Pardo D, Reggiani L (1999) Joint effect of Fermi and Coulomb correlations on shot-noise suppression in ballistic conductors. Physica B 272:285
Govoni M, Marri I, Ossicini S (2011) Auger recombination in Si and GaAs semiconductors: *Ab initio* results. Phys Rev B 84:075215
Grabert H, Devoret MH (eds) (1992) Single charge tunneling. Coulomb blockade in nanostructures. Plenum Press, New York
Gruhl H, Dorn H-P, Winzer K (1985) Calorimetric absorption spectroscopy of J-aggregate dye monolayers below 0.1 Kelvin. Appl Phys B 38:199
Gutsche E (1982) Non-Condon approximations and the static approach in the theory of non-radiative multiphonon transitions. Phys Stat Sol B 109:583
Hall RN, Fenner GE, Kingsley JD, Soltys TJ, Carlson RO (1962) Coherent light emission from GaAs junctions. Phys Rev Lett 9:366
Handel PH (1980) Quantum approach to 1/*f* noise. Phys Rev A 22:745
Handel PH (1982) Characteristic functional of quantum 1/*f* noise. Phys Rev A 26:3727
Handel PH (1994) Fundamental quantum 1/*f* noise in semiconductor devices. IEEE Trans Electron Devices 41:2023
Handel PH (1996) Coherent and conventional quantum 1/*f* effect. Phys Stat Sol B 194:393
Haug A (1972) Theoretical solid state physics. Pergamon Press, Oxford
Haug A (1980) Auger recombination with traps. Phys Stat Sol B 97:481
Haug A (1981) Auger recombination with deep impurities in indirect bandgap semiconductors. Phys Stat Sol B 108:443
Haug A (1988) Band-to-band Auger recombination in semiconductors. J Phys Chem Sol 49:599
Hayes W, Stoneham AM (1984) Defects and defect processes in nonmetallic solids. John Wiley, New York
Haynes JR, Hornbeck JA (1955) Trapping of minority carriers in silicon. II. *n*-type silicon. Phys Rev 100:606
Haynes JR, Nilsson NG (1965) Recombination radiation in semiconductors. Proceedings of the 7th symposium on radiative recombination, Paris 1964. Dunod, Paris, pp 21–31
Heitz R, Podlowski L, Böhrer J, Hoffmann A, Broser I, Bimberg D (1995) Calorimetric absorption spectroscopy of deep defects and quantum dots. Acta Physica Polonica A 88:619
Hensel JC, Phillips TG, Thomas GA (1978) The electron–hole liquid in semiconductors: experimental aspects. In: Ehrenreich H, Seitz F, Turnbull D (eds) Solid State Phys 32. Academic Press, New York, pp 87–314
Henry CH, Lang DV (1977) Nonradiative capture and recombination by multiphonon emission in GaAs and GaP. Phys Rev B 15:989
Hooge FN (1969) 1/*f* noise is no surface effect. Phys Lett A 29:139
Hooge FN, Kleinpenning TGM, Vandamme LKJ (1981) Experimental studies on 1/*f* noise. Rep Prog Phys 44:479
Iveland J, Martinelli L, Peretti J, Speck JS, Weisbuch C (2013) Direct measurement of Auger electrons emitted from a semiconductor light-emitting diode under electrical injection: Identification of the dominant mechanism for efficiency droop. Phys Rev Lett 110:177406
Johnson JB (1925) The Schottky effect in low frequency circuits. Phys Rev 26:71
Juhl A, Bimberg D (1988) Calorimetric absorption and transmission spectroscopy for determination of quantum efficiencies and characterization of ultrathin layers and non radiative centers. J Appl Phys 64:303
Kalt H (1994) The electron–hole plasma and liquid in confined semiconductor systems. J Lumin 60:262
Kapon E (1999) Semiconductor lasers, fundamentals. Academic Press, San Diego
Kayanuma Y, Fukuchi S (1984) Nonradiative transitions in deep impurities in semiconductors – study in a semiclassical model. J Phys Soc Jpn 53:1869
Kilmer J, Chenette EK, van Vliet CM, Handel PH (1982) Absence of temperature fluctuations in 1/*f* noise correlation experiments in silicon. Phys Stat Sol A 70:287

Kim S, Hood SN, Walsh A (2019) Anharmonic lattice relaxation during nonradiative carrier capture. Phys Rev B 100:041202
Kioupakis E, Steiauf D, Rinke P, Delaney KT, Van de Walle CG (2015) First-principles calculations of indirect Auger recombination in nitride semiconductors. Phys Rev B 92:035207
Knox RS (1963) Theory of excitons. Solid State Phys Suppl 5. Academic Press, New York
Kogan SM (1996) Electronic noise and fluctuations in solids. Cambridge University Press, Cambridge
Korotov AN (1994) Intrinsic noise of the single-electron transistor. Phys Rev B 49:10381
Kousik GS, van Vliet CM, Bosman G, Handel PH (1985) Quantum 1/*f* noise associated with ionized impurity scattering and electron-phonon scattering in condensed matter. Adv Phys 34:663
Kousik GS, van Vliet CM, Bosman G, Luo H-J (1989) Quantum 1/*f* noise associated with intervalley scattering in nondegenerate semiconductors. I. Analytical calculations. Phys Stat Sol B 154:713
Laks DB, Neumark GF, Hangleiter A, Pantelides ST (1988) Theory of interband Auger recombination in *n*-type silicon. Phys Rev Lett 61:1229
Lampert MA (1958) Mobile and immobile effective-mass-particle complexes in nonmetallic solids. Phys Rev Lett 1:450
Landsberg PT (1987) The band-band Auger effect in semiconductors. Solid State Electron 30:1107
Landsberg PT, Pimpale A (1976) Recombination-induced non-equilibrium phase transitions in semiconductors. J Phys C 9:1243
Landsberg PT, Willoughby AF (eds) (1978) Recombination in semiconductors. Solid State Electron 21
Levitov LS, Lesovik GB (1993) Charge distribution in quantum shot noise. JETP Lett 58:230
Li YP, Tsui DC, Heremans JJ, Simmons JA, Weimann GW (1990) Low-frequency noise in transport through quantum point contacts. Appl Phys Lett 57:774
Liefrink F, Scholten AJ, Dekker C, Eppenga R, van Houten H, Foxon CT (1991) Low-frequency noise of quantum point contacts in the ballistic and quantum Hall regime. Physica B 175:213
Lochmann W, Haug A (1980) Phonon-assisted Auger recombination in Si with direct calculation of the overlap integrals. Sol State Commun 35:553
Maiman TH (1960) Optical and microwave-optical experiments in ruby. Phys Rev Lett 4:564
Malvezzi AM (1987) Interaction of picosecond laser pulses with solid surfaces. Proc SPIE 793:49
McWorther AL (1955) 1/*f* noise and related surface effects in germanium. Lincoln lab report No 80, Boston
Mott NF (1974) Metal-insulator transitions. Barnes and Noble, New York
Nakamura A, Weisbuch C (1978) Resonant Raman scattering versus hot electron effects in excitation spectra of CdTe. Solid State Electron 21:1331
Nathan MI, Dumke WP, Burns G, Dill FH Jr, Lasher G (1962) Stimulated emission of radiation from GaAs *p-n* junctions. Appl Phys Lett 1:62
Nicolet MA, Bilger HR, Zijlstra RJJ (1975) Noise in single and double injection currents in solids (I). Phys Stat Sol B 70:9. And: Noise in single and double injection currents in solids (II). Phys Stat Sol B 70:415
Nyquist H (1928) Thermal agitation of electric charge in conductors. Phys Rev 32:110
Pellegrini B (1986) On mobility-fluctuation origin of $1/f$ noise. Solid State Electron 29:1279
Peuker K, Enderlein R, Schenk A, Gutsche E (1982) Theory of non-radiative multiphonon capture processes. Solution of old controversies. Phys Stat Sol B 109:599
Quist TM, Rediker RH, Keyes RJ, Krag WE, Lax B, McWhorter AL, Zeigler HJ (1962) Semiconductor maser of GaAs. Appl Phys Lett 1:91
Reinecke TL (1982) Electron–hole liquid condensation in semiconductors. In: Devreese JT, Peeters F (eds) Polarons and excitons in polar semiconductors and ionic crystals. Plenum Press, New York, pp 343–382

Rice TM (1977) The electron–hole liquid in semiconductors: theoretical aspects. In: Ehrenreich H, Seitz F, Turnbull D (eds) Solid State Phys 32. Academic Press, New York, pp 1–86
Reshchikov M, McNamara J, Toporkov M, Avrutin V, Morkoç H, Usikov A, Helava H, Makarov Y (2016) Determination of the electron-capture coefficients and the concentration of free electrons in GaN from time-resolved photoluminescence. Sci Rep 6:37511
Reynolds DC (1960) Temperature dependence of edge emission in cadmium sulfide. Phys Rev 118:478
Reynolds DC, Bajaj KK, Litton CW (1985) Study of shallow impurity states in compound semiconductors using high resolution photoluminescence spectroscopy. Solid State Commun 53:1061
Reznikov M, de Picciotto R, Heiblum M, Glattli DC, Kumar A, Saminadayar L (1998) Quantum shot noise. Superlattices Microstruct 23:901
Robbins DJ, Landsberg PT (1980) Impact ionisation and Auger recombination involving traps in semiconductors. J Phys C 13:2425
Ronda C (2008) Luminescence. Wiley-VCH, Weinheim
Rose A (1951) Concepts in photoconductivity and allied problems. RCA Rev 12:362
Rostworowski JA, Bergersen B (1978) Phonon replicas of the electron-hole liquid luminescence in intrinsic silicon. Solid State Commun 28:919
Schottky W (1918) Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. Ann Phys (Lpz) 57:541 (Regarding spontaneous current fluctuation in different electricity conductors, in German)
Shah J (1974) Distribution function of photoexcited carriers in highly excited GaAs. Phys Rev B 10:3697
Shah J (1978) Hot electrons and phonons under high intensity photoexcitation of semiconductors. Solid State Electron 21:43
Shah J, Leite RCC (1969) Radiative recombination from photoexcited hot carriers in GaAs. Phys Rev Lett 22:1304
Shah J, Combescot M, Dayem AH (1977) Investigation of exciton-plasma Mott transition in Si. Phys Rev Lett 38:1497
Shen J-X, Zhang X, Das S, Kioupakis E, Van de Walle CG (2018). Unexpectedly strong Auger recombination in halide perovskites. Adv Energy Mater 8:1801027
Sherif TS, Handel PH (1982) Unified treatment of diffraction and 1/*f* noise. Phys Rev A 26:596
Shionoya S (1991) A review of wide bandgap II-VI compounds. Mater Forum 15:132
Shockley W, Read WT Jr (1952) Statistics of the recombinations of holes and electrons. Phys Rev 87:835
Sikula J, Levinshtein M (2004) Advanced experimental methods for noise research in nanoscale electronic devices. Kluwer Academic Publishers, Dordrecht
Stierwalt DL, Potter RF (1967) Emittance studies. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 3. Academic Press, pp 71–90
Sumi H (1983) Nonradiative multiphonon capture of free carriers by deep-level defects in semiconductors: adiabatic and nonadiabatic limits. Phys Rev B 27:2374
Tacano M (1993) Hooge fluctuation parameter of semiconductor microstructures. IEEE Trans Electron Devices 40:2060
Thomas DG, Hopfield JJ, Frosch CJ (1965) Isoelectronic traps due to nitrogen in gallium phosphide. Phys Rev Lett 15:857
Trumbore FA, Gershenzon M, Thomas DG (1966) Luminescence due to the isoelectronic substitution of bismuth for phosphorus in gallium phosphide. Appl Phys Lett 9:4
Ulmer EA, Frankl DR (1968) Infrared emission from free carriers in germanium. Proceedings of the 9th international conference on the physics of semiconductors. Nauka, Moscow, pp 170–174
van der Ziel A (1950) On the noise spectra of semi-conductor noise and of flicker effect. Physica 16:359
van der Ziel A (1970) Noise: Sources, characterization, measurement. Prentice Hall Inc, Englewood Cliffs

van der Ziel A (1986) Noise in solid state devices and circuits. John Wiley & Sons, New York
van Kampen NG (1992) Stochastic processes in physics and chemistry. North Holland Publishing, Amsterdam
van Roosbroeck W, Shockley W (1954) Photon-radiative recombination of electrons and holes in germanium. Phys Rev 94:1558
van Vliet KM (1958) Irreversible thermodynamics and carrier density fluctuations in semiconductors. Phys Rev 110:50
van Vliet KM (1990) Quantum electrodynamical theory of infrared effects in condensed matter: II. Radiative corrections of cross sections and scattering rates and quantum 1/*f* noise. Physica A 165:126
van Vliet KM, Mehta H (1981) Theory of transport noise in semiconductors. Phys Stat Sol B 106:11
van Vliet KM, van der Ziel A (1958) On the noise generated by diffusion mechanisms. Physica 24:415
Voos M, Leheny RF, Shah J (1980) Radiative recombination. In: Moss TS, Balkanski M (eds) Handbook on semiconductors, Optical properties of solids, vol 2. North Holland Publishing Company, Amsterdam, pp 329–416
Voss RF, Clarke J (1976) Flicker (1/*f*) noise: equilibrium temperature and resistance fluctuations. Phys Rev B 13:556
Varshni YP (1967) Band-to-band radiative recombination in groups IV, VI, and III-V semiconductors (I). Phys Stat Sol B 19:459; (II) Phys Stat Sol B 20:9
Wagner J (1985) Heavily doped silicon studied by luminescence and selective absorption. Solid State Electron 28:25
Weisbuch C, Ulbrich RG (1978) Resonant interactions of LA-phonons with excitonic polaritons in GaAs. In: Balkanski M (ed) Proceedings of the international conference on lattice dynamics. Flammarion Sciences, Paris, pp 167–169
Wilken S (2020) Charge recombination in organic solar cells. In: Ren J, Kan Z (eds) Soft-matter thin film solar cells: physical processes and device simulation. AIP Publishing, Melville NY
William EW, Hall R (1978) Luminescence and the light emitting diode. Pergamon Press, New York
Yariv A (1975) Quantum electronics, 2nd edn. John Wiley & Sons, New York
Yen WM, Shionoya S, Yamamoto H (eds) Phosphor Handbook. CRC Press, Boca Raton 2007

Photoconductivity

31

Karl W. Böer and Udo W. Pohl

Contents

1 Basic Photoconductivity Processes 1300
1.1 Carrier Generation 1301
1.2 Intrinsic Photoconductivity 1306
2 Extrinsic Photoconductivity 1309
2.1 Effect of Traps on Photoconductivity 1310
2.2 Effect of Recombination Centers 1315
3 Persistent and Negative Photoconductivity 1317
3.1 Persistent Photoconductivity 1317
3.2 Negative Photoconductivity 1319
4 Summary 1323
References 1323

Abstract

Free carriers, causing an increase in electrical conductivity, can optically be generated either intrinsically by band-to-band absorption or extrinsically involving defect states in the bandgap. Photoconductivity provides information about carrier excitation and relaxation processes and hence about electronically significant imperfections. Photoconductors can be substantially sensitized by doping with slow recombination centers. An exceedingly long dwell time for carriers captured in traps may induce persistent photoconductivity. A related very small recombination cross-section occurs for deep impurities with a large lattice relaxation. Photoconductivity can be quenched (reduced) by a shift of minority carriers from predominantly slow to fast

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_31

recombination centers. Such a shift can be induced optically with additional long-wavelength light, as well as thermally or by an electric field.

Keywords

Activator · Capture cross-section · Extrinsic photoconductivity · Intrinsic photoconductivity · Negative photoconductivity · Persistent photoconductivity · Photoconductivity · Photo-ionization cross-section · Quenching · Recombination center · Sensitization · Trap

1 Basic Photoconductivity Processes

Photoconductivity denotes the increase of the electric conductivity due to an increased carrier density as a result of an optical excitation. With excitation from the valence to the conduction band, it is termed *intrinsic photoconductivity*; with excitation involving levels in the bandgap, it is called *extrinsic photoconductivity*; semiconductors that show strong photoconductivity (see below) are called *photoconductors*. Related phenomena are photoexcitation-caused changes in the Hall effect (photo-Hall effect), in the thermoelectric effect (photothermoelectric effect), and photo-induced capacitance changes of a semiconductor junction (photocapacitance); when additional carriers are photoexcited in a material containing an *internal* electric field, a current and voltage are generated (photovoltaic effect). For reviews on photoconductivity, see Bube (1992), Capasso (1990), Ryvkin (1964), and Stillman et al. (1977); perovskites are addressed by Savanije et al. (2020), methods are briefly reviewed in Brinza et al. (2005).

In this chapter, we will first review different intrinsic and extrinsic carrier-generation processes. Then, we will give an overview of a reaction-kinetic analysis, which is the basis for a quantitative description of the change in carrier distribution due to optical excitation, and discuss the effect of traps and recombination centers. Finally, we will provide typical examples for various processes of photon-induced changes in conductivity.

A general diagram for the main phenomena involved in photoconductivity is given in Fig. 1. In the bandgap, only localized energy states associated with crystal imperfections such as intrinsic defects or impurities exist. The major processes of interest are optical absorption creating free carriers, electrical transport by which free carriers contribute to the electrical conductivity, capture of free carriers leading to trapping or recombination, and thermal excitation of trapped carriers to the nearest band.

The response of a photoconductor upon irradiation with light is described by a set of equations, one for each of the electron densities in a given level or band; we will neglect a redistribution of carriers within the band and assume here that all transition coefficients are independent of such a redistribution. In a homogeneous semiconductor, the kinetic expressions are of the form

$$\frac{dn_i}{dt} = g_i - r_i, \quad i = 1, \ 2, \ \ldots, \tag{1}$$

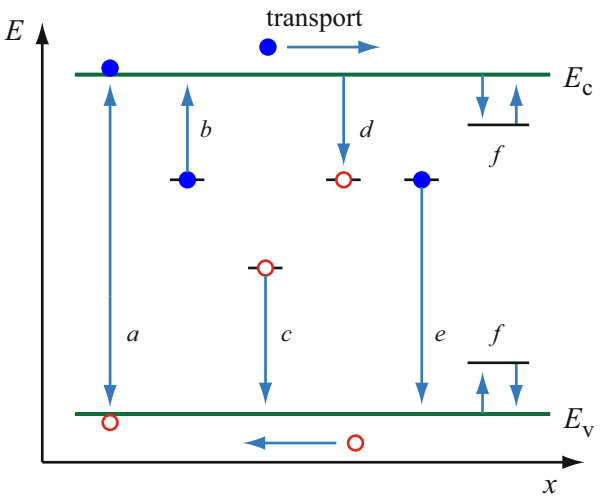


Fig. 1 Major transitions and phenomena associated with photoelectronic effects in homogeneous semiconductors: (*a*) intrinsic absorption, (*b*) and (*c*) extrinsic absorption, (*d*) and (*e*) capture and recombination, (*f*) trapping and detrapping. *Arrows* indicate electron transitions

where g_i and r_i are the generation and recombination rates for the ith energy level, respectively, defined in ▶ Chap. 29, "Carrier Generation". The generation term contains all contributions that cause an increase in the population of this level, while the recombination term combines all contributions that cause a decrease of the population. The rates can be written as

$$g_i = \sum\nolimits_j n_j e_{ji} \left(N_i - n_i\right) \text{ and } r_i = \sum\nolimits_j n_i c_{ij} \left(N_j - n_j\right) \tag{2}$$

where e_{ji} and c_{ij} are the transition coefficients between level i and level j, N_i is the density of defects with level i, and n_i is the density of electrons in the level, assuming single occupancy of each level; e transitions are upward (excitation) transitions within the band diagram, while c transitions are downward (capture) transitions with release of energy; see Fig. 1 and Fig. 1 of ▶ Chap. 29, "Carrier Generation".

Equation 1 represents a set of first-order, nonlinear differential equations, which usually cannot be solved in closed form except for some exceedingly simple cases, examples of which are given in Sects. 1.2 and 2.

1.1 Carrier Generation

Optical absorption due to band-to-band excitation is an *intrinsic photoexcitation* process creating equal densities of electrons and holes (Fig. 1a). At photon energies exceeding the bandgap energy, the optical absorption becomes large; consequently the carrier generation occurs closer to, and often within, the first 10–100 nm of the front surface. The deformed electron-energy distribution immediately after photoexcitation relaxes to approach a Boltzmann distribution, with most of the electrons near the lower edge of the conduction band and most of the holes near the upper edge of the valence band. Therefore most of the photoconductivity occurs near the band edges.

Photoconductivity with excitation energy *below* the band edge, i.e., *extrinsic photoexcitation*, is rather homogeneous throughout the volume of the semiconductor. Optical absorption results from transitions of electron-occupied imperfection states to the conduction band (Fig. 1b), producing only free electrons, or from transitions from the valence band to electron-unoccupied imperfection states (Fig. 1c), producing only free holes. Free carriers may also be produced by a ground-to-excited state excitation involving an imperfection, when a secondary process completes excitation into the band by a process such as tunneling, impact ionization, thermal ionization, or excitation by additional photons.

Optical absorption is described quantitatively through the optical absorption constant α_o. In the simplest case – that is, neglecting effects due to reflection or interference of light with intensity I_0 that is incident on a sample with thickness d and absorption constant α_o – the transmitted light intensity through the photoconductor is expressed by the *Beer-Lambert law*:

$$I = I_0 \exp(-\alpha_o d). \tag{3}$$

The magnitude of α_o for intrinsic photoabsorption is usually large compared to the thickness of the photoconductor and can be calculated by a quantum mechanical perturbation calculation in which the effect of the light is treated as a perturbation in the basic Schrödinger equation. We will briefly point out the energy dependence of α_o in the following section.

In the case of *extrinsic photoabsorption,* α_o is usually much smaller and can be simply expressed as

$$\alpha_o = s_{opt} N, \tag{4}$$

where N is the density of imperfections involved in the absorption process and s_{opt} is their photoionization cross-section, a quantity usually on the order of 10^{-16} cm^2 near the band edge, but smaller for transitions further from the band edge.

1.1.1 Wavelength Dependence of the Generation Rate

The generation rate of carriers depends on the energy of the impinging photons. *Band-to-band excitation* was discussed in Sect. 1 of ▶ Chap. 13, "Band-to-Band Transitions", with a focus on an excitation near the fundamental edge. Most intrinsic transitions are direct transitions (vertical in an $E(\mathbf{k})$ representation), which can proceed from any point in $\mathbf{k}$ space with the appropriate photon energy as illustrated in Fig. 2. The optical *generation rate* g_o introduced in Sect. 1.1 of ▶ Chap. 29, "Carrier Generation", is proportional to the impinging light flux ϕ_o (photons/cm^{2}s) normal to a sample surface (i.e., in z direction) and the optical absorption coefficient α_o, all of which are functions of the wavelength:

$$g_o(\lambda) = \eta \ \phi_o(\lambda)\, \alpha_o(\lambda) \exp(-\alpha_o z). \tag{5}$$

The *quantum efficiency* η describes the number of carriers (or e–h pairs) per absorbed photon. It is usually close to unity for photon energies below and near the bandgap energy; at very high photon energies (above $2E_g$), a gradual increase is

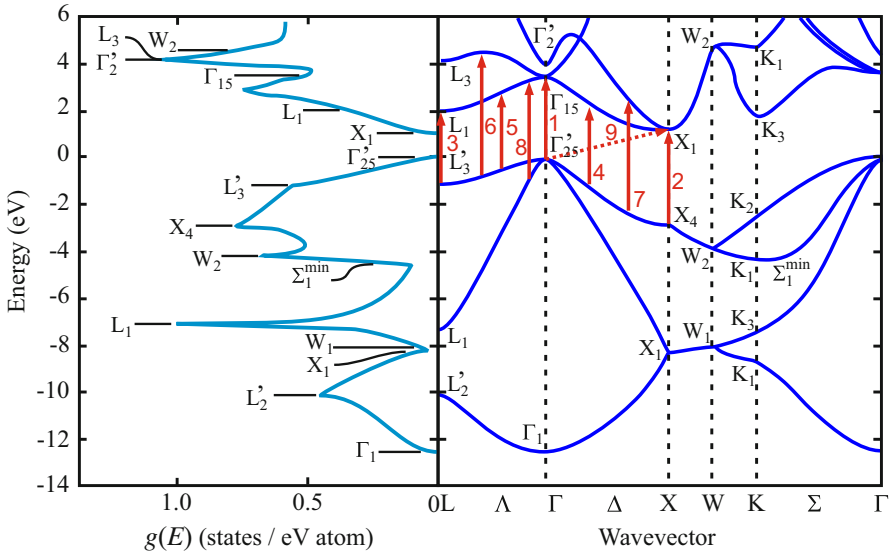


Fig. 2 Band structure and density-of-state distribution of Si with competing direct electron transitions (*arrows*) indicated at characteristic points (1, 2, 3) along $E(k)$ ranges with parallel bands and comparatively high joint density of states (4 and 5), transitions into higher conduction bands (6 and 7), and from the light-hole valence band (8). The indirect band-edge transition (9) is indicated by a *dotted arrow*

found due to impact ionization. Data of η for various photoconductors are reported by Marfaing (1980).

With high optical absorption, the reflection $r(\lambda)$ can no longer be neglected. We then need to replace Eq. 5 by

$$g_o(\lambda) = \eta \ \phi_o \alpha_o (1 - r) \frac{\exp(-\alpha_o z) + r \exp(\alpha_o z) \exp(-2\alpha_o d)}{1 - r^2 \exp(-2\alpha_o d)}, \tag{6}$$

which yields Eq. 5 for the limit of vanishing reflection r.

The optical absorption is proportional to the square of the matrix element for each of the transitions at a given photon energy and to the joint density of states between valence and conduction band $J_{c,v}$:

$$\alpha_o(\lambda) = \frac{e^2 \lambda}{6\varepsilon_0 n_r c^2 m_0^2 V} |\mathbf{M}_{fi}|^2 J_{c,v}. \tag{7}$$

Integration over the energy range, given by the specific optical excitation (e.g., a band-to-band excitation with a given spectral band width), yields the total generation rate with electrons distributed over the corresponding regions of the **k** space (see Fig. 6 of ▶ Chap. 13, "Band-to-Band Transitions"). From this deformed distribution,

the electrons will relax to approach the Boltzmann distribution, with most of the electrons near the lower edge of the conduction band and holes near the upper edge of the valence band.

Defect-to-band excitation was extensively discussed in Sects. 1.2.3 and 2.4.2 of ▶ Chap. 17, "Optical Properties of Defects"; here, again, excitations into higher states of the band need to be considered. In addition we must take the electron occupation probability f into account. If, e.g., the photo-ionization of an acceptor impurity is considered, we obtain an absorption coefficient:

$$\alpha_o = s_A N_A (1 - f), \tag{8}$$

where N_A is the acceptor concentration and s_A is the photo-ionization cross section (see below).

Excitation from deeper centers shows a wider distribution, since they are more localized in real space. Hence, their eigenfunctions are more spread out in **k** space, and they offer a broad range of vertical transitions, whereas for shallow-level centers, the direct transitions are restricted closer to $k = 0$.

Finally, an excitation of free carriers from states near the band edge to higher states influences the photoconductivity by changing the effective mass and relaxation time. Devices based on this effect are referred to as *free-electron bolometers.*

1.1.2 Photo-Ionization Cross Section

The excitation of electrons from deep-level centers into the conduction band can be described as transitions into resonant states. Such transitions are possible when the excited state contains Bloch functions of the same **k** value as the ground state (conservation of momentum). The cross section of localized deep centers extends over a large range in **k** space; it therefore extends to a wide energy range of the band continuum. This cross section can be approximated (see Sect. 2.4.1 of ▶ Chap. 17, "Optical Properties of Defects") by

$$s_o(h\nu) \propto \frac{h\nu - |E_I|}{(h\nu)^3} \; g(h\nu - |E_I|), \tag{9}$$

where E_I is the ionization energy and $g(h\nu - |E_I|)$ is the density of states, which, for a simple parabolic band near its edge, can be expressed as

$$g(h\nu - |E_I|) \propto \sqrt{h\nu - |E_I|}. \tag{10}$$

The cross section can be calculated when the deep center potential is known. When it is approximated by a square-well function (Lucovsky 1965), we obtain

$$s_o(h\nu) = \frac{1}{n_r}\left(\frac{F_{eff}}{F}\right)^2 \frac{32\pi^2 e^2 \hbar \sqrt{E_I}\ (h\nu - |E_I|)^{3/2}}{3m^* c\ (h\nu)^3}, \tag{11}$$

where F_{eff}/F is the ratio of the effective local field to the electromagnetic radiation field, which takes into consideration the local screening by charges near the defect center. This ratio is typically on the order of 1 for shallow levels; s_o is on the order of 10^{-16} cm^2 near the band edge. Further from the band edge, the expression becomes more complicated, causing a rapid reduction of the cross section (Jaros 1977). Equation 11 describes rather well the IR absorption of GaAs:Au or GaAs:Ag (Queisser 1971).

For excitation into an excited state, we obtain for the cross section

$$s_o(h\nu) = \frac{1}{n_r}\left(\frac{F_{eff}}{F}\right)^2 \frac{2\pi^2 e^2 \hbar}{3m^* e}\ f_{ba}, \tag{12}$$

with f_{ba} as the oscillator strength, which is given for a hydrogen atom in Table 1 and applies to a quasi-hydrogen (shallow) defect.

The relative cross section of different impurities can be easily obtained by a constant-photoconductivity method developed by Grimmeiss and Ledebo (1975). One can show from detailed balance[1] that the photoionization cross-section s_o is inversely proportional to the intensity $I(h\nu)$ of light required to produce, with light of different photon energies, *the same* photoconductivity:

$$s_o(h\nu) = \frac{\text{const}}{I(h\nu)}. \tag{13}$$

This assumes that, with changing photon energy, photo-ionization occurs from different deep defect levels into the same band and results in carriers with the same mobility. It also presumes that these levels do not communicate with other levels or bands.

An example of results obtained from this method is given in Fig. 3, which shows the photo-excitation from three deep impurity levels into the conduction band of GaAs:O. The deepest level occurs at an energy of 1.03 eV below the conduction-

Table 1 Oscillator strength (f_{ba}) for transitions of a hydrogen atom (After Bethe and Salpeter 1957)

$1s \rightarrow 2p$	0.4162	$1s \rightarrow 5p$	0.0139
$1s \rightarrow 3p$	0.0791	$1s \rightarrow 6p$	0.0078
$1s \rightarrow 4p$	0.0290	$1s \rightarrow$ contin.	0.436

[1]The principle of detailed balance states that in equilibrium all transitions into a level must equal all transitions out of this level between each group of two states.

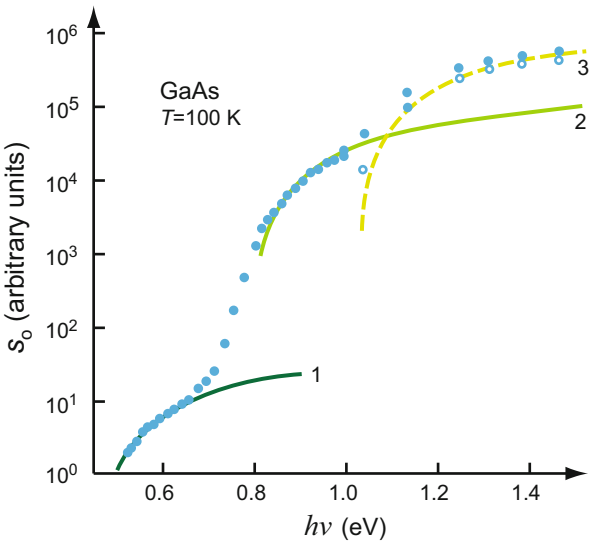


Fig. 3 Relative photo-ionization cross section for excitation of three deep impurity centers into the conduction band of GaAs (After Grimmeiss and Ledebo 1975)

band edge, with a cross section almost 10^5 times larger than the shallowest level at 0.46 eV below E_c. The middle level, with a cross section close to the deepest level, lies at 0.79 eV below E_c.

Careful checking as to whether all of the assumptions used are fulfilled is imperative. The *photo-Hall effect*[2] can be used to detect competing transitions involving the opposite carrier. For transient spectroscopy of the photo-Hall effect in semi-insulating GaAs, see Yasutake et al. (1997).

1.2 Intrinsic Photoconductivity

The density of carriers in the respective bands and in levels in the bandgap is changed from thermal equilibrium with optical excitation. These densities can be obtained from reaction-kinetic arguments. The carrier density in the respective bands is discussed for a number of simple reaction-kinetic models in order to identify the specific influence of certain defect levels.

In the following, we consider the effect of light impinging on a semiconductor on the conductivity. A semiconductor kept in the dark exhibits a *dark conductivity*:

$$\sigma_0 = |e| \ \left(n_0 \mu_n + p_0 \mu_p\right), \tag{14}$$

[2]That is, the Hall effect (Sect. 1.2.2 of ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients") measured for photogenerated majority carriers.

with carrier densities n_0 and p_0 originating from thermal generation of carriers. When light is irradiated on the semiconductor, electrons and holes are generated, and the conductivity is increased by

$$\Delta\sigma = |e| \ \left(\mu_n \Delta n + \mu_p \Delta p\right). \tag{15}$$

For intrinsic absorption, equal numbers of electrons and holes are generated: the intrinsic photoconductivity is *ambipolar*. In contrast, in the extrinsic photoconductivity with photon energies below the bandgap energy, only one kind of carriers may be generated by absorption at defects (see Sect. 2). From Eq. 15, we obtain the relative increase of conductivity:

$$\frac{\Delta\sigma}{\sigma_0} = \frac{\mu_n \Delta n + \mu_p \Delta p}{\mu_n n_0 + \mu_p p_0} = \frac{b\,\Delta n + \Delta p}{b\,n_0 + p_0}, \tag{16}$$

where $b = \mu_n/\mu_p$ is the ratio of the electron and hole mobilities.

In the following, we consider *high generation rates*. The incremental carrier densities Δn and Δp are then large compared to the thermally generated densities n_0 and p_0:

$$\Delta n = n - n_0 \cong n \text{ and } \Delta p = p - p_0 \cong p. \tag{17}$$

In the *intrinsic photoconductivity*, electrons and holes are always generated in pairs, and $n = p$; Eq. 16 hence yields for strong illumination:

$$\Delta\sigma/\sigma_0 \cong \Delta n/n_0. \tag{18}$$

When the intensity of the light is changed, the ensuing change in carrier densities is given by the difference between generation (g) and recombination (r) rates; with an optical generation rate g_o (much larger than thermal generation) we obtain

$$\frac{dn}{dt} = g - r = g_o - c_{c,v}\, np = \frac{dp}{dt}; \tag{19}$$

$c_{c,v}$ is the coefficient[3] for the transition from the conduction to the valence band (intrinsic recombination) introduced in Sect. 1.1 of ▶ Chap. 29, "Carrier Generation", In the intrinsic photoconductor $g_o = g_{v,c}$, and a similar equation as Eq. 19 applies for the kinetics of holes.

[3]The coefficients for downward transitions (capture c) or upward transitions (excitation e) used here are given in units of $cm^3\ s^{-1}$, yielding units of $cm^{-3}\ s^{-1}$ for generation or recombination rates. In literature symbols for capture or excitation are also defined differently, meaning transition probabilities (e.g., $c \times n$) measured in s^{-1}.

We consider boundary conditions with $n = g_o = 0$ for $t < t_0$ and $g = g_o$ for $t \geq t_0$, i.e., the illumination is turned on at $t = t_0$. The solution of Eq. 19 is then given by

$$n(t) = \sqrt{\frac{g_o}{c_{c,v}}} \tanh\left[\sqrt{g_o\, c_{c,v}}(t - t_0)\right]. \tag{20}$$

This function shows a nearly linear rise of n for $t_0 < t < t_0 + \tau_0$, with τ_0 as the photoconductive rise time

$$\tau_0 = 1/\sqrt{g_o\, c_{c,v}}. \tag{21}$$

According to the tanh function, the electron density approaches its *steady-state value*

$$n_{max} = \sqrt{\frac{g_o}{c_{c,v}}} \tag{22}$$

exponentially. When, after reaching a steady state, the optical excitation is switched off ($g = 0$ for $t > t_1$), the decay of the photoconductivity is described by (see Eq. 19)

$$\frac{dn}{dt} = -c_{c,v}\, n^2. \tag{23}$$

This decay of this bimolecular recombination is hyperbolic and given by

$$n(t) = \frac{1}{c_{c,v}(t - t_1)} = n_{max}\frac{\tau_1}{t - t_1} \tag{24}$$

with a decay time constant

$$\tau_1 = \left(n_{max}\, c_{c,v}\right)^{-1}. \tag{25}$$

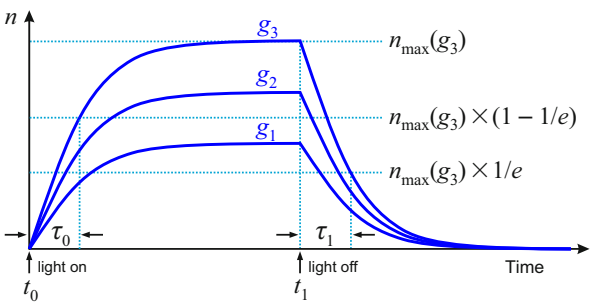


Fig. 4 Transient change of the carrier density n in intrinsic photoconductivity with the optical generation rate g as the family parameter and $g_3 > g_2 > g_1$. Rise time τ_0 and decay time τ_1 of $n(g_3)$ is indicated

Introducing n_{max} from Eq. 22, we see that both rise and decay time constants are equal:

$$\tau_1 = \tau_0. \tag{26}$$

Rise and decay of the carrier density n is illustrated in Fig. 4. Both processes proceed faster at higher light intensity and in semiconductors with larger recombination cross-sections $c_{c,v} = s_{c,v} v_{rms}$, see Sect. 1.2.2 of ▶ Chap. 29, "Carrier Generation"; the steady-state value n_{max} and the time constant τ_0 both depend on the intensity of the incident light (i.e., the generation rate g_0) and on $c_{c,v}$. Since the response of photoconductivity on switching of the illumination usually deviates from an exponential behavior, the time constant (or rise time τ_{rise} for light turned on) is specified in practice as the time required to increase the photocurrent from 10% to 90% of its final value.

2 Extrinsic Photoconductivity

Extrinsic photoconductivity involves levels in the bandgap. It is either *n*- or *p-type*. A defect center from which carrier generation occurs is called an *activator*. We can distinguish three types of extrinsic photoconductivities involving such activators:

- Excitation from an activator with direct recombination into the same type of level
- Excitation from an activator with carrier recombination through another type of level
- Excitation into a band from which major trapping occurs before recombination

Effect of Activators When generation and direct recombination involve the *same type of activator*, and this activator is separated far enough from the valence band so that thermal ionization of the optically generated hole is of minor importance, the carrier density follows the same relation as that for intrinsic carrier generation expressed by Eq. 19, with the coefficient $c_{c,v}$ replaced by the coefficient $c_{c,act}$ involving the activator state, and the density of holes in activators denoted p_{act}:

$$\frac{dn}{dt} = g_o - c_{c,act} n p_{act} = \frac{dp_{act}}{dt}. \tag{27}$$

The quasi-neutrality condition now reads $p_{act} = n$. In *steady state*, the carrier density hence attains a value:

$$n = \sqrt{\frac{g_o}{c_{c,act}}}. \tag{28}$$

The photoconductivity is then *n*-type with

$$\sigma_n = e n \mu_n. \tag{29}$$

At high intensities, depletion of these activators causes a saturation of the photoconductivity. Such a saturation has been observed, e.g., by Bube and Ho (1966) and Celler et al. (1975).

In a photoconductor which is sensitized with an activator, a small intensity of light causes a large change of carrier density. An example is an *n*-type photoconductor with doubly negatively charged activators, which become singly negatively charged when photo-ionized. Such a repulsive center has a very small recombination cross-section; in CdS a typical value of $s_{n\mathrm{rec}} \cong 10^{-22}$ cm^2 is observed, yielding $c_{\mathrm{c,act}} = s_{n\mathrm{rec}}\, v_{\mathrm{rms}} \cong 10^{-15}$ cm^3 s^{-1}; cross sections even down to 10^{-27} cm^2 were found, depending on defects related to sulfur deficiency (Yin et al. 2018). When doped with an activating recombination-center density of $N_{\mathrm{act}} = 10^{17}$ cm^{-3}, we obtain with a generation rate of 10^{18} cm^{-3}s^{-1} an increase in carrier density of $\Delta n = g_{\mathrm{o}}/(c_{\mathrm{c,act}} \times N_{\mathrm{act}}) \cong 10^{16}$ cm^{-3}. This increment is much larger than the carrier concentration in the dark n_0, which is typically below 10^{10} cm^{-3} for wide gap semiconductors. By contrast, insensitive photoconductors with a recombination cross-section of $s_{n\mathrm{rec}} \cong 10^{-13}$ cm^2 yield with the same density of centers and generation rate an increment of $\Delta n \cong 10^7$ cm^{-3}.

Sensitive photoconductors have a typical carrier lifetime $\tau_n = 1/(c_{\mathrm{c,act}} \times N_{\mathrm{act}})$ of 10^{-5} to 10^{-2} s, whereas the carrier lifetime in insensitive photoconductors is on the order of 10^{-12} to 10^{-8} s.

An example for a sensitizing center is the Zn impurity in Si. Zn introduces at room temperature a single-acceptor level at $E_{\mathrm{v}} + 0.31$ eV and a double-acceptor level at $E_{\mathrm{v}} + 0.57$ eV. When $E_{\mathrm{F}} > 0.57$ eV, the impurity is in the Zn^{2-} state and acts as a sensitizing center for *n*-type photoconductivity. At $T = 80$ K, the photoconductivity shows an optical quenching spectrum with a low-energy threshold at 0.58 eV, i.e., at the energy required to excite an electron from the valence band to the Zn^- level to form a Zn^{2-} center. At low temperature, Zn^{2-} changes to Zn^- as the result of hole capture; at these temperatures, an electron capture cross-section of 10^{-20} cm^2 is determined from the onset of optical quenching.

In the following sections, we analyze the two other types affecting extrinsic photoconductivity: traps and recombination centers.

2.1 Effect of Traps on Photoconductivity

Defects acting as traps provide levels in the bandgap, where carriers can be captured from the band and thermally be reemitted; see Sect. 2.1.1 of ▶ Chap. 30, "Carrier Recombination and Noise". When photogenerated carriers are trapped intermittently before they recombine, competing transitions exist and the carrier balance is shifted. We consider the effect of an electron trap on activated photoconductivity. The respective balance equations are

$$\frac{dn}{dt} = g_0 - \frac{dn_{\text{trap}}}{dt} - c_{\text{c, act}} n\, p_{\text{act}}, \quad \frac{dp_{\text{act}}}{dt} = g_0 - c_{\text{c, act}} n\, p_{\text{act}}, \tag{30}$$

$$\frac{dn_{\text{trap}}}{dt} = c_{\text{c,trap}}\, n \left(N_{\text{trap}} - n_{\text{trap}}\right) - e_{\text{trap,c}} N_{\text{c}}\, n_{\text{trap}}. \tag{31}$$

Here, $e_{\text{trap,c}}$ denotes thermal reemission of electrons from the trap into the conduction band. The quasi-neutrality condition now requires

$$p_{\text{act}} = n + n_{\text{trap}}. \tag{32}$$

The transitions of electrons into traps compete with transitions into the valence band.

Steady state In *steady state*, all time derivatives vanish. As a consequence, the two terms in Eq. 31 related to trapping drop out. The conditions then become identical to the balance equation 27, i.e., $g_{\text{o}} = c_{\text{c,act}}\ n\ p_{\text{act}}$. The influence of traps enters through the neutrality condition, yielding

$$g_{\text{o}} = c_{\text{c,act}}\, n \left(n + n_{\text{trap}}\right). \tag{33}$$

Deep traps tend to be completely filled, i.e.,

$$n_{\text{trap}} \cong N_{\text{trap}}. \tag{34}$$

Introducing this into Eq. 33, we obtain for the electron density:

$$n = \frac{1}{2}\left(-N_{\text{trap}} + \sqrt{N_{\text{trap}}^2 + 4\frac{g_{\text{o}}}{c_{\text{c,act}}}}\right). \tag{35}$$

For a small-optical generation rate with $g_{\text{o}} \ll N_{\text{trap}}^2 c_{\text{c,act}}/4$, Eq. 35 reduces to

$$n \cong \frac{g_{\text{o}}}{c_{\text{c, act}} N_{\text{trap}}}. \tag{36}$$

This approximation means that each electron finds a constant density of available recombination sites with $p_{\text{act}} \cong N_{\text{trap}}$. For high optical generation rates, Eq. 35 converts back to Eq. 28; p_{act} increases and becomes much larger than N_{trap}, and therefore $n \cong p$, yielding $n \propto \sqrt{g_{\text{o}}}$. Figure 5 shows the dependence of n on g_{o}, with a break between the linear and square-root branches at $n \cong N_{\text{trap}}$, permitting a determination of the trap density in a photoconductor with one dominating deep trap level.

Shallow traps have an incomplete degree of filling. Thus, Eq. 34 must be replaced by the equation obtained from Eq. 31 in steady state:

$$n_{\text{trap}} = \frac{c_{\text{c, trap}}\, n N_{\text{trap}}}{c_{\text{c, trap}}\, n + e_{\text{trap, c}} N_{\text{c}}}, \tag{37}$$

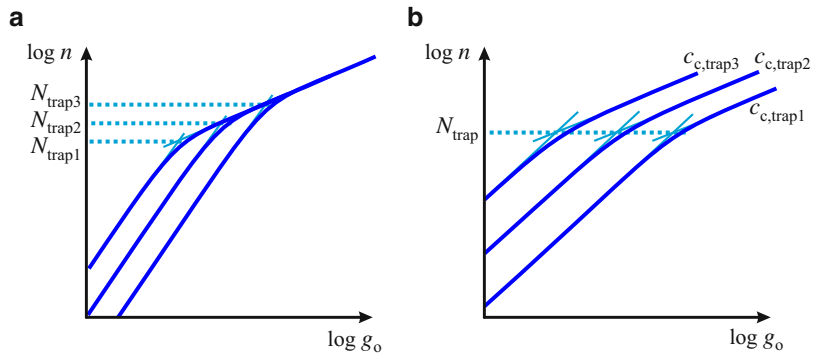


Fig. 5 Electron density n as a function of the optical generation rate with one deep trap level and (**a**) the trap density as a family parameter, or (**b**) the recombination coefficient as a family parameter

which yields Eq. 34 for high optical generation rates ($c_{c,trap}\ n \gg e_{trap,c}\ N_c$) and decreases proportionally to n for lower excitation rates:

$$n_{trap} \cong n \frac{c_{c,trap} N_{trap}}{e_{trap,c} N_c}. \tag{38}$$

The transitions rates $c_{c,trap}$ and $e_{trap,c}$ can be replaced from detailed balance arguments, yielding

$$n_{trap} \cong n \frac{N_{trap}}{N_c} \exp\left(\frac{E_c - E_{trap}}{kT}\right), \tag{39}$$

i.e., an exponential increase of electrons in traps with trap depth until saturation occurs. Then, Eq. 39 becomes identical to Eq. 34.

In the range of incomplete trap filling, we obtain by introducing Eq. 39 into Eq. 33 the relation

$$g_o = c_{c,act} n^2 \left(1 + \eta_{trap}\right), \tag{40}$$

with $\eta_{trap} = n_{trap}/n$ as an effective trap-availability factor:

$$\eta_{trap} \cong \frac{N_{trap}}{N_c} \exp\left(\frac{E_c - E_{trap}}{kT}\right); \tag{41}$$

Thus, for higher generation rates, we obtain

$$n = \sqrt{\frac{g_o}{c_{c,act}\left(1 + \eta_{trap}\right)}}. \tag{42}$$

Since $\eta_{\text{trap}} > 0$, we can see that the introduction of traps usually leads to a reduced photoconductivity[4]: part of the otherwise photoelectrically active electron population is stored in localized states, while the holes remain available for increased recombination.

Kinetic Processes When the intensity of the illumination changes, the derivatives in Eqs. 30 and 31 are finite. *Real* photoconductors require such reaction-kinetic equations for each type of level in the bandgap and their mutual redistribution. This set of equations (see Eqs. 1 and 2) can usually only be solved for exceedingly simple cases. We consider defects acting only as traps; capture $c_{\text{c,act}}$ in Eq. 30 is then replaced by $c_{\text{c,v}}$, and p_{act} in Eqs. 30 and 32 by $p = n + n_{\text{trap}}$. Similar to the intrinsic case of Fig. 4 we consider rise and decay of the photoconductivity when the light is turned on and off, and the now competing effect of electron transitions into either the traps or into the valence band. As long as traps are mostly empty, the first transition predominates; with $n_{\text{trap}} \ll N_{\text{trap}}$ we obtain

$$\frac{dn_{\text{trap}}}{dt} \cong c_{\text{c,trap}}\, n N_{\text{trap}}, \tag{43}$$

which leads to

$$\frac{dn}{dt} \cong g_{\text{o}} - c_{\text{c,trap}}\, n N_{\text{trap}} \tag{44}$$

as long as $c_{\text{c,trap}}\, N_{\text{trap}} \gg c_{\text{c,v}}\ \ p(n + n_{\text{trap}})$. After separation of variables, Eq. 44 can be solved, yielding

$$n(t) = n_{\max}\left[1 - \exp\left(-\frac{t - t_0}{\tau_{\text{trap}}}\right)\right] \text{ with } \tau_{\text{trap}} = \frac{1}{c_{\text{c,trap}} N_{\text{trap}}}, \tag{45}$$

and with τ_{trap} as the time constant for trap filling. Between carrier generation and trapping into nearly empty traps, there occurs a quasi-steady state (while the traps are being filled) with an electron density in the conduction band given by

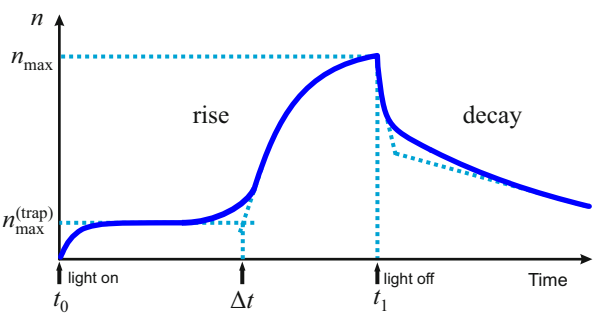


Fig. 6 Dynamics of the carrier density n in extrinsic photoconductivity. At medium intensity of the optical excitation, a plateau appears in the rise until deep traps are filled. In the decay, two slopes are indicated

[4]Except when carrier excitation occurs from filled trap levels.

$$n_{\max}^{(\mathrm{trap})} = \frac{g_\mathrm{o}}{c_{\mathrm{c,trap}} N_\mathrm{trap}}, \tag{46}$$

see Fig. 6. When the traps become filled, the trapping transition gets clogged. Consequently, n rises again until steady state between generation and recombination is reached. With $N_\mathrm{trap} \gg n$, i.e., for a low-generation rate, we obtain for the second rise

$$\frac{dn}{dt} \cong g_\mathrm{o} - c_{\mathrm{c,v}}\, n N_\mathrm{trap}. \tag{47}$$

The kinetics for trap filling is similar to the first rise of the electron density, except for a somewhat longer time constant, since usually $c_{\mathrm{c,trap}} \gg c_{\mathrm{c,v}}$, and correspondingly

$$\tau_1 = \left(c_{\mathrm{c,v}} N_\mathrm{trap}\right)^{-1}. \tag{48}$$

This behavior is presented in Fig. 6, showing first a rise to a plateau while trap filling occurs, and then another rise to reach a steady-state value at

$$n_\mathrm{max} = \frac{g_\mathrm{o}}{c_{\mathrm{c,v}} N_\mathrm{trap}}. \tag{49}$$

The ratio $n_\mathrm{max}/n_{\max}^{(\mathrm{trap})}$ can be used to determine the ratio of capture cross-sections of electron traps and band-to-band transitions:

$$\frac{n_{\max}^{(\mathrm{trap})}}{n_\mathrm{max}} = \frac{s_{\mathrm{c,trap}}}{s_{\mathrm{c,v}}}. \tag{50}$$

The length of the plateau Δt can be used to estimate the density of traps $N_\mathrm{trap} \cong g_\mathrm{o} \Delta t$, if recombination during trap filling can be neglected.

The decay after switching off the light proceeds inversely. First with a time constant given by the recombination transition obtained from

$$\frac{dn}{dt} \cong -c_{\mathrm{c,v}}\, n N_\mathrm{trap}, \tag{51}$$

which yields an exponential decay

$$n(t) = n_{\max}^{(\mathrm{trap})} \exp\left(-\frac{t - t_1}{\tau_1}\right), \tag{52}$$

where τ_1 is given by Eq. 48. Later, electrons are supplied by the emission from traps ($e_{\mathrm{trap,c}}\, N_\mathrm{c}\, n_\mathrm{trap}$), which determines the slow tail of the decay of n. The kinetics of trap filling of oxygen-related thermal donors in silicon was recently studied by the much slower *emission* curve, yielding capture cross sections for majority and minority carriers (Siriwardhana et al. 2021).

Evidence for the effect of traps on the measured carrier lifetime is also observed in spatially resolved measurements. In multicrystalline silicon, the lateral structure of the reduced excess-carrier lifetime is found to be clearly correlated to the local distribution of defects in the crystal structure (Schubert et al. 2006); simultaneous imaging of the recombination lifetime, the trap density, and the related etch-pit density of same region proves the direct relation of trap density and lifetime. In *n*-type Si the density of traps is related to interstitial oxygen and thermal donors (Hu et al. 2012). The parameters of the traps (trap density, energy level, capture cross-section, trapping time) could well be determined from a fit of kinetic data to the model reported by Hornbeck and Haynes (1955), assuming (after pulse excitation) quasisteady-state conditions with constant concentration of trapped holes; in the analysis the recombination of electrons and holes in the trap level and the emission of majority carriers from the trap level are neglected. In a more recent study indications for transient instead of steady-state conditions were observed, and the results were analyzed in terms of the actual measurement conditions (Zhu et al. 2019). Trapping effects were described also for complex defect models by applying *numerically* a system of generalized rate equations (Schüler et al. 2010).

2.2 Effect of Recombination Centers

Recombination of photogenerated carriers proceeds mostly not radiatively with the emission of photons but nonradiatively via recombination centers. The activators from which the conduction electrons are optically generated can act as such centers; for the demarcation between traps and recombination centers, see Sect. 2.1.2 of ▶ Chap. 30, "Carrier Recombination and Noise". The effectiveness of a recombination center depends on its capture cross-section s_{rec}, which is affected by the charge character before and after photoionization. Such a recombination can be very strong, e.g., into Coulomb-attractive centers, resulting in a low carrier lifetime; it may also be weak, e.g., for Coulomb-repulsive centers or centers with strong lattice coupling, resulting in a long lifetime. A photoconductor with a long carrier lifetime is called a *sensitive photoconductor*; see Sect. 2 (*Effect of Activators*) and the discussion following Eq. 58.

The rate of capture of photo-induced excess electrons Δn at a single type of recombination centers (acting as deep donors) is given by

$$\Delta n/\tau_n = c_{\mathrm{c,rec}} N_{\mathrm{rec}}\, n \left(1 - N_{\mathrm{rec}}^{(0)}/N_{\mathrm{rec}}\right), \tag{53}$$

where $c_{\mathrm{c,rec}} = s_{\mathrm{rec}} v_{\mathrm{rms}}$ is the capture coefficient of the recombination centers (see Eq. 15 of ▶ Chap. 29, "Carrier Generation"), N_{rec} is their density, τ_n is the electron lifetime, and $N_{\mathrm{rec}}^{(0)}/N_{\mathrm{rec}} \equiv f$ is the probability of finding the center in its neutral (occupied) state.

Without illumination and in thermal equilibrium, the electron capture is balanced by the thermal emission rate $g_{\mathrm{th}} = \Delta n/\tau_n$, with the electron lifetime τ_n. This condition yields

$$n \quad \frac{1-f}{f} = \nu^{-1} N_c \exp\left[-(E_{rec} - E_c)/(kT)\right], \tag{54}$$

where N_c is the effective density of states of the conduction band and ν is the dimensionless spin factor depending on the kind of the defect. If the Fermi energy coincides with the energy of the recombination center E_{rec}, Eq. 54 equals a particular electron density $\nu_{donor}^{-1} n_1$.

When light is impinging on this photoconductor, nonequilibrium conditions are established, and the recombination rate of electrons changes to

$$\Delta n/\tau_n = \left\{ n\,(1-f) - \nu_{donor}^{-1} n_1 f \right\} \quad /\tau_{n_0}, \tag{55}$$

where $\tau_{n_0} = \left(c_{c,rec} N_{rec}\right)^{-1}$. Such a recombination rate applies also for holes:

$$\Delta p/\tau_p = \left\{ pf - \nu_{acceptor} p_1 (1-f) \right\} \quad /\tau_{p_0} \tag{56}$$

with $\tau_{p_0} = \left(c_{v,rec}^{(p)} N_{rec}^{(p)}\right)^{-1}$. In a *steady-state illumination*, electrons and holes recombine in pairs, i.e., with equal quantities, yielding $\Delta n/\tau_n = \Delta p/\tau_p$. Denoting this by $\Delta n/\tau$ and using the intrinsic carrier concentration n_i for the condition of E_F coinciding with E_{rec} (index 1) and for the dark conditions (index 0), $n_i = \sqrt{n_1 p_1} = \sqrt{n_0 p_0}$, we obtain

$$\Delta n/\tau = \frac{n_0 \Delta p + p_0 \Delta n + \Delta p \Delta n}{\tau_{p_0}\left(n + \nu_{donor}^{-1} n_1\right) + \tau_{n_0}\left(p + \nu_{acceptor} p_1\right)}; \tag{57}$$

here $\Delta n = n - n_0$ and $\Delta p = p - p_0$ are the carrier concentrations in conduction and valence bands in excess of the dark concentrations, induced by the light irradiation.

At *low light intensity*, the photo-induced carrier concentrations are much smaller than the concentrations in the dark, i.e., $\Delta n = \Delta p \ll n_0$ and p_0; Eq. 57 then yields

$$\tau \cong \frac{n_0 + \Delta n + \nu_{donor}^{-1} n_1}{n_0 + p_0} \quad \tau_{p_0} + \frac{p_0 + \Delta p + \nu_{acceptor} p_1}{n_0 + p_0} \quad \tau_{n_0}. \tag{58}$$

In a p-type semiconductor with $p_0 \gg n_0$ and $\nu_{acceptor}\, p_1$, we obtain from Eq. 58 $\tau \cong \tau_{n_0} = \left(c_{c,rec} N_{rec}\right)^{-1}$; similarly we obtain in an n-type semiconductor with $n_0 \gg p_0$ and $\nu_{donor}^{-1} n_1$ from Eq. 58 $\tau \cong \tau_{p_0} = \left(c_{v,rec}^{(p)} N_{rec}^{(p)}\right)^{-1}$. In both cases, τ does not depend on the carrier densities in the dark p_0 or n_0. Note that this is in contrast to radiative recombination, where Eq. 19 yields a lifetime depending on carrier concentration. If $p_0 \gg n_0$ does not apply and p_0 in the p-type semiconductor is small and near the intrinsic carrier concentration n_i, the time constant τ gets maximum.

3 Persistent and Negative Photoconductivity

3.1 Persistent Photoconductivity

Current Continuity The generation, recombination, extraction, and replenishment of carriers require a more detailed analysis of the current within the photoconductor. When an electric field is acting, electrons generated in one volume element are carried into another one by the current drift, where they recombine. Consequently, the net change of the population in this volume element is given by the difference between birth (generation) and death (recombination) rates plus the net difference between the incoming and outgoing traffic (current density) from electrode to electrode in the direction of the electric field; therefore, we have for electron or holes

$$\frac{dn}{dt} = g_n - r_n - \frac{1}{e}\frac{dj_n}{dx} \text{ or } \frac{dp}{dt} = g_p - r_p - \frac{1}{e}\frac{dj_p}{dx}. \tag{59}$$

These are the basic equations[5] that permit an analysis of the reaction-kinetic behavior. For steady state, we obtain the *steady-state current continuity equations*:

$$\frac{1}{e}\frac{dj_n}{dx} = g_n - r_n \text{ and } \frac{1}{e}\frac{dj_p}{dx} = g_p - r_p. \tag{60}$$

In the previous sections, we assumed that only one of the photogenerated carriers is mobile and that the other is trapped. The charge neutrality requires current continuity of the photogenerated carriers; therefore, for an *n*-type photoconductor,

$$\frac{dj}{dx} = \frac{dj_n}{dx} \equiv 0. \tag{61}$$

Here, all generated carriers must recombine within the photoconductor, typically after passing several times through the external circuit. In actuality, some of the generated holes are mobile, e.g., between thermal ionization from activators and trapping in recombination centers, although with substantially different lifetimes for an *n*-type photoconductor with $\tau_p \ll \tau_n$. With any bias, some of these holes can be pulled into the electrode and recombine outside of the photoconductor in the external circuit with electrons supplied from the other electrode, thereby maintaining

[5]These relate to the basic Maxwell's equation with its condition for the conservation of electrons div $\mathbf{j} = -d\rho/dt$; for equilibrium, with $d\rho/dt \equiv 0$, it follows div $\mathbf{j} \equiv 0$. In semiconductors with the introduction of holes, we have two types of currents, j_n and j_p, and expect with $g_n = g_p$ and $r_n = r_p$ that div $(\mathbf{j}_n + \mathbf{j}_p) \equiv 0$; the sign dilemma in comparing this equation with Eq. 60 can be resolved by replacing the conventional $e = |e|$ with $-e$ for electrons and $+e$ for holes. This is the condition for the conservation of charges. In actuality, however, only a fraction of the electrons and holes are mobile, others are trapped and do not contribute to the currents while participating in the total neutrality account.

neutrality. This renders $dj_n/dx \neq 0$. In order to make the total current divergency-free, it requires a finite dj_p/dx of the same magnitude and opposite sign, yielding

$$\frac{dj_n}{dx} + \frac{dj_p}{dx} \equiv 0. \tag{62}$$

This situation is negligible in homogeneous semiconductors, but plays a major role in structures containing space-charge regions.

Persistent Photoconductivity An interesting consequence of current continuity in photoconductors with activators of extremely small capture cross-section is a *persistent photoconductivity* (Lang et al. 1979). At low temperature, photoconductivity has been observed which persists for days or even months after the optical excitation is terminated. The initially generated carriers are extracted from the photoconductor and are repeatedly replenished. The current continues in order to maintain neutrality, since recombination is extremely slow. A prominent example is a deep donor in $Al_xGa_{1-x}As$ with a thermal depth of 0.1 eV and an optical depth of 1.3 eV, thus with a giant lattice relaxation of 1.2 eV, which shows a recombination cross-section of $s_n < 10^{-30}$ cm^2 at 77 K (Lang and Logan 1977; Northrop and Mooney 1991).

Such extremely small recombination cross-sections are difficult to explain with Coulomb-repulsive centers. With large impurity-lattice coupling, however, a very large relaxation, e.g., for some *DX* centers (Sect. 2.7 of ▶ Chap. 19, "Deep-Level Centers"), is possible when the charge state of the impurity is changed; see Fig. 7. Such a relaxation causes a large effective barrier for carrier capture, which cannot be overcome at low temperatures by thermal excitation. The

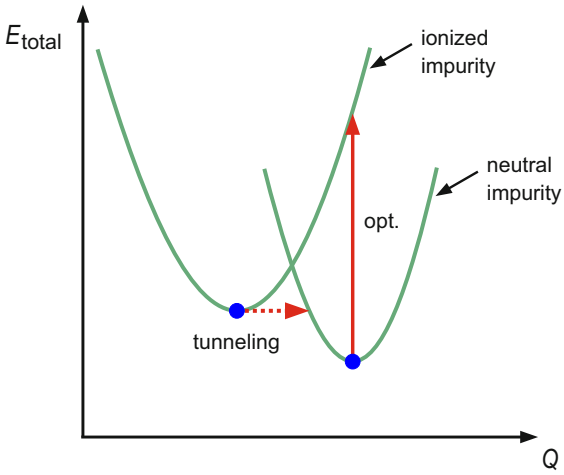


Fig. 7 Configuration-coordinate diagram for an impurity with a large lattice coupling, accounting for an extremely small recombination cross-section

tunneling probability through this barrier is sufficiently small to explain the exceedingly small recombination cross-section; for a review, see Bhattacharya and Dhar (1988).

First-principles electronic structure calculations indicate that such large lattice relaxations easily occur for anion vacancies in II-VI and chalcopyrite Cu-III-VI_2 semiconductors, inducing a charge-dependent metastable behavior similar to the DX center. A change of the charge state shifts the energy of the localized defect state beyond the band edge of the host and creates a resonant delocalized state; this induces, e.g., persistent *n*-type photoconductivity in *n*-ZnO,[6] and *p*-type photoconductivity in *p*-$CuInSe_2$ (Lany and Zunger 2005). Such *n*-type photoconductivity was experimentally also observed for CdS, being correlated with sulfur deficiency (Yin et al. 2018); photogenerated minority carriers are trapped at the intrinsic V_S defects, and a large lattice relaxation provides a kinetic barrier to recombination of free electrons in the conduction band.

The photo-induced persistent conductivity depends on the dose of irradiated photons. Measurements of the Hall effect in *n*-type GaAs:Cr show a linear increase of the electron concentration with the cumulative flux of above-bandgap photons irradiated on a sample, which was kept in the dark after illumination (Queisser and Theodorou 1979; Theodorou et al. 1982). After optical electron-hole pair generation, the deep Cr traps immobilize the holes and thereby preclude recombination. The number of remaining free electrons is hence proportional to the number of electron-hole pairs up to a maximum saturation level, provided the holes remained trapped. In CdS, a power law $\sigma \propto I^{\alpha}$ is found for the dependence of steady-state photoconductivity σ on light intensity I, with α (0.3 . . . 0.7) depending on temperature and on intensity (Yin et al. 2018). Such a behavior can be induced by a combined effect of linear and additional quadratic recombination.

3.2 Negative Photoconductivity

Negative photoconductivity is defined as a *decrease* in conductivity with additional light. It can be observed when two light beams of different wavelengths are applied and the second beam causes a reduction in carrier lifetime. The reduction of photoconductivity is called *quenching.*

The reduction in majority-carrier lifetime is induced by a redistribution of minority carriers over recombination centers of different capture cross-sections. This redistribution can also be induced by thermal or field ionization. Consequently, we can distinguish among *optical*, *thermal*, and *field quenching.* All quenching transitions can be regarded as a *desensitization,* i.e., a shift of minority carriers from slow

[6]Calculations yield a relaxed neutral oxygen vacancy V_O^0 in ZnO with a shallow occupied *s*-like donor state in the bandgap, and a broad unoccupied resonance in the conduction band for the charged V_O^{2+} vacancy.

to fast recombination centers. Therefore, quenching can only be observed in sensitized photoconductors.

3.2.1 Optical Quenching

The quenching transition is, for an *n-type* photoconductor, induced by the optical excitation of holes from the slow recombination centers into the valence band. Although in a sensitized photoconductor these holes tend to fall back into the Coulomb-attractive slow recombination centers, their population can be significantly reduced when the quenching light is sufficiently intense. The average dwell time τ_2 introduced for traps in Sect. 3.1 is hence substantially reduced. The fraction of holes that is captured by *fast* recombination centers then causes an increase in the recombination of electrons through these centers and thereby results in a decrease in photoconductivity.

The optically induced desensitization is illustrated in Fig. 8. An optical excitation with consequent sensitizing (S) is shown on the left side of Fig. 8a. The recombination occurs predominantly through slow centers (denoted *slow*) as illustrated in the middle. On the right side of this figure, the optical quenching transition (Q) with a long wavelength λ_2 is shown: it removes holes from slow centers and thereby reactivates some recombination traffic through fast centers (*fast*) illustrated at the right. The spectral distribution of the optical quenching is given in Fig. 8b. The first beam of light at a constant wavelength λ_1 produces a photocurrent j_1. The second beam with variable wavelength causes an increase of the photocurrent near λ_1, yielding the photocurrent j_2, and a decrease with a pronounced minimum below the photocurrent j_1 at λ_2, corresponding to the transition into the slow recombination center.

An example for photoconductivity quenching of an *n*-type GaN/SiC layer ($n = 2 \times 10^{18}$ cm^{-3}) is given in Fig. 9. A beam of radiation (λ_1) with a near-bandgap photon energy of 3.4 eV generates strong photoconductivity; a simultaneously irradiated second beam with smaller photon energy (λ_2), which provides

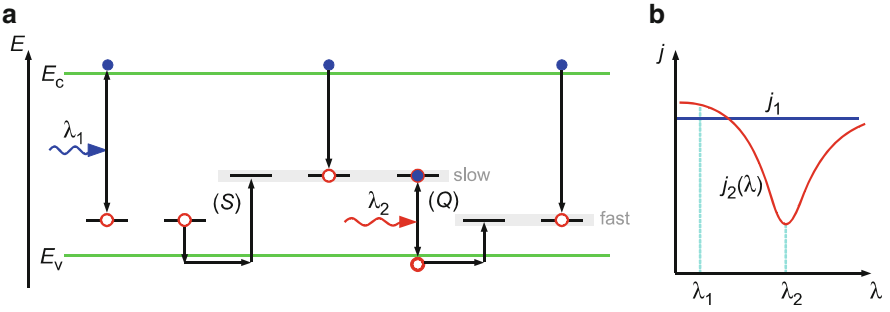


Fig. 8 (**a**) Sensitizing (S) and quenching (Q) transitions in a sensitized photoconductor. (**b**) Corresponding spectral distribution of the photoconductor excited with a second light beam of variable wavelength λ (*red line*) while the first light beam is kept constant; the *horizontal blue line* indicates the photocurrent j_1 excited with the first beam only

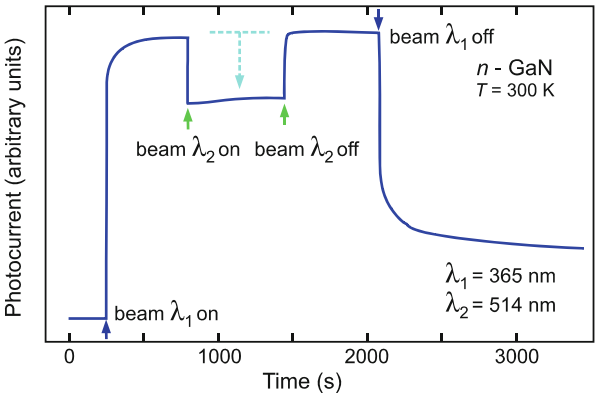


Fig. 9 Quenching of the photoconductivity of *n*-type GaN subjected to the excitation of a UV beam at λ_1 and temporarily an additional green beam at λ_2 (After Ursaki et al. 2003)

solely extrinsic excitation, produces a pronounced reduction of the photoconductivity. The quenching is assigned to the presence of hole traps with a broad distribution of the density of states peaking at E_v + 1.2 eV, tentatively attributed to gallium vacancies (Ursaki et al. 2003).

In some photoconductors, e.g., ZnS and CdS, the sensitizing center has more than one level; optical quenching can be observed from both of these levels at low temperatures, while at higher temperatures only the higher-energy transition is observed since holes from the lower level are already thermally depleted. An example is the Cu center in CdS, which forms a deep donor with an excited hole state; see Grimmeiss et al. (1981).

3.2.2 Thermal Quenching

The quenching transition can be thermally initiated. Sensitizing (S) and desensitizing (quenching, Q) steps are both thermal ionization steps and depend exponentially on the temperature. With little ionization from slow centers, holes are stored there until they are eliminated by recombination with electrons. When they are ionized faster than they can recombine with electrons, quenching is observed. A shift of the hole population occurs toward fast centers. This emphasizes the competitive nature of the slow and fast recombination processes and indicates the light-intensity dependency of thermal quenching.

Thermal quenching is shown in Fig. 10 as a steep decrease in photocurrent with increasing temperature. Thermal quenching needs higher temperatures to become dominant for higher-optical excitation.

3.2.3 Field Quenching

Since sensitizing centers are Coulomb attractive, they can be ionized by relatively low fields in the 10 kV/cm range due to the Frenkel-Poole effect as schematically illustrated in Fig. 11a; see also Sect. 2.1 of ► Chap. 29, "Carrier Generation". As a result, a reduction in photosensitivity can be observed when the recombination

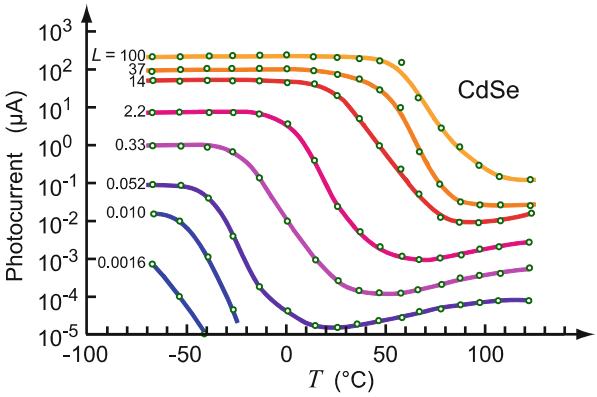


Fig. 10 Thermal quenching of the photocurrent measured at 10 V bias, shown as a steep reduction with increasing temperature as observed for CdSe; the light intensity L in relative units is the family parameter (After Bube 1992)

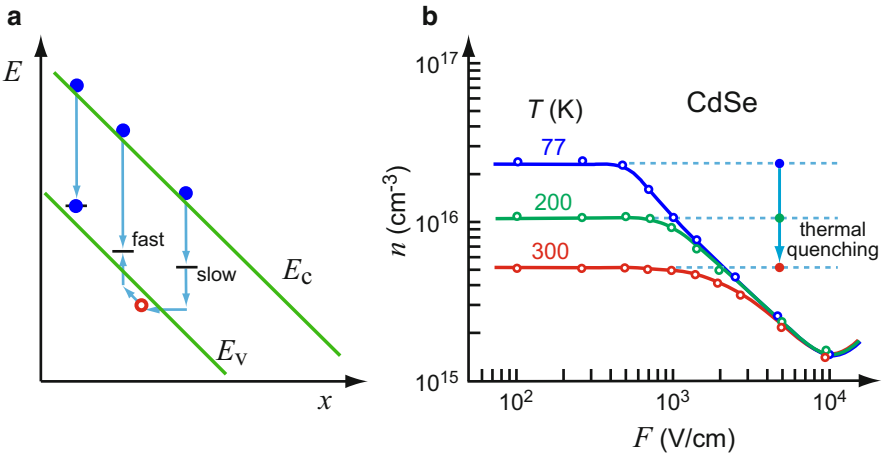


Fig. 11 (**a**) Schematic of the Frenkel-Poole effect causing field-quenching by a depletion of slow centers. (**b**) Dependence of the electron density n on the applied electric field F in CdSe with the temperature as the family parameter; the sample is irradiated by an intense beam of white light (After Yodogawa et al. 1973)

traffic is shifted from slow to fast recombination centers by the action of the electric field. If this field quenching is strong enough at sufficient light intensity and fields, a decrease in the photocurrent with increased bias is found (Dussel and Böer 1970).

The effect of the electric field on the electron density in the conduction band of a CdSe layer is shown in Fig. 11b (Yodogawa et al. 1973). At low electric field, the electron density decreases at decreased temperature due to thermal quenching. At higher electric fields the field-induced quenching leads to an additional decrease of the electron density; this effect is more pronounced at lower temperature.

4 Summary

Photoconductivity is initiated by the generation of free carriers either directly by photon-absorption from band-to-band or bound-defect states or indirectly with the assistance of phonons either for matching momenta or to supply additional energy. The resultant photoconductivity can be used to detect electromagnetic radiation (light, IR) and to measure the successful carrier generation to obtain information about the excitation process: the joint density-of-states distribution, the symmetry consideration distinguishing allowed and forbidden transitions, photon capture cross-section, and lifetime of the excited state. With excitation into higher states of the band, the relaxation of carriers and recombination from different energy distributions can be investigated.

By doping with slow recombination centers, photoconductors can be dramatically sensitized, although, usually at the expense of time response. A long dwell time of carriers in traps may induce persistent photoconductivity, being preserved for days or even months after termination of the optical excitation. The required exceedingly small recombination cross-section occurs for deep impurities with large lattice relaxation, providing large barriers for thermal excitation.

Photoconductivity can be quenched (reduced) by a shift of minority carriers from predominantly slow to fast recombination centers. Such a shift can be induced optically by a simultaneous irradiation with long wavelength light. In addition, quenching may be initiated thermally or by an electric field.

References

Bethe HA, Salpeter EE (1957) Mechanics of one and two-electron atoms. Springer, Berlin

Bhattacharya PK, Dhar S (1988) Deep levels in III-V compound semiconductors grown by molecular beam epitaxy. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 25. Academic Press, Boston, pp 143–228

Brinza M, Willekens J, Benkhedir ML, Emelianova EV, Adriaenssens GJ (2005) Photoconductivity methods in materials research. J Mater Sci Mater Electron 16:703

Bube RH (1992) Photoelectronic properties of semiconductors. Cambridge University Press, Cambridge UK

Bube RH, Ho CT (1966) Laser saturation of photoconductivity and determination of imperfection parameters in sensitive photoconductors. J Appl Phys 37:4132

Capasso F (ed) (1990) Physics of quantum electron devices. Springer, Heidelberg

Celler GK, Mishra S, Bray R (1975) Saturation of impurity photoconductivity in *n*-GaAs with intense YAG laser light. Appl Phys Lett 27:297

Dussel GA, Böer KW (1970) Field-enhanced ionization. Phys Status Solidi B 39:375

Grimmeiss HG, Ledebo L-Å (1975) Spectral distribution of photoionization cross sections by photoconductivity measurements. J Appl Phys 46:2155

Grimmeiss HG, Kullendorff N, Broser R (1981) Photocapacitance studies of CdS:Cu. J Appl Phys 52:3405

Hornbeck JA, Haynes JR (1955) Trapping of minority carriers in silicon. I. *p*-type silicon. Phys Rev 97:311

Hu Y, Schøn H, Øyvind N, Øvrelid EJ, Arnberg L (2012) Investigating minority carrier trapping in *n*-type Cz silicon by transient photoconductance measurements. J Appl Phys 111:053101

Jaros M (1977) Wave functions and optical cross sections associated with deep centers in semiconductors. Phys Rev B 16:3694

Lang DV, Logan RA (1977) Large-lattice-relaxation model for persistent photoconductivity in compound semiconductors. Phys Rev Lett 39:635

Lang DV, Logan RA, Jaros M (1979) Trapping characteristics and a donor-complex (DX) model for the persistent-photoconductivity trapping center in Te-doped $Al_xGa_{1-x}As$. Phys Rev B 19:1015

Lany S, Zunger A (2005) Anion vacancies as a source of persistent photoconductivity in II-VI and chalcopyrite semiconductors. Phys Rev B 72:035215

Lucovsky G (1965) On the photoionization of deep impurity centers in semiconductors. Solid State Commun 3:299

Marfaing Y (1980) Photoconductivity, photoelectric effects. In: Moss TS, Balkanski M (eds) Handbook on semiconductors, vol 2. Optical properties of solids. North Holland Publishing, Amsterdam, pp 417–495

Northrop GA, Mooney PM (1991) Confirmation of large lattice relaxation of the DX center by extended photo-ionization cross-section measurements. J Electron Mater 20:13

Queisser HJ (1971) Deep Impurities. In: Madelung O (ed) Festkörperprobleme/Advances in solid state physics, vol 11. Vieweg, Braunschweig, pp 45–64

Queisser HJ, Theodorou DE (1979) Hall-effect analysis of persistent photocurrents in *n*-GaAs layers. Phys Rev Lett 43:401

Ryvkin SM (1964) Photoelectric effects in semiconductors. Consultants Bureau, New York

Savenije TJ, Guo D, Caselli VM, Hutter EM (2020) Quantifying charge-carrier mobilities and recombination rates in metal halide perovskites from time-resolved microwave photoconductivity measurements. Adv Energy Mater 10:1903788

Siriwardhana M, Zhu Y, Hameiri Z, Macdonald D, Rougieux F (2021) Photoconductance determination of carrier capture cross sections of slow traps in silicon through variable pulse filling. IEEE J Photovoltaics 11:273

Schubert MC, Riepe S, Bermejo S, Warta W (2006) Determination of spatially resolved trapping parameters in silicon with injection dependent carrier density imaging. J Appl Phys 99:114908

Schüler N, Hahn T, Schmerler S, Hahn S, Dornich K, Niklas JR (2010) Simulations of photoconductivity and lifetime for steady state and nonsteady state measurements. J Appl Phys 107:064901

Stillman GE, Wolfe CM, Dimmock JO (1977) Far-infrared photoconductivity in high purity GaAs. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 12. Infrared detectors II. Academic Press, New York, pp 169–290

Theodorou DE, Queisser HJ, Bauser E (1982) Profiling of deep impurities by persistent photocurrent measurements. Appl Phys Lett 41:628

Ursaki VV, Tiginyanu IM, Ricci PC, Anedda A, Hubbard S, Pavlidis D (2003) Persistent photoconductivity and optical quenching of photocurrent in GaN layers under dual excitation. J Appl Phys 94:3875

Yasutake K, Kakiuchi H, Takeuchi A, Yoshii K, Kawabe H (1997) Deep level characterization in semi-insulating GaAs by photo-induced current and photo-Hall effect transient spectroscopy. Bull Jpn Soc Precis Eng 63:264

Yodogawa Y, Shimizu K, Kanamori H (1973) Field quenching in photoconductive CdSe films. Jpn J Appl Phys 12:711

Yin H, Akey A, Jaramillo R (2018) Large and persistent photoconductivity due to hole-hole correlation in CdS. Phys Rev Materials 2:084602

Zhu Y, Juhl MK, Coletti G, Hameiri Z (2019) Reassessments of minority carrier traps in silicon with photoconductance decay measurements. IEEE J Photovoltaics 9:652

Dynamic Processes

32

Karl W. Böer and Udo W. Pohl

Contents

1 Carrier Transit and Relaxation 1326
1.1 Transit Effects in Carrier Transport 1327
1.2 Relaxation of Carriers 1333
1.3 Recombination in Electron-Hole Plasmas and Liquids 1348
2 Phonon and Exciton Kinetics 1349
2.1 Relaxation of Phonon Distributions 1350
2.2 Exciton Kinetics 1356
3 Relaxation of the Spin Momentum 1364
3.1 Measurement of Spin Relaxation 1365
3.2 Spin-Relaxation of Free Electrons in Bulk Semiconductors 1367
3.3 Spin-Relaxation of Free Carriers and Excitons in Quantum Wells 1369
4 Summary 1372
References 1374

Abstract

When an external parameter such as an electric field or an optical generation rate is changed as a function of time, carriers in the semiconductor respond on this disturbance by a redistribution controlled by relaxation times. Relaxation proceeds by elastic or inelastic scattering with carriers, phonons, defects, or spin momenta, and respective time constants range from femtoseconds to years.

Relaxation of injected carriers is given by the carrier lifetime and related to their diffusion or drift length. Nonthermal excess energy of hot carriers is transferred to the lattice mostly by optical phonons. At high carrier density also plasmons, and at

Karl W. Böer: deceased.

U. W. Pohl (✉)
Institut für Festkörperphysik, EW5-1, Technische Universität Berlin, Berlin, Germany
e-mail: pohl@physik.tu-berlin.de

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0_32

high carrier-generation rates and low lattice temperature, condensation into electron-hole droplets with evaporation into excitons are involved. Optical phonons, excited by fast carriers or by an IR light pulse, relax their momenta by elastic scattering with phonons in the same branch, or by a decay into acoustical phonons.

Relaxation of excitons created by nonresonant optical excitation proceeds by inelastic scattering, eventually yielding radiative recombination for momenta near the zone center. The rise time in the luminescence after pulse excitation is controlled by the balance to uncorrelated electron-hole pairs. Resonantly excited excitons show a fast rise in the coherent regime and an exponential decay with an observed time constant depending on excitation density.

Carrier spin and orbital momenta are coherently aligned by excitation with polarized light. The subsequent relaxation can be detected by the degree of polarization of the radiative recombination. Holes in semiconductors with degenerate valence bands at the zone center have short spin-relaxation times in the sub-ps range; lifting this degeneracy slows relaxation down. Electrons have usually longer spin-relaxation times, limited by various mechanisms. In an exciton with weak electron-hole interaction the spin-relaxation time of the sequential spin flip of electron and hole is given by the slower particle, while at stronger interaction the faster simultaneous spin flip occurs.

Keywords

Carrier cooling · Carrier heating · Dephasing · Dispersive transport · Electron-hole plasma · Energy relaxation · Exciton relaxation · Hot carriers · Lattice temperature · Lifetime · Momentum relaxation · Phonon relaxation · Polarization · Recombination · Relaxation time · Scattering processes · Spin relaxation · Trapping

1 Carrier Transit and Relaxation

Relaxation Times in Dynamic Processes In previous chapters usually stationary conditions were assumed in the consideration of structural, optical, and electronic properties. There are, however, also various *kinetic* effects in semiconductors; these are effects in which any of the semiconductor variables change with time. These could be related to defect densities, carrier densities and mobilities, optical absorption, luminescence, or to the polarization. In order to induce such changes, external parameters are changed as a function of time, such as an external bias, magnetic field, optical generation rate, temperature, pressure, or other electromagnetic or particle irradiations.

Five major systems with vastly different relaxation times can be distinguished:

- Creation of atomic lattice defects
- Carrier-scattering relaxation
- Phonon relaxation
- Carrier redistribution in defect levels
- Electron spin relaxation

The corresponding relaxation times span time constants from geologic times down to the femtosecond range. In the relaxation of carrier distribution after *scattering* we have to distinguish relaxation times referring to momentum and to energy relaxation. In the case of *optical* excitation the incident electromagnetic field creates a polarization, which is initially in phase with the field; this phase coherence is destroyed by scattering processes and by recombination. There is hence a phase-relaxation time in addition to the relaxation time of the optical excitation.

The external changes of parameters are induced as part of the operation of a semiconducting device, e.g., the bias or changing light intensity, or as a means of obtaining specific information about certain semiconducting parameters. In this chapter, a few examples of both will be presented.

1.1 Transit Effects in Carrier Transport

We can divide kinetic studies of carriers into carrier relaxation within or between energy bands, and transit through a device. In addition, global effects, which deal with all carriers, and the detailed analysis of the carrier distribution have to be distinguished. First, we will analyze the transit of carriers from electrode to electrode, which causes a specific transient behavior of injected current pulses and reveals transport and trapping properties. In Sect. 1.2 we will discuss changes in carrier distribution within the band, which follow specific excitation pulses. The ensuing relaxation usually is analyzed by optical means, and reveals insight into the various relaxation mechanisms.

When electrons are injected at a certain position of the semiconductor, and a bias voltage is applied, the electrons drift toward the anode. Their *drift velocity* can be determined by observing the increase in current as the drifting carrier cloud arrives at the anode – see Sect. 1.1.2. This drift is slowed down by an intermittent trapping and release of these carriers, and can be used to obtain information about a number of typical relaxation mechanisms in the semiconductor.

When a bias pulse is applied and there is no trapping in a homogeneous semiconductor, the corresponding current pulse rises without delay, following the change in the electric field. However, if there is a space-charge region (a barrier at an electrode or a *pn* junction), the changes in field distribution require a change in carrier distribution in the space-charge region (without trapping); this change follows with a delay given by the dielectric relaxation time[1] τ_σ.

[1]The dielectric relaxation time takes polarization effects in the current after changing the bias into account. In absence of traps this quantity is given by $\tau_\sigma = \varepsilon\varepsilon_0/\sigma$. Simple polarization effects have a very short time constant; with σ in the 10^{-4} to $10^{+2}\,\Omega^{-1}\,\mathrm{cm}^{-1}$ range for typical semiconductors, τ_σ is on the order of 10^{-8} to 10^{-14} s.

1.1.1 Characteristic Transport Times and Lengths

When electrons are injected and drift in an electric field F without being trapped until they reach the anode at a distance Δx, we observe a *transit time*

$$\tau_{\text{trans}} = \frac{\Delta x}{\mu F}, \tag{1}$$

which is usually much longer than the dielectric relaxation time τ_σ. The injected electrons have a *lifetime* in the conduction band

$$\tau_n = \frac{1}{c_{\text{c,rec}}(N_{\text{rec}} - n_{\text{rec}})}, \tag{2}$$

where $N_{\text{rec}} - n_{\text{rec}}$ is the density of unoccupied recombination centers, and $c_{\text{c,rec}}$ is the recombination coefficient for the transition from the conduction band into the recombination centers. When this lifetime is smaller than the transit time, these carriers do not reach the electrode, but travel for a distance of the *drift length*

$$L_{\text{drift}} = \mu_n F \tau_n. \tag{3}$$

In absence of an electric field F, carriers diffuse in a random walk during their lifetime from the position of injection to a distance given by the *diffusion length*

$$L_n = \sqrt{\frac{\mu_n kT}{e}\,\tau_n}. \tag{4}$$

With an electric field applied, the distance traveled can be obtained by solving the transport and continuity equations for a homogeneous semiconductor with constant electric field, and is given by Smith (1978)

$$L_{n\,(\text{u,d})} = L_n \frac{2L_n}{\sqrt{4L_n^2 + L_{\text{drift}}^2} \pm L_{\text{drift}}}, \tag{5}$$

where $L_{n(\text{u})}$ and $L_{n(\text{d})}$ are the *upstream* and *downstream diffusion lengths*, for which the sign in the denominator is positive or negative, respectively. This modified diffusion length is determined by an electric field opposed to, or in the direction of, the diffusion of the carriers. Consequently, the diffusion profile is compressed or stretched as shown in Fig. 1. For high electric fields ($L_{\text{drift}} \gg L_n$), it yields for the *upstream diffusion length*

$$L_{n\,(\text{u})} = \frac{L_n^2}{L_{\text{drift}}}, \tag{6}$$

and for the *downstream diffusion length,* it yields the drift length (Eq. 3)

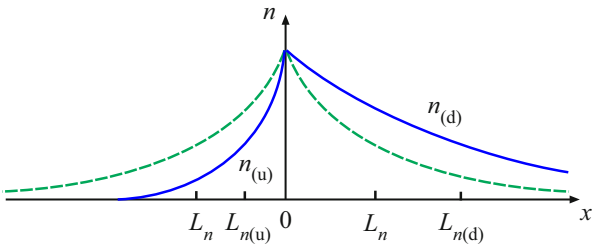


Fig. 1 Diffusion profile (schematic) of carriers injected at $x = 0$ in absence of an electric field ($F = 0$, *dashed curve*), and for a sufficiently large electric field F with upstream diffusion at the left and downstream diffusion at the right (*solid line*)

$$L_{n\,(\mathrm{d})} \cong L_{\mathrm{drift}}. \tag{7}$$

These distances are often compared to the characteristic length of a space-charge region, the *Debye length* L_D (Eq. 95 of ▶ Chap. 10, "Interaction of Light with Solids"). When caused by a cloud of free carriers, L_D can be written in a fashion similar to the diffusion length, but with the dielectric relaxation time τ_σ replacing the carrier lifetime:

$$L_\mathrm{D} = \sqrt{\frac{\mu k T}{e}\tau_\sigma} = \sqrt{\frac{\varepsilon \varepsilon_0 k T}{e^2 n}} = 1205\,\text{Å} \times \sqrt{\frac{\varepsilon}{10}\,\frac{T}{300\,\mathrm{K}}\,\frac{10^{15}\mathrm{cm}^{-3}}{n}}. \tag{8}$$

Often, however, the space charge is due to trapped carriers n_trap. The carrier density n in Eq. 8 must then be replaced by n_trap. The length that injected carriers can travel in a given time is given by the *drift velocity* $v_\mathrm{d} = \mu_n F$, with

$$L_n = \frac{v_\mathrm{d}}{\tau_n}; \tag{9}$$

the same drift velocity is used to describe the current:

$$j = e\,n\,v_\mathrm{d}. \tag{10}$$

This is justified when we consider that only while in the conduction band do the electrons contribute to the current. The density of these electrons in steady state is n. In the bulk of a semiconductor, quasineutrality forces the establishment of a constant electron density, which is given by the detailed balance between trapping and reemission of trapped carriers into the band.

When *intermittent trapping* and subsequent release from traps occur, the time required for a certain group of injected carriers to traverse a given distance becomes longer, since it includes the time of resting in the traps. This traversing time can be

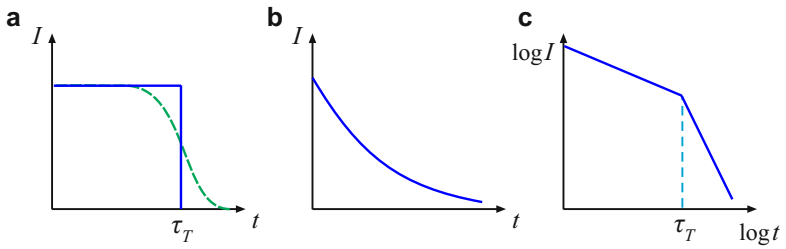


Fig. 2 Shape of a current transient caused by a pulse of carriers injected near one electrode. (**a**) Ideal train of carriers with a Gaussian spread (*dashed curve*) without trapping, (**b**) exponential response with carrier trapping in a time shorter than the transit time, (**c**) dispersive bilinear logarithmic response with a high degree of carrier dispersion

obtained by kinetic experiments – see below. Here, we account for this delayed arrival by defining a *modified drift velocity* according to

$$\tilde{v}_\mathrm{d} = \mu_\mathrm{d} F, \tag{11}$$

where μ_d is the *drift mobility,* which includes trapping.

Current Kinetics with Traps Trapping and carrier release from traps modify the current kinetics of injected carriers. When one type of trap is present, trapping and release time are represented by the time constant τ_1 (Eq. 26 of ▶ Chap. 31, "Photoconductivity"); an injected carrier pulse decays exponentially with τ_1.

When a trap *distribution* is present, a distribution of time constants determines the decay, which becomes *dispersive.* There are three types of current decays, which are shown in Fig. 2: one is box-like, with Gaussian spread shown as a dashed curve (see also Fig. 4 for carrier transport without trapping); another one is a simple exponential for a single trap level, with a well-defined decay time τ_1; still another one is dispersive, and usually has two straight-line segments in a double logarithmic plot with slopes <1 and >1 below and above the break, respectively – see Fig. 5 (Jonscher 1983; Tiedje 1984).

1.1.2 Shockley-Haynes-Type Experiment

The modified drift velocity, which includes trapping, can be measured when the movement of an injected carrier cloud can be followed. Haynes and Shockley (1951) used the injection of a carrier cloud from a point contact into a long semiconducting sample (Fig. 3), and detected, with an electric field F applied, the delayed arrival of this carrier cloud after a transit time $t_\mathrm{trans} = t_2 - t_1 \cong t_4 - t_3$. The drift mobility is then derived from

$$\mu_\mathrm{d} = \frac{d}{t_\mathrm{trans} F}. \tag{12}$$

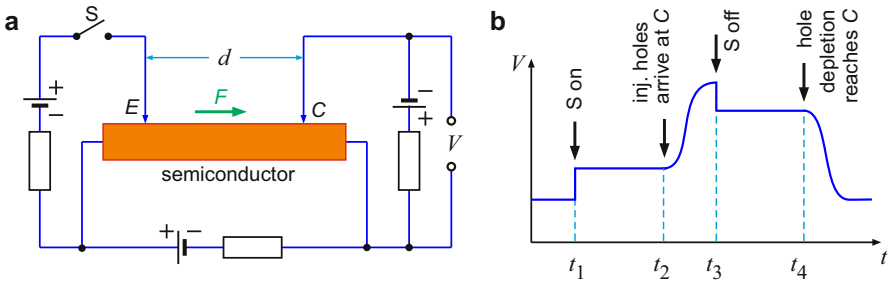


Fig. 3 (**a**) Shockley-Haynes experiment to measure at *point C* the drift time of minority carriers injected at *position E* in a long semiconductor sample. (**b**) Schematic representation of the voltage $V(t)$ when the *switch S* is closed at time t_1 and opened at time t_3

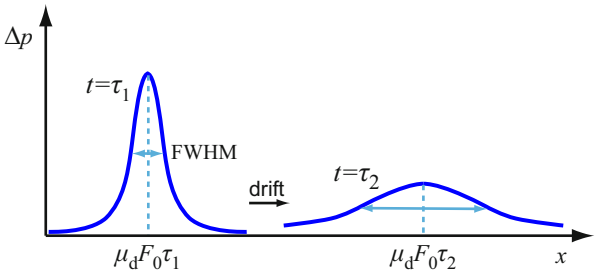


Fig. 4 Diffusion profile for minority carriers injected at $x = 0$, after elapsed time intervals τ_1 and τ_2

The solution of transport and continuity equations for Shockley-Haynes-type experiments can be written for the injected electron pulse as

$$n(t) = \frac{n_0 \exp(-t/\tau_n)}{\sqrt{4\pi D t}} \exp\left(-\frac{\Delta x^2}{4Dt}\right), \tag{13}$$

with $\Delta x = \mu_\mathrm{d} F t$, and the halfwidth of the Gaussian pulse can be written as

$$\mathrm{FWHM} = 2\sqrt{\ln 2}\sqrt{4Dt} = 3.33\sqrt{Dt}; \tag{14}$$

this permits us to determine the diffusion constant $D = \mu_\mathrm{d}\, k\, T/e$, the drift mobility, and the minority carrier lifetime τ_n. The spreading of the pulse due to out-diffusion of carriers from the original confines within the narrow injected cloud is shown in Fig. 4.

1.1.3 Dispersive Carrier Transport in Amorphous Semiconductors

A wide distribution of a large density of traps in amorphous semiconductors causes a highly dispersive type of carrier transport; see Sect. 4 of ▶ Chap. 28, "Carrier Transport Induced and Controlled by Defects". This can be measured by the current kinetics induced by a light pulse absorbed in a thin near-surface layer of a photoconductive platelet. The carriers are then driven by an external electric field across

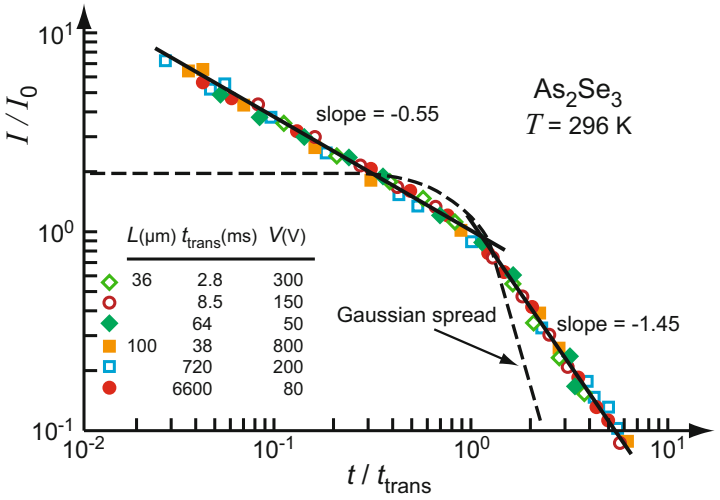


Fig. 5 Dispersive transport of injected hole transitions in amorphous As_2Se_3. Plots are obtained by parallel shifting of the curves along the time and current axis to match the kinks at (1,1). The kink identifies the transit time t_{trans}, which is listed together with the sample width L and bias V in the table insert of the figure. The *dashed curve* gives the expected Gaussian spread for a transit time of 6600 ms (After Pfister and Scher 1977)

the sample to the opposite electrode, in an arrangement similar to that shown in Fig. 3. In the example of amorphous As_2Se_3 given in Fig. 5 the measured current is represented by two slopes on the order of -0.5 and -1.5 in the log j versus log t diagram. These slopes remain unchanged, with changes in temperature (not shown), electric field, or sample thickness.

The dispersive carrier transport in an amorphous semiconductor originates from the exponentially tailing distribution of traps. Here, carriers that were generated near one electrode are rapidly trapped while diffusing toward the other electrode. With repeated retrapping, deeper traps will become progressively filled at the expense of the shallow traps.

It is observed that a drift-current pulse decays after injection according to Scher and Montroll (1975):

$$j(t) = \begin{cases} t^{\alpha-1} & \text{for } t < t_{trans} \\ t^{-\alpha-1} & \text{for } t > t_{trans} \end{cases}, \tag{15}$$

where α is the dispersion parameter and t_{trans} is the transit time. This behavior can be explained by a progressive redistribution of carriers in an exponential trap distribution (Tiedje 1984). The dispersion parameter is given by

$$\alpha = \frac{T}{T_c}, \text{with} \quad N(E) = N_0 \exp\left(-\frac{E}{kT_c}\right), \tag{16}$$

and provides a means to determine the slope of the trap distribution in the range where it is exponential.

In some amorphous semiconductors such as α-Se a transition between dispersive and nondispersive carrier transport becomes visible in a certain temperature range, where the trap distribution is not exponential and only few traps are active (Pfister 1976). A unified description of dispersive and nondispersive transport regimes in disordered and crystalline semiconductors in terms of kinetic equations with fractional derivatives is given by Sibatov and Uchaikin (2009).

1.2 Relaxation of Carriers

Free carriers in thermal equilibrium follow the Fermi-Dirac distribution within their respective bands. They are coupled with phonons, which, as bosons, follow Bose-Einstein statistics. In equilibrium, both systems are described by the *same* temperature. The average momentum of the two systems is zero; the average energy is related to the common temperature, e.g., $3kT/2$ per quasi-particle for a Boltzmann ensemble.

When external forces such as a field, light, etc., are applied, the thermal equilibrium is disturbed, thermal conditions of the subsystems become diverse, and the distributions can no longer be described by exact thermodynamic distribution functions. Depending on the nature of the forces, either the electronic or phononic subsystem is disturbed primarily, while the other subsystem reacts by their mutual coupling: the electron-phonon scattering. This interaction has a dominating influence on the carrier or phonon transport determining the respective mobilities. This was discussed for steady-state conditions in Sect. 3.2 of ▶ Chap. 5, "Phonon-Induced Thermal Properties", and in ▶ Chap. 23, "Carrier Scattering at Low Electric Fields", and ▶ 24, "Carrier Scattering at High Electric Fields". However, the addition of a multitude of effects often makes an unambiguous analysis in steady state difficult. A kinetic study permits differentiation of the involved subsystems when they have substantially different relaxation times. In addition, steady-state experiments relating to carrier transport do not permit a direct analysis of the distribution of carriers as a function of energy. Instead, an average of the contribution of *all* electrons is obtained and expressed by the carrier mobility, from which only indirect conclusions about the distribution function can be deduced.

Relaxation studies deal with an analysis of the distribution functions of quasi-particles of the different subsystems after external forces are switched off. The momentum and energy relaxation of electrons or holes will be discussed first. We need to distinguish two ranges, similar to the steady-state consideration in ▶ Chap. 23, "Carrier Scattering at Low Electric Fields": one of a small perturbation, with the system not far removed from thermal equilibrium; and one of a large perturbation, somewhat similar to the discussion of warm and hot electrons in ▶ Chap. 24, "Carrier Scattering at High Electric Fields".

1.2.1 Momentum Relaxation of Electrons

Under the influence of an external electric field F, the conduction-electron gas is moved through the semiconductor with a momentum given by the product of its effective mass and the drift velocity $\mu_n F$. When the field is switched off, this motion

of the entire electron gas will relax to zero with a characteristic time, the *momentum-relaxation time.* The momentum relaxation occurs through all types of scattering processes (see ▶ Chap. 23, "Carrier Scattering at Low Electric Fields"), i.e., the same types of scattering that are responsible for the steady-state distribution; however, not necessarily in the same sequence of importance. For instance, in Si high-energy electrons are preferably scattered by LO phonons or as optical intervalley scattering. When equilibrium is approached, the dominant scattering processes involve preferably acoustic deformation-potential scattering and piezoelectric scattering. At the end of the relaxation process, the momenta of the electrons are distributed equally in all directions, and the electron gas has come to a standstill relative to the lattice.

A simple method of measuring this relaxation, wherein we observe the decay of a *current* after the electric field (bias) is switched off, runs into difficulties because of the extremely short relaxation times, which are on the order of $\tau_m \cong \lambda/v_{rms} \cong 10^{-13}$ s (=100 femtoseconds), with λ as the mean free path. Therefore, we better work in the frequency domain than in the time domain, and measure the complex conductivity at low IR frequencies. Here, the semiconductor responds to a harmonic excitation, i.e., to an electromagnetic sinusoidal wave, and the time- and frequency-dependent behaviors are related to each other by the Fourier transformation.

The response can be analyzed by using the Boltzmann equation (Sect. 4 of ▶ Chap. 22, "Carrier-Transport Equations"),

$$\frac{\partial f}{\partial t} = -e\mathbf{F}\ \frac{\partial f}{\partial \mathbf{k}} - \frac{f - f_0}{\tau_m}, \tag{17}$$

which, with $\mathbf{F} = \mathbf{F}_0 \exp(i\ \omega\ t)$ and the ansatz $f = f_0 + \delta f \exp(i\ \omega\ t)$, yields

$$\delta f = -\frac{1}{i\,\omega}\left(e\mathbf{F}\ \frac{\partial}{\partial \mathbf{k}} f_0 \frac{-\delta f}{\tau_m}\right). \tag{18}$$

With $\partial f_0/\partial \mathbf{k} = v\, f_0/(k\ T)$ (see Sect. 4.3 of ▶ Chap. 22, "Carrier-Transport Equations"), we obtain the solution

$$\delta f = \frac{e\,(\mathbf{F}_0 \cdot \mathbf{v})\tau_m}{kT(1 + i\,\omega\tau_m)} f_0, \tag{19}$$

which yields for the current (see Eq. 108 of ▶ Chap. 22, "Carrier-Transport Equations")

$$\mathbf{j} = \sum_{\mathbf{v}} \frac{e^2\, \mathbf{v}\mathbf{F}_0 \cdot \mathbf{v}\tau_m}{kT(1 + i\,\omega\tau_m)} f_0 \exp(i\,\omega t) = \sigma(\omega)\, \mathbf{F}_0 \exp(i\,\omega t) \tag{20}$$

with a frequency-dependent conductivity. Applying the arguments of Sect. 4.4 in ▶ Chap. 22, "Carrier-Transport Equations" for averaging, we obtain for the complex conductivity

$$\sigma(\omega) = \frac{2}{3}\frac{ne^2}{m_n kT}\left[\left\langle\frac{E\tau_m}{1+\omega^2\tau_m^2}\right\rangle - i\omega\left\langle\frac{E\tau_m^2}{1+\omega^2\tau_m^2}\right\rangle\right], \tag{21}$$

showing that, for frequencies for which $\omega \ \tau_m \cong 1$, the electrons are no longer in phase with the electric field. This results in a dispersion of the dielectric constant (see Sect. 1.1 of ▶ Chap. 10, "Interaction of Light With Solids"),

$$\varepsilon(\omega) = \varepsilon_L - \frac{ne^2}{m_n}\left(\frac{3}{2}kT\right)^{-1}\left\langle\frac{E\tau_m^2}{1+\omega^2\tau_m^2}\right\rangle, \tag{22}$$

where ε_L is the dielectric constant of the lattice without free electrons. Here, $\varepsilon(\omega)$ shows a frequency dependence in the $\omega \geq \tau_m^{-1}$ range, i.e., for far IR radiation with $\lambda \leq 0.1$ mm, which can be measured. For reviews, see Conwell (1982) and Jonscher (1983).

This analysis assumes an independence of τ_m on E. Usually, this is not fulfilled, and significant errors are encountered when scattering is dominated by ionized impurities, where a substantial $\tau_m(E)$ dependence exists. Errors are small, however, when acoustic or optical phonon scattering predominates (Nag 1975).

Other means of measuring momentum-relaxation times relate to the width of a cyclotron-resonance peak, indicative of damping; also magneto-resistance or Faraday rotation measurements can be employed (Nag 1984). Further means evaluate the optical phase of a differential transmitted or reflected radiation (Funk et al. 2009), or the dephasing of the coupled LO phonon-plasmon mode (Chang and Gwo 2007). The observed values for the momentum-relaxation time in typical semiconductors lie in the 1 ps and 0.1 ps ranges at liquid nitrogen and room temperatures, respectively.

Ballistic Carrier Transport When the dimensions of a semiconductor device become comparable to the mean free path between scattering events, or the time of observation becomes comparable to the relaxation time, there is a *drift overshoot* of the carrier velocity if the applied electric field is high enough, as shown by Monte-Carlo calculation (Maloney and Frey 1977) and given in Fig. 6.

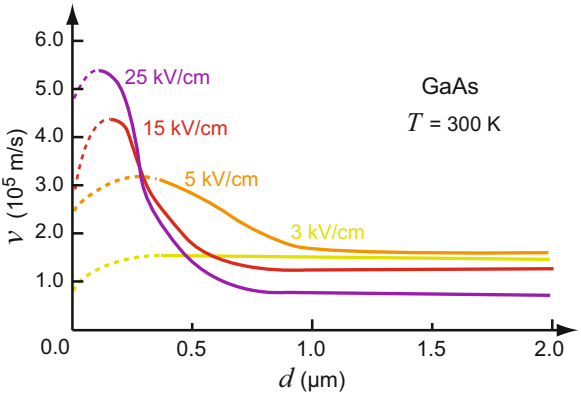


Fig. 6 Drift-velocity overshoot as a function of the platelet thickness with the (homogeneous) electric field as the family parameter; obtained by Monte-Carlo calculation for GaAs at 300 K (after Ruch 1972)

For times and distances that are *shorter* than those between scattering events, a *ballistic transport* occurs, which can be compared to the electron transport in a vacuum diode (Shur and Eastman 1981): the electron is accelerated in the electric field without scattering until it hits the anode. The current-voltage characteristic can be calculated similarly to that of a vacuum diode in a space-charge-limited case, using $j = e\ n\ v$ with $m_n v^2/2 = e\ V$, and V as the bias voltage. Integrating the Poisson equation (Eq. 79 of ▶ Chap. 22, "Carrier-Transport Equations") with $\rho = e\ [(n_0 - n(x)]$, we can approximate the current by

$$j_n \cong \frac{4}{9}\sqrt{\frac{2e}{m_n}}\frac{\varepsilon\,\varepsilon_0}{l^2}V^{3/2}, \tag{23}$$

where l is the length of the ballistic region.

Ballistic electron transport accounts for the improved high-frequency performance of small devices at high electric fields. In addition, it permits observation of quantum-mechanical interference phenomena, since the phase relation is not disturbed by scattering. Therefore, maxima and minima of the current flow are observed when the bias, and thereby the energy of the ballistic electrons, is varied. Ballistic transport was confirmed in GaAs by Heiblum et al. (1985) and may occur over mesoscopic length scales in the tenth of μm range (Gupta et al. 2021). For a review, see Eastman (1982). For diffraction of ballistic electrons in semiconductor gratings see Henderson et al. (1993). The distinction between ballistic and diffusive regimes is discussed in Sect. 3.1 of ▶ Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors".

Transient Carrier-Transport Analysis The transport dynamics in the ballistic overshoot regime are no longer described by the classical Boltzmann equation, which assumes simultaneous response of the carriers to an applied force (Barker and Ferry 1980). On a short time scale, there are memory effects from path to path. Retardation effects due to nonzero collision duration (Kreuzer 1981) must also be taken into consideration.

One conceptual approach is to replace the mean additional velocity

$$\langle \Delta v \rangle = \frac{e}{m_n} F\,t \tag{24}$$

with a more appropriate time-dependent velocity function

$$v(t) = \frac{e}{m_n} F \int_0^t \phi_v(t', \theta)\ dt', \tag{25}$$

by introducing an autocorrelation function ϕ_v, which describes the variation of v in time (Zimmermann et al. 1981). This velocity is linear in time only if $\phi \cong$ const. We can then use the corresponding balance equations

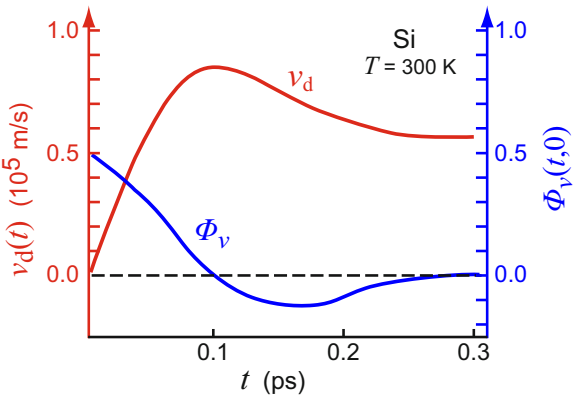


Fig. 7 Velocity autocorrelation function ϕ_v and drift velocity v_d as a function of the time after switching on the bias. Computed for an ensemble in Si at 300 K applying the Monte-Carlo method (After Ferry and Barker 1981)

$$m_n \frac{dv}{dt} = eF - m_n \int_0^t X_v(t')\, v_d(t - t')\ dt' \tag{26}$$

and

$$\frac{d\overline{E}}{dT} = eF\, v(t)[1 - \phi_v(t,0)] - \int_0^t \left[\overline{E}(t - t') - E_0\right] X_E(t')\ dt', \tag{27}$$

where X_E and X_v are the decay functions, which are intimately related to the energy and velocity autocorrelation functions ϕ_E and ϕ_v. After integration over a sufficiently long time, the time integrals over X_E and X_v yield the corresponding relaxation times τ_e and τ_m.

The correlation function is determined by a number of effects, including the influence of the electric field on the duration of each collision (Ferry 1980; Barker 1980) and on bandgap renormalization with injection of a large density of free carriers (Ferry 1978). A bandgap narrowing is determined by the self-energy of electron-hole pairs (Haug and Schmitt-Rink 1985), and by a change in phonon frequencies due to free-carrier bond weakening[2] (Brooks 1955).

The typical behavior of the velocity autocorrelation function is shown in Fig. 7, as calculated by a Monte-Carlo approach for Si (Ferry and Barker 1981; Ferry 1991). The resulting transient drift velocity shows the velocity overshoot when the correlation function decreases in time; for more detail, see the review of Ferry et al. (1984).

1.2.2 Energy Relaxation of Electrons

In contrast to the momentum relaxation, there is no direct means of measuring the *energy* relaxation of hot electrons directly from carrier transport. The average energy of carriers cannot be measured directly; it can be obtained indirectly from parameters that are a function of the carrier energy, such as the mobility. More direct information

[2]When light of sufficient energy is absorbed in a semiconductor, free carriers or excitons are produced; this reduces the bond strength of the lattice atoms from which ionization took place and changes related lattice parameters (i.e., elastic stiffness). The changes observed in mechanical and thermal properties are small, since only a very small fraction of the bonds are involved.

can be obtained from optical transmission, reflection, luminescence, or Raman scattering experiments after carrier heating in a pulsed electric field (Bauer 1978) or after an optical excitation pulse (Hearn 1980). Optical techniques allow for relaxation studies into the femtosecond range (Shah and Leheny 1984, Rullière 2005), and are reviewed in the following section.

Picosecond Spectroscopy Traditional electronic methods are limited in their high-frequency resolution in the 100 ps range. This range can be extended to the 100 fs range (in this time a light pulse travels only 10^{-3} cm) by optical means, with light pulses of this duration produced by mode-locked lasers (Valdmanis et al. 1985). These laser pulses have a large natural line width in the 10 meV range, requiring a laser cavity that permits an optical gain in a rather large band of frequencies. If the loss or gain of the cavity is modulated in this frequency range, an optical *wave packet* can be created, which produces a train of light pulses of extremely short duration (Laubereau and Kaiser 1974).

Detection of light signals in the picosecond range requires fast shutters (*Kerr cells*) or a *streak camera.* In the latter, the light is focused onto an image converter, and the electrooptical image is swept by electronic means across the screen with a speed near light velocity. The streak image, when it passes a narrow slit, is then read by a videcon coupled to a multichannel analyzer, yielding the light intensity as a function of time and photon energy. A streak camera covers the range from micro-seconds down to a few picoseconds; the direct *electronic sequencing analysis*, however, is limited in time resolution by the electronic circuit.

The detection limit can be extended from some (tens) of picoseconds down to a few femtoseconds by applying correlation techniques and nonlinear optical means. The pulse of the excitation laser can be analyzed by autocorrelation: the pulse is split into two parts of half intensity by a beam splitter, and one part is variably delayed as illustrated in Fig. 8a. Both beams are then focussed on a nonlinear crystal, which produces the second harmonic. Only while both pulses are simultaneously present within the mixing crystal is the harmonic produced, which can then be detected by a properly tuned light sensor. The amplitude of this signal is proportional to the intensity product of both beams while overlapping. The *pulse shape* of a longer pulse can be analyzed by using a shorter pulse as a gate, and varying its time delay by variation of the relative optical path length (Fig. 8b), as shown by Mahr and Hirsch (1975). The nonlinear crystal produces a sum frequency of the two beams that can be detected as a function of the mutual time delay.

In another setup the temporal evolution of a luminescence or a transmitted signal is measured by correlating this signal with the trigger pulse of a Kerr cell as shown in Fig. 8c. The crossed polarizers and a Kerr cell in between transmit the signal only when the trigger pulse induces a birefringence in the Kerr cell; the pulse shape is analyzed by shifting the delay of the trigger pulse.

Periodic pulse repetition permits detection above background noise. For reviews, see von der Linde (1979) and Shah (1999).

Optical Studies of Carrier Distribution Information about the occupation of states within the bands can be obtained by using a high-intensity optical excitation from valence-band to conduction-band states, and consequent scanning of the optical absorption spectrum, which yields the energy distribution of carriers in

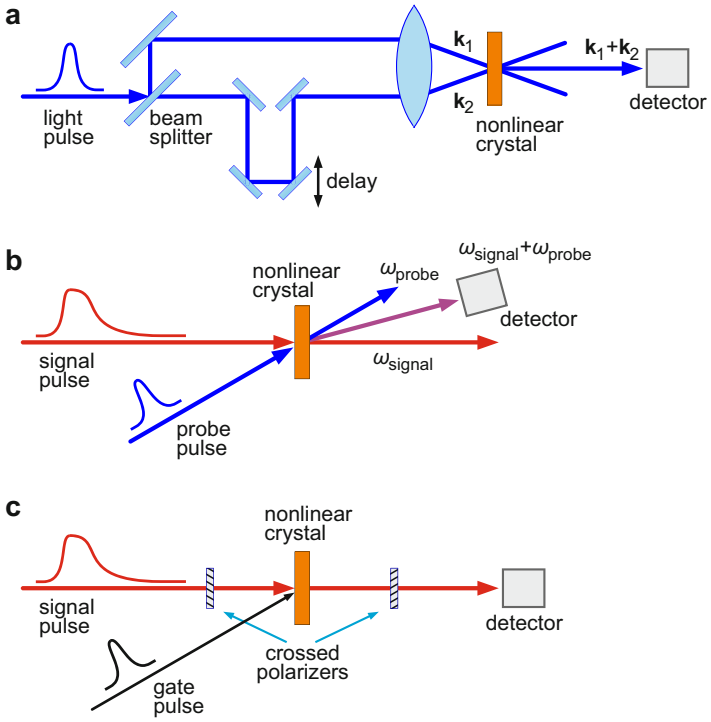


Fig. 8 Signal analysis using nonlinear optics in (**a**) autocorrelation, (**b**) upconversion by correlating signal and probe pulses, (**c**) gating with an optical Kerr-cell shutter. The detector in (**a**) and (**b**) is sensitive only to the mixed signal

these bands when significant filling is attained. Such a filling causes a bleaching of the corresponding optical absorption. The kinetics of such a distribution is then observed by picosecond laser spectroscopy.

Optical absorption spectra yield information on the evolution of the carrier distribution near the band edge, from which the carrier relaxation can be derived (Haug and Koch 1990). An example is given in Fig. 9 for GaAs. It shows three major features that are clearly visible in the difference spectra shown in curves *b* to *e*, compared to the initial absorption given in curve *a*:

1. 1 ps after optical excitation with a 500 fs pulse, we observe an increase in absorption below the bandedge energy ($E < 1.52$ eV) because of a renormalization of the gap, i.e., a decrease of the bandgap energy due to exchange and correlation effects of the high-density free-carrier plasma (Sect. 2.3 of ▶ Chap. 8, "Bands and Bandgaps in Solids"), as summarized by Arya and Hanke (1981), and Vashishta and Kalia (1982). These new states are rapidly filled by relaxing electrons; consequently, this enhanced absorption below the original band edge vanishes after 10 ps (curves *b* and *c*).

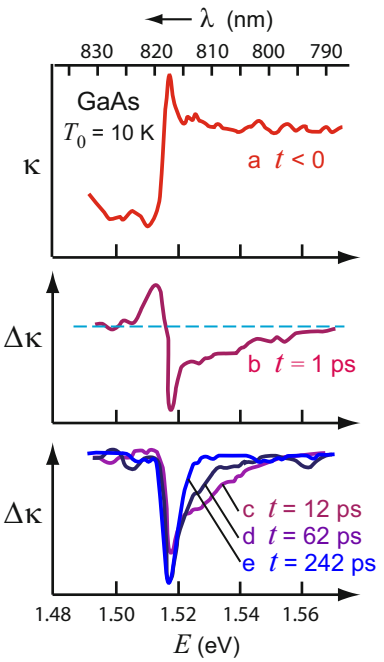


Fig. 9 (**a**) Optical absorption spectrum of GaAs near the band edge at 10 K during excitation with a 500 fs broadband light pulse, and difference spectra after various delays in subfigures (**b**) to (**e**), respectively (After Leheny et al. 1979)

2. A sharp decrease in absorption at the exciton peak ($E \cong 1.52$ eV) is observed, due to the screening of the exciton states by free electrons, which persists for the length of the carrier lifetime (>250 ps, curves *b* to *e*).
3. The main feature indicating the dynamics of the energy distribution of the carriers is the tail of the bleaching within the band for $h\nu > 1.53$ eV, which decreases with a short relaxation time at higher energies (curves *b* and *c*) and a somewhat longer time at energies close to the band edge near 1.52 eV (curves *d* and *e*). This *relaxation of hot carriers* involves the effective scattering with longitudinal optical phonons and will be discussed in the following section.
4. Later, and not clearly discernible in Fig. 9, further relaxation takes place via scattering with acoustic phonons, a substantially slower process; carrier recombination also becomes effective as seen by the area reduction under the difference-of-absorption curves.

Luminescence Studies Luminescence provides information about the *electron temperature* from the slope of the *high-energy tail* of the emission peak. Use of a streak camera shows a successive sharpening of the emission peak, with proceeding energy relaxation of the excited electron gas (Graudszus and Göbel 1981; Tanaka et al. 1980); this is shown in Fig. 10 for GaAs in a somewhat longer time frame, and clearly indicates the successive cooling of the electron gas that was heated by the exciting light pulse.

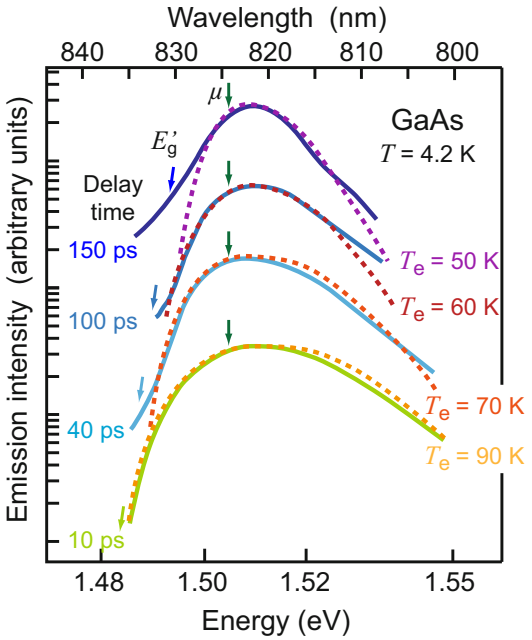


Fig. 10 Low-temperature spectra of spontaneous band-to-band luminescence of GaAs with 4 mW cm^{-2} excitation at 5320 Å, and the time delay after cessation of the excitation as the family parameter. Theoretical curves (*dashed*) are given for the listed electron temperatures, indicating a cooling from 90 K to ~50 K within 150 ps delay (After Tanaka et al. 1980)

When carriers are confined within *quantum-well structures*, many electronic properties change, such as their density of states, carrier screening (it becomes weaker), and plasmon properties – they have zero energy and zero wavevector. However, the interaction of electrons with LO phonons in 2D structures has shown rather similar behavior as in 3D bulk semiconductors, as discussed in several reviews (Lyon 1986; Shah 1986). The average energy-loss rate from heated carriers to the lattice can be expressed much like that in a bulk semiconductor; it is given by

$$\left\langle \frac{dE}{dt} \right\rangle_{e,\mathrm{LO}} = -\frac{\hbar\omega_{\mathrm{LO}}}{\tau_{\mathrm{LO}}} \exp\left(-\frac{\hbar\omega_{\mathrm{LO}}}{kT}\right), \tag{28}$$

where τ_{LO} is an effective electron-LO phonon relaxation time (Shah and Leheny 1984). A more advanced analysis shows only a negligible increase in this loss rate compared to that of bulk GaAs.[3] Hot-phonon modes, however, may cause a more significant reduction in the energy-loss rate, as shown by Price (1985).

Optical probing by luminescence while heating in an electric field shows larger energy-loss rates for holes than for electrons in multiquantum wells (MQWs), as

[3]Modification of the simple model with parabolic minibands, infinite potential steps, nondegenerate electrons, nonscreening to include degeneracy (insignificant up to 2D densities of 10^{12} cm^{-2}), slab modes (small effect, see Shah et al. 1985), plasma effects, and screening (less than 40% influence, see Das Sarma and Mason 1985) has shown little effect on the energy transfer.

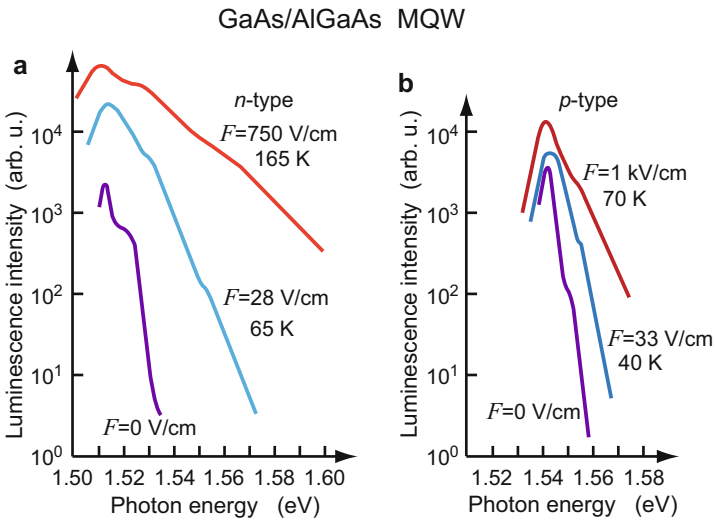


Fig. 11 Photoluminescence of modulation-doped GaAs/$Al_{0.20}Ga_{0.80}As$ quantum wells for various electric fields in a semilogarithmic plot of intensity. (**a**) *n*-type ($n_{2D} = 7 \times 10^{11}$ cm^{-2}) 262 Å wide QWs, (**b**) *p*-type ($p_{2D} = 3.5 \times 10^{11}$ cm^{-2}) 94 Å wide QWs. Temperatures at the curves refer to $T_{carrier}$, the lattice temperature is $T_{Lattice} = 1.8$ K (After Shah 1999)

shown in Figs. 11 and 12. The samples were modulation-doped to achieve high carrier mobility and structured for Hall measurements (Shah et al. 1985). A DC current was applied, and the electric field was determined from the voltage drop across two additional adjacent contacts. The photoluminescence (PL) spectra were excited by a laser beam which was kept weak (~10 mW cm^{-2}) to prevent *optical* carrier heating. The PL spectra given in Fig. 11 show the effect of carrier heating by the electric field F via the high-energy tail, which has an exponentially decreasing intensity at higher photon energy. This behavior proves that the carriers are characterized by a Fermi-Dirac distribution function with a temperature $T_{carrier}$ higher than the lattice temperature $T_{Lattice}$.

By combining the electrical and optical measurements the energy-loss rate of the carriers can be determined as a function of the carrier temperature. The steeper slope of the high-energy tail in the *p*-type sample for a given field F indicates a larger energy-loss rate for holes than for electrons. The reduction of energy losses from electrons could be explained by significant heating of LO phonons: a heated LO phonon gas is less effective in cooling than a cold phonon gas (Shah et al. 1985). Consequently, the electron temperature remains higher than the hole temperature. The electron-hole scattering rate is ~10^{-13} s^{-1} as shown by Höpfel et al. (1986). Typically, carriers thermalize within less than 100 fs (see review by Göbel et al. 1986). The most direct information is obtained by time-resolved bleaching experiments (Knox et al. 1986), which give direct evidence of *spectral hole burning* in GaAs/AlGaAs MQWs and show electron relaxation with a relaxation time of ~50 fs

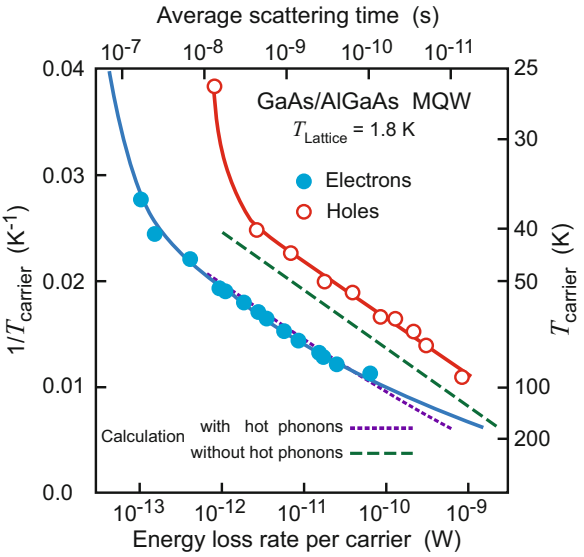


Fig. 12 Inverse carrier temperature $T_{carrier}$ for electrons and holes as a function of the energy-loss rate in GaAs/AlGaAs quantum wells at 1.8 K lattice temperature. *Solid curves* guide the eye, the *dashed line* is calculated for a nondegenerate 3D electron gas (After Shah et al. 1985)

at a carrier density of $\sim 10^{18}$ cm^{-3}. A review of the field is given by Fouquet and Burnham (1986).

1.2.3 Energy-Relaxation Mechanisms

The time response of the different scattering mechanisms of carriers depends on the type of carrier, its density, and its initial energy distribution (Lyon 1986). A large amount of data has been reviewed by Luzzi and Vasconcellos (1984) and discussed in terms of a nonequilibrium thermodynamics approach, using generalized rate equations. The different contributions depend on the material, experimental setup, and progressing relaxation, when one or the other transition rate becomes dominant, as will be discussed below. In addition, we must distinguish between hole and electron relaxation.

Hole Cooling Mechanisms Holes usually have a larger effective mass than electrons, and follow different selection rules for the excitation of phonons: holes excite mainly TO phonons, while electrons excite predominately LO phonons (for a review, see Wiley 1975 and Srivastava 1990). In Fig. 13 the different relaxation processes are summarized: when electrons and holes are produced simultaneously by absorbing a photon (vertical transition), the electron obtains a much higher energy (typically 400 meV) before the hole is energetic enough (~35 meV) to excite an optical phonon; see also Fig. 6 of ▶ Chap. 13, "Band-to-Band Transitions". Consequently, most of the *hole cooling* is caused by scattering with *acoustic phonons*, which is a relatively slow process.

Electron Cooling Mechanisms Electron cooling, starting from sufficiently high energy, can be divided into four regimes:

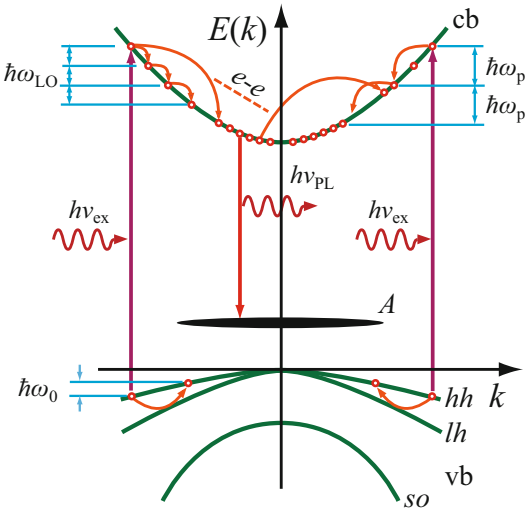


Fig. 13 Schematics of different carrier-relaxation processes after optical excitation (see text). The *line denoted A* is an acceptor level with the varied width indicating the spreading in E and k space

(a) Electron-electron interaction thermalizes the electron ensemble if the density of free carriers is high enough (Shah 1978). The critical electron density at which electron-electron scattering exceeds interaction with phonons can be estimated from plasmon emission (Quinn 1962), which yields for the energy loss

$$\left[\frac{dE}{dt}\right]_{e\text{-}e} \cong \frac{2\pi n e^4}{\varepsilon_{\text{opt}}^2 m_n v_{\text{rms}}^2}. \tag{29}$$

At high electron densities, preferred energy loss occurs by hopping down in units of the plasmon energy (Lyon 1986) when the plasmon energy exceeds the LO-phonon energy. Therefore, from Eq. 4 of ▶ Chap. 12, "Photon–Free-Electron Interaction", we obtain as the critical electron density for preferred electron-electron interaction:

$$n_{e\text{-}e} \cong \frac{\omega_{\text{LO}}^2 \varepsilon_{\text{opt}} \varepsilon_0 m_n}{e^2} = 3.14 \times 10^{17}\text{cm}^{-3} \times \left(\frac{\omega_{\text{LO}}}{10^{13}}\right)^2 \frac{\varepsilon_{\text{opt}}}{10} \frac{m_n}{m_0}. \tag{30}$$

This thermalization from a spike-like initial excitation takes place in the 100 fs range. Experimentally, we obtain as critical density a somewhat lower value of $n_{e\text{-}e} \cong 8 \times 10^{16}$ cm^{-3} (Kash 1989).

(b) Electron-LO-phonon interaction (Fröhlich interaction, see Fröhlich 1937) is very strong.[4] This interaction produces step-like reductions of excess energy (Mirlin 1984), and proceeds in the Γ valley of GaAs with a time constant of 180 fs (Levi et al. 1986). An observed *reduced* rate of cooling via LO phonons at higher carrier densities (Leheny et al. 1979) can be explained by carrier screening of the Fröhlich interaction (Graudszus and Göbel 1983; Yoffa 1981). Another reason for such a reduced cooling rate may be the creation of "hot phonons" – specifically, heating the LO mode, which consequently reduces the cooling efficiency of this mode until it has time to relax (Lyon 1986).
(c) Final cooling involves acoustic phonons and is a relatively slow process, extending into the nanosecond range, due to the small amount of energy transferred in any one transition (Ulbrich 1978).
(d) Recombination of electrons with holes depends on the product of their densities for intrinsic recombination, and on the defect density for extrinsic recombination (see Dymnikov et al. 1978); it can extend from the nanosecond to the microsecond range or longer.

High-Energy Relaxation Mechanisms When an optical excitation creates electrons with sufficient energy, intervalley scattering becomes important. Such a scattering can be investigated with the help of the free-to-bound (e,A^0) luminescence as a probe (Fasol and Hughes 1986); it is very fast.

An overview of the relaxation of higher-energy electrons in GaAs can be obtained from Fig. 14. When excitation occurs below 1.9 eV, only the main valley (Γ) is involved with relaxation via LO phonons or electron-plasma interaction, as discussed in the previous section. The corresponding curve (*a*) shows the luminescence spectrum after excitation from heavy (*hh*) and light (*lh*) hole bands, with additional maxima given by LO-phonon relaxation. At slightly higher excitation energies (1.9 eV $< h\nu <$ 2.3 eV), electrons can rapidly scatter from the Γ- into the *L*-valley with a time constant of ∼540 fs. At energies above ∼2.3 eV, an *intervalley scattering* from the Γ- into the *X*-valley now becomes dominant, with a time constant of ∼180 ps. These time constants can be obtained from competition with the LO relaxation, and from the Γ valley, with consequent (e,A^0) luminescence, acting as an *internal clock* with 180 fs time constant. When electrons are scattered into side valleys, such a luminescence does not occur: the extent of the A^0 level in k does not reach to the minima of the side valleys. This is seen from curves (*b*) and (*c*) in Fig. 14 which are taken with excitation at energies at which intervalley scattering is important. Consequently, the direct *hh* and *lh* signatures are much reduced, giving a measure of the relative population of Γ and side valleys.

In addition, a substantially shifted reentry signal (*R*), after back-scattering from the *X*- or *L*- into the Γ-valley, and its LO-relaxation peaks, can be observed, again

[4]TO phonons can also interact with electrons and are coupled through their deformation potential. They are, however, forbidden to do so with carriers in *s*-like states (Wiley 1975); such forbidden transitions have a factor of only 3 reduced probability, and are important for holes.

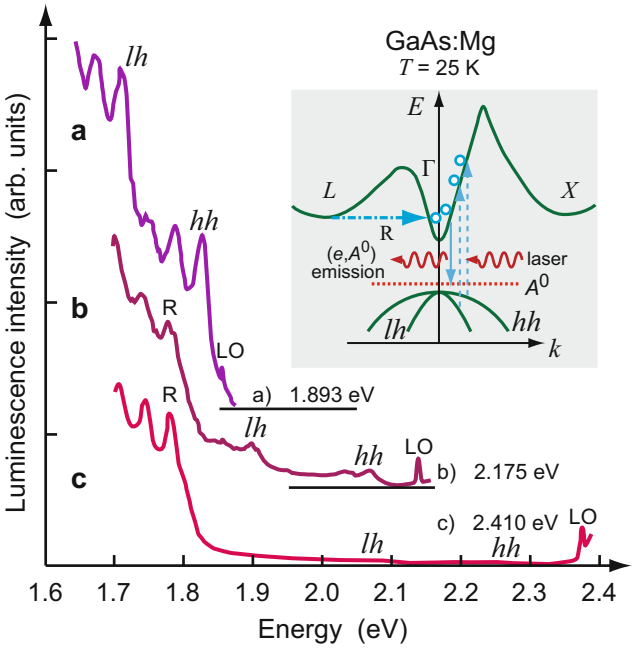


Fig. 14 (e,A^0) luminescence in GaAs:Mg at 25 K ($p = 1.2 \times 10^{17}$ cm^{-3}). Laser excitation at 1.893, 2.175, and 2.410 eV for *curves* (*a*) *to* (*c*), respectively. The peaks identified as LO are Raman scatterings. Inset: $E(k)$ with relevant transitions (After Ulbrich et al. 1989)

with much lower intensities (Ulbrich et al. 1989). The time constant for the reentry (from L or X to Γ) is longer because of the larger effective mass in the satellite valleys. It is on the order of 2.5 ps, and can be obtained from the slow rise of the luminescence – e.g., after excitation above 2 eV. Here, fast scattering into the L valley provides a carrier storage, and consequently a slow supply path via reentry into the Γ valley (Shah et al. 1987). When no reentry into the Γ valley occurs, as, e.g., in InP, the rise of the luminescence is much faster (Shah et al. 1987).

At higher energy, excitation takes place further away from $k = 0$, and *warping of the valence bands* becomes marked. Consequently, there is a wider spread of initial electron energies from which LO-phonon relaxation starts, resulting in a broader luminescence peak: 1.5 meV at 1.57 eV, and 8 meV at 1.85 eV excitation (Ulbrich et al. 1989).

The scattering of holes between different valence bands becomes important for hole-distribution relaxation. This distribution relaxes in the 10 ... 100 ps range, as shown in Fig. 15 for wurtzite CdSe at 4.2 K. With sufficient energy, both the A band and the split-off B valence band are populated. Changes in population occur through inter-valenceband relaxation and via recombination of holes with electrons. Recombination with electrons occurs from both bands with a similar time constant of about 300 ps in the given example, as long as both are in quasithermal equilibrium. Inter-valenceband relaxation from $B_\perp$ to $A_\perp$ is much faster (30 ps). Since the split-off

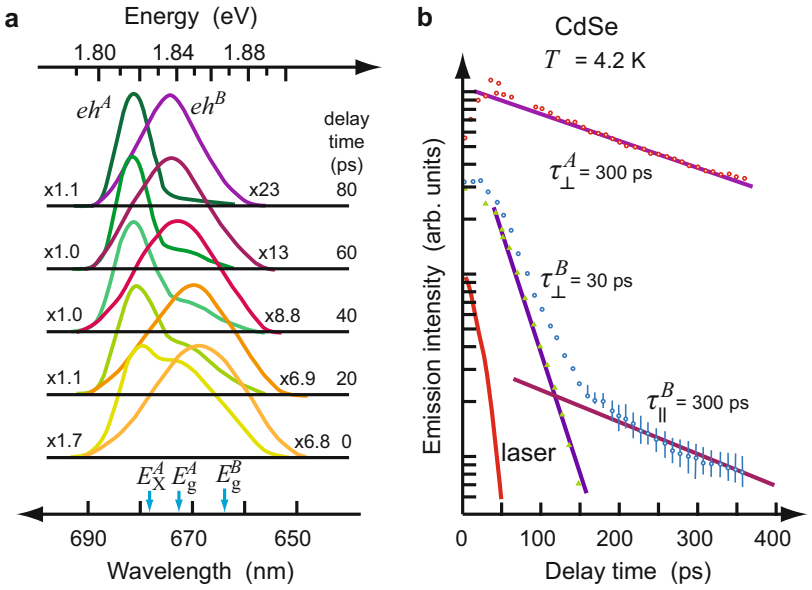


Fig. 15 (**a**) Time-resolved spectral distribution of the spontaneous luminescence of electron-hole pairs (*eh*) in CdSe at 4.2 K excited by a 532 nm picosecond laser pulse of 50 MW cm^{-2} and ~20 ps duration (see *red curve* in *b*). *Green curves* $E \perp c$, *orange/violet curves* $E \| c$ (superscript *A* and *B* in *eh* refers to the *A* and *B* valence bands). E_X^A gives the energy of the exciton associated with the *A* band; E_g^A and E_g^B indicate the two respective bandgap energies. (**b**) Time decay of the three luminescence peaks shown in (*a*). The straight line for $\tau_\perp^B$ is obtained after subtracting the eh^A signal (after Yoshida et al. 1981)

energy (26.3 meV) is sufficient to accommodate an LO phonon with an energy of 26.1 meV, such a scattering, in addition to LA-phonon scattering, is probably responsible for the relaxation.

In the optical study of the initial carrier distribution the coherent coupling between the laser pulse and the interband polarization strongly influences the luminescence spectra; an observed nearly constant linewidth of the successively created phonon replicas is explained by a broadening occurring in the ultrafast generation process (Leitenstorfer et al. 1994).

Lattice Heating Through Hot Electrons The electron-electron scattering does not remove energy from the electron ensemble, but causes thermalization within this ensemble: it smoothes out the distribution, which initially is shaped by the exciting laser pulse, and results in a Maxwell-type distribution with a well-defined electron temperature $T_e > T_{\text{Lattice}}$. A substantial amount of energy can be stored in the electron ensemble before interaction with phonons cools the electron plasma. In fast-rising, high-energy laser pulses, the electron plasma can be heated to more than 10^3 K in the 100 fs range before the lattice temperature increases in the low ps range (Malvezzi 1987).

1.3 Recombination in Electron-Hole Plasmas and Liquids

The energy relaxation and recombination of an electron-hole plasma are highly density-dependent. When the optical excitation occurs at an energy slightly *below* bandgap energy, initially excitons, or exciton-polaritons, are formed. When the generation rate is high enough so that the created excitons closely fill the semiconductor and their orbits start to overlap, the exciton state becomes unstable and an electron-hole plasma is created – see Sect. 3 of ▶ Chap. 21, "Equilibrium Statistics of Carriers". In typical indirect-bandgap semiconductors, this plasma condenses at low temperatures to an *electron-hole liquid* of droplet shape (Rice 1977); see Fig. 20 in ▶ Chap. 21, "Equilibrium Statistics of Carriers".

Kinetics of Electron-Hole Liquids The electron-hole liquid and electron-hole plasma in indirect-gap materials have a rather long recombination lifetime, typically in the microsecond range. The low probability of finding a phonon of proper wavenumber at the same position with an electron and a hole causes the low probability for recombination in indirect-gap semiconductors. In direct-bandgap semiconductors, however, no evidence of droplet formation is yet obtained (Saito and Göbel 1985), while a high-density electron-hole plasma is formed (Haug et al. 1980), but with a substantially shorter recombination lifetime in the nanosecond range.

Electron-hole droplets evaporate, thereby creating excitons, which in turn recombine. After cessation of the optical excitation which created the droplets, droplet evaporation continues until all droplets have evaporated. This process can take much longer than the lifetime of excitons. Evidence for the slow evaporation of electron-hole drops, which replenishes the exciton gas, is shown in Fig. 16. When all the drops are evaporated, the density of the exciton gas decreases with its much faster decay time. When the exciting light pulse is stronger, more droplets are created, and it takes longer until all of them are evaporated (Manenkov et al. 1976).

During a short and intense light pulse, a sufficiently large density of electron-hole pairs is created near the surface, and a liquid layer can be formed. Consequent instabilities occur which complicate the kinetics: the phonon wind generated within the layer creates capillary waves at the liquid surface, from which droplets of critical size (Sect. 3.2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers") break off and are then driven away (Keldysh 1986). This can be experimentally observed since the index of refraction in droplets is different from that in a normal crystal, which makes the droplets visible (Worlock et al. 1974). For electron-hole plasma and liquid in confined semiconductors see Keldysh (1997) and Kalt (1994).

Electron-Hole Pair Dephasing and State Filling When the optical excitation occurs with an energy *exceeding* the bandgap energy, *electron-hole pairs* are formed, as opposed to excitons, which exist only at energies below the bandgap. These optically coupled electron-hole pairs, when scattered, undergo a *dephasing*, i.e., a change of momentum for at least one of the carriers. The dephasing time is extremely short. It has been observed by following the time evolution of optical bleaching in a thin GaAs layer after excitation with a 100 fs high-power laser pulse (Oudar et al. 1985). The intense laser pulse bleaches out the corresponding states in the valence

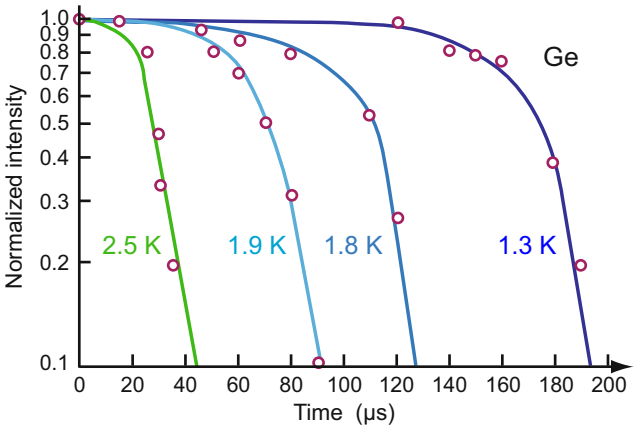


Fig. 16 (**a**) Decay of the exciton density in Ge after cessation of an optical pulse excitation. While electron-hole drops slowly evaporate, the exciton gas is replenished and its density remains constant; thereafter it decays with the exciton lifetime of ~20 μs. Family parameter is the temperature, which delays the evaporation at cooled conditions (after Manenkov et al. 1976)

and conduction bands because of *state filling.* This is to be distinguished from *band-filling*, which is observed after thermalization, e.g., in steady state, and is known as the *Burstein-Moss effect* – see Sect. 3.2 of ▶ Chap. 17, "Optical Properties of Defects". State filling is also referred to as *optical* (or *spectral*) *hole burning within the bands*, since a very *narrow* wavelength range of optical excitation is preferentially bleached. These states have an inverted population.

Figure 17 shows the transmission spectrum of GaAs and its changes over time at 15 K. Bleaching (arrow) occurs initially in the wavelength range of the laser pulse shown in the lower inset. The bleached maximum at 1.54 eV then broadens and disappears when the bleaching pump pulse has ceased, indicating a dephasing time of the electron-hole pairs of ~300 fs. With the progression of time, the bleaching spectrum becomes smooth, indicating thermalization after ~4 ps with $T_e \cong 120$ K in the given example.

2 Phonon and Exciton Kinetics

Quasiparticles can be excited by short light pulses and, following this excitation, reveal their relaxation behavior. These quasiparticles include excitons, phonons, and the corresponding polaritons. With fast detection techniques, the relaxation of simple lattice polarization is accessible to observation. In the following we will give examples of these relaxation processes.

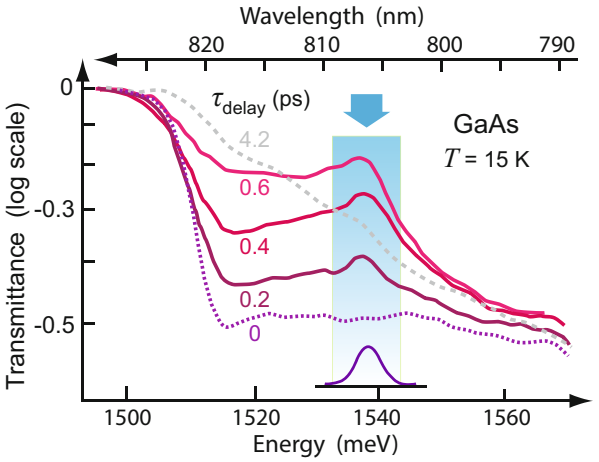


Fig. 17 Time-resolved transmittance (log scale) of a 0.75 μm thick GaAs layer after excitation with a 100 MW cm^{-2} 100 fs laser pulse at 806 nm (*lower inset*). The *shaded area* indicates spectral hole burning (After Oudar et al. 1985)

2.1 Relaxation of Phonon Distributions

Nonthermal changes in phonon distributions can be achieved by selectively heating a specific phonon mode. This can be accomplished most directly by optical (IR) excitation of a specific atomic vibration in the reststrahlen range, and by controlling the frequency and phase relation of the stimulated phonon branch. Another efficient method applies mixing of lasers beams with photon energies differing by the phonon energy; this creates coherent phonons, which can be probed by a delayed pulse measuring the coherent anti-Stokes Raman scattering (CARS). An indirect mode of phonon generation is that of inelastic scattering by hot electrons, which, when energetic enough, interact preferably with one phonon branch – the LO phonons.

The relaxation of a deformed phonon spectrum into its equilibrium distribution proceeds via inelastic three-phonon collisions: the optical phonon, following selection rules, decays into other phonons of lower energy, e.g., two acoustic phonons (Orbach 1967); various decay schemes for diamond-structure, zincblende, and wurtzite semiconductors are discussed by Barman and Srivastava (2004) considering anharmonic interaction between phonons.

In semiconductors with sufficient electron density, scattering with electrons provides an additional relaxation mechanism (von der Linde et al. 1980). When heated, electrons interact with each other; they thermalize with a relaxation time that can be shorter than the electron-phonon relaxation at sufficiently high electron densities; see Sect. 1.2.3. Interaction with LO phonons accelerates their relaxation.

2.1.1 TO-Phonon Relaxation

Due to their weak coupling with carriers TO phonons are usually less important in carrier thermalization. When, however, carrier–LO-phonon interaction is damped due to screening in highly doped semiconductors, TO phonons dominate in the redistribution of energy. Optical pumping of energy into a specific branch of the

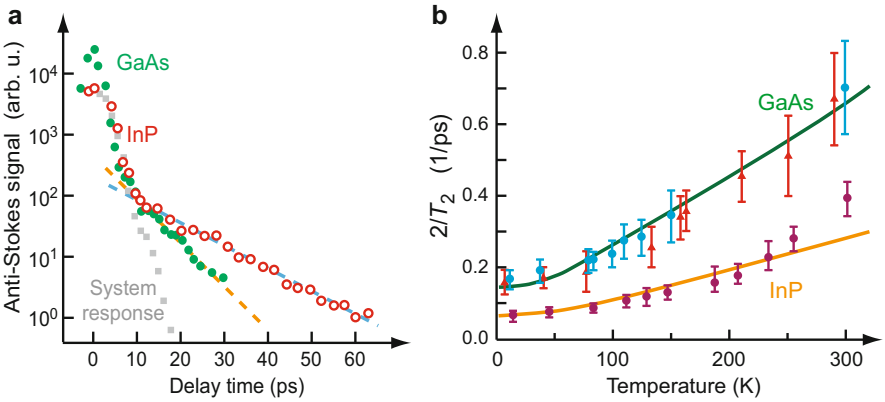


Fig. 18 (**a**) Coherent TO anti-Stokes signal, stimulated by a short laser-pulse pair terminating at $t = 0$ in GaAs and InP at low temperature. *Gray squares* denote the system-response function. (**b**) Measured temperature dependence of the TO phonon dephasing rate $\Gamma = 2/T_2$. *Solid lines* are calculated, in GaAs for a phonon decay into a TA and a LA phonon, and in InP into two equal lower-energy phonons and an additional up-conversion into a LO phonon (After Ganikhanov and Vallée 1997)

phonon spectrum requires intense illumination and can induce coherent oscillations due to nonlinear effects (Sect. 3 of ▶ Chap. 10, "Interaction of Light With Solids"), or, below threshold, incoherent oscillations. Excitation can be accomplished by two synchronized laser pulses of slightly different energy, so that the energy difference is equal to the energy of the phonon branch which is to be stimulated. The energy of such laser pulses is selected preferably below the bandgap energy, and away from any other resonances. Usually, only one phonon mode, the one with the larger Raman cross-section, is stimulated (Gale and Laubereau 1983).

When the intensity of the exciting light is high enough, and the crystal geometry is appropriate for the stimulated phonon branch so that a gain in excess of unity is achieved (Laubereau 1984), amplification arises from the input Stokes signal, and lasing occurs. Transient excitation of a specific mode of lattice oscillations appears when the duration of the pumping pulse is comparable to the lifetime of these phonons, typically in the 10 ps range. For transient pumping, a higher gain is needed to initiate stimulated scattering, starting from quantum noise (Penzkofer et al. 1979).

The kinetics of the phonon population can be measured directly with a delayed probing light beam, or indirectly from the *line width* of the Raman signal, using the Heisenberg uncertainty relation. A typical time evolution of the coherent anti-Stokes signal for hot TO phonons in GaAs and InP is shown in Fig. 18. A theoretical analysis of the zone-center TO mode in GaAs yields a lifetime similar to that of the LO mode, with low- and high-temperature values of 8.3 and 2.5 ps, respectively, with processes TO→LA + LA and TO→LA + TA involved (Barman and Srivastava 2004). These values are in good agreement with experimental results obtained by Ganikhanov and Vallée (1997) shown in Fig. 18 and lifetimes derived from the linewidths of *spontaneous* Raman scattering by Irmer et al. (1996). In InP the decay

of the TO mode can only occur via the channel TO→LA + LA due to energy conservation; the calculated lifetimes of 11.0 ps ($T = 6$ K), 9.7 ps (78 K), and 3.9 ps (300 K) agree well with the Raman measurements (Barman and Srivastava 2004; Debernardi 1998).

Similar measurements in GaP (Kuhl and von der Linde 1982) yield lifetimes for LO phonons at room temperature of 6.7 ± 0.3 ps. The lifetime increases with decreasing temperatures, and at 5 K a value of 26 ± 2.5 ps is measured. Measurements of the decay of *phonon-polaritons* in GaP, which are connected with the TO mode, have also been performed by Kuhl and von der Linde (1982), with apparent lifetimes below 1.3 ps.

Phonon Dephasing Phonon scattering, as discussed above, is inelastic. In addition, *elastic* scattering events ("dephasing") change the momentum of the heated phonons but not their energy. The *dephasing time* is equal to the *momentum-relaxation time.* An example is the phonon scattering with impurities or phonon-phonon interaction within the same phonon branch, which results in a loss of phase information. This elastic scattering, following a preceding excitation, will produce a heated phonon ensemble distributed over a wide range within the Brillouin zone. The evolution of the signal from spontaneous anti-Stokes scattering provides a possibility of directly measuring the phonon distribution and its time evolution.

The time for an incoherent phonon ensemble to approach its thermal distribution is the *energy-relaxation time.* This should not be confused with the lifetime of a mono-energetic phonon mode at a given wavevector.

In studies of ultrafast phenomena the dephasing time is usually referred to as T_2, while the lifetime of an excited state is referred to as T_1. The phase-relaxation time T_2 is connected to the decay of the polarization **P** produced in the semiconductor by an intense electromagnetic field according to $\mathbf{P}_{\text{coherent}} = \mathbf{P}_0 \exp(-t/T_2)$. On the other hand, the decay of an excited ensemble is described by the number density $N = N_0 \exp(-t/T_1)$. This number density N is proportional to the squared polarization, yielding $T_2 = 2\ T_1$; this equality holds if the decay of the excitation is the only phase-destroying process. Usually there are further processes disturbing the phase coherence, leading to the general relation $T_2 \leq 2T_1$; this is conventionally written as

$$\frac{1}{T_2} = \frac{1}{2T_1} + \frac{1}{T_2'}, \tag{31}$$

where T_2' is the pure dephasing time.

The exponential decay of the coherent anti-Stokes Raman signal shown in Fig. 18a translates to a homogeneous Lorentzian broadening of the TO phonon line. The linewidth is given by the dephasing rate $\Gamma = 2/T_2$ and shown in Fig. 18b for the measurements of panel *a* for various temperatures. The data points are well described by the decay of the TO phonons into lower-energy acoustic phonons. In the case of the phonon-dispersion curves in InP only few relaxation channels are allowed by energy and momentum conservation, and an additional up-conversion TO → LO is required to reproduce the measured data (Ganikhanov and Vallée 1997).

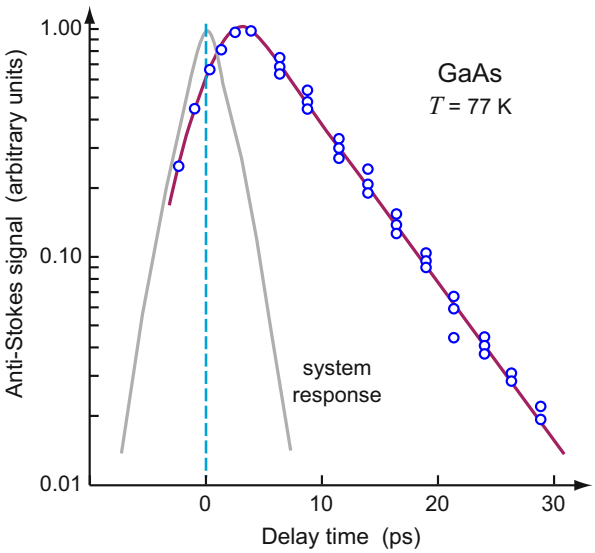


Fig. 19 Spontaneous anti-Stokes scattering signal for nonthermal LO phonons stimulated by fast electrons in GaAs at low temperature. The *gray curve* shows the time dependence of the exciting laser pulse (After von der Linde et al. 1980)

2.1.2 LO-Phonon Relaxation

Optically excited electrons at sufficient energy within the conduction band stimulate mainly LO phonons during scattering. Their heated distribution can be analyzed by the spontaneous (incoherent) Raman scattering induced by LO phonons with a time-delayed probing pulse.

Figure 19 shows the time evolution of the Raman signal in GaAs after cessation of the optical pulse (colored curve) that causes the generation of hot electrons. The measured signal is proportional to the density of LO phonons, and rises steeply with the integrated energy input from the exciting light pulse (von der Linde et al. 1980). The energy-relaxation time of the LO phonon is obtained from the exponential decay of the Raman signal, and yields 7 ± 1 ps in GaAs at 77 K (see also Kash et al. 1985). This value agrees with the lifetime of LO phonons in GaAs obtained from the Raman linewidth of 0.1 meV. The relaxation time of optical phonons depends generally only weakly on temperature; it increases by a factor of ~2–4 between 0 K and room temperature; see also Sect. 2.1.1 and the dependence for TO phonons shown in Fig. 18b. The lifetime of GaAs LO phonons calculated from anharmonic phonon interaction yields 8.7 and 2.3 ps at low and room temperatures, respectively; due to the downward dispersion of the LO branch, a dominant decay of the zone-center LO phonon into a zone-edge LO phonon and a TA phonon is concluded (Barman and Srivastava 2004).

In polar zincblende materials the lifetime of the TO mode compared to the LO mode depends on the cation-anion mass ratio; a ratio larger unity leads to a shorter lifetime and vice versa (Barman and Srivastava 2004). In GaAs with a ratio near 1 both modes have hence similar lifetimes.

The distribution of LO phonons generated by the relaxation of photoexcited electrons is usually not in equilibrium because the energy distribution of the

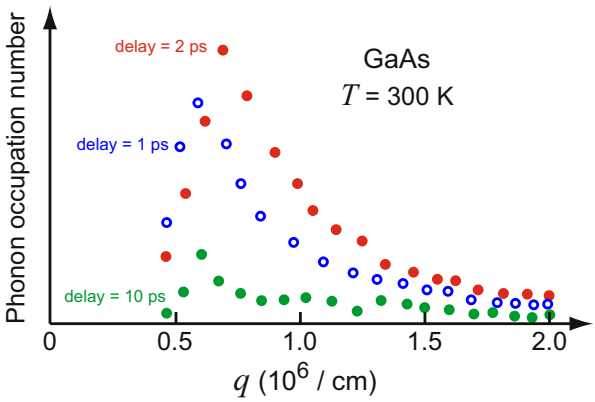


Fig. 20 Calculated phonon occupation in GaAs at 300 K lattice temperature as a function of the phonon wavevector q for different time delays after excitation by a delta-function laser pulse above the bandgap energy of GaAs (After Shah 1999)

photoexcited electrons is usually nonequilibrium. Depending on the electron concentration, the electrons can reach a quasi-thermal equilibrium in times as short as 100 fs or less in GaAs for an electron concentration of $\sim 10^{17}$ cm^{-3}. However, it typically takes the phonon distribution about 1 ps or longer to relax to equilibrium via elastic and inelastic scattering.

Since Raman scattering can probe only a limited region of the phonon-momentum space, only a small part of the relaxation of the nonequilibrium phonon distribution can be measured; relaxation processes, time range, and limits of experimental techniques are reviewed by Ferry (2021). Advances in computer simulation, however, allow for calculating the time evolution of the phonon distribution, e.g., via the ensemble Monte-Carlo simulation. Figure 20 shows the LO phonon-occupation number in GaAs at room temperature calculated by Lugli et al. (1989) as a function of the wavevector q for various delay times after excitation of the sample by a 400 fs laser pulse of above-bandgap photon energy. The calculation assumes that the LO phonon has a q-independent anharmonic decay time of 3.5 ps at room temperature. The calculated distribution shows a peak near 8×10^5 cm^{-1} and a sharp cutoff at the small-wavevector side near 5×10^5 cm^{-1}. Both features are in agreement with the results of a simple two-band parabolic model in GaAs calculated by Collins and Yu (1984). The fortuitous coincidence of this peak in the nonequilibrium LO-phonon distribution and the wavevectors which can be probed by Raman scattering in GaAs is responsible for the success of this technique as a probe of nonequilibrium LO phonons. Using their simulation Lugli et al. (1989) were able to obtain the time evolution of the anti-Stokes Raman signal after excitation. Their results are in good agreement with the data of von der Linde et al. (1980) shown in Fig. 19, as well as with the data of Kash et al. (1985).

One interesting result of the calculation of Lugli et al. (1989) is that the LO phonons with a wavevector near the peak of the distribution decay much faster than phonons with wavevectors twice as large. This finding can be understood by the coupling of the larger-wavevector phonons only to electrons of highest energy. When the kinetic energy of these electrons falls below a certain level they can no longer couple to the large-wavevector phonons, because the simultaneous

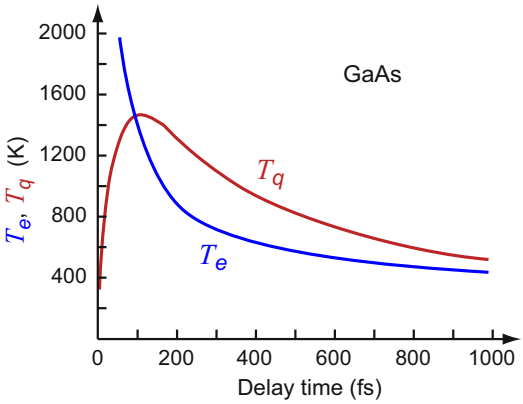


Fig. 21 Effective electron (T_e) and phonon (T_q) temperature of GaAs calculated as a function of time delay after excitation by a fs laser pulse. Intervalley scattering of excited electrons has been included in the calculation (After Kim and Yu 1991)

conservation of both energy and wavevector cannot be satisfied; the large-wavevector LO phonons are hence decoupled from partially cooled electrons.

On the other hand, LO phonons with smaller wavevectors can continue to interact with the lower-energy electrons. In fact, a simulation by Kim and Yu (1991) found that the "Raman-observable" phonons with $q \sim 8 \times 10^5$ cm^{-1} achieve thermal equilibrium with electrons only after about 1 ps. The net effect of this is twofold: first, the cooling of hot electrons slows down because their cooling by LO phonon becomes less effective. Second, the reabsorption of LO phonons by the electrons speeds up the LO-phonon energy relaxation. The first effect is essentially the "hot phonons" effect mentioned in the electron-cooling mechanisms (Sect. 1.2.3). The second effect is most prominent in times shorter than 1 ps when the electron temperature can be *lower* than the phonon temperature. Figure 21 shows the calculated "effective" electron temperature (T_e) and "effective" phonon temperature (T_q) for $q \sim 8 \times 10^5$ cm^{-1}. For delay times between 100 and 1000 fs the temperature of phonons is higher than that of the electrons (Kim and Yu 1990). The main reason for this phenomenon is the strong cooling rate of the high-energy electrons via intervalley scattering in GaAs from the Γ valley to the *L* valleys mentioned in Sect. 1.2.3. Kim and Yu (1991) demonstrated that this phenomenon would be absent if intervalley scattering is neglected.

2.1.3 Phonon Relaxation in Quantum Wells

Phonon properties are modified in quantum wells and superlattices (Sect. 3.3.1 of ► Chap. 4, "Elasticity and Phonons"). These modifications affect the dynamics of the phonon modes. Interface (IF) mode and confined LO phonons couple most strongly to electrons via their longitudinal electric fields. These modes are equivalent to the LO phonon in bulk semiconductors.

Most information on the time-dependent properties of phonons in quantum wells and superlattices is obtained by time-resolved Raman scattering. Fig. 22 shows the time dependence of the anti-Stokes Raman intensity of the IF and confined phonon modes in a 56 Å wide GaAs quantum well clad by AlAs barriers, excited by a 4 ps

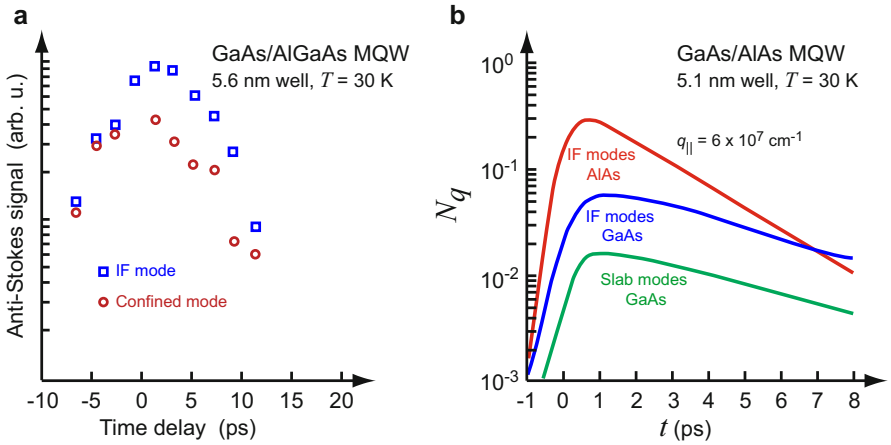


Fig. 22 (**a**) Time evolution of the anti-Stokes Raman intensity of the interface (IF) and confined LO phonon modes in a GaAs/AlGaAs multiple quantum well. (**b**) Nonequilibrium population of phonons with wavevector parallel to the surface in a GaAs/AlAs multiple quantum well, given as a function of time after generation by hot photoexcited carriers (After Shah 1999)

laser pulse (Ryan and Tatham 1992). There are two IF modes in such QW structures. The AlAs IF mode studied in Fig. 22 has a frequency between those of TO and LO phonons in AlAs; its electric field is strongest in the GaAs QW, producing a strong interaction with the confined electrons. The GaAs IF mode with a frequency between those of TO and LO phonons in GaAs couples only weakly to electrons and is difficult to measure.

The time dependence of the phonon population of the modes shown in Fig. 22a was simulated by the Monte-Carlo method (Lugli et al. 1992). The results given for 51 Å wide GaAs quantum wells in Fig. 22b reproduce the exerimentally observed faster decay of the AlAs IF mode compared to the GaAs IF mode. The decay of the AlAs IF mode was also measured as a function of the GaAs well thickness (Ruf et al. 1993; Tsen 1993). The decay rate decreases with increase in well thickness. This decrease is theoretically well described if only phonons with a limited range of wavevectors $q_{||}$ centered around $q_{||} = 6 \times 10^7$ cm^{-1} is assumed to be sampled in Raman scattering (Tsen 1993).

2.2 Exciton Kinetics

The exciton is a two-particle system consisting of an electron and a hole. Their charges give rise to an electric dipole moment, and the excitation of an exciton ensemble can produce a macroscopic polarization **P**. When excitons are excited by a laser beam with field $\mathbf{F}(\mathbf{q}, \omega)$ the resultant exciton polarization-wave has a well-defined energy and wavevector. This electric-dipole wave radiates in turn an electromagnetic wave; the exciton and the photon fields are hence coupled to form the

exciton polariton introduced in Sect. 1.3 of ▶ Chap. 14, "Excitons". The subtle nature of exciton-polaritons complicates their relaxation.

Free excitons in semiconductors can recombine radiatively only close to the center of the Brillouin zone due to momentum conservation. The dispersion curve of a photon is given by

$$E = h\nu = \hbar 2\pi\nu = \hbar ck/n_{\mathrm{r}}, \tag{32}$$

where k is the (small) photon wavevector and n_{r} is the refractive index. The dispersion of an exciton with mass $M = m_n + m_p$ and kinetic energy $E = \hbar^2K^2/(2M)$ reads

$$E = E_{\mathrm{g}} - E_{\mathrm{B,X}} + \hbar^2K^2/(2M); \tag{33}$$

here K is the wavevector of the exciton and $E_{\mathrm{B,X}}$ is the binding energy of the exciton in the ground state. These dispersion curves intersect at a wavevector K_0 given by

$$E_{\mathrm{g}} - E_{\mathrm{B,X}} = \hbar ck/n_{\mathrm{r}} - \hbar^2K_0^2/(2M). \tag{34}$$

The degeneracy of the photon and exciton energies is removed by the exciton-photon interaction, yielding the dispersion of the coupled quasi-particle *excitonic polariton* described in Sect. 1.3 of ▶ Chap. 14, "Excitons". There are several polariton branches, and depending on the wavevector the polariton has a more excitonic or photonic character. Radiative recombination within the exciton-polariton picture means the conversion of the exciton-polariton near the sample surface to a photon which exits to the outside of the semiconductor.

If excitons are nonresonantly created by hot carriers, inelastic scattering processes are required to obtain radiative recombination with a wavevector near the center of the Brillouin zone. In bulk semiconductors such a scattering is provided by phonons, but also by impurities or surfaces. The exciton dynamics in bulk semiconductors differs from that in low-dimensional structures of the same material. Moreover, the processes depend on excitation energy (resonant or nonresonant), excitation density, and lattice temperature. We first consider bulk excitons and discuss quantum-well excitons in Sect. 2.2.2; the spin dynamics is discussed separately in Sect. 3.

2.2.1 Dynamics of Bulk Excitons

Resonant Excitation The experimental study of free excitons requires generally samples with very low defect densities to prevent the formation of bound excitons,[5]

[5]The probability of finding an impurity within the volume of the exciton is proportional to the number of unit cells in the excitonic volume $(a_{\mathrm{X}}/a)^3$. If impurities are located in the volume of excitons, only bound-exciton recombination is observed. For experiments with GaAs high-purity thick epitaxial layers are used with a residual impurity concentration on the order of only 10^{12} cm^{-3}, additionally clad by AlGaAs barriers to prevent outdiffusion and surface recombination of optically excited excitons.

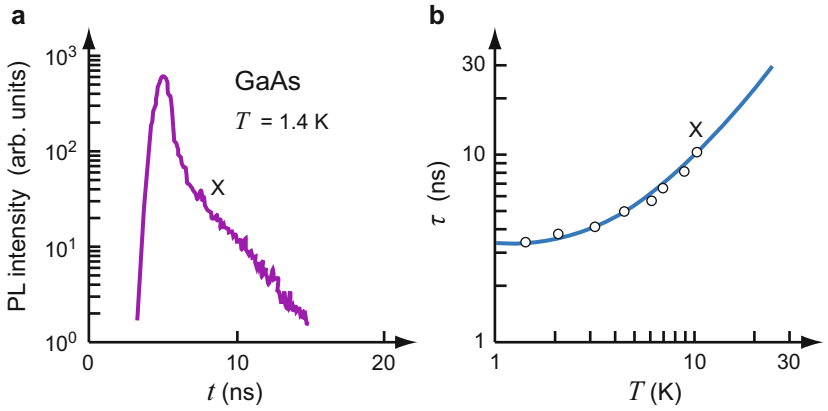


Fig. 23 (**a**) Photoluminescence decay (*tail*) in ultrapure GaAs after resonant pulsed excitation at 8182 Å with a low density of 0.1 W cm^{-2}. (**b**) Radiative lifetime of free excitons as function of the sample temperature (After t'Hooft et al. 1988)

since usually free excitons are rapidly trapped by impurities, where they recombine with a short lifetime of typically 0.5 or 1 ns for donor- or acceptor-bound excitons (Henry and Nassau 1970).

We focus on GaAs with an exciton binding energy of ~3.6 meV and an exciton Bohr-radius of ~155 Å. In ultrapure GaAs recombination paths of *bound* excitons can be saturated; in addition, resonant excitation into the 1*S* or 2*S* state is applied to directly address the creation and recombination of free excitons. The photoluminescence decay then yields a direct measure of the radiative lifetime T_1 of free excitons of about 3.3 ns at 1.4 K, increasing to 10 ns at 10 K, as shown in Fig. 23. This corresponds to an oscillator strength of nearly unity, indicating an oscillator strength per exciton volume rather than per unit cell volume (t'Hooft et al. 1987). Such lifetimes of free excitons can be explained by assuming a conversion of the originally generated exciton-like polaritons into photon-like polaritons via polariton – acoustic-phonon scattering and consequent recombination (Rappel et al. 1988).

The dephasing of the coherent macroscopic polarization introduced by the excitation on an ultrashort time scale is not resolved in Fig. 23a. A dephasing time T_2 of 7 ps was measured in GaAs, corresponding to a homogeneous Lorentzian linewidth of $\Gamma = (h/e)(2/T_2) = 0.18$ meV (Schultheis et al. 1986). The spectral width of ultrashort laser pulses may be much larger, e.g., 20 meV for a 100 fs pulse, leading to different population and polarization decay within and outside the exciton linewidth. Dephasing is almost exclusively studied with quantum wells and discussed in Sect. 2.2.2.

Nonresonant Excitation The dynamics for nonresonant excitation of excitons differs substantially from that for resonant excitation discussed above. Nonresonant excitation into continuum band states creates in a first stage hot carriers, which thermalize and cool down, accompanied by the formation of nonthermal excitons. Carrier cooling initially produces optical phonons, followed by the emission of acoustic phonons at lower carrier temperature. In a second stage the exciton

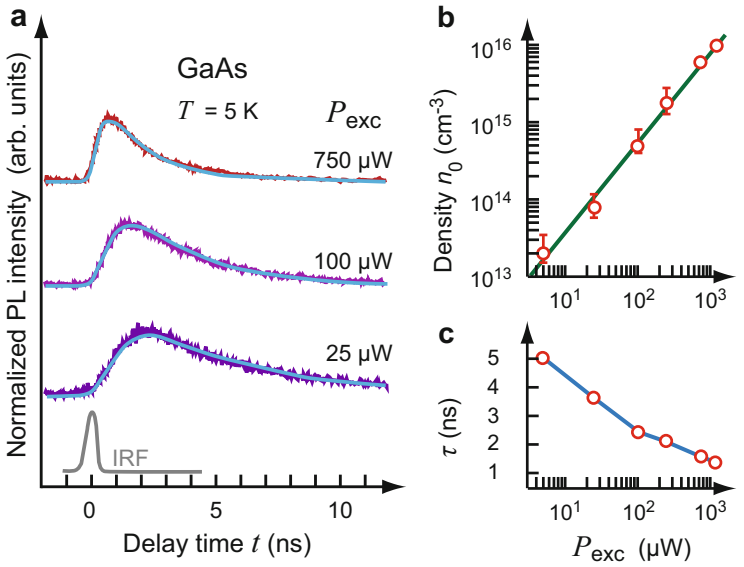


Fig. 24 (**a**) Temporal evolution of the nonresonantly excited free-exciton photoluminescence in pure bulk GaAs, detected on the second LO-phonon replica for various excitation densities. The rise time depends on the excitation density; *solid blue lines* are model calculations, the *gray curve* indicates the instrument response function (*IRF*). (**b**) Linear dependence of the total photocarrier density n_0 on the optical excitation density P_{exc}. (**c**) Dependence of the decay time τ on the excitation density (After Beck et al. 2016)

ensemble thermalizes by mutual interaction and by quasi-elastic interaction with acoustic phonons, until the excitons eventually cool to the range of the lattice temperature due to inelastic acoustic-phonon interaction.

Relaxation after nonresonant optical excitation of excitons in pure bulk GaAs is slow and leads to a pronounced rise time of the exciton luminescence on the order of 1 ns, which is not observed with resonant excitation; the subsequent decay time is similar to that observed with resonant excitation (Elsaesser et al. 1996; Gurioli et al. 1998); see Fig. 24. The long PL rise time was modeled by Monte-Carlo simulation with a set of coupled Boltzmann equations for electrons, holes, and excitons, and was assigned to the exciton relaxation to radiative states via acoustic-phonon emission. Studies at higher lattice temperature ($kT > E_{B,X}$) or high excitation densities ($>10^{15}$ cm^{-3} in bulk GaAs) indicate that the dynamics is dominated by coexisting contributions of exciton and electron-hole pair recombination and relaxation (Amo et al. 2006).

PL studies provided more recently clear evidence that the long-discussed slow rise time and the power-dependent decay time originate from a time-dependent balance between free excitons and the uncorrelated electron-hole plasma. Beck et al. (2016) monitored the second LO-phonon replica of the exciton recombination instead of the zero-phonon line; the strict K-vector selection rule of radiative relaxation only close to the Brillouin zone center is fully relaxed for this emission, because wavevectors of two LO phonons can always add up such that they

compensate for the center-of-mass momentum of the recombining exciton (Segall and Mahan 1968). Thereby detecting a signal of the *total* free-exciton density the rather slow buildup of the free-exciton population was measured.

The balance between the population of free excitons and that of the uncorrelated electron-hole plasma is described by the Saha equation

$$\frac{np}{n_X} = \left(\frac{kT_X}{2\pi\hbar^2}\right)^{3/2} \left(\frac{m_n m_p}{M}\right)^{3/2} \exp\left(\frac{E_{B,X}}{kT_X}\right), \tag{35}$$

where n_X, T_X, $E_{B,X}$, and M are respectively the density, temperature, binding energy, and mass of the free excitons. The time-dependent fraction of free excitons $f_X = n_X/n_0$ in a photo-generated population of carriers with a total electron-hole pair density n_0 can be determined from Eq. 35 by the condition

$$n = p = (1 - f_X)n_0. \tag{36}$$

The exciton fraction f_X following from Eq. 35 is given in the plot Fig. 25 for various total carrier densities n_0 and exciton temperatures T_X. A high density of photo-generated carriers n_0 produces a high density of excitons particularly at low exciton temperatures. At elevated exciton temperature thermal breakup of free excitons into pairs of unbound carriers leads to a decrease of the exciton population.

The modeled traces for the cooling of excitons shown in Fig. 25 indicate that the exciton ensemble does not attain the lattice temperature $T_L = 5$ K even for long delay times (Beck et al. 2016). Since only the subset of excitons at $K \cong 0$ recombines radiatively, low-energy excitons are selectively removed from the entire ensemble and the leftover exciton population has an increased kinetic energy distribution; such *recombination heating* compensates the cooling rate resulting from acoustic-phonon emission. This effect is potentially detrimental for creating excitonic Bose-Einstein condensates, which require ultracold free-exciton ensembles (Sect. 3.2 of ▶ Chap. 21, "Equilibrium Statistics of Carriers").

2.2.2 Dynamics of Quantum-Well Excitons

The interaction of excitons confined in a quantum well (QW) with electromagnetic radiation leads to two kind of modes, radiative polariton modes and surface polariton modes. The latter are stationary modes like those of bulk polaritons, but decay exponentially outside the well; they are only excited with special techniques. Most studies focus on the radiative modes.

Radiative modes are not stationary states of the QW system. Unlike bulk modes, which after nonresonant excitation need the interaction with phonons or defects for radiative recombination, radiative polariton modes of quantum wells have a finite *intrinsic* recombination rate. This effect originates from a coupling of radiative QW polaritons to a photon mode for an in-plane momentum (parallel to the interfaces) $K_\parallel < K_0$, with the momentum K_0 of the crossing point given by Eq. 34 (Hanamura 1988). In a quantum well, the absorption of an incident photon creates an exciton with well-defined $K_\parallel$, which is a good quantum number. In absence of scattering, such excitons can only radiate in the direction of transmission or reflection.

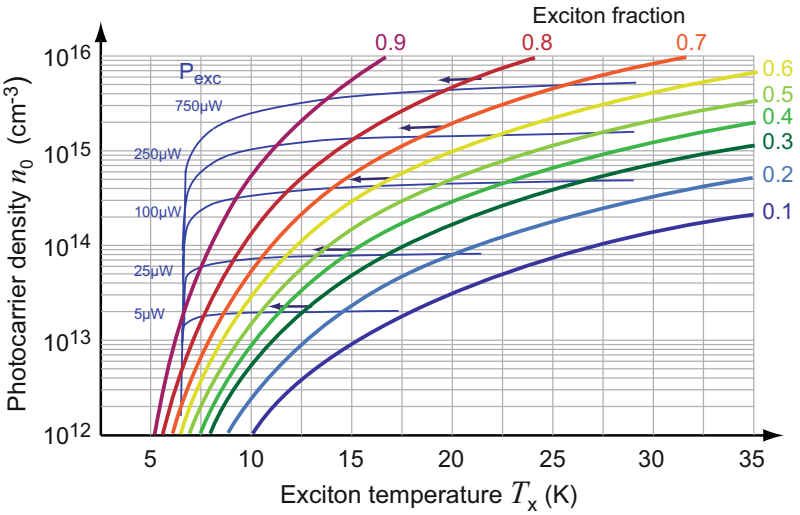


Fig. 25 Exciton fraction f_X (*colored curves*) in the thermodynamic equilibrium of the population of free excitons and that of the uncorrelated electron-hole plasma described by Eq. 35. The *thin blue curves* indicate the cooling of the exciton ensemble for the excitation densities applied in the luminescence transients shown in Fig. 24a (After Beck et al. 2016)

In past decades, most investigations on the recombination dynamics were performed applying time-resolved luminescence or optical pump-and-probe techniques; the recently available sources for terahertz radiation (see, e.g., Sect. 2.3.2 of ► Chap. 27, "Carrier Transport in Low-Dimensional Semiconductors") are meanwhile a valuable additional tool for such studies (Ulbricht et al. 2017). Optical-pump terahertz-probe experiments yield both real and imaginary parts of transient response functions. Terahertz radiation can probe intraexcitonic transitions largely independent of the momentum K, independent of the interband dipole moment; since such a probe is also sensitive to free carriers, i.e., unbound *e-h* pairs, the complex dielectric function derived from the terahertz response comprises both intraexcitonic and Drude features. A description in terms of a two-component dielectric function provides access to the transient density of both excitons and unbound *e-h* pairs (Kaindl et al. 2009).

The study of *intrinsic* relaxation mechanisms requires QW samples with well-defined interfaces to the barriers. Such samples have a basically homogenous linewidth of excitonic emission and a negligible Stokes shift between absorption and emission spectra.[6] Optical-pump terahertz-probe was used to probe transitions between the 1*S* and higher-energy exciton or continuum levels of either resonantly or nonresonantly optically excited excitons.

[6]QW samples with rough interfaces exhibit a large Stokes shift due to fluctuation in the well width. Exciton relaxation to $K_{||} = 0$ is then accompanied by the relaxation of excitons from narrow well regions to regions of larger well width. Luminescence intensity, energy, and shape consequently vary during relaxation in a way depending on the particular sample.

Resonant Excitation We consider results of the transient terahertz response[7] after resonant excitation at the 1*S*–heavy-hole exciton (Kaindl et al. 2009). The studied excitons in a GaAs/$Al_{0.3}Ga_{0.7}As$ MQW sample with 14 nm wide wells and 10 nm thick barriers were optically excited at 1.540 eV and probed by 7 meV THz excitation into the 2*P* level; the dielectric function comprising excitonic and free-carrier signatures was derived from the Fourier-transformed transmitted THz response. The density n_X of two-dimensional excitons derived from the fitted excitonic and Drude models is shown in Fig. 26a (dots). We observe the fast rise at zero time delay within the coherent regime and the long decay of the incoherent exciton population, described by an exponential decay (dashed line). The incoherent decay agrees well with time-resolved PL measurements recorded with a streak camera (solid line); deviations at the coherent regime shortly after excitation are attributed to insufficient suppression of perturbing effects.

The measured decay of ~1 ns is quite long. The exciton binding energy in a 2D well is expected to be significantly larger than in 3D bulk; see Eq. 27 of ▶ Chap. 14, "Excitons". The 2D exciton Bohr radius is consequently smaller and the electron-hole overlap larger, leading to an increased oscillator strength and faster decay (Andreani and Pasquarello 1990; Citrin 1993). The quite slow decay is assigned to the large well width and the high exciton density.

At increased lattice temperature the resonantly excited excitons are thermally ionized and unbound *e-h* pairs are generated, leaving a fraction f_X after pulse excitation. The temporal dynamics of f_X obtained from the dielectric function derived from the THz response for various lattice temperatures shown in Fig. 27a is well described by exponential fits with shorter time constants at higher temperature. The temperature dependence $f_X(T)$ agrees with a Boltzmann quasithermal equilibrium between excitons and *e-h* pairs given by the Saha model for a total pair density near 2×10^{10} cm^{-2} and long delay times after pulse excitation, showing the carrier cooling-dynamics (Fig. 27b).

Nonresonant Excitation After nonresonant pulse excitation the exciton luminescence given in Fig. 26b displays a slow rise with a maximum 1000 ps after excitation. The lack of the fast exciton formation indicates the dominant initial population of states with high momenta *K*. The figure shows for the given excitation density an initial exciton fraction of $f_X \cong 40\%$, followed by a slower binding of *e-h* pairs into excitons within several 100 ps and the eventual decay due to recombination.

Cavity Effects on Dynamics The relaxation dynamics of excitons, which are confined in low-dimensinal semiconductors, is substantially altered if the optical density of states in the environment is modified by an optical cavity, as pointed out in Sect. 2.2 of ▶ Chap. 10, "Interaction of Light With Solids". When the energy of the

[7] In this study the THz radiation was generated from a part of the pump pulse by optical rectification (Sect. 3.1.3 of ▶ Chap. 10, "Interaction of Light With Solids") in a ZnTe crystal.

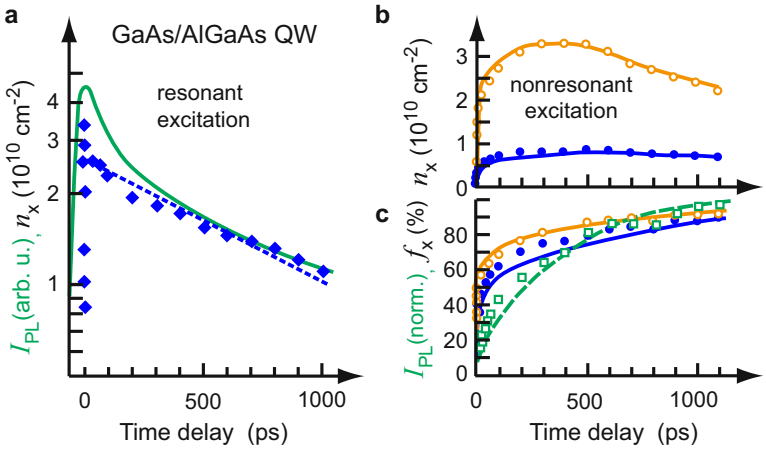


Fig. 26 (**a**) Transient exciton density n_X, derived from optical-pump terahertz-probe measurements after resonant 1*S*–*hh* excitation (*symbols*); the *dashed line* refers to an exponential decay with $\tau = 1087$ ps. The *solid line* represents the PL intensity after resonant excitation for $n \cong 2 \times 10^{10}$ cm^{-2} and $T_{Lattice} = 10$ K. (**b**) Exciton density n_X and (**c**) exciton fraction f_X during exciton formation after nonresonant excitation at $T_{Lattice} = 6$ K. *Filled dots* refer to a density of $n = 1 \times 10^{10}$ cm^{-2}, *open dots* to a 3.5 fold density; the *solid lines* are obtained from a Saha quasi-equilibrium model. The *dashed line* in (*c*) represents the normalized PL intensity after nonresonant excitation; *squares* are PL intensities calculated from $n_X(t)$ (After Kaindl et al. 2009)

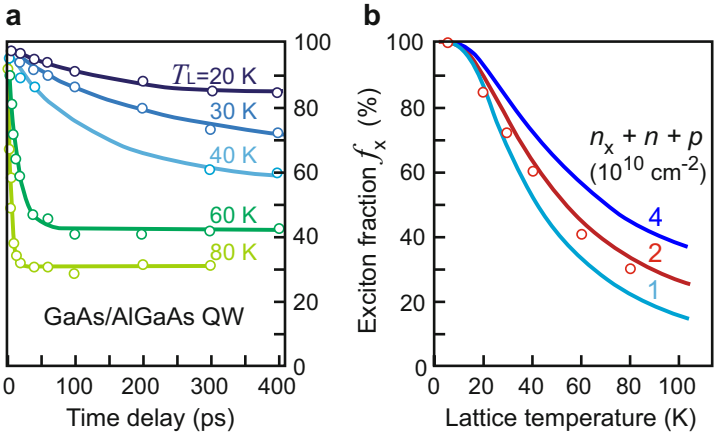


Fig. 27 *Left*: Exciton fraction f_X (*symbols*), derived from terahertz measurements with resonant optical excitation at various lattice temperatures T_L; *lines* are exponential fits with offsets. *Right*: Temperature dependence of the exciton fraction f_X derived from the measurements (*red dots*) compared to a Saha equilibrium model for various densities ($n_X + n + p$) (After Kaindl et al. 2009)

spontaneous emission lies within the optical bandgap of the cavity, the radiation can be suppressed; on the other hand the oscillator strength can be enhanced for a resonance to a cavity mode (*Purcell effect*). In addition, for a strong coupling between cavity mode and the embedded emitter an emitted photon can be reabsorbed and reemitted, giving rise to *Rabi oscillations*. For a review, see Yamamoto et al. (2000).

3 Relaxation of the Spin Momentum

When a photon creates an electron and a hole in a direct transition, the total spin and angular momentum of these two carriers equals the angular momentum of the absorbed photon. In right- or left-polarized light, the photons have a projection of their angular momentum in the direction of their propagation of $+\hbar$ or $-\hbar$, respectively.[8] This momentum is distributed between the electron and the hole, according to the selection rules within the band structure of the given semiconductor. The coherently aligned orientations of angular or spin momentum of the created carriers usually relax rapidly due to a variety of processes. Spin relaxation can occur for several reasons:

- Elliot-Yafet mechanism: momentum relaxation during electron scattering through spin-orbit coupling; this mechanism is strong in semiconductors with a small direct bandgap energy, large spin-orbit splitting, and high hole concentration (Elliot 1954; Yafet 1963).
- Dyakonov-Perel mechanism: relaxation *between* scattering events through a momentum-dependent spin precession with an effective Larmor frequency corresponding to the spin splitting of the conduction band; this mechanism is effective on electrons in noncentrosymmetric semiconductors like III-V and II-VI compounds (not in Si, Ge), particularly at high temperature (Dyakonov and Perel 1971, 1972).
- Bir-Aronov-Pikus mechanism: exchange interaction between electrons and an unpolarized population of holes; the mechanism contributes to spin relaxation particularly in highly doped semiconductors (Bir et al. 1976).
- Coupling of the electronic spin to the spins of lattice nuclei (Dyakonov and Perel 1984).
- Reabsorption of recombination luminescence (Kleinman and Miller 1981).

The relaxation times of coherent spin states are rather short, when the coupling to other degrees of freedom listed above is strong. Observation of the time evolution of such randomization may employ diffraction on a bleached-out grating, time-resolved Faraday rotation (transmission) or magneto-optical Kerr rotation (reflection) (see Baumberg et al. 1994; Zheludev et al. 1994; Awschalom and Samarth 2002) or four-

[8] Linearly polarized photons are a superposition of these two states.

wave mixing, in which nonlinear interaction between pump and test beams creates additional photons (Lyon 1986; Leo 1993). On the other hand, if the spin states are well separated from other degrees of freedom (e.g., the spin of an electron at a single defect), relaxation times can be rather long. Reviews are given by Pikus and Titkov (1984), Dyakonov (2008), and Wu et al. (2010). Calculated spin dynamics of III-V semiconductors based on kinetic spin Bloch equations and including the mechanisms listed above are reported by Jiang and Wu (2009). The dominant mechanism in spin relaxation was found to depend on type and density of doping, temperature, and photoexcitation density, with a strong Dyakonov-Perel contribution in many cases. The results basically coincide with those of simplified calculations based on the momentum relaxation rate (Song and Kim 2002) and largely agree with experimental data.

We focus on the dynamics of spin relaxation, introducing the Hanle effect as a means to measure the decoherence time, and then discuss a few examples for spin relaxation of single carriers and of excitons in semiconductors.

3.1 Measurement of Spin Relaxation

Spin-polarized conduction-band electrons can efficiently be generated by optical excitation with circularly polarized light of above-bandgap energy. In zincblende semiconductors the application of right (σ^+) circular light along the x direction generates preferentially electrons with spin $s_x = -1/2$ (in units of $\hbar$), because the dipole interband-transition probability from heavy holes ($j_{hh} = -3/2 \rightarrow s_x = -1/2$) is three times that from light holes ($j_{lh} = -1/2 \rightarrow s_x = +1/2$); see Eq. 23 of ▶ Chap. 9, "Magnetic Semiconductors", and Dyakonov and Perel (1984). Due to spin-orbit coupling the holes lose their spin-polarization generally rapidly,[9] in contrast to electrons (Sham 1993). The fraction of photo-excited carriers, which have *not* lost their original spin orientation, will cause a circularly polarized luminescence when recombining (Parsons 1969); thereby the spin-relaxation time can be extracted from the degree of polarization, when the recombination time is known and exponential decay of recombination and spin relaxation is assumed.

The method is improved by applying an external magnetic field **B** perpendicular to the direction of the excitation, which coherently aligns the spins of the electrons; the spins then perform Larmor precessions in a plane perpendicular to the **B** axis; see Fig. 28. The dependence of the luminescence on a transverse magnetic field is known from atoms as the *Hanle effect* (Hanle 1924) and is also observed in semiconductors (Lampel 1974). The precessions cause an additional depolarization of the luminescence, which now depends on the spin-relaxation time (τ_s), the electron-recombination time (τ_n), and the Larmor frequency ω_L; when the polarization has a

[9]For bulk GaAs a spin-relaxation time of 110 fs was measured for heavy holes (Hilton and Tang 2002).

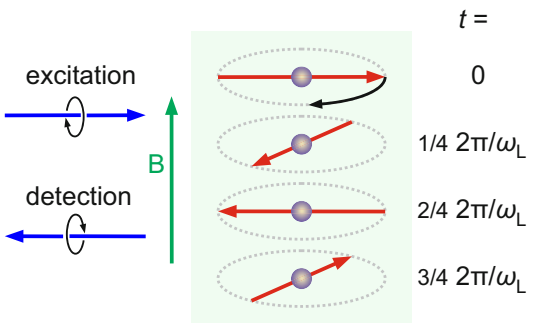


Fig. 28 Schematic illustrating the circular polarizarion of excitation and detection beams (*blue arrows*) to measure spin dephasing, and four stages of the precession of a spin (*red arrows*) in an externally applied magnetic field **B** (*green arrow*)

Lorentzian dependence on $|\mathbf{B}|$, then two of the three quantities can be determined if one is independently measured (Snelling et al. 1991).

When the magnetic field is perpendicular to the exciting light beam, i.e., $\mathbf{B} \perp \mathbf{s}_0$ (the initial spin), the component of the spin in beam direction (x) is given by

$$s_x(B) = s_0 \cos \omega_L \tau \quad \text{with} \quad \omega_L = \mu_B \, g B/\hbar, \tag{37}$$

if all electrons in a photo-excited ensemble have approximately the same descent time τ; μ_B is the Bohr magneton,[10] and g is the Landé factor. The value of τ and hence $s_x(B)$ determines the polarization of the luminescence. The average of the s_x component depends on the spin lifetime, i.e., what fraction of a Larmor revolution or how many revolutions can be completed before spin relaxation (τ_s) or electron recombination (τ_n) occurs:

$$\tau^{-1} = \tau_n^{-1} + \tau_s^{-1}. \tag{38}$$

Consequently, we obtain for the component of the electron spin

$$s_x(B) = \frac{s_x(B=0)}{1 + (\omega_L \tau)^2}, \text{with} \;\; s_x(B=0) = \frac{s_0}{1 + \tau_n/\tau_s}. \tag{39}$$

Since the degree of circular polarization ρ of the luminescence is equal to the component of the average spin in the direction of the exciting beam, we can obtain the *electron lifetime* from Eqs. 37 and 39 as

$$\tau_n = \frac{\rho_0}{\rho(B=0)} \; \frac{\hbar}{g \mu_B B_{1/2}}, \tag{40}$$

[10]The Larmor frequency is equivalent to the cyclotron frequency for free electrons, with $\omega_c = (2/g)\mu_L$ and g as the Landé factor (g factor), which is given for isolated atoms by Eq. 4 of ▶ Chap. 9, "Magnetic Semiconductors"; for electrons in a semiconductor the g factor is influenced by the spin-orbit splitting of the valence band, and can have substantially different values – see in ▶ Chap. 25, "Carriers in Magnetic Fields and Temperature Gradients", Sect. 2.2 and Eq. 53.

and the *spin-relaxation time* as

$$\tau_s = \frac{\rho_0}{\rho_0 - \rho(B=0)} \quad \frac{\hbar}{g\mu_B B_{1/2}}, \tag{41}$$

where $B_{1/2}$ is the value of the magnetic induction at which the degree of polarization is reduced to one half of its original value at $B = 0$, and with the degree of circular polarization under σ^+ excitation

$$\rho = \frac{I(\sigma^+) - I(\sigma^-)}{I(\sigma^+) + I(\sigma^-)}; \tag{42}$$

here $I(\sigma^+)$ and $I(\sigma^-)$ are the intensities of the right and left polarized luminescence. This is an interesting example of obtaining important relaxation times from steady-state experiments. The Hanle effect has been used to study the spin relaxation of electrons in a variety of semiconductors (Dyakonov and Perel 1984). More recent measurements of electron-spin relaxation-times in *n*-type GaAs using nearly perturbation free spin-noise spectroscopy obtained larger values (up to factor 2) than Hanle data; this is attributed to an enhanced averaging of the hyperfine interaction by additional free carrier excitation in Hanle experiments, and hence smaller values (Römer et al. 2010).

3.2 Spin-Relaxation of Free Electrons in Bulk Semiconductors

Relaxation in *p*-Type Semiconductors When exciting a *p*-type semiconductor with circularly polarized light, practically all conduction-band electrons are created by this light. By measuring the degree of circular polarization of the band-to-band (edge) luminescence, the average spin of these electrons in the conduction band is measured. In a cubic semiconductor, the maximum value of the degree of circular polarization of this luminescence in the absence of electron-spin relaxation for transitions between a Γ_8 valence band (heavy and light holes) and the Γ_7 conduction band is equal to the initial average value of electron spin in such transitions (see Dyakonov and Perel 1984).

$$\rho_{\max} = 0.25. \tag{43}$$

With spin relaxation, the degree of circular polarization is reduced to

$$\rho = \rho_0 \frac{1}{1 + \tau_n/\tau_s}, \text{where } \rho_0 = \rho_{\max}. \tag{44}$$

The measured electron lifetime and spin-relaxation time in the *p*-type zincblende semiconductors AlGaAs and GaSb are shown in Fig. 29 as a function of the temperature and acceptor density (at 4.2 K). These relaxation times vary widely,

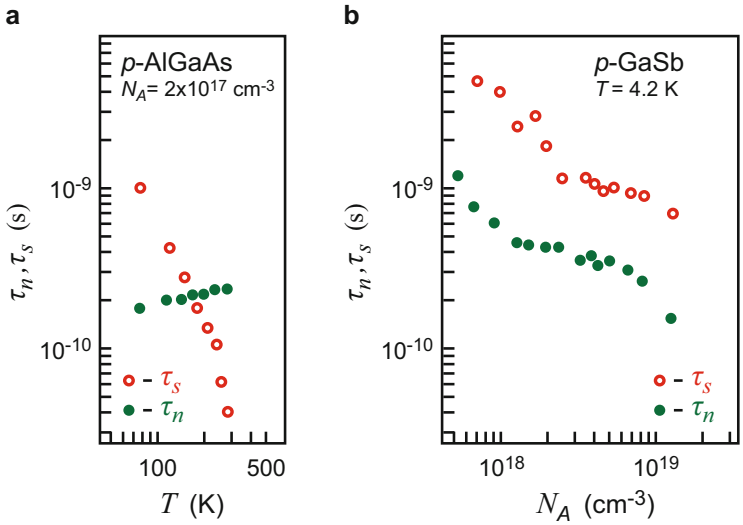


Fig. 29 Electron lifetime and spin relaxation time (**a**) as a function of the temperature in *p*-type AlGaAs for an acceptor concentration $N_A = 2 \times 10^{17}$ cm^{-3} (after Garbuzov et al. 1971) and (**b**) as a function of the doping density in GaSb at $T = 4.2$ K (after Titkov et al. 1981)

depending on the type of the most active relaxation mechanism, and lie typically between 10^{-8} and 10^{-11} s.

Spin relaxation can be used to show that in highly doped degenerate *p*-type semiconductors the interaction of electrons with thermalized, nonpolarized holes dominate: here, the spin relaxation has a temperature dependence consistent with the theory of electron-hole relaxation (Pikus and Titkov 1984).

Relaxation in *n*-Type Semiconductors Long spin-relaxation times up to 200 ns were observed in bulk *n*-type GaAs at low-temperatures, where electrons are bound to their donor dopants (Dzhioev et al. 2002a). At lowest doping levels the spin lifetime is limited by stochastic precession of electron spins in random magnetic fields created by the spins of lattice nuclei. The isolated donor-bound electrons interact independently with a large number of nuclei ($\sim 10^5$), which create a "hyperfine" field with a mean-squared fluctuation corresponding to the effect of $\sim\sqrt{10^5} \cong 300$ spins, amounting in GaAs to a magnetic field of ~4.3 kA/m (Dzhioev et al. 2002b). Larmor precession in this fluctuating field relaxes the spin. At higher donor concentration (still in the low-doping regime) the electron wavefunctions start to overlap, resulting in a dynamical averaging of the hyperfine interaction and consequently to a smaller effect of nuclear-spin fluctuations; the spin-relaxation time hence increases; see Fig. 30.

When at higher donor concentrations the overlap of electron wavefunctions becomes stronger, the anisotropic exchange interaction couples electrons of different donors which are randomly distributed. The related distribution of spin-rotation axes

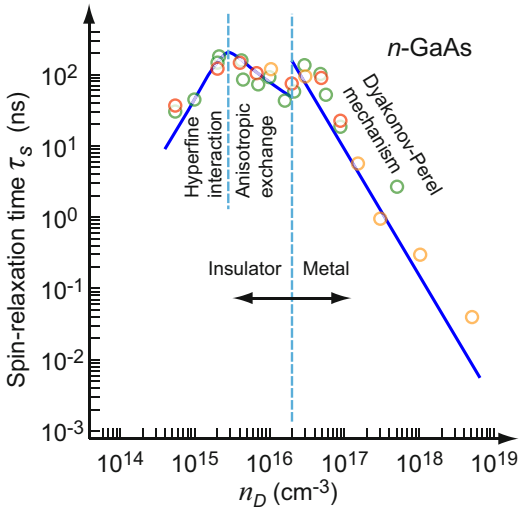


Fig. 30 Spin-relaxation time τ_s as a function of donor concentration in *n*-type GaAs. The *solid lines* refer to theory (After Dzhioev et al. 2002a and Kikkawa and Awschalom 1998)

leads to the relaxation of the total spin of the donor-bound electrons and a consequent decrease of the relaxation time for increased doping.

At high donor concentrations N_D exceeding 2×10^{16} cm^{-3} GaAs gets degenerate producing delocalized electrons. Due to Fermi statistics only electrons at the Fermi edge within the conduction band have a nonzero average spin, as observed in the polarization of the photoluminescence (Dzhioev et al. 2002a). In this metallic region an increase of the donor and related electron concentration leads to an increase of the Fermi wavevector, and consequently to a stronger effect of the Dyakonov-Perel mechanism decreasing the spin-relaxation time (Fig. 30).

3.3 Spin-Relaxation of Free Carriers and Excitons in Quantum Wells

The spin-relaxation in two-dimensional systems differs from that of bulk semiconductors. Localization in one direction alters the wavefunctions and energies of carriers and the related transitions. In addition, quantum wells are generally strained; in cubic semiconductors the degeneracy of heavy and light holes at the zone center is consequently lifted, even if the effective masses were equal. Usually the quantum well (QW) is compressively strained, yielding an upper heavy-hole and a lower light-hole valence band (in the electron coordinate system). Localization effects due to interface roughness of thickness and composition fluctuations add additional constraints.

Using doped quantum wells by applying either *p*-type or *n*-type modulation doping, the spin-relaxation dynamics of electrons and holes can be studied separately. The dynamics of excitons is investigated using *undoped* quantum wells; often GaAs (multi-) QWs clad by $Al_{0.3}Ga_{0.7}As$ barriers are used, since this model system

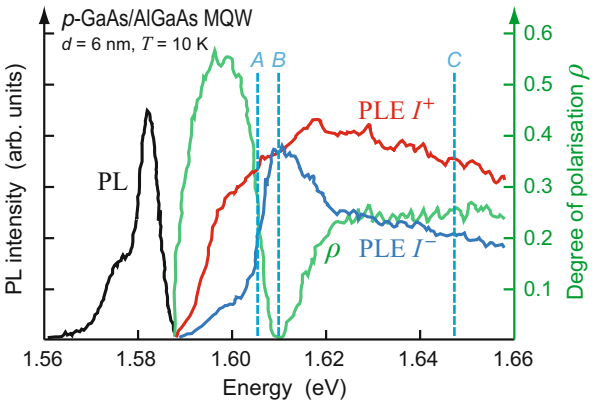


Fig. 31 σ^+ Photoluminescence (*PL*), PL excitation (*PLE*) with σ^+ light (I^+) or σ^- light (I^-), and polarization (ρ) spectra of a *p*-type GaAs/AlGaAs QW. Energy *A* refers to the Fermi edge of the highly doped sample, *B* indicates the energy of the light-hole exciton; in the range of energy *C* the heavy-hole exciton dominates (after Viña et al. 1992).

can be fabricated with exceptionally sharp interfaces preventing unintended lateral localization. For reviews see Viña (1999), Harley (2008), and Amand and Marie (2008).

Spin-Relaxation of Electrons The optical low-temperature spectra of a modulation-doped MQW sample with 6 nm wide *p*-type quantum wells given in Fig. 31 show a strong circular σ^+ luminescence for excitation with σ^+ light (I^+ curve) near the Fermi edge marked *A*. The σ^+ excitation near energy *A* creates heavy holes with $-3/2$ spin and electrons with $-1/2$ spin. The optically created holes (excitation density is 2×10^{10} cm^{-2}) are negligible compared to the holes produced by doping (4×10^{11} cm^{-2}); the polarization ρ of the band-to-band luminescence given by Eq. 42 hence reflects the spin of the electrons (minority carriers), and the decay of ρ their spin relaxation. A purely exponential decay with a lifetime of 150 ps for σ^+excitation at energy *A* was observed for this sample (Damen et al. 1991a). This is about four times shorter than in bulk *p*-type GaAs of this doping level; the Bir-Aronov-Pikus mechanism was assigned to this reduction.

At energy *B* there is no spin polarization ρ observed, i.e., $I(\sigma^+) = I(\sigma^-)$ in Eq. 42. The reason is a nearly equal population of spin $+1/2$ and $-1/2$ electrons at this light-hole exciton excitation energy.

Spin-Relaxation of Holes The investigation of the hole-spin relaxation is analogous to that for electrons described above. The optical low-temperature spectra of a single modulation-doped *n*-type 7.5 nm wide GaAs/AlGaAs QW implemented in a *pin* diode are given in Fig. 32 (Roussignol et al. 1992). The diode bias allows for an independent control of the electron sheet density in the quantum well. The excitation spectrum (PLE) of the band-to-band emission shows resonances of the e_1-hh_1, e_2-hh_1, and e_1-lh_1 transitions between light (*lh*) and heavy holes (*hh*) in the ground state and electrons (*e*) in the ground or excited states. The two heavy-hole resonances appear also in the cw polarization ρ_{cw}, while a steep decrease is observed at the e_1-lh_1 light-hole transition. Optical σ^+ excitation at the e_1-hh_1 resonance creates spin $-1/2$

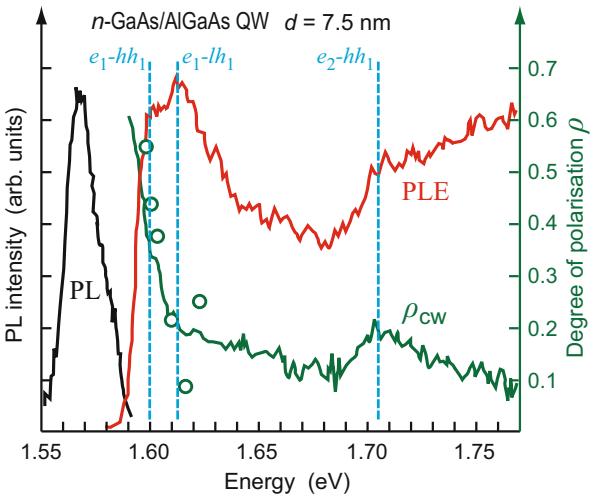


Fig. 32 PL, PLE, and cw polarization spectra of a 7.5 nm *n*-type GaAs/AlGaAs quantum well in a 6 V biased *pin* diode at $T = 2 \ldots 4$ K; *circles* indicate initial polarization at pulsed excitation. Energies of relevant optical transitions are indicated by *dashed vertical lines* (After Roussignol et al. 1992)

electrons and $-\frac{3}{2}$ heavy holes with a wavevector equals the Fermi wavevector, $k_{\|} = k_{\mathrm{F}}$, since conduction-band states with smaller wavevectors are occupied in the 2D electron gas; no light holes are created at this energy. The optically created electrons are negligible compared to those produced by doping, so that the polarization ρ reflects the spin of the minority holes, and the decay of ρ their spin relaxation. In pulsed measurements the observed initial polarization $\rho_{\mathrm{pulse}}(t = 0)$ is similar to the steady-state polarization ρ_{cw}. These measurements yield a hole-spin relaxation time $\tau_s^{(\mathrm{hole})}$ of 1 ns, in the investigated range being independent on the electron density; this time is much longer than the bias-dependent recombination time τ_{r} (300 . . . 100 ps). Finite hole-relaxation times, albeit mostly smaller, were also reported from other studies on QWs (typically on the order of a few tens of picoseconds; see Bar-Ad and Bar-Joseph 1992; Koudinov et al. 2016). These relaxation times are in clear contrast to the sub-picosecond hole-spin relaxation in cubic bulk semiconductors with their degenerate valence band in the zone center.

At increased electron density the polarization ρ decreases substantially. This is due to the creation of heavy holes with increasingly larger wavevector $k_{\|}$ at higher electron density, while always *detecting* the hole population at $k_{\|} \cong 0$; a significant fraction of the polarization is hence lost by scattering during relaxation to the band edge (Roussignol et al. 1992). In addition, spin states in the valence band are generally an admixture of various spin states for $k \neq 0$; any momentum or energy relaxation will hence relax the hole spins.

Spin-Relaxation of Excitons Spin relaxation of an exciton is affected by the spin relaxation of both, the electron and the hole, which are mutually bound by Coulomb interaction. For weak electron-hole interaction, electron and hole basically conserve their individual relaxation times. Since spin flip of *both* particles is required for observing an optically active (bright) exciton with total angular momentum $\langle J_z = \pm 1|$, the particle with slower spin relaxation – typically the electron – determines

the measured spin-relaxation time (indirect mechanism with sequential spin flips). In case of strong electron-hole interaction, both electron and hole flip simultaneously (direct mechanism), yielding an average relaxation time between that of the faster and of the slower particle.

For resonant excitation a slower spin-relaxation is observed than for nonresonant excitation at the same exciton density (50 ps vs. ∼10 ps; Damen et al. 1991b). This effect is assigned to the nonresonant formation of excitons at large wavevectors $K_{\|}$ as discussed in Sect. 2.2. Large $K_{\|}$ mean particularly large (heavy-) hole-wavevectors due to their larger effective mass. These holes have a strong spin-orbit mixing, leading to a high spin-relaxation rate. In contrast, excitons created resonantly have wavevectors close to $K_{\|} = 0$, and their holes have a weaker spin-orbit mixing.

Nonresonant excitation of GaAs/AlGaAs quantum wells shows that the spin-relaxation time of excitons increases with increasing density, varying from ≤ 20 ps for a sheet density of 5×10^8 cm^{-2} to 150 ps for 1×10^{10} cm^{-2} (Damen et al. 1991a, b; Viña et al. 1992). The increase of the relaxation time is due to the decrease of exciton-binding energy at higher density and the related reduced overlap of the electron and hole wavefunctions, leading to a reduced exchange interaction.

An additional lateral confinement of excitons creates *quantum dots*. Such zero-dimensional structures may be formed from unintentional lateral inhomogeneities of a quantum well, leading to weakly confining interfacial quantum dots, or from fabricated more strongly confining structures applying self-organization or other techniques; see Sect. 2.2.3 of ▶ Chap. 3, "The Structure of Semiconductors". The discrete energy spectrum as well as absence of translational motion of the particles and related relaxation mechanisms lead to longer spin lifetimes than in 2D and 3D semiconductors. The lateral confinement increases the exchange interaction of electron and hole, and leads thereby to a dependence of spin-relaxation times on the size of the confinement potential. The effect is pronounced for strong lateral confinement. The relaxation rates increase severely upon reduction in the dot size; for typical InAs/GaAs quantum dots, numerical estimates range from hundreds of microseconds for large dots to few tens of nanoseconds for small dots at low temperatures (Tsitsishvili and Kalt 2010; Wei et al. 2014). Besides size also the shape and composition of the quantum dots has a substantial influence, and experimental data show a large scatter; also a strong effect of dark-exciton states is observed (Johansen et al. 2010). Details on the complex spin dynamics of excitons are discussed in the review by Marie et al. (2008).

4 Summary

When any external parameter such as the temperature, an electric or magnetic field, or the optical generation rate is changed as a function of time, carriers in the semiconductor respond on this disturbance by a redistribution controlled by relaxation times. Relaxation times refer to elastic or inelastic scattering of carriers with other carriers, with phonons, with defects, or with spin momenta, and span time constants from femtoseconds up to geologic times.

When carriers are injected by a bias, their relaxation time is given by the carrier lifetime within the band, and relates to a corresponding diffusion or drift length. The lifetime of a carrier cloud depends on trapping, yielding an exponential decay of a current pulse for a single type of trap and the more common dispersive decay (usually following a power law) for a distribution of traps. When carriers are heated – e.g., by intense light pulses – the energy exceeding the thermal energy is transferred from electrons to the lattice mostly by LO phonons, and from holes primarily by TO phonons. At high optical excitation rates other coupling mechanisms become operative, and involve plasmons for a rapid first cooling step before LO phonons become dominant. The close interaction between electrons and holes in such a plasma has an influence of cooling the hotter electron gas because of $m_n < m_p$. Most of these fast processes are accessible either through ultrafast laser spectroscopy in the sub-ps range using correlation techniques, or from the broadened spontaneous luminescence signal. At very high carrier-generation rates and low lattice temperature, condensation into electron-hole droplets takes place, with droplet evaporation into an exciton gas and exciton annihilation determining the two relaxation mechanisms.

Optical phonons, excited directly by an IR light pulse (TO phonons) or indirectly by fast electrons (LO phonons) or holes (TO phonons), relax their momenta by elastic scattering with phonons in the same branch, or undergo an energy relaxation by scattering with a decay into two acoustical phonons. Energy relaxation of optical phonons can be measured by Raman spectroscopy, and is typically on the order of 10 ps.

Free excitons, observed in bulk semiconductors of very high purity, recombine radiatively only close to the center of the Brillouin zone. Resonant optical excitation creates exciton-polaritons, which convert to photon-like polaritons by phonon scattering before recombination. Nonresonant excitation creates nonthermal excitons via hot carriers; the balance between excitons and the electron-hole plasma leads to a slow rise time after pulsed excitation and an observed decay time depending on the excitation density. In quantum wells optical excitation creates excitons with well-defined in-plane wavevectors. Time constants of the radiative decay are usually smaller than those of respective bulk excitons. Nonresonant pulse excitation creates initially predominantly excitons with high momenta and leads to a slow rise time in the PL given by the balance to uncorrelated electron-hole pairs.

The spin of electrons, holes, and electron-hole pairs and their orbital momenta are coherently aligned by excitation in polarized light; the photon angular momentum is conserved by the excited electron-hole pair. With subsequent scattering, relaxation occurs, which can be detected by the degree of polarization of the ensuing recombination luminescence. Orbital and spin relaxation have short relaxation times, typically in the sub-ns range and decreasing at higher temperature and on doping, since they are sensitive to any of the various scattering events and wavefunction overlap. Holes in semiconductors with degenerate valence bands in the zone center have particularly short relaxation times in the sub-ps range; lifting this degeneracy in quantum wells yields larger times on the order of some tens of ps. Spin relaxation of an exciton occurs either indirectly with a sequential spin flip of electron and hole or

directly by a simultaneous flip; for weak electron-hole interaction the relaxation time is controlled by the slower particle, while an average relaxation time in the direct mechanism results for strong interaction.

References

Amand T, Marie X (2008) Exciton spin dynamics in semiconductor quantum wells. In: Dyakonov MI (ed) Spin physics in semiconductors. Springer, Berlin/Heidelberg, pp 55–89

Amo A, Martín MD, Viña L, Toropov AI, Zhuravlev KS (2006) Interplay of exciton and electron-hole plasma recombination on the photoluminescence dynamics in bulk GaAs. Phys Rev B 73:035205

Andreani CL, Pasquarello A (1990) Accurate theory of excitons in GaAs-$Ga_{1-x}Al_x$ quantum wells. Phys Rev B 42:8928

Arya K, Hanke W (1981) Many-body coulomb effects on the gain and absorption line shapes of the electron-hole plasma in GaAs. Phys Rev B 23:2988

Awschalom DD, Samarth N (2002) Optical manipulation, transport, and storage of spin coherence in semiconductors. In: Awschalom DD, Loss D, Samarth N (eds) Semiconductor spintronics and quantum computation. Springer, Berlin, pp 147–193

Bar-Ad S, Bar-Joseph I (1992) Exciton spin dynamics in GaAs heterostructures. Phys Rev Lett 68:349

Barker JR (1980) Quantum transport theory. In: Ferry DK, Barker JR, Jacoboni C (eds) Physics of nonlinear transport in semiconductors. Plenum Press, New York, pp 126–152

Barker JR, Ferry DK (1980) On the physics and modeling of small semiconductor devices I. Solid State Electron Dev 23:519

Barman S, Srivastava GP (2004) Long-wavelength nonequilibrium optical phonon dynamics in cubic and hexagonal semiconductors. Phys Rev B 69:235208

Bauer G (1978) Experimental aspects of hot electron distribution functions. Solid State Electron 21:17

Baumberg JJ, Awschalom DD, Samarth N, Luo H, Furdyna JK (1994) Spin beats and dynamical magnetization in quantum structures. Phys Rev Lett 72:717

Beck M, Hübner J, Oestreich M, Bieker S, Henn T, Kiessling T, Ossau W, Molenkamp LW (2016) Thermodynamic origin of the slow free exciton photoluminescence rise in GaAs. Phys Rev B 93:081204

Bir GL, Aronov AG, Pikus GE (1976) Spin relaxation of electrons due to scattering by holes. Sov Phys JETP 42:705

Brooks H (1955) Theory of the electrical properties of germanium and silicon. Adv Electr Electron Phys 7:85

Chang Y-M, Gwo S (2007) Direct measurement of momentum relaxation time in wurtzite InN. Applied Phys 102:083540

Citrin DS (1993) Radiative lifetimes of excitons in quantum wells: localization and phase-coherence effects. Phys Rev B 47:3832

Collins CL, Yu PY (1984) Generation of nonequilibrium optical phonons in GaAs and their application in studying intervalley electron-phonon scattering. Phys Rev B 30:4501

Conwell EM (1982) Transport: the Boltzmann equation. In: Moss TS, Paul W (eds) Handbook of semiconductors, vol 1. Band theory and transport properties. North Holland Publ, Amsterdam, pp 513–561

Damen TC, Viña L, Cunningham JE, Shah J, Sham LJ (1991a) Subpicosecond spin relaxation dynamics of excitons and free carriers in GaAs quantum wells. Phys Rev Lett 67:3432

Damen TC, Leo K, Shah J, Cunningham JE (1991b) Spin relaxation and thermalization of excitons in GaAs quantum wells. Appl Phys Lett 58:1902

Das Sarma S, Mason BA (1985) Screening of polar interaction in quasi-two-dimensional semiconductor microstructures. Phys Rev B 31:5536
Debernardi A (1998) Phonon linewidth in III-V semiconductors from density-functional perturbation theory. Phys Rev B 57:12847
Dyakonov MI (ed) (2008) Spin physics in semiconductors. Springer, Berlin/Heidelberg
Dyakonov MI, Perel VI (1971) Spin orientation of electrons associated with the interband absorption of light in semiconductors. Sov Phys JETP 33:1053
Dyakonov MI, Perel VI (1972) Spin relaxation of conduction electrons in noncentrosymmetric semiconductors. Sov Phys Solid State 13:3023
Dyakonov MI, Perel VI (1984) Theory of optical spin orientation of electrons and nuclei in semiconductors. In: Meier F, Zakharchenya BP (eds) Optical orientation. North Holland, Amsterdam, pp 11–72
Dymnikov VD, Mirlin DN, Perel VI, Reshina II (1978) Linear polarization of hot photoluminescence of gallium arsenide crystals. Sov Phys Sol State 20:1250
Dzhioev RI, Kavokin KV, Korenev VL, Lazarev MV, Meltser BY, Stepanova MN, Zakharchenya BP, Gammon D, Katzer DS (2002a) Low-temperature spin relaxation in *n*-type GaAs. Phys Rev B 66:245204
Dzhioev RI, Korenev VL, Merkulov IA, Zakharchenya BP, Gammon D, Efros AL, Katzer DS (2002b) Manipulation of the spin memory of electrons in *n*-GaAs. Phys Rev Lett 88:256801
Eastman LF (1982) Very high electron velocity in short gallium arsenide structures. In: Grosse P (ed) Festkörperprobleme, vol 22. Advances in solid state physics. Vieweg, Braunschweig, pp 173–187
Elliot RJ (1954) Theory of the effect of spin-orbit coupling on magnetic resonance in some semiconductors. Phys Rev 96:266
Elsaesser T, Leitenstorfer A, Kuhn T, Rossi F (1996) Ultrafast dynamics of electronic excitations in semiconductors. Prog Crystal Growth Character Mater 33:41
Fasol G, Hughes HP (1986) Band-structure determination of GaAs from hot-electron luminescence. Phys Rev B 33:2953
Ferry DK (1978) Energy-gap narrowing and state filling in semiconductors under intense laser irradiation. Phys Rev B 18:7033
Ferry DK (1980) Modeling of carrier transport in the finite collision duration regime: effects in submicron semiconductor devices. In: Ferry DK, Barker JR, Jacoboni C (eds) Physics of nonlinear transport in semiconductors. Plenum Press, New York, pp 577–588
Ferry DK (1991) Semiconductors. Macmillian, New York
Ferry DK (2021) Non-equilibrium longitudinal optical phonons and their lifetimes. Appl Phys Rev 8:021324
Ferry DK, Barker JR (1981) Generalized diffusion, mobility, and the velocity autocorrelation function for high-field transport in semiconductors. J Appl Phys 52:818
Ferry DK, Grubin HL, Iafrate GJ (1984) Transient transport in semiconductors and submicron devices. In: Alfano RR (ed) Semiconductors probed by ultrafast laser spectroscopy, vol 1. Academic Press, New York, pp 413–447
Fouquet JE, Burnham RD (1986) Recombination dynamics in GaAs/$Al_xGa_{1-x}As$ quantum well structures. IEEE J Quantum Electron QE 22:1799
Fröhlich H (1937) Theory of electrical breakdown in ionic crystals. Proc R Soc Lond A 160:280
Funk S, Acuna G, Handloser M, Kersting R (2009) Probing the momentum relaxation time of charge carriers in ultrathin layers with terahertz radiation. Opt Express 17:17450
Gale GM, Laubereau A (1983) Direct measurement of picosecond and sub-picosecond phonon lifetimes in α-quartz. Opt Commun 44:273
Ganikhanov F, Vallée F (1997) Coherent TO phonon relaxation in GaAs and InP. Phys Rev B 55:15614
Garbuzov DZ, Ekimov AI, Safarov VI (1971) Measurement of the lifetime and of the spin-relaxation time of electrons in semiconductors by the optical-orientation method. Sov Phys JETP Lett 13:24

Göbel EO, Kuhl J, Hoger R (1986) Short pulse physics of quantum well structures. J Lumin 30:541
Graudszus W, Göbel EO (1981) Picosecond luminescence studies of hot carrier relaxation in pure and highly doped GaAs. J Physique 42(Colloq C7):437
Graudszus W, Göbel EO (1983) Free carrier screening of the Fröhlich interaction in GaAs. Physica B 117:555
Gupta A, Heremans JJ, Kataria G, Chandra M, Fallahi S, Gardner GC, Manfra MJ (2021) Hydrodynamic and ballistic transport over large length scales in GaAs/AlGaAs. Phys Rev Lett 126:076803
Gurioli M, Borri P, Colocci M, Gulia M, Rossi F, Molinari E, Selbmann PE, Lugli P (1998) Exciton formation and relaxation in GaAs epilayers. Phys Rev B 58:13403
Hanamura E (1988) Rapid radiative decay and enhanced optical nonlinearity of excitons in a quantum well. Phys Rev B 38:1228
Hanle W (1924) Über magnetische Beeinflussung der Polarisation der Resonanzfluoreszenz. Z Physik 30:93. (On the magnetic influence of polarization of the resonance fluorescence, in German)
Harley RT (2008) Spin dynamics of free carriers in quantum wells. In: Dyakonov MI (ed) Spin physics in semiconductors. Springer, Berlin/Heidelberg, pp 29–54
Haug H, Schmitt-Rink S (1985) Basic mechanisms of the optical nonlinearities of semiconductors near the band edge. J Opt Soc Am B 2:1135
Haug H, Tran Thoai DB, Schmitt-Rink S, Bohnert K, Klingshirn C, Blattner G (1980) The electron-hole plasma in direct II-VI compounds, Proc 15th Int Conf Phys Semicond Kyoto. J Phys Soc Jpn Suppl A 49:503
Haug H, Koch SW (1990) Quantum theory of optical and electronic properties of semiconductors. World Scientific, Singapore
Haynes JR, Shockley W (1951) The mobility and life of injected holes and electrons in germanium. Phys Rev 81:835
Hearn CJ (1980) Physics of nonlinear transport in solids. Plenum Press, New York
Heiblum M, Nathan MI, Thomas DC, Knoedler CM (1985) Direct observation of ballistic transport in GaAs. Phys Rev Lett 55:2200
Henderson GN, Gaylord TK, Glytsis EN (1993) Diffraction of ballistic electrons by semiconductor gratings: rigorous analysis, approximate analyses, and device design. IEEE J Quantum Electron QE 29:121
Henry CH, Nassau K (1970) Lifetimes of bound excitons in CdS. Phys Rev B 1:1628
Hilton DJ, Tang CL (2002) Optical orientation and femtosecond relaxation of spin-polarized holes in GaAs. Phys Rev Lett 89:146601
Höpfel RA, Shah J, Gossard AC (1986) Nonequilibrium electron-hole plasma in GaAs quantum wells. Phys Rev Lett 56:765
Irmer G, Wenzel M, Monecke J (1996) The temperature dependence of the LO(Γ) and TO(Γ) phonons in GaAs and InP. Phys Stat Sol B 195:85
Jiang JH, Wu MW (2009) Electron-spin relaxation in bulk III-V semiconductors from a fully microscopic kinetic spin Bloch equation approach. Phys Rev B 79:125206
Johansen J, Julsgaard B, Stobbe S, Hvam JM, Lodahl P (2010) Probing long-lived dark excitons in self-assembled quantum dots. Phys Rev B 81:081304
Jonscher AK (1983) Dielectric relaxation in solids. Chelsea Dielectric Press, London
Kaindl RA, Hägele D, Carnahan MA, Chemla DS (2009) Transient terahertz spectroscopy of excitons and unbound carriers in quasi-two-dimensional electron-hole gases. Phys Rev B 79:045320
Kalt H (1994) The electron-hole plasma and liquid in confined semiconductor systems. J Lumin 60:262
Kash JA (1989) Carrier-carrier scattering in GaAs: quantitative measurements from hot (e,A^0) luminescence. Phys Rev B 40:3455
Kash JA, Tsang JC, Hvam JM (1985) Subpicosecond time-resolved Raman spectroscopy of LO phonons in GaAs. Phys Rev Lett 54:2151

Keldysh LV (1986) The electron-hole liquid in semiconductors. Contemp Phys 27:395
Keldysh LV (1997) Excitons in semiconductor-dielectric nanostructures. Phys Stat Sol A 164:3
Kikkawa JM, Awschalom DD (1998) Resonant spin amplification in n-Type GaAs. Phys Rev Lett 80:4313
Kim D-S, Yu PY (1990) Phonon temperature overshoot in GaAs excited by subpicosecond laser pulses. Phys Rev Lett 64:946
Kim D-S, Yu PY (1991) Hot-electron relaxations and hot phonons in GaAs studied by subpicosecond Raman scattering. Phys Rev B 43:4158
Kleinman DA, Miller RC (1981) Relaxation of optically pumped electron spins through a virtual photon: experimental evidence in heavily Zn-doped GaAs. Phys Rev Lett 46:68
Knox WH, Hirlimann C, Miller DAB, Shah J, Chemla DS, Shank CV (1986) Femtosecond excitation of nonthermal carrier populations in GaAs quantum wells. Phys Rev Lett 56:1191
Koudinov AV, Dzhioev RI, Korenev VL, Sapega VF, Kusrayev YG (2016) Optical spin orientation of minority holes in a modulation-doped GaAs/(Ga,al)as quantum well. Phys Rev B 93:165301
Kreuzer HJ (1981) Non-equilibrium thermodynamics and its statistical foundation. Claredon, Oxford
Kuhl J, von der Linde D (1982) Picosecond phenomena vol III. Springer, Berlin
Lampel G (1974) Optical pumping in semiconductors. In: Pilkuhn MH (ed) Proc 12th Int conf phys semicond. Teubner Verlag, Stuttgart, pp 743–750
Laubereau A (1984) Semiconductors probed by ultrafast laser spectroscopy vol I. Academic Press, New York
Laubereau A, Kaiser W (1974) Generation and applications of passively mode-locked picosecond light pulses. Optoelectronics 6:1
Leheny RF, Shah J, Fork RL, Shank CV, Migus A (1979) Dynamics of hot carrier cooling in photoexcited GaAs. Sol State Commun 31:809
Leitenstorfer A, Lohner A, Elsaesser T, Haas S, Rossi F, Kuhn T, Klein W, Boehm G, Traenkle G, Weimann G (1994) Ultrafast coherent generation of hot electrons studied via band-to-acceptor luminescence in GaAs. Phys Rev Lett 73:1687
Leo K (1993) Quantum beats in quantum wells. In: Henneberger F, Schmitt-Rink S, Göbel EO (eds) Optics of semiconductor nanostructures. Akademie Verlag, Berlin, pp 127–148
Levi AFJ, Hayes JR, Bhat R (1986) "ballistic" injection devices in semiconductors. Appl Phys Lett 48:1609
Lugli P, Bordone P, Reggiani L, Rieger M, Kocevar P, Goodnick SM (1989) Monte Carlo studies of nonequilibrium phonon effects in polar semiconductors and quantum wells. I. Laser photoexcitation. Phys Rev B 39:7852
Lugli P, Bordone P, Molinari E, Rücker H, de Paula AM, Maciel AC, Ryan JF, Shayegan M (1992) Interaction of electrons with interface phonons in GaAs/AlAs and GaAs/AlGaAs heterostructures. Semicond Sci Technol B 7:116
Luzzi R, Vasconcellos AR (1984) Relaxation processes in non-equilibrium semiconductor plasma. In: Alfano RR (ed) Semiconductors probed by ultrafast laser spectroscopy, vol 1. Academic Press, Orlando, pp 135–169
Lyon SA (1986) Spectroscopy of hot carriers in semiconductors. J Lumin 35:121
Mahr H, Hirsch MD (1975) An optical up-conversion light gate with picosecond resolution. Opt Commun 13:96
Maloney TJ, Frey J (1977) Transient and steady-state electron transport properties of GaAs and InP. J Appl Phys 48:781
Malvezzi AM (1987) Interaction of picosecond laser pulses with solid surfaces. Proc SPIE 793:49
Manenkov AA, Milyaev VA, Mikhailova GN, Sanina VA, Seferov AS (1976) High-frequency breakdown of excitons and kinetics of free carriers and excitons in germanium in the presence of electron-hole drops. Sov Phys JETP 43:359
Marie X, Urbaszek B, Krebs O, Amand T (2008) Exciton-spin dynamics in semiconductor quantum dots. In: Dyakonov MI (ed) Spin physics in semiconductors. Springer, Berlin/Heidelberg, pp 91–113

Mirlin DN (1984) Optical alignment of electron momenta in GaAs-type semiconductors. In: Meier F, Zakharchenya BP (eds) Optical orientation. Elsevier Science, New York
Nag BR (1975) Microwave magnetoconductivity of polar semiconductors. J Appl Phys 46:4819
Nag BR (1984) Relaxation of momentum and energy of carriers in semiconductors. In: Alfano RR (ed) Semiconductors probed by ultrafast laser spectroscopy vol 1. Academic Press, Orlando, pp 3–44
Orbach R (1967) Phonon breakdown. IEEE Trans Sonics Ultrasonics 14:140
Oudar JL, Hulin D, Migus A, Antonetti A, Alexandre F (1985) Subpicosecond spectral hole burning due to nonthermalized photoexcited carriers in GaAs. Phys Rev Lett 55:2074
Parsons RR (1969) Band-to-band optical pumping in solids and polarized photoluminescence. Phys Rev Lett 23:1152
Penzkofer A, Laubereau A, Kaiser W (1979) High intensity Raman interactions. Prog Quantum Electron 5:55
Pfister G (1976) Dispersive low-temperature transport in α-selenium. Phys Rev Lett 36:271
Pfister G, Scher H (1977) Time-dependent electrical transport in amorphous solids: As_2Se_3. Phys Rev B 15:2062
Pikus GE, Titkov AN (1984) Spin relaxation under optical orientation in semiconductors. In: Meier F, Zakharchenya BP (eds) Optical orientation. North Holland, Amsterdam, pp 73–132
Price PJ (1985) Hot phonon effects in heterolayers. Physica B & C 134:164
Quinn JJ (1962) Range of excited electrons in metals. Phys Rev 126:1453
Rappel WJ, Feiner LF, Schuurmans MFH (1988) Exciton-polariton picture of the free-exciton lifetime in GaAs. Phys Rev B 38:7874
Rice TM (1977) The electron-hole liquid in semiconductors: theoretical aspects. In: Ehrenreich H, Seitz F, Turnbull D (eds) Solid state physics, vol 32. Academic Press, New York, pp 1–86
Römer M, Bernien H, Müller G, Schuh D, Hübner J, Oestreich M (2010) Electron-spin relaxation in bulk GaAs for doping densities close to the metal-to-insulator transition. Phys Rev B 81:075216.
Roussignol P, Rolland P, Ferreira R, Delalande C, Bastard G, Vinattieri A, Martinez-Pastor J, Carraresi L, Colocci M, Palmier J, Etienne B (1992) Hole polarization and slow hole-spin relaxation in an *n*-doped quantum-well structure. Phys Rev B 46:7292
Ruch JG (1972) Electron dynamics in short channel field-effect transistors. IEEE Trans Electron Devices ED 19:652
Ruf T, Belitsky VI, Spitzer J, Sapega VF, Cardona M, Ploog K (1993) Disorder-induced Raman scattering of folded phonons in quantum wells and superlattices. Sol State Electron 37:609
Rullière C (ed) (2005) Femtosecond laser spectroscopy, 2nd edn. Springer, New York
Ryan JF, Tatham MC (1992) Time-resolved Raman measurements of electron-phonon interactions in quantum wells and superlattices. In: Shah J (ed) Hot carriers in semiconductor nanostructures. Academic Press, San Diego, pp 345–378
Saito H, Göbel EO (1985) Picosecond spectroscopy of highly excited CdS. Phys Rev B 31:2360
Scher H, Montroll EW (1975) Anomalous transit-time dispersion in amorphous solids. Phys Rev B 12:2455
Schultheis L, Kuhl J, Honold A, Tu CW (1986) Picosecond phase coherence and orientational relaxation of excitons in GaAs. Phys Rev Lett 57:1797
Segall B, Mahan GD (1968) Phonon-assisted recombination of free excitons in compound semiconductors. Phys Rev 171:935
Shah J (1978) Hot electrons and phonons under high intensity photoexcitation of semiconductors. Solid State Electron 21:43
Shah J (1986) Hot carriers in quasi-2-D polar semiconductors. IEEE J Quantum Electron QE 22:1728
Shah J (1999) Ultrafast spectroscopy of semiconductors and semiconductor nanostructures, 2nd edn. Springer, Berlin

Shah J, Leheny RF (1984) Hot carriers in semiconductors probed by picosecond techniques. In: Alfano RR (ed) Semiconductors probed by ultrafast laser spectroscopy vol 1. Academic Press, Orlando, pp 45–75
Shah J, Pinczuk A, Gossard AC, Wiegmann W (1985) Energy-loss rates for hot electrons and holes in GaAs quantum wells. Phys Rev Lett 54:2045
Shah J, Deveaud B, Damen TC, Tsang WT, Gossard AC, Lugli P (1987) Determination of intervalley scattering rates in GaAs by subpicosecond luminescence spectroscopy. Phys Rev Lett 59:2222
Sham LJ (1993) Spin relaxation in semiconductor quantum wells. J Phys Condens Mater 5:A51
Shur MS, Eastman LF (1981) Near ballistic electron transport in GaAs devices at 77°K. Solid State Electron 24:11
Sibatov RT, Uchaikin VV (2009) Fractional differential approach to dispersive transport in semiconductors. Uspekhi Fizicheskikh Nauk 179:1079
Smith RA (1978) Semiconductors. Cambridge University Press, Cambridge, UK
Snelling MJ, Flinn GP, Plaut AS, Harley RT, Tropper AC, Eccleston R, Phillips CC (1991) Phys Rev B 44:11345
Song PH, Kim KW (2002) Spin relaxation of conduction electrons in bulk III-V semiconductors. Phys Rev B 66:035207
Srivastava GP (1990) The physics of phonons. Hilger, Bristol
Tanaka S, Kobayashi H, Saito H, Shionoya S (1980) Luminescence of high density electron-hole plasma in GaAs. J Phys Soc Jpn 49:1051
t'Hooft GW, van der Poel WAJA, Molenkamp LW, Foxon CT (1987) Giant oscillator strength of free excitons in GaAs. Phys Rev B 35:8281
t'Hooft GW, van der Poel WAJA, Molenkamp LW, Foxon CT (1988) True radiative lifetime of free excitons in GaAs. In: Del Sole R, D'Andrea A, Lapiccirella A (eds) Excitons in confined systems. Springer, Berlin, pp 59–62
Tiedje T (1984) Information about band-tail states from time-of-flight experiments. In: Willardson RK, Beer AC, Pankove J (eds) Semiconductors and semimetals, vol 21C. Academic Press, New York, pp 207–238
Titkov AN, Chaikina EI, Komova EM, Ermakova MG (1981) Low-temperature luminescence of degenerate *p*-type crystals of direct-gap semiconductors. Sov Phys Semicond 15:198
Tsen KT (1993) Electron-optical phonon interactions in polar semiconductor quantum wells. Int J Mod Phys B 7:4165
Tsitsishvili E, Kalt H (2010) Exciton spin relaxation in strongly confining semiconductor quantum dots. Phys Rev B 82:195315
Ulbrich RG (1978) Low density photoexcitation phenomena in semiconductors: aspects of theory and experiment. Sol State Electron 21:51
Ulbrich RG, Kash JA, Tsang JC (1989) Hot-electron recombination at neutral acceptors in GaAs: a cw probe of femtosecond intervalley scattering. Phys Rev Lett 62:949
Ulbricht R, Hendry E, Shan J, Heinz TF, Bonn M (2017) Carrier dynamics in semiconductors studied with time-resolved terahertz spectroscopy. Rev Mod Phys 83:543
Valdmanis JA, Fork RL, Gordon JP (1985) Generation of optical pulses as short as 27 femtoseconds directly from a laser balancing self-phase modulation, group-velocity dispersion, saturable absorption, and saturable gain. Opt Lett 10:131
Vashishta P, Kalia RK (1982) Universal behavior of exchange-correlation energy in electron-hole liquid. Phys Rev B 25:6492
Viña L (1999) Spin relaxation in low-dimensional systems. J Phys Condens Matter 11:5929
Viña L, Damen TC, Cunningham JE, Shah J, Sham LJ (1992) Spin relaxation dynamics in GaAs quantum wells: free carriers and excitons. Superlatt Microstruct 12:379
von der Linde D (1979) Picosecond spectroscopy: methods and applications. In: Treusch J (ed) Festkörperprobleme, vol 19. Advances in solid state physics. Vieweg, Braunschweig, pp 387–402

von der Linde D, Kuhl J, Klingenberg H (1980) Raman scattering from nonequilibrium LO phonons with picosecond resolution. Phys Rev Lett 44:1505

Wei H, Guo G-C, He L (2014) Slow exciton spin relaxation in single self-assembled $In_{1-x}Ga_xAs$/GaAs quantum dots. Phys Rev B 89:245305

Wiley JD (1975) Mobility of holes in III-V compounds. In: Willardson RK, Beer AC (eds) Semiconductors and semimetals, vol 10. Academic Press, New York, pp 91–174

Worlock JM, Damen TC, Shaklee KL, Gordon JP (1974) Determination of the optical properties and absolute concentrations of electron-hole drops in germanium. Phys Rev Lett 33:771

Wu MW, Jiang JH, Weng MQ (2010) Spin dynamics in semiconductors. Physics Reports 493:61

Yafet Y (1963) *g* factors and spin-lattice relaxation of conduction electrons. In: Seitz F, Turnbull D (eds) Solid state physics, vol 14. Academic Press, New York, pp 1–98

Yamamoto Y, Tassone F, Cao H (2000) Semiconductor cavity quantum electrodynamics. Springer, Berlin/New York

Yoffa EJ (1981) Screening of hot-carrier relaxation in highly photoexcited semiconductors. Phys Rev B 23:1909

Yoshida H, Saito H, Shionoya S (1981) Luminescence and inter-valence band hole relaxation in high density electron-hole plasma in CdSe and CdS. J Phys Soc Jpn 50:881

Zheludev NI, Brummell MA, Harley RT, Malinowski A, Popov SV, Ashenford DE, Lunn B (1994) Giant specular inverse faraday effect in $Cd_{0.6}Mn_{0.4}Te$. Sol State Commun 89:823

Zimmermann J, Lugli P, Ferry DK (1981) Non-equilibrium hot carrier diffusion phenomena in semiconductors I, a theoretical non-Markovian approach. J Physique 42(Colloq C7):95

Appendix: High-Field Domains

Karl W. Böer

Abstract
In regimes of over-linear negative differential conductivity of a bulk semiconductor, the field distribution must get inhomogeneous and divides into two parts: a domain with low conductivity and a high field, and the other part with a low field; this field distribution can be stationary or nonstationary. Both regions are joined by a narrow layer of only a few Debye width. The high-field domain can be imaged by applying the Franz-Keldysh effect. CdS has proved an ideal model semiconductor, where high-field domains can sweep the bulk free of space charges and connect the work functions of cathode and anode. The electrode-adjacent high field can set CdS into different thermodynamic stable states being *n*- or *p*-type, depending on whether the domain is attached to the cathode or anode. Within the domains, the crystal is free of space charges and permits to investigate sharp electronic transitions to defect levels around the quasi-Fermi levels without the interference from broadening electric fields.

Inhomogeneous Field or Current Distributions

Homogeneously doped semiconductors develop inhomogeneous distributions of carrier densities when this density at the metal-semiconductor boundary deviates from the density in the semiconductor bulk. Then, a space-charge layer is formed which creates inhomogeneous field and potential distributions near the electrode. When the deviation of the boundary electron density n_c from the bulk density n_{10} is significant, it influences the current through the semiconductor, reducing it by creating a Schottky barrier when $n_c \ll n_{10}$, and increasing it via carrier injection when $n_c \gg n_{10}$.

A different kind of inhomogeneous distribution of the electric field or the current density may occur in the regime of *nonlinear transport far from thermodynamic equilibrium*. In that case, the spatial patterns arise in a *self-organized* way, even in homogeneous semiconductors. Two types of nonlinear spatial patterns that are associated with specific field or current density inhomogeneities, respectively, are the topic of this appendix.

© Springer Nature Switzerland AG 2023

K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0

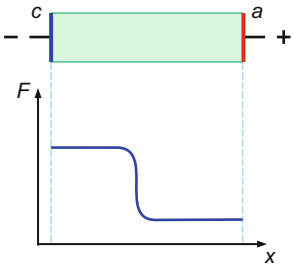


Fig. A.1 *Top*: Semiconductor between anode *a* and cathode *c* electrodes. *Bottom*: Electric field distribution in a semiconductor with a cathode-adjacent high-field domain

We will first ignore the boundary condition and deal with the semiconductor bulk. Close to thermodynamic equilibrium – that is, at sufficiently low bias voltage, the current density j as a function of the local electric field F in a homogeneous semiconductor is linear (*Ohm's law*), but at larger bias the $j(F)$ relation will generally become nonlinear and may display a regime of *negative differential conductivity*

$$\sigma_{\text{diff}} = \frac{dj}{dF} < 0. \tag{A.1}$$

Thus, the current density decreases with increasing field and vice versa, which in general corresponds to an unstable situation. Here, the originally homogeneous field and current density distributions may evolve into an inhomogeneous pattern, referred to as a *high-field domain* sketched in Fig. A.1. The high-field domains can be stationary, when they are attached to an electrode, or they can move from one electrode to the other.

Negative Differential Conductivity

In bulk semiconductors, the drift-current density of electrons can generally be written as

$$j = e\, n(F)\ \ v(F). \tag{A.2}$$

Here e is the electron charge, and $n(F)$ and $v(F)$ are the field-dependent electron concentration and drift velocity, respectively. The differential conductivity, assuming a homogeneous electric field, is expressed as

$$\sigma_{\text{diff}} = \frac{dj}{dF} = e\left(n\frac{\partial v}{\partial F} + v\frac{\partial n}{\partial F}\right). \tag{A.3}$$

There are then two distinct ways to achieve $\sigma_{\text{diff}} < 0$ – that is, negative differential conductivity:

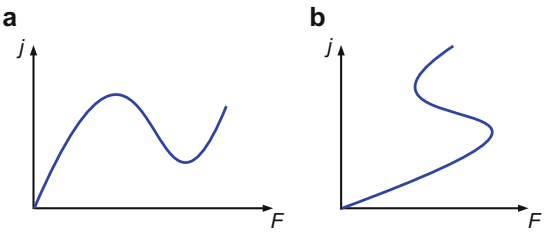


Fig. A.2 Typical current density-field characteristics that show a negative differential conductivity range. (**a**) *N*-shaped characteristics leading to high-field domain formation. (**b**) *S*-shaped characteristics leading to current filament formation

1. $dv/dF < 0$, which corresponds to negative differential mobility (*drift instability*)
2. $dn/dF < 0$, which corresponds to negative differential carrier concentration (*generation-recombination instability*)

We explore some typical example systems below.

Two important classes of negative differential conductivity (*NDC*) are described by an *N*-shaped or an *S*-shaped $j(F)$ characteristic and are denoted by *NNDC* and *SNDC,* respectively (see Fig. A.2). However, more complicated forms like *Z*-shaped, loopshaped, or disconnected characteristics are also possible and may be treated analogously (Wacker and Schöll 1995). *NNDC* and *SNDC* are associated with voltage- or current-controlled instabilities, respectively. In the *NNDC* case the current density is a single-valued function of the field, but the field is multivalued: the $F(j)$ relation has three branches in a certain range of j. The *SNDC* case is complementary in the sense that F and j are interchanged. This duality is in fact far-reaching (Shaw et al. 1992).

The *global* current-voltage characteristic $I(V)$ of a semiconductor can, in principle, be calculated from the *local* $j(F)$ relation by integrating the current density j over the cross section A of the current flow

$$I = \int_A \mathbf{j}\ d\mathbf{a}, \tag{A.4}$$

and the electric field F over the length L of the sample

$$V = \int_0^L F\ dx. \tag{A.5}$$

Unlike the $j(F)$ relation, the $I(V)$ characteristic is not only a property of the semiconductor material, but also depends on the geometry, the boundary conditions, and the contacts of the sample. Only for the idealized case of spatially homogeneous states, the $j(F)$ and the $I(V)$ characteristics are identical, up to rescaling. The $I(V)$ relation is said to display *negative differential conductance* if

$$\frac{dI}{dV} < 0. \tag{A.6}$$

In case of *NNDC,* the *NDC* branch is often but not always – depending upon external circuit, contacts, and boundary conditions – unstable against the formation of electric field domains, while in the *SNDC* case current filamentation generally occurs (Ridley 1963). These primary self-organized spatial patterns may themselves become unstable in secondary bifurcations leading to periodically or chaotically breathing, rocking, moving, or spiking filaments or domains, or even solid-state turbulence and spatiotemporal chaos (Schöll et al. 1998; Schöll 2001).

At this point, a word of warning is indicated. First, negative differential conductivity does not always imply instability of the steady state, and positive differential conductivity does not always imply stability. For example, *SNDC* states can be stabilized (and experimentally observed!) by a heavily loaded circuit, and, on the other hand, the bifurcation of a self-sustained oscillation can occur on a $j(F)$ characteristic with positive differential conductivity. Second, there is no one-to-one correspondence between *SNDC* and filaments or between *NNDC* and domains. Finally, it is important to distinguish between the local $j(F)$ characteristic and the global $I(V)$ relation, which may exhibit negative differential conductance even if the $j(F)$ relations do not show *NDC* and vice versa (Schöll 2001).

Domains and filaments can be theoretically described as special nonuniform solutions of the basic semiconductor transport-equations, subject to appropriate boundary conditions. A linear stability analysis around the spatially homogeneous steady state (singular point) for small space- and time-dependent fluctuations of the electromagnetic field and the relevant transport variables (e.g., carrier concentrations or charge densities) yields conditions for the onset of domain-type or filamentary instabilities (Schöll 1987).

Domain Instability

The formation of field domains in a sample with *NNDC* can be understood from a simple argument. By way of example, we consider the case where the negative differential conductivity results from negative differential mobility, as in the Gunn effect (Gunn 1963, 1964), so that a range of fields exists where $dv/dF < 0$. The case of negative differential carrier concentration will be analyzed in Sect. Phase-Space Analysis below.

To illustrate the response of an *NNDC* element to a charge fluctuation, we consider a uniform field with a domain of increased field in the center of the sample as shown in Fig. A.3a. The charge distribution that produces this field is shown in Fig. A.3b. There is a net accumulation of charge on the left side of the domain and a depletion layer on the right. If we consider *negatively* charged carriers, the carriers and hence the domain will be moving to the right (cathode at left, anode at right).

Assuming that the field within the domain is within the *NDC* range and the field outside the domain is within the ohmic range, but close to the field of peak velocity, then it is clear that the field fluctuation will initially grow with time. This happens

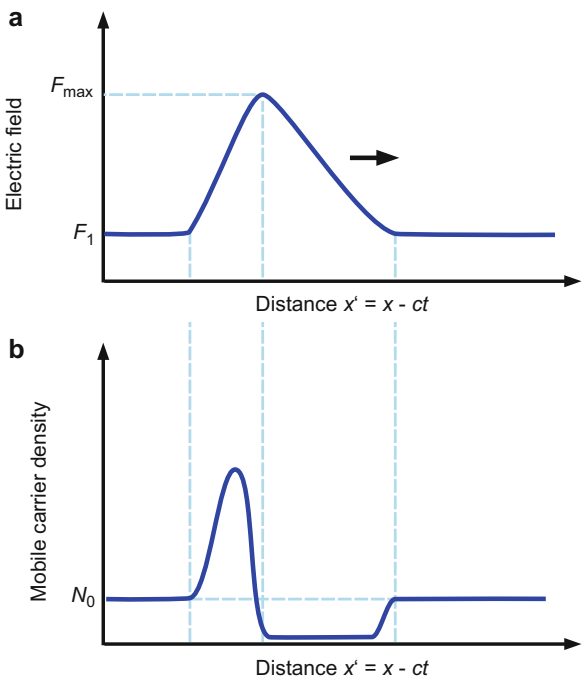


Fig. A.3 (**a**) Electric field profile, (**b**) carrier density profile of a high-field Gunn domain. The domain is moving with velocity v in the positive x direction (schematic) (After Shaw et al. 1992)

because the higher upstream field in the center of the domain results in carriers moving more slowly than those at the edges, where the field is lower. Charge will therefore deplete on the right (leading) edge of the domain and accumulate at the left (trailing) edge. This charge will add to what is already there, increasing the field in the domain. If the element is in a resistive circuit, the increasing voltage across the domain will cause the current to decrease in the circuit and lower the field outside the domain. The field in the domain will continue to grow in the interior of the domain and drop outside. Thus, an instability of the uniform field with respect to domain formation results.

The constitutive equations in one spatial dimension are the carrier continuity-equation, neglecting recombination processes,

$$\frac{\partial n(x,t)}{\partial t} + \frac{\partial}{\partial x}\left(n(x,t)\, v(F) - D\frac{\partial n(x,t)}{\partial x}\right) = 0 \tag{A.7}$$

and Gauss' law

$$\frac{\partial F(x,t)}{\partial x} = \frac{e}{\varepsilon\,\varepsilon_0}\left(n(x,t) - N_0\right), \tag{A.8}$$

where eN_0 is the negative uniform background charge density, and D is the diffusion constant. Equations A.7 and A.8 can be combined by differentiating Eq. A.8 with

respect to time and substituting into Eq. A.7. Integration over x gives the total current density $j(t)$, which is composed of displacement current, drift, and diffusion contributions, respectively:

$$\varepsilon\,\varepsilon_0 \frac{\partial F(x,t)}{\partial t} + e\left(n(x,t)\,v(F) - D\frac{\partial n(x,t)}{\partial x}\right) = j(t), \tag{A.9}$$

with the external current density $j(t)$. Substituting Eq. A.8 into Eq. A.9 to eliminate n yields the governing nonlinear transport equation

$$\varepsilon\,\varepsilon_0\left(\frac{\partial F(x,t)}{\partial t} + v(F)\frac{\partial F(x,t)}{\partial x} - D\frac{\partial^2 F(x,t)}{\partial x^2}\right) + eN_0\,v(F) = j(t). \tag{A.10}$$

The stability of the uniform steady state F^* can be tested by linearizing Eq. A.10 around F^* for small space- and time-dependent fluctuations

$$\delta F(x,t) = \exp(\lambda t)\ \ \exp(ikx). \tag{A.11}$$

Substituting this ansatz into Eq. A.10 for fixed external current density j determines the time increment λ for any given wave vector k – that is, the dispersion relation

$$\lambda = -\frac{1}{\tau_M} - Dk^2 - ikv(F^*). \tag{A.12}$$

Here the effective differential dielectric relaxation time $\tau_M \equiv (eN_0(dv/dF)/\varepsilon\ \varepsilon_0)^{-1}$ has been introduced. For negative differential mobility ($dv/dF < 0$), τ_M is negative, and small fluctuations

$$\delta F(x,t) = \exp\left(-t\left(\frac{1}{\tau_M} + Dk^2\right)\right)\ \ \exp(ik(x - vt)) \tag{A.13}$$

grow in time for $k^2 < (D\,|\tau_M|)^{-1}$ – that is, for long wavelengths. Equation A.13 describes an undamped traveling wave propagating in the positive x direction with velocity $v_0 = v(F^*)$.

This instability leads to the bifurcation of moving field domains from the uniform steady state, which depends, however, sensitively upon the sample length, the boundary conditions at the cathode contact, and the external circuit (Shaw et al. 1979, 1992). A simple criterion for the onset of a moving field domain can be derived from Eq. A.13 by neglecting diffusion and assuming that the linearization remains valid throughout its transit across the element of length L. The growth at the end of the transit – that is, after the transit time $t = L/v_0$ – is then given by

$$G = \frac{\delta F(t)}{\delta F(0)} = \exp\left(\frac{L}{v_0\,|\tau_M|}\right) = \exp\left(\frac{N_0 L e\,\frac{dv}{dF}}{\varepsilon\,\varepsilon_0\,v_0}\right). \tag{A.14}$$

Substantial growth, i.e., a moving domain instability, occurs if the exponent in Eq. A.14 is larger than unity or

$$N_0 L > \frac{\varepsilon \varepsilon_0 v_0}{e \frac{dv}{dF}}. \tag{A.15}$$

This is the N_0L product stability-criterion for the Gunn effect derived by McCumber and Chynoweth (1966) and by Kroemer (1968).

The fully developed field domains must be calculated from the full nonlinear transport equation (Eq. A.10), subject to appropriate boundary conditions. A simple visualization of the moving domain solution can be obtained by transforming Eq. A.10 to the comoving frame $x' = x - ct$, where c is the domain velocity, yielding

$$\varepsilon \varepsilon_0 \left((v(F) - c) \frac{\partial F}{\partial x'} - D \frac{\partial^2 F}{\partial x'^2} \right) + e N_0 (v(F) - v_\infty) = 0. \tag{A.16}$$

Here $j = eN_0\, v_\infty$ with $v_\infty = v(F_1)$ is the current density in the neutral material outside the domain. The result can be written in the form of a "dynamic system" of two first-order differential equations for F and $n = (\varepsilon\varepsilon_0/e)\partial F/\partial x + N_0$, where the parameter corresponding to "time" is given by x'.

The topology of the solutions may be conveniently discussed in terms of a *phase-space analysis* or *field-of-directions analysis* (Shaw et al. 1992), which gives qualitative insight into the nature of the solutions and the effect of boundary conditions. Such methods have been extensively used by Böer and Voss (1968a) and Böer (1969) in the early investigations of Gunn domains and recombination domains, and they will be reviewed in the next section.

Phase-Space Analysis

The analysis of (stationary) high-field domains can be performed by inserting the field dependence of the relevant parameters (e.g., of the generation or recombination rates) into the set of governing transport equations; the solutions then can be obtained numerically.

Another type of analysis is suited to obtain more general information; it shows the solution type more easily and can be used to obtain quantitative information in a simple but satisfactory approximation. This analysis is known as *field-of-directions analysis* or as *phase-space analysis* and can be used easily in a *one-carrier model* in which the set of governing equations (transport, Poisson, and continuity equations) may be written as[1]

[1] For an n-type semiconductor, and assuming $p \ll n$ and $j_p \ll j_n \cong j$.

$$\frac{\partial n}{\partial x} = \frac{e}{kT}\left(\frac{j}{e\,\mu_n} - nF\right) \tag{A.17}$$

and

$$\frac{\partial F}{\partial x} = \frac{e}{\varepsilon\,\varepsilon_0}(p_d - n - n_a). \tag{A.18}$$

For definiteness we will assume first that the field dependence enters via reaction kinetics and makes the carrier density n, the density of electrons in acceptors n_a, and of holes in donors p_d implicit functions of F.

The solution curves of interest are $n(x)$ and $F(x)$ and, derived from the latter, the electron potential distribution $\Psi_n(x)$. Because Eqs. A.17 and A.18 are an autonomous system of differential equations that does not contain x explicitly, and fulfills the Lipschitz conditions of having continuous partial derivatives with respect to F and n, the solution curves can be projected into any nF plane, yielding a single-valued curve. That is, through every point of the nF plane traverses one and only one solution curve or every point in this plane uniquely defines *one* slope, except for *singular points* or *fixed points* which are defined by

$$\frac{dF}{dx} = \frac{dn}{dx} \equiv 0. \tag{A.19}$$

Solutions cannot leave or enter singular points that represent solutions in which $n(x) = \text{const}$ and $F(x) = \text{const}$. One therefore concludes that no solutions can cross each other in the entire nF plane.

When identifying the direction of increasing x from the cathode toward the anode, one can attach to each point in the nF plane an arrow indicating this direction as illustrated in Fig. A.4. This field of arrows is known as a *field of directions*.

When the boundary values [$n(x = 0)$ and $F(x = 0)$] are given, one can start from this point and follow along the directions from point to point to obtain the solution curve $n(F)$. Because an x value is attached to every point in the nF plane, one finally obtains the desired $n(x)$ and $F(x)$. Such a procedure seems to be tedious. However, with the help of two auxiliary curves that can easily be identified, a simple method can be derived, which permits immediate classification of the possible set of solution curves. These auxiliary curves (*null-isoclines*) are those for which (Böer and Wilhelm 1963)

$$\frac{dF}{dx} \equiv 0, \text{identified as } n_1(F), \tag{A.20}$$

and

$$\frac{dn}{dx} \equiv 0, \text{identified as } n_2(F), \tag{A.21}$$

and are shown in Fig. A.5.

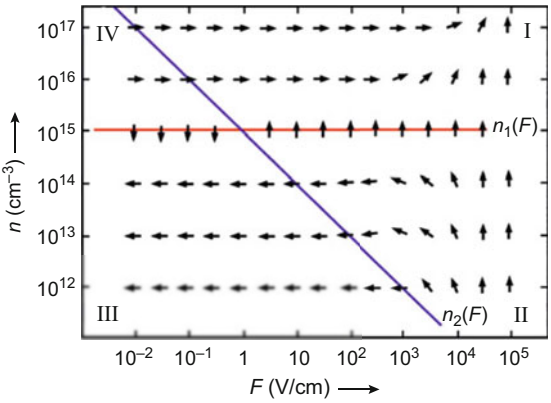


Fig. A.4 Field of directions computed for a typical homogeneous semiconductor, with the neutrality electron density $n_1(F)$ of vanishing space charge, and the drift-current curve $n_2(F)$ of vanishing diffusion current. $n_1(F)$ and $n_2(F)$ divide the nF plane into four quadrants I to IV, which indicate by the *arrows* the different signs of dF/dx and dn/dx

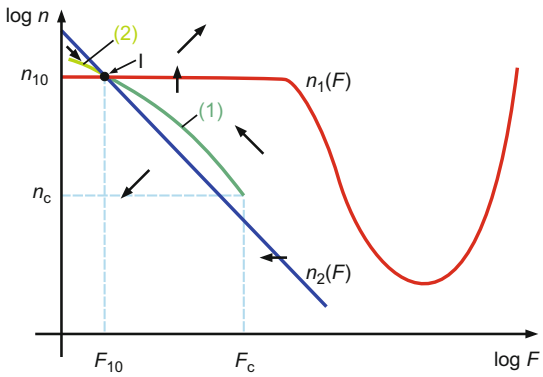


Fig. A.5 Symbolized field of directions with quasi-neutrality curve $n_1(F)$ and drift-current curve $n_2(F)$; *solid arrows* indicate the four quadrants of possible directions of the solution curve. *I* indicates the singular point; *curves* (1) and (2) show typical solutions for a Schottky barrier and an injection contact, respectively

The first curve, $n_1(F)$, is known as the *quasi-neutrality curve* because, following the Poisson equation, the space charge vanishes for any point on this curve. It can be crossed only vertically in the nF plane because, according to Eq. A.20, F remains constant with changing x. If F changes in x, the $n_1(F)$ curve is the locus of field extrema.

The second curve, $n_2(F)$, is known as the *drift-current curve* (see Eq. A.17), since for any point on this curve the diffusion current vanishes. It is therefore also referred to as *diffusion-neutrality curve.* This curve can only be crossed horizontally in the nF plane because n remains constant with changing x according to Eq. A.21.

Below $n_1(F)$, one obtains[2] $dF/dx < 0$; above $n_1(F)$, one has $dF/dx > 0$, as seen from Eq. A.18. In a similar fashion, one sees that $dn/dx < 0$ below and >0 above $n_2(F)$. Therefore, the direction of any solution curve $n(F)$ points into different

[2] Observe that the field F is negative here.

quadrants below or above the quasi-neutrality and the drift-current curves, as symbolized by the arrows[3] in Fig. A.5. This figure also shows typical shapes for the $n_1(F)$ and $n_2(F)$ curves.

The $n_1(F)$ curve shows the field dependence of the electron density under the neutrality condition, that is, in the given one-carrier model $n_1(F)$ is the same as $n(F)$ in a homogeneous and space-charge-free semiconductor. It is constant for low fields and usually increases at sufficiently high fields because of field excitation. With optical carrier generation it may have an intermediate field range in which n decreases with increasing F because of field quenching (Dussel and Böer 1970). Such a decreasing branch is included in the example given in Fig. A.5.

The $n_2(F)$ curve is a hyperbola for a given j, as long as the mobility is field independent. In the double logarithmic representation of Fig. A.5, it is given as a straight line with a slope of -1. The crossing point of curves $n_1(F)$ and $n_2(F)$ is the singular point I. Solution curves that are long compared to the Debye length must approach this singular point. In the bulk of a semiconductor, which is usually several orders of magnitude thicker than the Debye length, $n(x)$ and $F(x)$ change very little with x; hence, the solution curve "approaches" the singular point. It should, however, be remembered that no solution can leave or cross a singular point. Thus, a physically meaningful solution will always start or end in one of the quadrants, but slightly removed from the singular point.

One value of the field-of-direction analysis lies in the fact that one can immediately identify what types of solutions are possible. These are given as curves 1 and 2 and are the only types possible as stationary solutions for a long (compared to the Debye length) device.[4]

A typical solution curve for a Schottky barrier is shown as curve 1 in Fig. A.5. It starts at the metal-semiconductor interface ($x = 0$) at a lower density n_c and a higher field F_c than in the bulk (given by the singular point) and follows the field of directions (see Fig. A.4) toward the singular point I. Another solution is shown for an injecting contact as curve 2 in Fig. A.5. It starts at a higher electron density and consequently at a lower field at $x = 0$, and it also moves toward the singular point I.

From Fig. A.5 one sees that it is possible to approach the singular point I from the second or fourth quadrants only; it is impossible to enter any other quadrants from the first or third quadrant. Namely, crossing the $n_1(F)$ or $n_2(F)$ curve from the second or fourth quadrant means irrevocably turning away from the singular point, yielding physically meaningless solutions. In mathematical terms the singular point I is a saddle point, i.e., it is impossible to circle it.

[3] The arrow in the upper right quadrant, for example, indicates that through any point in this quadrant the solution curve must have a direction between 0° and 90°; this holds correspondingly for the other quadrants.

[4] These solutions represent a device with one electrode. When including the other electrode one must be careful with the signs in the transport equations which result in a reversal of the arrows in the field of directions when approaching the second electrode; this reversal permits the traverse of the projection of the solution curve in the opposite direction in part of the solution.

We will now apply this field-of-direction analysis to include a range of negative differential conductivity. This can be achieved by a decrease of the carrier density due to field quenching or a decrease of the mobility, for example, by carrier excitation into a sidevalley in GaAs, with increasing field.

Decrease of Carrier Density with Field

The carrier density in a homogeneous semiconductor can decrease with increasing field only in nonequilibrium conditions. In steady state, this means that the field causes an increase of the recombination (a reduction of the generation is unlikely). This can be achieved in a number of ways; most researched are quenching transitions in which the field activates fast recombination centers.

Field Quenching

Field quenching is a means to reduce the space charge by field-enhanced ionization (Frenkel-Poole ionization) of minority carriers from compensating deep hole traps. When incorporating fast recombination centers (N_r), the model can explain field quenching of photogenerated carriers in homogeneous photoconductors.

Figure A.6 shows such a model. Light generates electron-hole pairs; when the holes are captured more readily by Coulomb-attractive hole traps (N_a) that have a very small capture cross section for electrons, the electron density in the conduction band increases substantially and n-type photoconductivity is observed: The lifetime of photogenerated electrons is much larger than that of holes, thus $n \gg p$. When a sufficient field is applied, however, the trapped holes can be freed from these Coulomb-attractive hole traps. A fraction of them will be captured by fast recombination centers (Fig. A.6b), which have a much larger cross section for electrons (typically 10^{-16} cm^2) than the deep hole traps (typically 10^{-21} cm^2). This causes a reduction of the electron density as a function of the field. Such a reduction of $n(F)$ is pictured in the quasi-neutrality curve $n_1(F)$ of Fig. A.5 and immediately precedes the steeply rising branch of carrier generation by field ionization.

Cathode- and Anode-Adjacent Domains

When increasing the bias to the device, the current increases, thereby causing a parallel shift of the drift current curve toward higher values – for example, from curve (1) to curve (2) in Fig. A.7. If the field-quenching branch of the quasi-neutrality curve decreases more than linearly with the field, then $n_1(F)$ and $n_2(F)$ will cross again, thereby creating two more singular points (II and III in Fig. A.7).

Assuming for simplicity that the boundary density n_j does not change with increasing bias and current ($n_j = n_c$), the field at the boundary increases until the starting point of the solution has moved close to the second singular point. Then the drift-current curve cannot rise further, because it would force the solution curve to cross $n_1(F)$ while maintaining the boundary condition at n_c. Such a crossing of $n_1(F)$ from the fourth quadrant is impossible. This means that with further increasing bias, $n_1(F)$ can no longer rise, the current must saturate, and the solution expands in x near

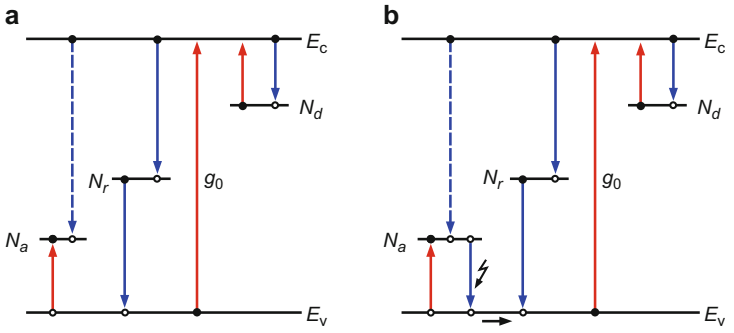


Fig. A.6 (**a**) Band model of an *n*-type photoconductor with hole traps N_a and fast recombination centers N_r. (**b**) Field-induced transition from slow (hole traps) to fast recombination centers causes field quenching of the photoconductivity

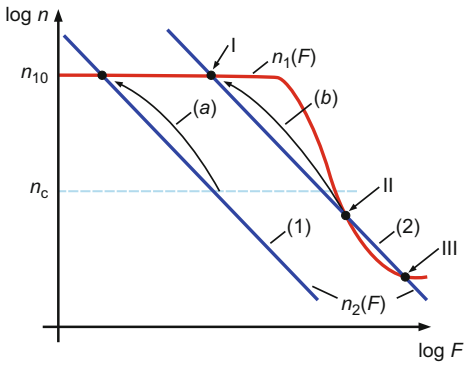


Fig. A.7 Field of directions representation with quasi-neutrality curve (n_1) and two drift-current curves (n_2) for lower (1) and higher (2) current with corresponding solution curves (*a*) and (*b*)

the second singular point. This solution (2) becomes *step like*, i.e., $n(x)$ and $F(x)$ are almost constant near $x = 0$, rather than $n(x)$ steeply increasing and $F(x)$ linearly decreasing near $x = 0$ in a Schottky barrier. This distinctly different, step-like behavior of $n(x)$ and $F(x)$ is referred to as a *high-field domain* (Böer et al. 1958; Böer 1959; Böer and Wilhelm 1963).

With increased bias, the high-field domain widens without increasing the field at the cathode (in *n*-type materials), while the current remains constant. In the field-of-directions representation, no changes are visible, except for a minute extension of the solution closer to the singular point II.

As long as the field remains constant within the domain, here also the space charge vanishes. When the field is increased to initiate field quenching, localized holes in deep hole traps, which produce this space charge in compensated photoconductors, are freed and can either move into the adjacent electrode or recombine,

thus reducing the positive space charge. With sufficient depletion of these hole traps, the space charge is reduced to zero.[5]

When the bias is increased sufficiently so that the high-field domain has expanded to fill the entire device, the solution remains at the singular point II. A further increase in bias causes a minute increase of the current [a slight shift upward of $n_2(F)$] that brings the solution into the quadrant between the singular points II and III and between $n_1(F)$ and $n_2(F)$; it extends the solution from II toward the singular point III. A still further increased bias causes part of the solution (toward the anode) to approach III, now forming an anode-adjacent domain of even higher fields (F_{III}), which then expands from the anode toward the cathode until it fills the entire device.

In the next section, we will show that a major part of this domain behavior can be experimentally confirmed before electric breakdown phenomena set in. Without a field-of-direction analysis, this domain behavior would be rather difficult to explain. The given discussion, however, even though qualitative makes such analysis transparent. It is for this reason that we have included the more qualitative field-of-directions method with the otherwise preferred numerical or analytical discussion.

We will also show in the next sections that in a number of realistic cases this method provides enough information to yield unambiguously the carrier density and mobility as a function of the electric field from rather simple experiments. The key to the success of this simple analysis is the fact that the field is essentially constant within a domain, and the transition region between bulk and domain has a thickness of only a few Debye lengths; that is, its width can usually be neglected compared to the thickness of the device and the domain. Then a simple step approximation is sufficient to describe the voltage drop across the device:

$$V = F_{II} x_1 + F_I (d_1 - x_1). \tag{A.22}$$

With changing bias the width x_1 of the high-field domain changes linearly. This permits the determination of F_I and F_{II} when the domain width is measured as function of the bias.

Franz-Keldysh Effect to Directly Observe High-Field Domains

The domain width can be measured directly with probing light by using the Franz-Keldysh effect (Franz 1958; Keldysh 1958), which causes the absorption edge to shift toward longer wavelengths with increasing field. When observing the light transmitted through the semiconductor platelet,[6] one detects the domain as region of increased absorption when using light at the absorption edge (Böer et al. 1958, 1959), as shown in Fig. A.8. Such high-field domains were first seen in CdS (Böer et al. 1958) and have since been extensively investigated with this method.

[5] Overcompensation is avoided because a reversal of space charge would reduce the field that was the cause for field quenching and would automatically cause an adjustment to yield $\rho \equiv 0$.

[6] For reasons of optical observation, the geometry of the device was chosen so that the thin platelet dimension is parallel to the probing light and the field direction is perpendicular to it.

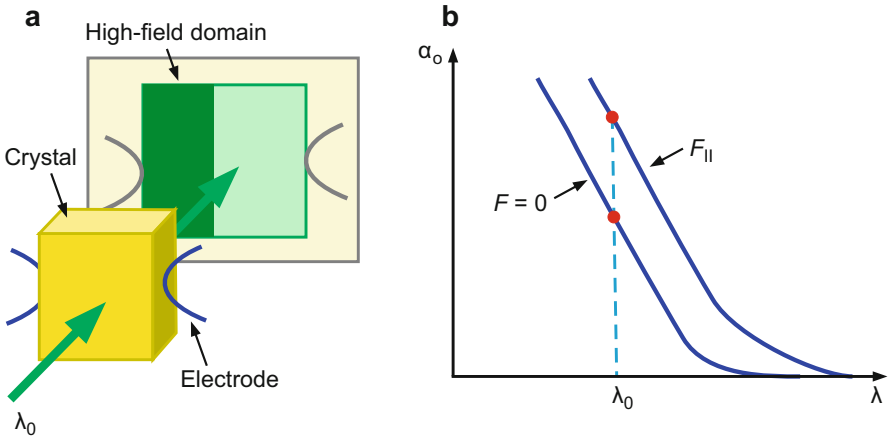


Fig. A.8 (**a**) Experimental setup to use the Franz-Keldysh effect for observing high-field domains. (**b**) Shift of the absorption edge due to the electric field. The crystal platelet is viewed in transmitted monochromatic light (λ_0) that shows the high-field domain as a region of increased absorption

The high-field domains are directly visible in the shadow projection given in Fig. A.9a for a CdS sample, showing the widening of the cathode-adjacent domain (upper row, at left) with increasing bias until the domain fills the entire device at 1750 V (upper row at right). With further increased bias, an anode-adjacent domain of higher field strength (with an increased band-edge shift, hence darker shadow) emerges from the anode and broadens toward the cathode with further increased bias (lower row of Fig. A.9a). The current-voltage characteristics show no break in the saturation behavior during the transition from cathode- to anode-adjacent domain as indicated by the arrow in Fig. A.9b, in agreement with the prediction from the field-of-direction analysis (Böer and Voss 1968a, b).

As long as the domain is attached to the cathode, the electron density at the cathode and the work function of the cathode are measurable. When, on the other side, the domain is attached to the anode, the hole density at the anode is measurable with the work function of holes from the anode. All this occurs with the same current that has not changed: at the cathode-adjacent domain it is a drift current of electrons, and at the anode-adjacent domain it is again a drift current, but now from the holes:

$$j = j_n = j_p = e n \mu_n F_{\text{cathode}} = e p \mu_p F_{\text{anode}}. \tag{A.23}$$

This is very remarkable because the carrier densities and mobilities are different; this holds, however, only as long as high-field domains are present and they adjust in width and field to permit the equality of Eq. A.23.

When plotting the domain width as a function of the bias for the cathode-adjacent (A) and anode-adjacent (B) high-field domains, as shown in Fig. A.10a, one obtains from the slopes and intersections with the abscissa according to Eq. A.22 the fields of the three singular points. These are shown in Fig. A.10b. With the electron mobility

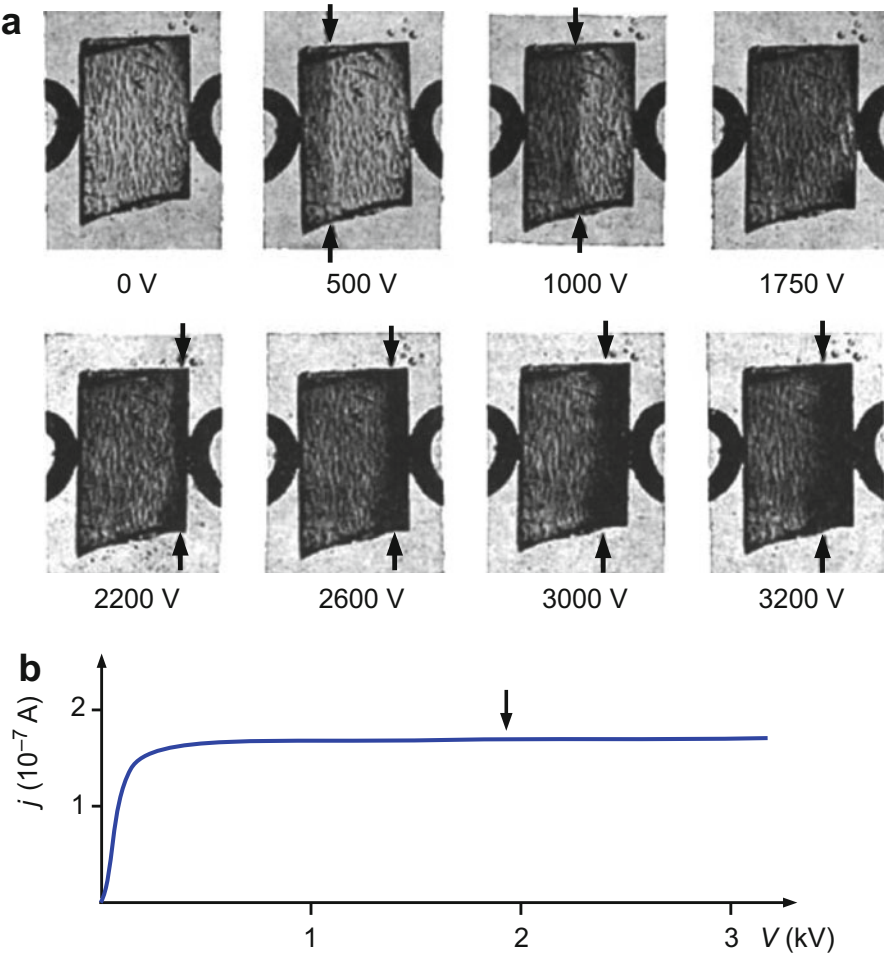


Fig. A.9 (**a**) Shadow pictures of a CdS platelet (at 210 K) between two electrodes at a distance of 0.36 mm, in monochromatic light (10^{12} photons/cm^2 s) at the band edge (at 505 mm). The cathode is at the *left* side, the anode at *right*; *arrows* indicate the edge of the high-field domain. The anode-adjacent domain broadens in the opposite direction from the anode toward the cathode indicating that the CdS has changed from *n*-type to *p*-type. (**b**) Current-voltage characteristics showing current saturation in the entire bias range. The *arrow* indicates the transition from a cathode- to an anode-adjacent domain and the change of the CdS from *n*-type to *p*-type (After Böer and Voss 1968a)

μ_n and field in the cathode-adjacent domain F_{cathode} or the hole mobility μ_p and field in the anode-adjacent F_{anode}, the carrier densities are then directly obtained from the drift-current curves. Most remarkable is the fact that the drift current at the transition point from cathode- to anode-adjacent domains is also the transition from *n*-type to *p*-type conductivity. With only drift currrent in the two cases of cathode- and anode-adjacent domains this, hence, requires the equality of Eq. A.23.

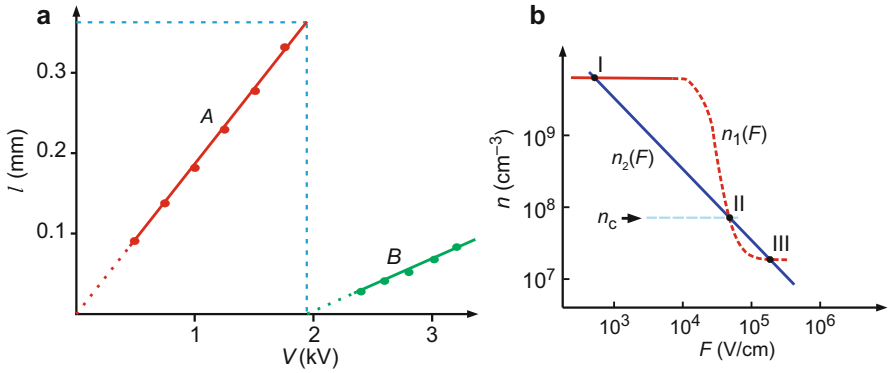


Fig. A.10 (**a**) Domain width l as function of bias in CdS at 230 K, obtained from the experiment described in Fig. A.9. (**b**) Singular points obtained from the slopes of the curves A and B in panel (a) (After Böer and Voss 1968a)

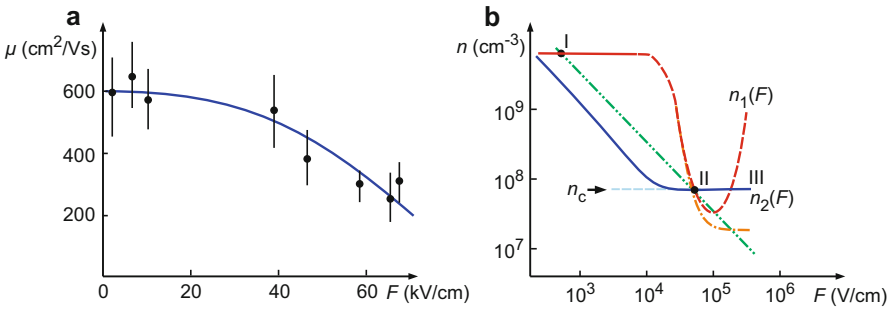


Fig. A.11 (**a**) Hall mobility in CdS as function of the electric field measured in a high-field domain at 230 K (After Böer and Bogus 1968). (**b**) Field of directions with corrected $n_2(F)$ and values for the three singular points obtained by the corrected experimental values. Uncorrected $n_1(F)$ and $n_2(F)$ shown as *dash-dotted curve* segments

Drift-Velocity Saturation

The field-dependent Hall mobility was measured using high-field domains (Böer and Bogus 1968) with Hall electrodes close to the cathode and applying a sufficient bias so that the cathode-adjacent domain extends well beyond the Hall electrodes. The low-field mobility can be checked by reversing the polarity, thereby placing the Hall probes in the low-field bulk. Different pairs of experimental points were obtained by using different metal electrodes yielding different values of n_c.

The results of these experiments are shown in Fig. A.11. Starting near 20 kV/cm, the mobility decreases with increasing field. Such a decrease follows a $1/F$ behavior at higher fields and causes a saturation of the drift velocity (μF = const) as it approaches the rms velocity of the electrons. This results in a bending of the drift-current curve $n_2(F)$ in the field of directions and results in corrected values for the

first and third singular point (Fig. A.11b). The figure also reproduces as dash-dotted curves the results given in Fig. A.10b for comparison.[7]

Workfunction Dependence on Photoconductivity

The field near the metal-semiconductor interface cannot exceed the domain field which is usually <100 kV/cm, thereby excluding tunneling through the barrier. The carrier density in the domain is therefore a reasonable measure of the metal-semiconductor workfunction according to

$$n_{\mathrm{II}} \cong n_{\mathrm{c}} = N_{\mathrm{c}} \exp\left(-\frac{\psi^*_{\mathrm{MS}}}{kT}\right), \tag{A.24}$$

where ψ^*_{MS} is the effective workfunction, reduced from the workfunction by the image potential

$$\psi^*_{\mathrm{MS}} = \psi_{\mathrm{MS}} - \sqrt{\frac{eF_{\mathrm{II}}}{4\pi\varepsilon\,\varepsilon_0}}, \tag{A.25}$$

with n_{II} and F_{II} the carrier density and electric field, respectively, at the second singular point (Fig. A.7). The image force lowering with fields of $F_{\mathrm{II}} \cong 10^5$ V/cm is on the order of 50 meV.

The current over the top of the barrier is carried by Richardson-Dushmann emission

$$j = e\,n_{\mathrm{c}}\,v^*_{\mathrm{rms}} \cong e n_{\mathrm{II}}\,\mu F_{\mathrm{II}}, \tag{A.26}$$

with $\mu F_{\mathrm{II}} \cong v^*_{\mathrm{rms}}$ in the drift-velocity saturation range, this permits the identification of n_{c} with n_{II}.

When this method is applied to vacuum-cleaved CdS with immediately thereafter vapor-deposited electrodes, one observes with optical excitation rather high saturation currents in reverse bias. This yields workfunctions that are reduced to about one half of their values for vanishing currents which, for a gold electrode or other blocking metals, are on the order of 1 eV.

Figure A.12 shows the linear dependence of domain width as a function of bias, and it indicates a slight increase in domain fields with increasing optical generation rate (from 83 to 122 kV/cm for optical generation rates g_{o} changing from 4×10^{14} to 5×10^{16} cm^{-3} s^{-1}, respectively). Because the carrier density within the domain also increases with increasing optical generation, one thereby observes a trend of a *decreasing workfunction with increasing optical excitation* for all investigated contact metals (Fig. A.12b).

These findings indicate that holes can significantly change the dipole layer at the metal-semiconductor interface. These holes are created by excitation and released by

[7] The shift of the first singular point is not resolved experimentally because of the limited accuracy of the method to determine this point.

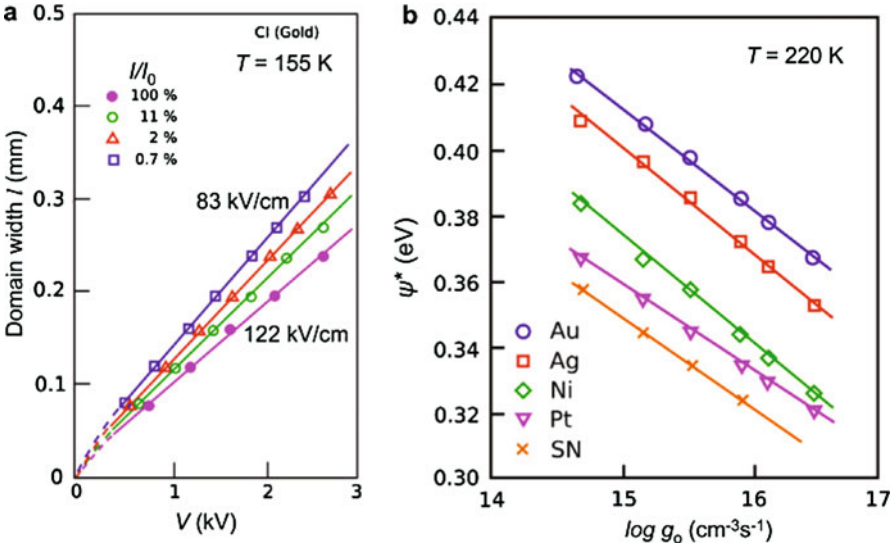


Fig. A.12 (**a**) Domain width in CdS as a function of the applied voltage with light intensity as family parameter. Slopes yield the domain fields. (**b**) Effective workfunction as a function of the optical generation rate (After Stirn et al. 1971)

field quenching within the high-field domain directly adjacent to the interface and thereby cause the corresponding lowering of the workfunction.

Moving High-Field Domains

In the previous sections, we discussed stationary high-field domains. Under certain boundary conditions the high-field domains become unstable after formation and start to break up and move through the bulk. Moving field domains due to the Gunn effect (Shaw et al. 1979; Shaw et al. 1992) or a recombination instability (Bonch-Bruevich et al. 1975) have been extensively investigated experimentally and theoretically in the 1960s and 1970s. The earliest studies of such *NNDC* and *SNDC* instabilities in semiconductors were published by Böer et al. (1961) for CdS. They discovered moving highfield domains in CdS single crystals (Böer 1959), long before the observation of Gunn domain oscillations in GaAs was reported (Gunn 1963, 1964), and developed the electro-optical method described in Sect. Franz-Keldysh Effect to Directly Observe High-Field Domains above to measure these field inhomogeneities (Böer et al. 1959) based on the first experimental evidence of the Franz-Keldysh effect.

Well-known mechanisms for field-domain formation in bulk semiconductors are provided, for example, by field-quenching in CdS (Böer 1959) and in semi-insulating GaAs (Northrop et al. 1964; Samuilov 1995; Piazza et al. 1997), intervalley transfer (Ridley and Watkins 1961; Gunn 1963; Kroemer 1964),

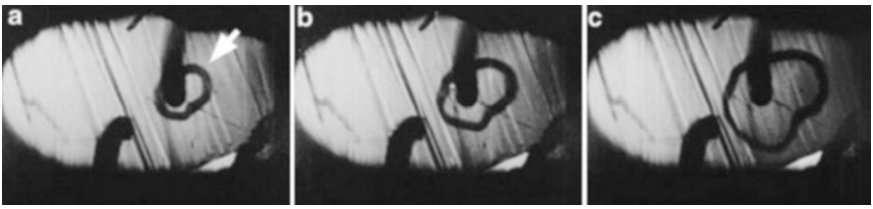


Fig. A.13 Ring-shaped high-field domain (*arrow*) enclosing a small circular cathode and expanding (**a**)–(**c**) in time (pictures taken in 1 s time intervals). After the high-field domain touches the anode at the end of the contact wire at the lower middle left, the ring disappears, a new ring develops at the cathode, and the cycle repeats itself (period length 5 s, contact distance ~2 mm).

nonlinear generation-recombination kinetics in the regime of impurity impact-ionization breakdown (Kahn et al. 1991, 1992a, b; Cantalapiedra et al. 1993; Bonilla et al. 1994, 1997; Bergmann et al. 1996) in *p*-Ge, and the photorefractive Gunn effect (Segev et al. 1996; Bonilla et al. 1998).

The Franz-Keldysh effect can most effectively be used to study kinetic effects. Under certain conditions domains can separate from the electrode and move through the device, accompanied by current oscillations, as observed first by Böer et al. (1959) in any solid; see also Böer and Rompe (1960), Böer and Dussel (1967), and Böer and Voss (1968a).

Moving high-field domains that enclose a small circular electrode as a ring indicate the two-dimensional equipotential distribution following the contour of the ring (with equicurrent lines as orthogonal trajectories of the equipotential lines). In the example given on Fig. A.13 (CdS at 250 K) the speed of expansion of the ring is $\sim 10^{-2}$ cm/s. When the ring reaches the anode at one point, it disappears altogether with a sharp increase in current. A new ring then emerges from the cathode with a corresponding current decrease. This motion repeats itself with almost perfect periodicity.

In the stationary high-field domains in *n-type* semiconductors discussed in previous sections the domains either extended from the cathode or, at higher applied voltages with substantially higher fields, emerged from the anode. Under boundary conditions set for the sample shown in Fig. A.13 the high-field domain breaks up and moves toward the anode. Depending on boundary conditions and applied voltage, three kinds of moving high-field domains were observed:

1. Unstable domains that appear near the cathode of an otherwise observed stationary cathode-adjacent domain and move toward the anode while disappearing
   (a) Before they reach the edge of the previous domain
   (b) After passing through the edge of this domain
2. Undeformed domains moving from cathode to anode

When moving domains of type 1a or 1b appear, the current oscillates around the value of its stationary saturation, exceeding this value when the moving domain reaches its anode limit, and decreasing below this value after starting near the

cathode. Free-moving domains of type 2 have a much larger field and are maintained by a substantially lower current. For a comprehensive analysis see Böer (2010).

Summary and Emphasis

Whenever the conductivity decreases more than linearly with increasing electric field, a new type of solution of transport and Poisson equations is attained which divides the semiconductor into two parts, one with a high field (the high-field domain) and the other part with a low field. This field distribution can be stationary or nonstationary. In stationarity, the field is essentially constant in each of these regions. Both regions are joined by a narrow step of only a few Debye lengths width.

When the high-field domain moves from one electrode to the other, it disappears when it reaches the electrode, while a new domain evolves from the first electrode and the process repeats itself periodically. Stationary domains are usually observed when the negative differential conductivity is caused by a field-dependent decrease of the carrier density, while no stationary solution can be maintained when the change in conductivity is caused by a mobility that decreases more than linearly with increasing field. Other examples of stationary or moving domains are provided by resonant tunneling in superlattices, real-space transfer in modulation-doped heterostructures, or nonlinear generation-recombination kinetics in *p*-type Ge or semiinsulating GaAs. With increasing bias, stationary high-field domains expand at the expense of the low-field region, until the entire semiconductor is filled by the high-field domain. With further increased bias, a second type of domain with a substantially higher field develops from the opposite electrode and expands as required for absorbing the entire applied voltage.

A relatively simple field-of-direction analysis of the family of solution curves of Poisson and transport equations permits us to obtain a general overview of the different types of domains and to deduce unambiguous information on the field dependence of carrier densities and mobilities, as well as on the carrier density and field at the metal-semiconductor boundary. Stationary high-field domains can be used to measure the workfunction of a metal-semiconductor contact and its dependence on various parameters, providing direct evidence of a reduction of this workfunction with increasing optical generation rates in photoconductors.

The use of a pseudoelectrode by a band of reduced illumination in photoconductors permits the direct measurement of the field-dependent carrier density in semiconductors. Moving high-field domains are slow when related to carrier redistribution over traps and can be very fast when trapping can be neglected in domains caused by a field induced decrease of carrier mobilities. The Gunn effect is an example where such steeper-than-linear decrease of the mobility with increasing field in GaAs is used for high frequency ac generation. High-field domains can be visually observed by the shift of optical absorption edge due to the Franz-Keldysh effect.

High-field domains are a valuable tool to unambiguously investigate the field dependence of important parameters in semiconduetors, and as moving domains they are of technical interest for ac generation and the detectieon of small lattice defects.

References

Bergmann MJ, Teitsworth SW, Bonilla LL, Cantalapiedra IR (1996) Solitary-wave conduction in p-type Ge under time-dependent voltage bias. Phys Rev B 53:1327

Böer KW (1959) Inhomogene Feldverteilung in CdS Einkristallen im Bereich hoher Feldstärken. Z Phys 155:184. (Inhomogeneous field distribution in CdS single crystals at high fields, in German)

Böer KW (1969) Light-induced modulation of absorption (LIMA) of CdS single crystals. Z Naturforsch 24a:1306

Böer KW (2010) Electro-optical effects to visualize field and current distributions in semiconductors. Springer, Berlin

Böer KW, Bogus K (1968) Electron mobility in CdS at high electric fields. Phys Rev 176:899

Böer KW, Dussel GA (1967) Uniformly propagating solutions of transport and poisson equations for periodic field domains. Phys Rev 154:292

Böer KW, Rompe R (1960) Raumladungsschwingungen in Halbleitern bei höheren Feldstärken. Ann Phys (Leipzig) 5:200. (Oscillation of space charges in semiconductors at high fields, in German)

Böer KW, Voss P (1968a) Stationary high-field domains in the range of negative differential conductivity in CdS. Phys Rev 171:899

Böer KW, Voss P (1968b) Stationary anode-adjacent high-field domains in cadmium sulfide. Phys Stat Sol B 28:355

Böer KW, Wilhelm WE (1963) Characteristic layer-like field inhomogeneities in homogeneous photoconductors in the pre-breakdown range. I. Steady-state case. Phys Stat Sol B 3:1704

Böer KW, Hänsch H-J, Kümmel U (1958) Methode zum Sichtbarmachen von Leitfähigkeitsinhomogenitäten in Halbleitern. Naturwissenschaften 45:460. (Method to image inhomogeneities of the conductivity in semiconductors, in German)

Böer KW, Hänsch H-J, Kümmel U (1959) Anwendung elektro-optischer Effekte zum Nachweis elektrischer Inhomogenitäten. Z Phys 155:170. (Application of electro-optical effects to detect electrical inhomogeneities, in German)

Böer KW, Hänsch H-J, Obernick H (1961) Elektrolumineszenz von CdS Einkristallen. Phys Stat Sol B 1:352. (Electroluminescence of CdS single crystals, in German)

Bonch-Bruevich VL, Zvyagin IP, Mironov AG (1975) Domain electrical instabilities in semiconductors. Consultants Bureau, New York

Bonilla LL, Galán J, Cuesta JA, Martínez FC, Molera JM (1994) Dynamics of electric-field domains and oscillations of the photocurrent in a simple superlattice model. Phys Rev B 50:8644

Bonilla LL, Cantalapiedra IR, Gomila G, Rubí JM (1997) Asymptotic analysis of the Gunn effect with realistic boundary conditions. Phys Rev E 56:1500

Bonilla LL, Kindelan M, Hernando PJ (1998) Photorefractive Gunn effect. Phys Rev B 58:7046

Cantalapiedra IR, Bonilla LL, Bergmann MJ, Teitsworth SW (1993) Solitary-wave dynamics in extrinsic semiconductors under dc voltage bias. Phys Rev B 48:12278

Dussel GA, Böer KW (1970) Field quenching as mechanism of negative differential conductivity in photoconducting CdS. Phys Stat Sol B 39:391

Franz W (1958) Influence of an electric field on an optical absorption edge. Z Naturforsch 13A:484

Gunn JB (1963) Microwave oscillations of current in III–V semiconductors. Sol State Commun 1:88

Gunn JB (1964) Instabilities of current in III-V semiconductors. IBM J Res Dev 8:141

Kahn AM, Mar DJ, Westervelt RM (1991) Spatial measurements of moving space-charge domains in p-type ultrapure germanium. Phys Rev B 43:9740

Kahn AM, Mar DJ, Westervelt RM (1992a) Dynamics of space-charge domains in ultrapure Ge. Phys Rev Lett 68:369

Kahn AM, Mar DJ, Westervelt RM (1992b) Spatial measurements near the instability threshold in ultrapure Ge. Phys Rev B 45:8342

Keldysh LV (1958) The effect of a strong electric field on the optical properties of insulating crystals. Sovjet Phys JETP 34:1138
Kroemer H (1964) Theory of the Gunn effect. Proc IEEE 52:1736
Kroemer H (1968) The Gunn effect under imperfect cathode boundary conditions. IEEE Trans Electron Devices 15:819
McCumber DE, Chynoweth AG (1966) Theory of negative-conductance amplification and of Gunn instabilities in "two-valley" semiconductors. IEEE Trans Electron Devices 13:4
Northrop DC, Thornton PR, Tresize KE (1964) Electrical transients in high resistivity gallium arsenide. Solid State Electron 7:17
Piazza F, Christianen PCM, Maan JC (1997) Propagating high-electric-field domains in semi-insulating GaAs: experiment and theory. Phys Rev B 55:15591
Ridley BK (1963) Specific negative resistance in solids. Proc Phys Soc Lond 82:954
Ridley BK, Watkins TB (1961) The possibility of negative resistance effects in semiconductors. Proc Phys Soc Lond 78:293
Samuilov VA (1995) Nonlinear and chaotic charge transport in semi-insulating semiconductors. In: Niedernostheide F-J (ed) Nonlinear dynamics and pattern formation in semiconductors and devices. Springer, Berlin, pp 220–249
Schöll E, Schwarz G, Wacker A (1998) Nonlinear and oscillatory electronic transport in superlattices as a probe of structural imperfections. Physica B 249–251:961
Schöll E (2001) Spatio-temporal dynamics and chaos in semiconductors. Cambridge University Press
Schöll E (1987) Non-equilibrium phase transitions in semiconductors. Springer, Berlin
Segev M, Collings B, Abraham D (1996) Photorefractive Gunn effect. Phys Rev Lett 76:3798
Shaw MP, Grubin HL, Solomon P (1979) The Gunn-Hilsum effect. Academic Press, New York
Shaw MP, Mitin VV, Schöll E, Grubin HL (1992) The physics of instabilities in solid state electron devices. Plenum Press, New York
Stirn RJ, Böer KW, Dussel GA (1971) CdS – metal contact at higher current densities. Phys Rev B 7:1433
Wacker A, Schöll E (1995) Criteria for stability in bistable electrical devices with S- or Z-shaped current voltage characteristic. J Appl Phys 78:7352

Index

A
Abarenkov-Heine potential, 771, 833
Ab initio calculations, 226, 674
Absorption coefficient
  indirect exciton, 387, 541, 546, 735, 739
  optical, 537
Absorption constant, 716, 718
Acceptor
  energies, 786
  ground and excited states of, 773
  hydrogen-like, 771–772
  levels, 772, 773
Acceptor-bound exciton (ABE), 786
Acenes, 79, 213
Acoustic branch, 133, 137
Acoustic deformation-potential scattering, 480, 991
Acoustic phonon scattering, 993
Activator, 1278, 1309, 1318
Adamantine, 76
Adiabatic approximation, 221, 715, 731
Adiabatic bandgap, 214
Adiabaticity parameter, 715
Air band, 406
Alkanes, 78
Alkyls, 28
Alloys, 52, 76
  scattering, 1011
Amorphous chalcogenides, 869, 890
Amorphous semiconductors, 91, 840, 1166, 1175, 1178–1180, 1197
  bands in, 310
  density-of-states, 310
  electronic structure, 197
  extrinsic absorption, 523–524
  intrinsic absorption, 521–523
  specific heat of, 164
Amorphous silicon, 99
Amphoteric behavior, 831
Anderson localization, 865–867
Anderson model, 662–664, 863–866
Anderson-Mott localizations, 865–867
Anderson-Mott transition, 739
Anharmonic oscillations, 414
Anion vacancy, 824
Anisotropic lattices, 515
Anisotropic shielding, 763
Anthracene, 149, 248
Antibonding state, 40, 197, 340
Antiferromagnetic interaction, 324, 332
Antiferromagnetic ordering, 1102
Antiferromagnetic susceptibility, 329
Antiphase defect, 635
Antisite defects, 611, 830–831
Anti-Stokes scattering, 451, 453
Anti-Stokes lines, 550
Anti-Stokes Raman scattering, 471
Approaches containing electron–electron interaction, 234
APW, *see* Augmented plane-wave (APW) method
Athermal diffusion, 627
Atomic basis, 60
Atomic electronegativity, 816, 853
Atomic polarizability, 431
Atomic radii, 48
Auger-recombination rate, 1251, 1252
Augmented plane-wave (APW) method, 232
Autocorrelation function, 572, 576
  coherent radiation, 573
  thermal radiation, 573
Avalanche current, 1221
Avalanche formation, 1215
Average surplus energy, 978
Axial quasicrystals, 101

© Springer Nature Switzerland AG 2023
K. W. Böer, U. W. Pohl, *Semiconductor Physics*,
https://doi.org/10.1007/978-3-031-18286-0

B
Ballistic phonon transport, 171
Ballistic transport, 1134–1136, 1157
Band alignment
  Anderson model, 662–664
  classification of interfaces, 661
  Frensley-Kroemer model, 669
  linear models, 664–665
  space-charge region, 658–661
  transition-metal levels, 666–668
  valence-band offsets (*see* Valence-band offsets)
Band bending, 660
Band conductance, 1188–1191, 1197
Band-edge fluctuation, 205
Band filling, 746, 1275
Bandgap, 284
  of alloys, 285
  energy of organic solids, 309
  at high doping level, 291
  pressure dependence, 290
  temperature dependence, 290
Band model, 196, 204, 258, 600
Band offset, 296
Band of localized states, 1169
Band parameters, 272, 275
Bands in amorphous semiconductors, 310
Band structure
  calculation, 226
  of organic crystals, 247
  of three-dimensional lattices, 240
  of typical semiconductors, 243
Band tails, 739–744, 862, 867, 871
Band-to-band luminescence, 1266
Band-to-band recombination, 1247, 1250, 1264, 1268
Band-to-band transitions, 1274
  absorption coefficient and dielectric function, 504–508
  allowed and forbidden transitions, 513–515
  amorphous semiconductors, optical bandgap of (*see* Amorphous semiconductors)
  fundamental absorption edge, 509–511
  indirect transitions, 511–513
  joint density of states, 500–503
  magneto-absorption, 516–518
  in quantum wells (*see* Quantum wells)
  valence bands, 516
Bardeen-Cooper-Shrieffer (BCS) theory, 1090, 1094, 1095, 1106, 1107
Bardeen model, 688–690
BCS theory, *see* Bardeen-Cooper-Shrieffer (BCS) theory
Beer-Lambert law, 1302
Bell state, 576
Bethe lattice, 98
Biaxial stress, 794
Biexcitons, 530, 551, 554, 560
Binding energy, 536, 782
  biexciton, 554
  0D exciton complexes, 564
  1D exciton, 562
  2D exciton, 556
  2D trion, 560
  negative, 564
  trion, 553
Bipolaron, 943
Birefringence, 419
Blackbody radiation, 1263
Bloch electron, 261, 298, 935
Bloch function, 200, 223, 757
Bloch oscillations, 211, 949, 1127–1128
Bloch sphere, 569
Bloch theorem, 200, 223
Bohr magneton, 322, 798, 1072
Bohr radius, 531, 758
Boltzmann distribution, 509, 903
Boltzmann equation, 965–979, 1034
Bond-alternation defect, 886
Bond angle, 93
Bonding in organic crystals, 54
Bonding states, 40, 197, 340
Bond length, 41, 52
Born forces, 37
Born-Haber cyclic process, 38
Born-Oppenheimer approximation, 221
Bose-Einstein condensation, 927, 1090, 1092, 1109
Bose–Einstein distribution function, 159, 512, 547, 904
Bound-exciton lines, 785
Bound-exciton luminescence, 1276
Bound excitons, 782, 784, 786, 794
Bound states, 297, 758
Bourgoin and Corbett mechanism, 627
Bowing parameter, 286
Bragg reflections, 211, 402, 474
Bravais lattices, 62
Breathing mode, 143, 707, 708, 711, 726
Bremsstrahlung photons, 1288
Brewster angle, 391
Bridgman method, 19
Brillouin and Raman scattering, 449–462
Brillouin function, 322

Brillouin scattering, 452–454, 550
resonant, 453
stimulated, 454
Brillouin zone (BZ), 70, 142, 207, 248, 266, 299, 501
Broken-gap band alignment, 294
Brower approximation, 617–620
Bruggemann effective-medium model, 393
Buckingham potential, 55
Building blocks, 80, 93
Built-in field, 963
Bulk-boundary correspondence, 1140, 1143, 1152
Burgers vector, 630
Burstein-Moss effect, 745, 1275

C

Calorimetric absorption spectroscopy, 1250
Capture cross-sections, 1243, 1248, 1258, 1310, 1314, 1315, 1318, 1319, 1323
Carrier, 4, 261
carrier-carrier scattering, 1006
concentration, 963
current, 1058
density, 917, 1115, 1116
distribution, 961–965, 1030, 1038, 1050, 1338
drag, 1006
heating, 962, 1030, 1034, 1051, 1338, 1342
mobility, 950, 964, 986, 1114, 1119–1121, 1157
scattering, 970
temperature, 1035
Casimir length, 179
Cation interstitials, 781
Cation vacancy, 825
Cauchy relation, 122
Cavity-polariton, 411
Cellular method, 231
Central-cell correction, 765, 766
Central-cell potential, 768, 815, 834
Ceramic superconductor, 1099, 1102, 1104
Cern number, 1143
Chalcogens in Si, 816, 835–837
Chalcopyrite, 74, 287
Charge-fluctuation models, 1108
Charge of a hole, 261
Charge-transfer complexes, 1189
Charge-transfer transitions, 844, 851, 853
Chemical potential, 9
Chemical shift, 835
Chemical trend, 816, 818, 832
Chemical valence force, 42
Chemical vapor transport, 22
Chemisorption, 27
Circular polarizarized light, 485
Classes of superconducting semiconductors, 1088
Clausius-Mossotti relation, 431
Cluster calculations, 818
Coherence length, 419
Coherent quantum 1/f effect, 1289
Coherent tunneling, 1125
Collision integral, 967, 971
Colloidal quantum dots, 90
Color center, 826, 827
Compensation, 840, 849, 914, 1176–1178
Complex conductivity, 378
Complex dielectric constant, 375, 439
Complex index of refraction, 374, 393
Composite fermion, 1083
Compositional disorder, 724
Conductance, 352, 1134
Conductance quantization, 1135–1138, 1157
Conduction band, 213, 260, 268, 280
Conductivity, 378, 948, 1134
Configurational specific heat, 163
Configuration-coordinate diagram, 711, 726
Configuration-coordinates, 708
Confined states, 305
Conjugated π bonds, 79
Conjugated π electrons, 54
Conservation laws, 1058
Constant-energy surface, 265
Constraints per atom, 94
Contact noise, 1291
Conversion efficiencies, 420
Conwell-Weisskopf dependence, 1001
Cooling carriers, 1358, 1362
Cooling electron, 1343, 1345
Cooling holes, 1343, 1344
Cooper pairs, 839, 1090–1092, 1104, 1107
Coordination number, 37, 45, 71, 93, 863, 865, 866, 870–872
Core electron, 224
Core potential, 756, 770, 771, 787, 802, 813, 816
Coulomb blockade, 1152, 1153, 1155, 1156
Coulomb diamonds, 1156
Coulomb tail, 815–816, 833, 853
Coupling constant, 939
Covalent bonding, 40
Covalent semiconductors, 624
Critical cluster size, 16
Critical current density, 1092

Critical energy of nucleation, 13
Critical points, 142, 244, 390, 408
Critical temperature, 1088, 1089, 1092–1093, 1102, 1103, 1108
Crossover pressure, 791
Cross section
  capture, 1314, 1315, 1318, 1319
  excited state, 1305
  photo-ionization, 1304
Crystal
  potential, 224
  structure, 60
  systems, 62
Crystal defects
  defect notation, 599–601
  extrinsic donors and acceptors, 598–599
  lattice defects, diffusion of, 620–627
  line defects (*see* Line defects)
  planar defects (*see* Planar defects)
  point defects (*see* Point defects)
  vacancies and interstitials, 601–602
Crystal field
  splitting, 274, 324, 340, 845, 851
  theory, 820–823, 850
Crystal interfaces
  heterointerfaces (*see* Heterointerfaces)
  metal-semiconductor interfaces (*see* Metal-semiconductor interfaces)
Crystallographic notations, 66
Cubic semiconductors, valence band, 269
Cuprates, 1099, 1101–1104, 1106–1108
Curie constant, 324
Curie's law, 324
Curie temperature, 335
Curie-Weiss law, 328
Curie-Weiss temperature, 329
Current continuity, 1317–1318
Current density, 948
Cyclotron frequency, 278, 797, 1072
Cyclotron resonance, 277, 1070
  absorption, 489
Czochralski method, 20

D

Damped wave equation, 374
Dangling bonds, 668, 826, 834, 840, 869, 871–873, 876, 891
Dark conductivity, 1306
Davidov splitting, 535
DBR mirror, 407
de Broglie wavelength, 85, 199, 210
Debye frequency, 160
Debye length, 658, 660, 687
Debye model, 160
Debye temperature, 160, 161
Deep-center spectra, 715
Deep-level defects, 706–710
Deep-level transient spectroscopy (DLTS), 639
Deep traps, 812, 813, 853, 1311
Defect-center recharging, 1037
Defect-center relaxation, 708–710
Defect level, 206
Defect types, in amorphous semiconductors, 871–876
Deformation potential, 793, 995, 996, 1015
Deformed Boltzmann distribution, 970
Degenerate semiconductor, 922
DeHaas-van Alphen effect, 1074
Delta symbol, 69
Demarcation line, 1257
Density-functional theory (DFT), 235, 676–677
Density of spin-polarized states, 239, 340
Density-of-state mass, 907
Density of states, 246, 742, 902, 908, 1070, 1074, 1084, 1167, 1169, 1176, 1183, 1192
  amorphous semiconductor, 310
  dimensionality, 301
  effective, 901, 902
  quantum wire, 304
  two-dimensional, 303
Dephasing, 1348, 1349, 1352, 1358, 1366
Detailed-balance principle, 1211
Dexter-Klick-Russell rule, 1249
Diamagnetic materials, 321
Diamagnetic shift, 799
Diamond lattice, 72
Dielectric band, 406
Dielectric breakdown, 1223
Dielectric constant, 397
Dielectric constant tensor, 380
Dielectric function, 393, 394, 396
Dielectric matrix, 484
Dielectric polarization, 379, 380
Dielectric response function, 433
Dielectric screening, 398, 400
Difference mixing, 418, 419
Diffusion
  carrier transport, 1181–1182
  coefficient, 955
  conduction, 1136
  current, 954–955, 958
  current saturation, 1028
  length, 1328
  measurements, 1030

potential, 686
thermal transport, 170
transport, 1134–1135
Diluted magnetic semiconductors (DMS), 48, 330
exchange mechanisms, 339
III–V, 335
II–VI, 331
miscibility, 331
Dirac cone, 1150
Dirac point, 1148
Dirac relativistic equation, 238
Direct exchange, 325
Direct-gap excitons, 537
Direct gap semiconductors, 512
Direct optical absorption edge, 509
Dislocations, 628
Disordered crystals, 738–747
Disorder-induced gap states (DIGS) model, 672–673
Dispersion
carrier transport, 1184–1185
equation, 130, 133
relations, 123, 202, 240
transport, 1185, 1332
Distributed Bragg reflectors, 404
Distribution function, 903, 904, 1034, 1058
Donor, 756, 760
donor-acceptor pairs, 775, 776, 1276
donor-bound exciton, 785, 794, 800
ionization-energy, 759
Doping, 862, 870, 874–875, 883, 1102
Double-barrier quantum well, 1123–1125
Double donor, 779
Double exchange, 342
energy gain, 343
Drift current, 951–954
Drift length, 1328
Drift measurements, 1027
Drift velocity, 948, 964, 1329, 1337
saturation, 1025
Driving force for crystallization, 11
Drude approximation, 947, 950
Drude model, 476
Dulong-Petit law, 159
DX center, 837–838, 841, 1318

E

Edge states, 1080, 1141, 1143
helical, 1144, 1147, 1148
spin-textured, 1150
Effective charges, 442
Effective dielectric constant, 400, 401
Effective electron mass, 210, 268, 279
Effective impurity charge, 770
Effective impurity potential, 769
Effective ion charge, 43, 49
Effective level density, 905
Effective mass, 210, 229, 261, 782, 993
anisotropy of, 543
at higher energies, 282
density-of-state, 906
measurement, 279
in real bands, 268
Efficiency droop, 1254
Effusion rate, 26
Einstein model, 159
Einstein relation, 956
EL2 defect, 831, 837, 841
Elastic and inelastic scattering, 451–452
Elastic compliance constants, 117
Elastic constant, higher-order, 168
Elastic moduli, 115
Elastic scattering, 989
Elastic stiffness constants, 117, 121
Elastic waves, 122
Electrical field, 964
Electric current density, 477
Electric displacement, 379
Electric field
built-in, 952
external, 951
Electrochemical potential, 960
Electromagnetic eigenmode, 404
Electron affinity, 650, 663
Electron density, 905, 906
Electron distribution, 904
Electron drag, 979
Electron energy-distribution, 778
Electron energy-loss spectroscopy (EELS), 472
Electron-hole condensation, 923–928
Electron-hole drag effect, 1006
Electron-hole droplets, 927
Electron-hole liquid, 925, 1272, 1348
Electron-hole plasma, 924, 1348, 1359, 1361, 1373
Electronic conductivity, 947–949
Electronic noise, 1283
Electronic polarization, 439
Electron-lattice coupling, 710, 749, 1246, 1249
Electron lifetime, 1252, 1286
Electron paramagnetic resonance (EPR), 1073
See also Electron spin resonance (ESR)
Electrons, 1355
temperature, 1026, 1034, 1040–1043
traps, 1256
tunneling, 1225
turnstile, 1155
Electron wave, 199

Electron spin resonance (ESR), 759, 800–802, 1073
Electro-optical effects, 421, 422
Electrostriction, 432
Ellipsometric angles, 392
Ellipsometry, 390, 394
Emission probability, 1212
Empirical pseudopotential method (EPM), 225
Empty band, 260
Empty lattice, 208, 241, 405
Empty shells, 777
Energy band, 195
Energy density, 373
Energy flow, 373
Energy flux, 383
Energy relaxation, 970–973, 977, 1037, 1327
Entanglement, 576
  entanglement swapping, 579
  fidelity, 581
  generation of entangled pairs, 580
Entropy contribution, 605–607
Envelope function, 758
Epitaxy, 24
EPM, *see* Empirical pseudopotential method (EPM)
Equi-energy surface, 266
Equilibrium crystal shape, 17
Equilibrium distance, 36
Equilibrium noise, 1283
Equivalent noise resistor, 1284
Equivalent noise temperature, 1284
Evanescent mode, 410
Exact-exchange (EXX) method, 237
EXAFS, *see* Extended x-ray absorption fine structure (EXAFS)
Exchange-correlation potential, 235, 239
Exchange interaction, 234
Exchange mechanisms in DMS, 339
Exchange parameter, 325
Excited donor states, 761
Exciton Bohr-radius, 85, 531, 536
Exciton luminescence
  in direct-gap semiconductors, 1270
  in indirect-gap semiconductors, 1271
Exciton–polariton equation, 548
Excitons, 531, 557, 561, 563
  in Cu_2O, 540
  energy, 536, 556
  exciton-polaritons, 547–553
  free biexcitons, 554–555
  free trions, 553
  Frenkel excitons (*see* Frenkel excitons)
  kinetics, 1356
  molecules, 1272
  polariton, 530, 1270, 1357
  in QDs, 563
  in quantum wells, 555–561
  in QWRs, 561–562
  Schrödinger equation, 535
  Wannier-Mott excitons (*see* Wannier-Mott excitons)
Expansion coefficient, 165
Extended x-ray absorption fine structure (EXAFS), 97
External field
  external, 961
  internal, 961
External modes, 149
External surfaces, 1009
Extinction coefficient, 375
Extrinsic luminescence, 1274
Extrinsic photoconductivity, 1300, 1309
Extrinsic semiconductor, 910
Extrinsic transitions, 1266
EXX method, *see* Exact-exchange (EXX) method

F

Faraday angle, 490
Faraday effect, 490–491
F center, 813, 819, 824–825
Fermi-Dirac distribution, 903
Fermi energy, 246, 909, 913, 915, 916, 964
Fermi glass, 1180
Fermi integral, 908
Fermi level, 293
Fermi's golden rule, 501
Ferromagnetic domains, 326
Ferromagnetic interaction, 336
Ferromagnetic stabilization energy, 337
Ferromagnetic state, stability, 337
Ferromagnetic susceptibility, 328
Fibonacci sequence, 104
Fick's laws of diffusion, 620–622
Field ionization, 1213
Filling factor, 1078, 1081
Flicker noise, 1287, 1291
Floating bond, 872–873, 875
Flux-line lattice, 1096, 1105
Flux quantum, 1096, 1098
Fock contribution, 234
Fock state, 574
Folded branches, 146

Forbidden gap, 195
Formation energy, 828, 830
Fractional quantum Hall effect, 1081–1083
Franck-Condon principle, 726–728
Franck-Condon shift, 727
Frank-Read source, 638
Frank's rule, 631
Frank-van der-Merwe growth mode, 14
Franz-Keldysh effect, 1234
Free-electron resonance absorption
  charge-density waves, 474–475
  electron-plasma absorption, 469–472
  valence-electron plasma absorption, 472–473
Free-to-bound transitions, 775–779, 1275
Frenkel defects, 607
Frenkel disorder, 602, 603
Frenkel excitons, 529, 532, 545
  in alkali halides, 533
  in organic crystals, 534–535
Frenkel pairs, 602, 605
Frenkel-Poole effect, 1213
Frensley-Kroemer model, 669
Frequency mixing, 416
Fresnel coefficients, 392
Fresnel formula, 487
Fresnel's equations, 382
Fröhlich coupling constant, 940
Fröhlich interaction, 545, 997
Frozen-in carrier densities, 920, 921
Frustration mechanism, 333

G

Galvanomagnetic effects, 1060, 1062, 1084
Gapless semiconductor, 264
Gaussian distribution, 728
Gaussian line shape, 720
Geminate recombination, 1245
Generalized gradient approximations (GGA), 236, 678
Generation processes, 1206
Generation rate, 1302–1304
Generation-recombination noise, 1286
GGA, *see* Generalized gradient approximations (GGA)
Gibbs energy, 9
Glasses, 91
Glass formation, 92
Glide set, 632
Goldschmidt radii, 48
Gradient freezing method, 19
Grain boundary, 881, 882, 887
Green's-function calculations, 817
Green's function (KKR) method, 232
Ground-state energy, 297
Group velocity, 135, 209
Growth, 9
  affinity, 11
  habit, 17
  kinetics, 15
  modes, 14
  of organic crystals, 23
  from a solution, 20, 23
Grüneisen parameter, 168
Grüneisen rule, 166
Gunn-effect, 1049
GW approximation (GWA), 238, 676

H

Hall angle, 1066
Hall coefficient, 1066, 1103
Hall effect, 1057, 1060, 1065–1067, 1077–1083
Hall insulator, 1080
Hall mobility, 1066
Hall-resistance step, 1079
Ham effect, 820
Hanle effect, 1365
Hardness, 50
Hartree approximation, 222, 229
Hartree-Fock approximation (HFA), 234
Hayne's rule, 783
Heat capacity, 158
Heat current, 1058
Heated electron gas, 481
Heavily doped semiconductor, 739
Heavy doping, 725, 744–747, 1175–1178
Heavy-hole band, 270, 557
Heisenberg model, 325
Helmholtz free energy, 603
HEMT, *see* High-electron-mobility transistor (HEMT)
Hermann-Mauguin symbol, 65
Herringbone alignment, 80
Herringbone packing, 78
Heterogeneous nucleation, 13
Heterointerfaces
  band alignment (*see* Band alignment)
  DFT, 676–677
  DIGS model, 672–673
  heterovalent interfaces, 669–670

Heterointerfaces (*cont.*)
interface-dipole theory, 670–672
model-solid approach, 675–676
pseudomorhic layers, 651–652
strain relaxation, 652–656
HFA, *see* Hartree-Fock approximation (HFA)
High-electron-mobility transistor (HEMT), 1116
Higher charged coulomb centers, 779
Highest occupied molecular orbital (HOMO), 212
High-field domain, 1131–1132
High-field transport in built-in fields, 1031
High-T_c superconductor
ceramic, 1099–1102
mediating-partner models, 1106–1108
normal-state properties of, 1102–1103
superconductive state of, 1104–1106
Historic development, 4
Hole, 260
density, 906
traps, 1257
Homogeneous lines, 719–721
Homogeneous nucleation, 11
HOMO, *see* Highest occupied molecular orbital (HOMO)
Hooke's law, 115
Hopping, 1182
conductance in organic crystals, 1192–1196
conduction, 215, 1166, 1167, 1174, 1184
conductivity, 1186
mobility, 1182, 1184
Hot carriers, 1340, 1347, 1357
Hot electrons, 1026, 1121
Huang-Rhys factor, 707, 710, 728, 1248, 1250
Hubbard band, 1166
Hubbard correlation energy, 838
Hund's rules, 323, 843
Hybridization, 42
Hydrogenation, 99
Hydrogen
bonding, 47
impurities, 884
like model, 705
in semiconductors, 831–832, 840
states, 531
Hydrostatic pressure, 790
Hydrothermal method, 20
Hyperfine splitting, 802
Hyperpolarization, 413
Hyper-Raman scattering, 551, 552
Hysteresis curve, 328
Hysteresis loop, 337

I

Icosahedral quasicrystals, 100
Impact ionization, 1215, 1220
Implementations of qubits, 570
Impurity
bands, 864, 866, 1166–1172, 1196
conductivity, 1185
potential, 816, 832, 833
Index of refraction, 373
Indirect bandgap semiconductors, 512
Indirect semiconductors, 762
Induced level splitting, 793
Inelastic scattering, 990
Inhomogeneous line broadening, 722
Insulator, 260
Interaction parameter, 324
Interatomic forces, 36, 114
Interatomic potential, 114
Intercalated compounds, 84
Internal modes, 148
Intervalley interaction, 767
Intervalley scattering, 1013, 1047
Intracenter transitions, 842, 844, 848, 853
Intraionic transitions, 844, 850, 851
Intravalley scattering, 1041
Intrinsic carrier density, 910, 918
Intrinsic carrier generation, 910
Intrinsic level, 1261
Intrinsic luminescence, 1265
Intrinsic photoconductivity, 1300, 1306
Intrinsic semiconductor, 900
Intrinsic transitions, 1265
Inverse-opal structure, 409
Inversion-domain boundaries (IDBs), 641
Ioffe-Regel rule, 1171
Ionic bonding, 37
Ionicity, 43
Ionic plasma frequency, 440
Ionic polarizability, 431
Ionic radii, 48
Ionization
energy, 212
field, 796
rate, 1217, 1218
Ionized impurity scattering, 482, 999
Isoelectronically bound exciton, 787
Isoelectronic centers, 787
Isoelectronic defects, 834–835
Isoelectronic pairs, 788
Isomorph, 68
Isotope effect, 142, 838, 1091, 1093

J
Jahn-Teller distortion, 813, 818, 827, 853
Jahn-Teller effect, 819–820, 842
Jahn-Teller stabilization energy, 820
Johnson-Nyquist noise, 1283
Josephson junction, 571, 1098
Josephson tunneling, 1097–1099
Joule's heating, 964, 978

K
K•p method, 227
Keating's relation, 122
Kelvin relation, 1060
Kerr effect, 422
Kinetically limited growth, 30
Kinetic processes, 1313–1315
Kink site, 15
KKR method, *see* Green's function (KKR) method
Knudsen cells (K-cells), 26
Knudsen equation, 26
Kramers-Heisenberg dielectric function, 396
Kramers-Kronig analysis, 389
Kramers-Kronig interaction, 501
Kramers-Kronig relations, 398, 433–437
Kröger-Vink notation, 600
Kronig-Penney model, 201, 299
Kubo-Greenwood formula, 1167

L
LACO, *see* Linear combination of atomic orbitals (LCAO)
Landauer-Büttiker formalism, 1138–1140
Landauer formula, 1140, 1148
Landau levels, 516–518, 1070, 1072, 1076, 1078, 1080, 1084
Landé g factor, 321, 798, 1072
Langmuir-Blodgett films, 84
Laplace transformation, 621
LAPW, *see* Linearized augmented plane-wave (APW) method
Larmor frequency, 1073
Lasing, 1281
Lattice
  constant, 62
  coupling, 167, 726, 728, 936
  oscillation, 167
  polarization, 439
  relaxation, 605, 726–728
  temperature, 1026, 1031, 1033, 1035, 1037, 1043–1046, 1342
Laughlin state, 1082
Layered electron gas (LEG), 491–493
LDA, *see* Local density approximation (LDA)
Level splitting, 821, 823
Lifetime
  electron, 1328, 1331, 1366, 1367
  electron-hole plasma, 1348
  exciton, 1358
  LO phonons, 1352, 1353
  spin, 1366, 1368, 1370, 1372
  TO-phonon, 1351, 1353
Lifshitz model, 1168–1169
Ligand hyperfine splitting, 802
Light-hole band, 270
Light-hole exciton, 557
Lindemann relation, 166
Linear combination of atomic orbitals (LCAO), 230
Linearized augmented plane-wave (LAPW) method, 232
Linearized muffin-tin orbital method, 233
Line defects, 597
  Burgers vector, 629–631
  compound semiconductors, dislocations in, 632–635
  dislocation counting, 631–632
  dislocation kinks and jogs, 635–637
  dislocations and electronic defect levels, 638–639
  dislocation velocity, 637–638
  edge dislocations, 628
  screw dislocation, 628–629
Line shape, 716, 719–738
Liquid encapsulation, 20
Liquid-phase epitaxy (LPE), 24
Liquid phase growth, 18
Local density approximation (LDA), 235, 506
Local equilibrium, 27
Localization energy, 782, 865–867
Localized state, 206
Localized vibrational mode (LVM), 142
Local phonon modes, 142
Local pseudopotential, 769
Local-spin-density approximation (LSDA), 239
London penetration depth, 1094
Longitudinal dielectric constant, 394
Longitudinal lattice oscillation, 129
Longitudinal mode, 129
Lorentzian line shape, 719
Lorentzian line-shape function, 441
Lorentz number, 178

Lowest unoccupied molecular orbital (LUMO), 213, 248, 308
LSDA, *see* Local-spin-density approximation (LSDA)
Luminescence, 1265, 1279, 1340
Luminescence
  centers, 1278
  efficiency, 1265
  photoflux, 1268
Luminescent intensity, 1279
LUMO, *see* Lowest unoccupied molecular-orbital (LUMO)
Luttinger parameter, 272, 543, 772
Lyddane-Sachs-Teller, 442
Lying-down configuration, 878

M
Madelung constant, 38
Magnetic fields
  ballistic transport in, 1077–1083
  quantum effects in, 1070–1076
  transport equations and thermoelectric effects, 1056–1070
Magnetic moment, 321
Magnetic ordering, 324, 1104, 1109
Magnetic susceptibility, 321
Magnetization density, 321
Magneto-electric effects, 1064–1070
Magneto-phonon effects, 1076
Magnetoplasma reflection, 487–488
Magnetoresistance effect, 1057, 1067–1070, 1081
Magnetoresistance coefficient, 1068
Majority carriers, 911
Manley-Rowe equations, 418
Many-body problem, 220
Mathiessen's rule, 988
Matrix glasses, 99
Maxwell relation, 389
Maxwell's equations, 371
Mean free path, 864, 865, 946, 949, 977
Meissner-Ochsenfeld effect, 1094–1095
Meissner phase, *see* Meissner-Ochsenfeld effect
Metal, 262
Metal-induced gap states (MIGS) model, 694–695
Metal-ion vacancies, 601
Metallic bonding, 46
Metallic conductivity, 1170, 1171, 1190
Metalorganic chemical vapor deposition (MOCVD), 24
Metalorganic precursors, 28
Metalorganic vapor-phase epitaxy (MOVPE), 28
Metal-oxide semiconductor FET (MOSFET), 1115, 1116
Metal-semiconductor interfaces
  Bardeen model, 688
  classification of, 680–681
  defect model, 695
  MIGS model, 694–695
  Schottky barriers, chemical trends for, 691–694
Metamaterials, 373
Metastability, 837, 840, 841
Microcavity, 410, 412
Microcrystallite boundaries, 1010
Miller indices, 67
Mini-bandgap, 298
Mini-band structure, 299
Mini-Brillouin zone, 82
Minority carriers, 911, 918
Miscibility, diluted magnetic semiconductors, 331
Mobility, 262, 949, 993, 995, 997, 1001, 1003, 1007, 1014, 1041, 1043, 1047, 1181
  anisotropy, 1004
  in disordered organic semiconductors, 1193
  edge, 867, 883, 885, 886, 890, 1180–1181, 1182, 1192, 1197
  effective mass, 974–975
  gap, 1181
  in organic crystals, 1192
  in polymers, 1195–1196
  tensor, 974
Model potentials, 771
Modes of oscillations, 129
Modulation doping, 1115
Modulation spectroscopy, 387, 390
Mohs hardness, 51
Molecular beam epitaxy (MBE), 25
Molecule crystals, 47
Momentum effective mass, 281
Momentum relaxation time, 969, 975, 993, 1001, 1012
Monte Carlo method, 1038
Morphotrop, 68
MOSFET, *see* Metal-oxide semiconductor FET (MOSFET)
Mott density, 747, 924, 929
Mott transition, 922, 1169–1172
Muffin-tin potential, 224
Multiphonon absorption, 448–449
Multiphonon process, 1209
Multiplication factor, 1222

N
Nanocrystal nucleation, 90
Nanocrystals, 306
Nanowires, 87
Narrow-gap semiconductor, 262
Near-band-edge emission, 1275
Nearest-neighbor tight-binding model, 231
Nearly empty lattice, 242
Néel temperature, 329
Negative differential resistance (NDR), 1049, 1124, 1125, 1157
Negative photoconductivity, 1319
Negative refractive index, 373
Negative-*U* behavior, 827, 840, 841
Negative-*U* center, 837–840, 873
Network structures, 98
Neutrality condition, 909
Newtonian particle, 209
Noise, 1281
  contact, 1291
  electronic, 1283
  equilibrium, 1283
  flicker, 1287, 1291
  generation-recombination, 1286
  nonequilibrium, 1284
  quantum 1/f, 1288
  shot, 1284
  thermal, 1283
Nonequilibrium noise, 1284
Non-interacting electrons, 227
Nonlinear optical effects, 412, 422
Nonmetal-ion vacancies, 601
Nonradiative multiphonon recombination, 1250
Nonradiative recombination, 1243
  at Coulomb-attractive centers, 1244
  at deep centers, 1246
Nonresonant free-carrier absorption
  dispersion relation for, 475–478
  free-hole absorption, 483–484
  scattering mechanisms, 479–483
Nonvertical transitions, 511
Novotny-Juza compounds, 75
Nucleation, 11
Numerical solution, 1038

O
Ohm's law, 948
Oligoacenes, 78, 308
Oligomers, 79
Oligothiophenes, 78
One-dimensional photonic crystal, 404, 407
One-dimensional transport, 1132–1140
One-electron approximation, 222
One-mode behavior, 143
One-phonon spectrum, 448–449
Optical absorption, 1302, 1303
Optical absorption coefficient, 375
Optical bandgap, 214, 309
Optical branch, 133, 138
Optical carrier generation, 1207
Optical carrier heating, 1038
Optical cooling, 1037
Optical cross section, 716, 718
Optical defects, 410
Optical deformation potential scattering, 481
Optical generation rate, 1208
Optical phonons, 135
Optical phonon scattering, 995, 997
Optical rectification, 420
Ordered compounds, 84
Ordering, 288
Organic interlayer, 880, 881
Organic semiconductors, 54, 78, 308, 1168, 1187–1196
Organic superconductor, 1089–1090
Organic superlattices, 84
Orgel diagram, 821
Oscillator strength, 557, 716–719
Oscillatory luminescence, 1270
Oscillatory strength, 785
Over-charged donors, 780
Overcoordination, 868, 872
Oxygen impurities, 883
Oxygen in GaP and GaAs, 836–837

P
Pair distance, 776
Paracrystallinity, 887, 889
Paramagnetic ion, 321
Paramagnetic materials, 321
Partial dislocations, 631
Passivation, 831
Pauli operator matrix, 238
p-d exchange interaction, 334, 343
Peltier effect, 1060, 1084
Penetration depth, 1095
Penrose tiling, 104
Pentacene, 877–882, 884, 888
  on insulating substrates, 879–880
  on metals, 878–879
  on organic substrates, 880–882
Pentacenequinone, 883
Percolation, 1174, 1178, 1180, 1182, 1193

Periodicity approach, 199, 204
Periodic lattice structure, 204
Perpendicular carrier transport, 2D, 1122–1132
Persistent photoconductivity, 1317
Perturbative methods, 817
Phase matching, 419
Phase space, 966
Phase velocity, 136
Phillips cancellation theorem, 224
Phonons, 128, 137, 511
  activated conduction, 1172–1175
  in amorphous semiconductors, 150
  broadening, 736–738
  density of states, 140
  dispersion, 149
  distribution function, 162
  drag, 979
  dressed electrons, 938
  gas, 171
  hot, 1345, 1355
  modes, 143, 1355
  in organic semiconductors, 148
  Phonon-phonon scattering, 173
  polariton, 445–447
  replica, 729
  scattering rate, 1219
  spectra, 128
  in superlattices, 145
  velocity, 135
Phonon scattering, 177
  in amorphous semiconductors, 178
  at lattice defects, 174
  mechanisms, 172
Phosphorescence, 1280
Phosphors, 1278
Photoconductivity, 1300–1301
  carrier generation, 1301–1306
  extrinsic, 1309–1316
  intrinsic, 1306–1309
  main processes, 1300
  negative, 1319–1322
  persistent, 1317–1319
  response, 1309
Photo ionization, 734, 1304–1306
Photon
  entangled photon pairs, 576
  indistinguishable photons, 576
  single photons, 575
Photon-free-electron interaction
  electric and magnetic fields, carrier dispersion in, 484–491
  free-electron resonance absorption, 468–475
  nonresonant free-carrier absorption, 475–484
  2D semiconductors, plasmon dispersion in, 491–494
Photonic bandgaps, 402, 406, 408, 412
Photonic crystals, 402
  one-dimensional, 404
  scaling property, 404
  three-dimensional, 409
  two-dimensional, 407, 409
Photonic density of states, 408
Photon number distribution
  coherent radiation, 573
  nonclassical radiation, 574
  thermal radiation, 573
Photon-phonon interaction
  dielectric function in IR range, 438–445
  lattice polarization, 430–437
  scattering of photons with phonons, 445–462
Photon scattering at crystal boundaries, 179
Photothermal expansion, 170
Physical vapor deposition, 22
Physisorption, 27
Picosecond spectroscopy, 1338
Piezoelectric interaction, 994
Piezoelectricity, 432
Planar defects, 598, 644
  antiphase domains, 640–641
  large-angle grain boundaries, 642
  low-angle grain boundaries, 641–642
  stacking faults, 639–640
Planck's formula, 1263
Plasma frequency, 469
Plasmon energy, 395, 399
Plasmon-induced recombination, 1255, 1291
Pnictides, 75, 1100, 1102, 1106, 1109
Pockels effect, 421, 422
Point-charge Coulomb behavior, 770
Point defects, 597, 882–885
  Brouwer approximation, 617–620
  defect-chemistry, 611–614
  defect-formation energy, 609–611
  frozen-in intrinsic defect density, 607–609
  intrinsic point defects, density of, 602–607
  stoichiometry and compensation, changing of, 615–617
Point groups, 65
Poisson's equation, 380, 469, 1011
Poisson's ratio, 86
Polariton, 396
  dispersion, 549
  modes of quantum wells, 1360

Polarization energy, 213
Polaron, 214, 936–944
  existence criteria, 944
  large, 938–941
  mass, 940
  radius, 939
  self-energy, 937, 939
  small, 942–944
Polar optical scattering, 482
Polyacetylene, 886
Polymers, 80, 877, 886–890, 1194–1196
Polymorph, 68
Polythiophene, 886, 888
Polytypes, 73, 640
Poynting vector, 373, 383
Pressure coefficient, 144
Primitive unit cell, 64
Probability amplitude, 569, 576
Proximity approach, 205
  band structure, 194
  in organic crystals, 212
Pseudo Brewster angle, 384
Pseudodielectric function, 393
Pseudogap, 1102
Pseudopotentials, 224
  charge, 770
  first-principles, 226
Purcell effect, 411, 581, 1364

Q
QCL, *see* Quantum-cascade laser (QCL)
Quadratic Stark effect, 797
Quantization energy, 295, 306
Quantized state, 305
Quantum 1/f noise, 1288
Quantum-cascade laser (QCL), 1128–1131
Quantum dots (QDs), 88, 305, 563
Quantum efficiency, 1302
Quantum Hall effect, 1077–1083, 1142
Quantum information processing, 571
Quantum processing
  based on photons, 572
  based on superconduction, 571
Quantum spin, 1080
  Hall insulator, 1144
  Hall phase, 1080, 1084, 1148
Quantum wells (QW), 86, 295
  absorption in, 520–521
  energy levels in, 518–519
Quantum wire (QWR), 86, 304, 561–562, 1132, 1134, 1135, 1137, 1157
Quasi-Bohr radius, 705
Quasicrystals, 100
Quasi-Fermi levels, 921, 959, 1036
Quasi-free electron, 208
Quasi-hydrogen energy, 705, 758
Quasi-hydrogen radius, 531, 539, 758
Quasi-neutrality equation, 617
Quasiparticle GW calculations, 237
Quasi-particles, 1091, 1106
Quasiperiodicity, 101
Qubit, 568
  implementation, 570
  pure state, 570
Quenching
  field-induced, 1321–1322
  optical, 1320–1321
  thermal, 1321
  transition, 1320

R
Rabi frequency, 411
Rabi oscillation, 579
Rabi splitting, 411
Radiative recombination, 1249, 1262, 1267, 1268
Radical ion salts, 1189
Raman scattering, 449–462, 470, 551, 552
  first-and higher-order, 457
  in glasses, 461–462
  from Local Modes, 459
  polar and nonpolar, 456
  resonant, 458
  in superlattices, 459–461
Random distribution, 332
Rare earth ion, 322
Rayleigh scattering, 452
Reciprocal lattice, 68
Recombination, 1256
  of free excitons, 1270
  centers, 714, 813, 841, 853, 1257, 1261, 1315–1316
Recombination coefficient, 1267
Recombination cross-section, 1245, 1267
Recombination heating, 1360
Reduced effective mass, 529
Reduced exciton mass, 531
Reduced k vector, 207
Reduced mass, 546
Reflectance, 381, 384
Reflection, 1262
Refractive index, 387

Relaxation
electron, 1345
energy, 1337–1339
LO-phonon, 1353, 1354
mechanisms, 1343
momentum, 1333–1335
parameter, 52
phonon, 1350
phonon in quantum wells, 1355, 1356
process, 710–716
spin, 1364, 1365, 1367, 1370
spin exciton, 1371
spin holes, 1370
TO-Phonon, 1350
Relaxation time, 973, 1327
in dynamic processes, 1326
energy, 977, 1037, 1040, 1352
momentum, 968, 975, 1333
phonon, 1353
time approximation, 968–970, 1057
spin, 1364, 1367
Resonance scattering, 791
Resonant state, 733, 760
Resonant tunneling diode, 1123
Resonating-valence-bond (RVB) theory, 1107
Reststrahlen method, 444
Reststrahl range, 397
Richardson-Dushman current, 956
Roosbroeck-Shockley relation, 1267
Root mean square velocity, 945
Rubrene, 79, 249, 877, 885
Rutherford scattering, 1224
Rydberg constant, 529
Rydberg energy, 531, 536, 556, 759
Rydberg series, 529, 562

S
Saha model, 1360, 1362
Sasaki-Shibuya effect, 1048
Satellite conduction band, 268
Satellite minima, 244
Saturation drift-velocity, 1044
Saturation velocity, 1026
Scaling of photonic structures, 404
Scattering
carrier-carrier, 1006
centers, 988
at dislocations, 1005
elastic, 972, 989
inelastic, 972
by intrinsic point defects, 1008
ionized-impurity, 999
by neutral lattice defects, 1007
optical phonon, 997
processes, 1119, 1129, 1327, 1334, 1357
radius, 1001
surface-induced, 1009
with phonons, 989
Schönflies symbol, 65
Schottky disorder, 603
Schottky energy, 603
Schottky-Mott model, 682–688
Schrödinger equation, 220, 531, 535
Screened Coulomb potential, 400, 401, 1176
Screening function, 399
Screening parameters, 400
Second harmonic generation, 415, 416
Seebeck coefficient, 1063
Seebeck effect, 1084
Self-activated conductivity, 919
Self-energy, 237
Self-interstitials, 829–830
Self-trapped electron, 937, 942
Semiconducting glasses, 872, 874–875, 890
Semiconductors, 243, 260
Semimetal, 263
Sensitive photoconductor, 1315
Sensitization, 1310, 1320
Sequential tunneling, 1125, 1126
Seraphin coefficients, 389
Shallow-center spectra, 715
Shallow acceptor, 771
Shallow defects in compound semiconductors, 772–775
Shallow-level defect, 704–706
Shallow traps, 1311
Shell number, 776
Shockley approximation, 1042
Shockley-Haynes experiment, 1330, 1331
Shockley-Read-Hall Center, 1260
Short-range order, 94
Shot noise, 1284
Shubnikov-DeHaas effect, 1075
Shuffle set, 632
Sidevalley anisotropy, 763
Sign convention, 947
Single-component organic crystals, 1188
Single-electron transistor, 1153
Single-electron tunneling, 1153–1156
Single-photon emission
from bright excitons, 577
from dark excitons, 579
Singlet exciton, 535
Singlet–singlet transition, 534
Singlet state, 212

Site-percolation, 333
Size quantization, 85
Slater determinant, 230
Sliding-boat technique, 25
Small-molecule crystals, 876–885, 891, 1188
Snell's law, 382
Solvothermal method, 20
Sound damping, 127
Sound velocities, 125, 136
Sound waves, 123
Space groups, 65
Spatially indirect transitions, 792
Specific heat of amorphous semiconductors, 164
Specific resistivity, 948
Spectral hole burning, 722, 1342, 1349
Spectroscopic ellipsometry, 393
Spin degeneracy, 1136, 1138
Spin density, 874, 876
Spin-flip resonances, 801
Spin-glass, 333
Spin-orbit splitting, 238, 270, 274
Spin-polarized state density, 340
Spin relaxation, 1364, 1365, 1367, 1368, 1370, 1371, 1373
Split-gate technique, 1132
Split-off band, 270
Spontaneous emission, 1266
Square potential, 295
Square-well potential, 814–815
SQUID, *see* Superconducting quantum-interference device (SQUID)
Standing-up phase, 879, 891
Stark effect, 722, 796
Stark shift, 558
State filling, 1339, 1349
Steady state, 1259
Step-flow growth, 15
Sticking coefficient, 27
Stimulated emission, 1280
Stirling approximation, 604
Stokes line, 550
Stokes scattering, 451
Stokes shift, 453, 728
Stop band, 407
Strain coefficients, 116
Stranski-Krastanow growth, 14, 89
Stress-strain relations, 115
Strong-coupling regime, 411
Strong-field schemes, 822
Subband, 296, 1116, 1118, 1129, 1133, 1137, 1139, 1157
Substitutional defects, 832–835
Supercell technique, 818
Superconducting quantum-interference device (SQUID), 1097–1099
Superconduction energy-gap, 1104, 1109
Superconductive materials, 1088
Supercooling, 10
Superexchange, 325, 340, 341
Super-hyperfine splitting, 802
Superlattice, 81, 298
  ultrathin, 301
  thermal transport in, 179
Supersaturation, 10
Superspace approach, 103
Surface-induced scattering, 1009
Surface phonons, 146
Surface-polariton modes, 553
Surface states, 1141
Susceptibility tensor, 413
Symmetry, 71
Symmetry of E(k), 244
Symmetry points, 71

T

Tailing of states, 521, 739, 745, 862–869, 885
Tail states, 742, 744
Tanabe-Sugano diagram, 822
Tetrahedrally bonded amorphous semiconductors, 870, 875, 890
The binding energy, 775
Thermal conductivity, 170
  alloy, 175
  cross-plane, 185
  modeling, 184
Thermal energy flux, 171
Thermal equilibrium, 1259
Thermal excitation probability, 1210
Thermal expansion, 165, 169
Thermal ionization, 1209
Thermal noise, 1283
Thermal radiation, 1263, 1264
Thermal resistivity, 175
Thermal transport in superlattices, 179
Thermal velocity, 945
Thermodynamic equilibrium, 900
Thermoelectric effects, 1056–1070
Thermoelectric power, 1063, 1064
Thermomagnetic effects, 1062, 1084
Thin-film phase, 876, 878
Thiophenes, 79
Third harmonic generation, 416
Third-order elastic constants, 118

Third-order susceptibility, 416
Thomas-Fermi screening length, 399, 401
Three-dimensional lattices, band structure of, 240
Three-dimensional photonic crystal, 407, 409
Threshold energy for impact ionization, 1217
Tight-binding approach, 197, 230
Tight-binding approximation, 818, 841
Time reversal
  invariance, 1144
  symmetry, 1149
Topological band theory, 1141
Topological insulator, 1080, 1084, 1140
  higher order, 1152
  second generation, 1149
  three-dimensional, 1148
Total carrier mobility, 988
Total current, 957
Transfer integral, 863, 866, 877
Transistor, 5, 1115
Transition
  coefficient, 1207
  metal impurities, 813, 819, 821, 832, 841–853
  metal ion, 322
  metals, 330
  probability, 1208
  rate, 1207
  temperature, 92
Transit time, 1328
Transmission coefficient, 1123, 1139, 1157
Transmittance, 381, 384, 386
Transmon, 571
Transport
  analysis, 1336
  ballistic, 1335
  bandgap, 309
  coefficients, 1059
  dispersive, 1330–1333
  equations, 1056–1070
  limited growth, 30
Transverse and longitudinal excitons, 545
Transverse dielectric constant, 394
Transverse-electric (TE) modes, 407
Transverse lattice oscillation, 131
Transverse-magnetic (TM) modes, 407
Transverse oscillation in a diatomic lattice, 132
Transverse polarizations, 131
Trapping, 206, 1256, 1310–1315, 1327, 1329, 1330, 1373
Trap states, 877, 879, 882, 883, 885, 887
Trions, 530, 553
  2D, 558
  free, 553
Triplet excitons, 535
Triplet-singlet transition, 534
Triplet state, 212
Tunneling, 300, 1227
  assisted by phonons, 1232
  assisted by traps, 1233
  band-to-band, 1228
  coherent and sequential, 1125–1126
  current, 1230
  definition, 1122
  double-barrier quantum well, 1123–1125
  induced transport, 1197
  parabolic barrier, 1228
  planar barrier, 1122–1123
  rectangular Barrier, 1225
  single-electron, 1153–1156
  and size quantization, 1128
  spectroscopy, 1231
  superlattices, 1125
  three-dimensional, 1229
  triangular barrier, 1227
  wells, variable width, 1126
Two-component organic semiconductors, 1189–1191
Two-dimensional electron gas (2DEG), 491, 1115, 1132–1136
  carrier mobility, 1119–1121
  energy levels, 1117–1118
  heterointerface, 1116–1117
  inversion layer, 1115–1116
Two-dimensional hole gas (2DHG), 1115, 1132
Two-dimensional photonic crystal, 407, 409
Two-electron transitions, 785
Two-fluid model, 1093
Two-hole transitions, 785
Two-level acceptors, 780
Two-mode behavior, 143
Type I and II band alignment, 294
Type I and II pairs, 776
Type I and II superconductors, 1095–1097, 1099, 1109

U

Ultrathin superlattices, 82
Umklapp process, 173, 989
Umklapp processes, 1014
Uniaxial stress, 793
Unit cell, 60, 69
Up-conversion, 418, 419
Urbach tail, 740, 744

V
Vacancy, 819, 824–825
in compound semiconductors, 828–829
in covalent crystals, 826–828
in GaAs, 828
vacancy states, 832
Valence band, 259
cubic semiconductors, 269
quantum well, 297
wurtzite semiconductors, 273
Valence-band offsets
common-anion rule, 657
commutativity for, 657
crystal face orientation, 657
experimental values for, 658
Harrison valence-band offsets, 666
transitivity of, 657
Van der Waals, 78
bonding, 46
interaction, 47, 55
Van Hove singularity, 502
Vapor-liquid-solid (VLS), 87, 562
Vapor-phase epitaxy (VPE), 24
Vapor-phase growth, 21
Variable-range hopping, 1183–1184, 1186, 1193
V center, 825
Vegard's rule, 52
Verdet coefficient, 490
Vertical gradient freeze (VGF), 19
Voigt configuration, 488
Voigt notation, 117
Volmer-Weber growth mode, 14
von-Klitzing constant, 1079
von Neumann noncrossing principle, 547, 558
Vortices, 1095, 1096, 1109

W
Wannier function, 223
Wannier-Mott excitons
band continuum, 537
band degeneracies, lifting of, 543
effective masses, anisotropy of, 543
electron-hole exchange interaction, 544
exciton-phonon interaction, 545
indirect-gap excitons, 546–547
reduced exciton mass, 543
Wannier-Mott excitons, 529, 535
Wannier-Stark ladder, 1127
Warm carrier, 1034
Warm electrons, 1035
Water contamination, 884–885
Wave equation, 372, 373, 403
Waveguide, 410
Wave packet of Bloch functions, 757
d-wave pairing, 1104
Wave vector, 68, 123, 199
Weak coupling, 411
Weak-field scheme, 822
Wetting angle, 13
Wetting layer, 89
Wiedemann-Franz law, 178
Wigner-Seitz cell, 70
Woodpile structure, 409
Wulff's theorem, 17
Wurtzite lattice, 72
Wurtzite-type semiconductors, 273, 543

Y
Young's relation, 13
Yukawa potential, 400

Z
Zeeman effect, 797
Zeeman splitting, 798
Zeeman transitions, 801
Zero-dimensional transport, 1152–1156
Zero-field cooling, 334
Zero-phonon transitions, 820, 850, 851, 853
Zero-point energy, 132
Zincblende lattice, 72